Applying Design of Experiments to Evaluate the Influence of Parameters on the Economic Feasibility of the Eco-Industrial Parks

Kien Cao-Van(✉)

Faculty of Information Technology, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam
cvkien@ntt.edu.vn

Abstract. This research employs the design of experimental (DoE) to examine how various parameters impact the economic feasibility and overall satisfaction of enterprises operating within eco-industrial parks (EIPs). A full factorial design is constructed, using economic feasibility and overall satisfaction as response variables, and experimental data is generated by simulating diverse scenarios. Each iteration of the experiment utilizes a single-leader multi-follower (SLMF) game optimization model, focusing on designing water exchange networks within EIPs. The investigation encompasses several parameters in a case study involving ten follower enterprises aiming to minimize their annual operational costs. Concurrently, the EIP authority assumes the leader role with the objective of reducing the collective freshwater consumption of the EIP. Furthermore, this study employs binary logistic and multi-linear regressions to establish causal relationships. These relationships link input parameters with economic feasibility and overall satisfaction of operating businesses within EIPs. Ultimately, the reliability of the DoE methodology is showcased, offering valuable insights into enterprise parameters, EIP design, economic feasibility, and overall satisfaction.

Keywords: Eco-industrial park · Design of experiments · Single-leader multi-follower game · Logistic regression · Multi-linear regression

1 Introduction

Design of Experiments (DoE) serves as a systematic and efficient approach, empowering researchers and engineers to comprehend the impact of experimental parameters (x_i, independent variables) on response variables (y, dependent variables). This is achieved by constructing mathematical models ($y = f(x_i)$). Originally formulated in the early 1920s, the research planning strategy, DoE, was pioneered by Sir Ronald Fisher in the field of agricultural research [6]. About 50 years for the DoE technique has been widely applied in different fields of medicine, biology, marketing research, and industrial production. We refer

P. Cong Vinh and H. Mahfooz Ul Haque (Eds.): ICTCC 2023, LNICST 586, pp. 148–172, 2024.
https://doi.org/10.1007/978-3-031-59462-5_11

the reader to references [9,16,17,19] for a survey on different applications of DoE. DoE allows using a minimum experiments, in which several experimental parameters are systematically varied to determine their effects on the response variables. Utilizing the acquired data, we construct a regression model for the analyzed procedure. This regression model serves the purpose of comprehending how the experimental parameters impact the response variables and identifying the optimal conditions for the process.

We refer the reader to [12] for general guidelines and procedures for implementing DoE. In fact, the practical steps required to plan and conduct a DoE include: stating the objectives, defining response variables, determining factors and levels, determining experimental design type, performing the experiment, data analysis, and practical conclusions and recommendations. Furthermore, the selection of factors and levels in the DoE usually depends on the type of investigation, the type of process, and the available resources. Thus, different selections of levels or factors will result in different DoEs that could theoretically be appropriate for the type of investigation. In a very recent publication [8], the authors recommend procedures for preparing input data for various types of experimental designs to select the most successful and the most efficient designs.

Due to the growing concern about the limitation of the available resources (steam, water, electricity, ...) and the need for more sustainable industrial development, the concept of Industrial Ecology (IE) has emerged [4]. IE is described as a group of interconnected industries in a certain region where waste production and resource consumption are reduced by allowing waste products from one sector to be used as raw materials for another. One of its main practical applications is the design of eco-industrial parks (EIPs) and a definition commonly admitted was given by Lowe [10] as: "A community of manufacturing and service businesses seeking enhanced environmental and economic performance through collaboration in managing environmental and resource issues including energy, water, and materials. By working together, the community of businesses seeks a collective benefit that is greater than the sum of the individual benefits each company would realize if it optimized its individual performance only". As we can see, enterprises operating in eco-industrial parks not only gain economic benefits but also achieve environmental benefits. To achieve these benefits, it is necessary to design an optimal resource exchange network between enterprises.

Based on a literature review, Boix et al. [4] classified resource exchanges in EIPs into three types: water, energy, and material. Moreover, water exchange constitutes the predominant form of interaction within EIPs, where the optimization of water distribution and the reduction of freshwater consumption often revolve around the strategic inclusion of a wastewater treatment facility alongside its associated network for the discharge of treated water. Most of the resource exchange network models proposed in the EIP have been modeled as an SLMF game [1,14,15]. In the SLMF game, the upper-level optimization subproblem represents the leader's decision-making viewpoint in order to minimize the consumption of natural resources while the lower-level optimization subproblems represent the followers' viewpoint in order to minimize their operating cost. Figure 1 shows the general scheme of such a model.

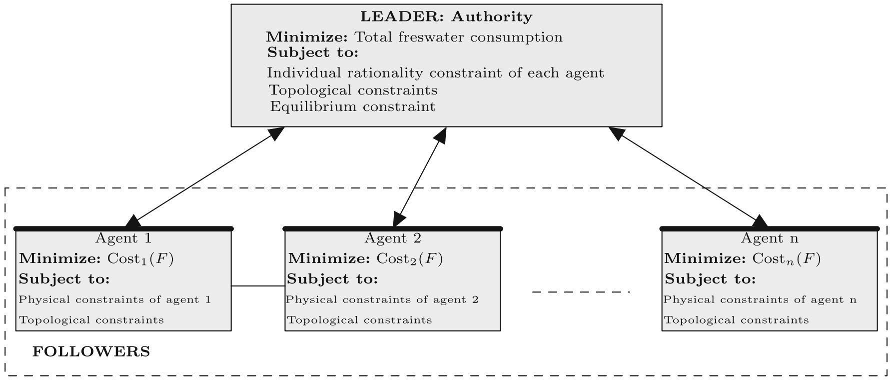

Fig. 1. General scheme of SLMF Game

In SLMF game model, the system variables are divided into those controlled by the leader and by the followers. The objective functions of both upper- and lower-level subproblems involve the controlled variables of both leader and followers. The constraint of the upper-level subproblems contains the lower-level subproblem. At the upper-level subproblem, the authority of the park chooses the connections of the exchange network, whereas at the lower-level subproblem, each enterprise manages the distribution of its output flux and natural resource consumption. In accordance with determinations by the EIP authority, all enterprises engage in a parametric non-cooperative generalized Nash game, where the strategies of the EIP authority are treated as external parameters. As exemplified in prior works [1,14,15], the robustness of this approach is evident, as it allows for the comprehensive assessment of multiple criteria within the model. Of paramount significance, the equilibrium solution ensures economic gains for each constituent enterprise engaged in the EIP.

The primary objective of this study is to combine the concepts of experimental design and the SLMF game model proposed in [15]. This fusion aims to assess how distinct parameters affect both the economic feasibility and the overall satisfaction of operational enterprises in eco-industrial parks.

The forthcoming sections of this paper are structured as follows: Sect. 2 explores the water exchange network model within eco-industrial parks, utilizing a single-leader multi-follower game framework. The conversion of the eco-industrial park modeling challenge into a mixed integer linear programming formulation is covered in Sect. 3. Moving on to Sect. 4, an empirical investigation is carried out using statistical experimental design to analyze the influence of various parameters on the economic feasibility and overall satisfaction of operating enterprises in eco-industrial parks. Finally, concluding thoughts and prospects for the future are presented in Sect. 5.

2 EIP Problem Statement and Model

In this section, we follow the ideas from papers [1,14,15].

2.1 EIP Problem Statement

From a number of specific enterprises in the EIP, it is possible to create a suitable network of connections between them. Each enterprise is required to comply with predefined parameters for the quantity and quality of incoming water, as well as the quantity and quality of the output wastewater. For each enterprise, the water input may include fresh water and/or water supplied from other enterprise units. In fact, wastewater from one enterprise can be transferred to other enterprise and/or released into the environment immediately. The main goal of the EIP model is to define an efficient water connection system between units, ensuring both reduction of global freshwater consumption and cutting operating costs for individual companies on the industrial premises, while still complying with all process and environmental constraints.

2.2 Enterprise's Problem

Assume n is the fixed number of firms in the EIP. We use the index set $I := \{1, \ldots, n\}$ to represent enterprises and 0 to represent a sink node representing a polluted water treatment pit. Importantly, we determine that $I_0 = \{0\} \cup I$. Each enterprise $i \in I$ receives water throughput from partners in the EIP. However, due to technical limitations on the P_i process, the pollutant concentration delivered by the other enterprises is limited to a maximum value here denoted by $C_{i,\text{in}}$ [ppm]. In contrast, each firm $i \in I$ produces a fixed amount of pollutant M_i [g/h] which, from its in-house production process, needs to be diluted before discharge. To do this, enterprise i must consume a sufficient amount of fresh water z_i [T/h] to ensure that the concentration of pollution in the exit flux after dilution is always lower than $C_{i,\text{out}}$. This means that, when considering enterprise i in I, they will optimize their processes to ensure that the output contaminant concentration always reaches the $C_{i,\text{out}}$ value. Obviously we have that $C_{i,\text{in}} \leq C_{i,\text{out}}$. This structure is illustrated in Fig. 2.

Within eco-industrial parks (EIPs), enterprises engage in the exchange of materials. Conversely, an exchange network within the EIP context takes the form of a simple directed graph denoted as (I_0, E), where the presence of a link $(i, j) \in E$ signifies that enterprise i has the capability to direct its output water to enterprise j. In a similar vein, the utilization of the connection $(i, 0)$ by enterprise i signifies the disposal of water from the park, releasing it into the environment.

The *stand-alone* and *complete* configurations are defined as follows:

$$E_{\text{st}} := \{(i, 0) \,:\, i \in I\} \quad \text{and} \quad E_{\max} := \{(i, j) \,:\, i \in I,\, j \in I_0\},$$

a valid exchange network E must satisfy that $E_{\text{st}} \subset E \subset E_{\max}$. This definition yields that:

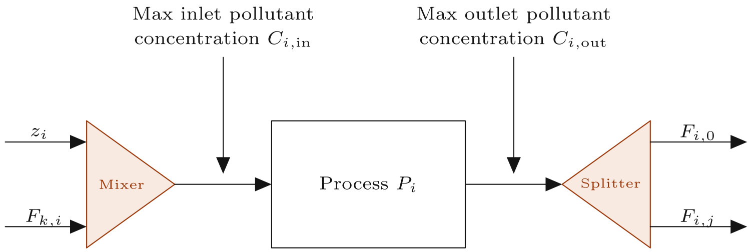

Fig. 2. Water mixture description for a given enterprise. Here $C_{i,\text{in}} \leq C_{i,\text{out}}$.

1. Every enterprise $i \in I$ maintains a connection to the sink node, i.e., $(i, 0) \in E$.
2. The sink node doesn't possess any outgoing connections, signifying that $(0, i) \notin E$ for all $i \in I$.

The collection of permissible networks within the EIP framework is represented as $\mathcal{E}$. Subsequently, for any E belonging to $\mathcal{E}$, the collection of unutilized connections, distinct from those in E, is denoted as E^c, where $E^c = E_{\max} \setminus E$.

Let $F_{i,j}$ [T/h] symbolize the water flow across the link (i, j) for every $(i, j) \in E_{\max}$. Furthermore, we establish $F = (F_{i,j} \ : \ (i, j) \in E_{\max})$ as the comprehensive flux vector across the network.

Moreover, considering each enterprise denoted as $i \in I$, we represent F as (F_i, F_{-i}), with $F_i = (F_{i,j} \ : \ j \in I_0)$ and $F_{-i} = (F_{k,j} : k \in I \setminus i)$. This representation underscores the array of flows associated with enterprise i. Consequently, given a fixed network E, a permissible flow vector F must adhere to the subsequent set of constraints:

1. **Use of connections in** E**:** Given that E denotes the feasible connections, it becomes imperative to enforce the following constraint:
$$\forall (i, j) \in E^c, \, F_{i,j} = 0. \tag{1}$$
2. **Water Mass Conservation:** Considering enterprise $i \in I$, the following equation holds true for water mass balance:
$$z_i + \sum_{(k,i) \in E} F_{k,i} = \sum_{(i,j) \in E} F_{i,j}. \tag{2}$$
3. **Contaminant Mass Conservation:** In the context of enterprise $i \in I$, the ensuing equation reflects the equilibrium of contaminant mass:
$$M_i + \sum_{(k,i) \in E} C_{k,\text{out}} F_{k,i} = C_{i,\text{out}} \sum_{(i,j) \in E} F_{i,j}. \tag{3}$$
4. **Inlet/outlet concentration constraints:** for an enterprise $i \in I$ we have
$$\sum_{(k,i) \in E} C_{k,\text{out}} F_{k,i} \leq C_{i,\text{in}} \left(z_i + \sum_{(k,i) \in E} F_{k,i} \right). \tag{4}$$

5. **Positivity of fluxes:** It is a requisite that all fluxes within the park maintain a positive value:

$$\forall (i,j) \in E,\ F_{i,j} \geq 0 \qquad \text{and} \qquad \forall i \in I,\ z_i \geq 0. \tag{5}$$

Note that by amalgamating Eqs. (2) and (3), we derive:

$$M_i + \sum_{(k,i)\in E} C_{k,\text{out}} F_{k,i} = C_{i,\text{out}} \left(z_i + \sum_{(k,i)\in E} F_{k,i} \right), \quad \forall i \in I, \tag{6}$$

Thus, the quantity of fresh water procured by enterprise $i \in I$ is determined by the flows originating from other enterprises, indicating:

$$z_i(F_{-i}) = \frac{1}{C_{i,\text{out}}} \left(M_i + \sum_{(k,i)\in E} \left(C_{k,\text{out}} - C_{i,\text{out}} \right) F_{k,i} \right). \tag{7}$$

Each enterprise i strives to minimize its operational cost, as defined by the following expression:

$$\text{Cost}_i(F_i, F_{-i}, E) = A \left[c_f \cdot z_i(F_{-i}) + \delta \sum_{(k,i)\in E} F_{k,i} + \delta \sum_{\substack{(i,j) \in E \\ j \neq 0}} F_{i,j} + \beta F_{i,0} \right]. \tag{8}$$

In this context, A [h] represents a time constant that characterizes the park's lifecycle analysis, while c_f [\$/T] pertains to the procurement cost of freshwater. Additionally, δ [\$/T] accounts for the expenditure associated with pumping contaminated water between enterprises, and β [\$/T] denotes the expense linked to discharging polluted water. It is noteworthy that the model adopts an assumption wherein the cost β significantly outweighs δ.

With all of these considerations, the problem of each enterprise i, for a given network $E \in \mathcal{E}$, is given by:

$$P_i\left(F_{-i}, E\right) = \begin{cases} \min\limits_{F_i} \text{Cost}_i\left(F_i, F_{-i}, E\right) \\ s.t. \begin{cases} z_i + \sum\limits_{(k,i)\in E} F_{k,i} = \sum\limits_{(i,j)\in E} F_{i,j}, \\ \sum\limits_{(k,i)\in E} C_{k,\text{out}} F_{k,i} \leq C_{i,\text{in}} \left(z_i + \sum\limits_{(k,i)\in E} F_{k,i} \right), \\ z_i(F_{-i}) \geq 0, \\ F_i \geq 0, \\ F_i\big|_{E^c} = 0. \end{cases} \end{cases} \tag{9}$$

In the context of a network $E \in \mathcal{E}$, we define a vector F as an equilibrium for the enterprises at the lower-level problem if and only if

$$\forall i \in I,\ F_i \text{ solves the problem } P_i\left(F_{-i}, E\right).$$

The collection of such equilibria for network E is symbolized as $\mathrm{Eq}(E)$.

Remark 1. *Note that in cases where enterprise i does not receive contaminated water, its freshwater consumption z_i is determined as follows:*

$$z_i = \frac{M_i}{C_{i,\mathrm{out}}}.$$

This subsequently leads to the formulation of the standalone operational cost, represented as STC_i [\$/h], according to the equation:

$$\mathrm{STC}_i = A \cdot (c_f + \beta) \frac{M_i}{C_{i,\mathrm{out}}}. \tag{10}$$

2.3 Authority's Problem

At the upper-level problem, the EIP authority wants to reduce the entire use of natural resources, thus he attempts to reduce the function:

$$Z(F) = \sum_{i \in I} z_i(F_{-i}). \tag{11}$$

In order to incentivize an enterprise $i \in I$ to participate in the EIP, the EIP authority needs to offer a proportional cost reduction of $\alpha \in]0,1[$ compared to the expenses of standalone operation. This entails ensuring that:

$$\mathrm{Cost}_i(F_i, F_{-i}, E) \leq \alpha \cdot \mathrm{STC}_i. \tag{12}$$

The concept of this minimal relative enhancement was initially introduced in [15]. Subsequently, the problem of the EIP authority is

$$\begin{aligned} &\min_{F \in \mathbb{R}^{|E_{\max}|}, E \in \mathcal{E}} Z(F) \\ &s.t. \begin{cases} F \in \mathrm{Eq}(E), \\ \mathrm{Cost}_i(F_i, F_{-i}, E) \leq \alpha \cdot \mathrm{STC}_i, \qquad \forall i \in I. \end{cases} \end{aligned} \tag{13}$$

Interpreting optimization problem (13) reveals the following scenario: The EIP authority will offer enterprises a feasible exchange network denoted as E and a corresponding operation $F \in \mathbb{R}^{|E^{\max}|}$ that adheres to all physical constraints. Importantly, this operation ensures that no enterprise has the motivation to independently deviate from the proposal due to constraint $F \in \mathrm{Eq}(E)$, while simultaneously securing the participation of all enterprises. This participation is effectively guaranteed by the precondition established in constraint (12).

3 Solution Methodologies

The authority's problem (13) is formulated in the form of mathematical programming with equilibrium constraints (MPEC) (see, e.g., [2,11,18]). This section is dedicated to illustrating the transformation of the challenging-to-solve MPEC formulation into a single Mixed-Integer programming problem.

3.1 Characterization of Equilibria

Given a topology E, we will write

$$E_{i,\text{in}} := \{(k,i) : (k,i) \in E\} \quad \text{and} \quad E_{i,\text{out}} := \{(i,j) : (i,j) \in E\}.$$

We define the set of active and inactive of i, denoted by I_{act} and I_{inact} respectively as follows:

$$\begin{aligned} I_{act} &:= \{i \in I \mid \exists j \in I \text{ with } (i,j) \in E_{i,\text{ out}}\} \\ I_{inact} &:= \{i \in I \mid E_{i,\text{ out}} = \{(i,0)\}\}. \end{aligned}$$

If each enterprise $i \in I_{\text{act}}$, we define the set of active arcs of i, denoted by $E_{i,\text{act}}$, that is,

$$E_{i,act} := \{(i,j) \in E_{i,\text{ out}} : j \neq 0\}.$$

If each enterprise $i \in I_{\text{inact}}$, we define the set of inactive arcs of i, denoted by $E_{i,\text{ inact}}$, that is,

$$E_{i,\text{ inact}} := \{(i,0)\}.$$

Theorem 2. *For any valid exchange network $E \in \mathcal{E}$ and denoting $S(E)$ by the set*

$$S(E) = \left\{ F \; : \forall i \in I, \begin{array}{ll} \begin{cases} z_i(F_{-i}) + \sum\limits_{(k,i)\in E} F_{k,i} = \sum\limits_{(i,j)\in E_{i,\text{act}}} F_{i,j}, \\ F_{i,0} = 0, \end{cases} & \text{if } i \in I_{\text{act}} \\ z_i(F_{-i}) + \sum\limits_{(k,i)\in E} F_{k,i} = F_{i,0}, & \text{if } i \in I_{\text{inact}} \\ \sum\limits_{(k,i)\in E} C_{k,\text{out}} F_{k,i} \le C_{i,\text{in}} \left(z_i + \sum\limits_{(k,i)\in E} F_{k,i} \right) & \\ z_i(F_{-i}) \ge 0 & \\ F_i \ge 0 & \end{array} \right\} \tag{14}$$

then, one has $S(E) = \text{Eq}(E)$. Additionally, any optimal solution (F,E) of the mathematical programming problem

$$\begin{aligned} &\min_{F\in\mathbb{R}^{|E_{\max}|}, E\in\mathcal{E}} Z(F) \\ &s.t. \begin{cases} F \in S(E), \\ \text{Cost}_i(F_i, F_{-i}, E) \le \alpha \cdot \text{STC}_i, \qquad \forall i \in I. \end{cases} \end{aligned} \tag{15}$$

constitutes an optimal solution for the SLMF problem (13).

Proof. The proof follows a similar path as outlined in [15, Theorem 4.1]. Our focus centers on demonstrating the equivalence $S(E) = \text{Eq}(E)$, as the latter aspect naturally emerges by substituting the constraint "$F \in \text{Eq}(E)$" with "$F \in S(E)$".

Hence, our objective is to demonstrate the inclusion $S(E) \subseteq \mathrm{Eq}(E)$. Let's consider an arbitrary $F \in S(E)$. Because $E_{i,\mathrm{act}} \subset E$ for all $i \in I_{\mathrm{act}}$ and $E_{i,\mathrm{inact}} \subset E$ for all $i \in I_{\mathrm{inact}}$, it follows that F_i serves as a feasible solution for $P_i(F_{-i}, E)$ across all $i \in I$.

Now, fix $i \in I$ and let F_i' be another feasible point of $P_i(F_{-i}, E)$. Then, $F_i' \geq 0$ and the water mass balance constraint (2) is satisfied. Therefore,

- if $i \in I_{\mathrm{act}}$, one has

$$\begin{aligned}\mathrm{Cost}_i(F_i', F_{-i}, E) - \mathrm{Cost}_i(F_i, F_{-i}, E) &= A\left[\delta \sum_{(i,j)\in E} F_{i,j}' + \beta F_{i,0}' - \delta \sum_{(i,j)\in E_{i,\mathrm{act}}} F_{i,j}\right] \\ &\geq A\delta\left[\sum_{(i,j)\in E} F_{i,j}' + F_{i,0}' - \sum_{(i,j)\in E_{i,\mathrm{act}}} F_{i,j}\right].\end{aligned}$$

  Moreover, the mass balance constraint (2) is satisfied for F_i' and F_i for any $i \in I_{\mathrm{act}}$, thus

$$\sum_{(i,j)\in E, j\neq 0} F_{i,j}' + F_{i,0}' = z_i(F_{-i}) + \sum_{(k,i)\in E} F_{k,i} = \sum_{(i,j)\in E} F_{i,j} = \sum_{(i,j)\in E_{i,\mathrm{act}}} F_{i,j}.$$

  Hence, $\mathrm{Cost}_i(F_i', F_{-i}, E) \geq \mathrm{Cost}_i(F_i, F_{-i}, E)$ for any $i \in I_{\mathrm{act}}$.
- if $i \in I_{\mathrm{inact}}$, one has

$$\begin{aligned}\mathrm{Cost}_i(F_i', F_{-i}, E) - \mathrm{Cost}_i(F_i, F_{-i}, E) &= A\left[\beta F_{i,0}' - \beta F_{i,0}\right] \\ &= A\beta\left[F_{i,0}' - F_{i,0}\right].\end{aligned}$$

  Moreover, the mass balance constraint (2) is satisfied for F_i' and F_i for any $i \in I_{\mathrm{inact}}$, thus

$$F_{i,0}' = z_i(F_{-i}) + \sum_{(k,i)\in E} F_{k,i} = F_{i,0}.$$

  Hence, $\mathrm{Cost}_i(F_i', F_{-i}, E) = \mathrm{Cost}_i(F_i, F_{-i}, E)$ for any $i \in I_{\mathrm{inact}}$.

Thus, F_i solves $P_i(F_{-i}, E)$, and since this holds for every $i \in I$, we conclude that $F \in \mathrm{Eq}(E)$.

Now, let us prove that $\mathrm{Eq}(E) \subseteq S(E)$. Let $F \in \mathrm{Eq}(E)$, and assume that $F \notin S(E)$. Since for each $i \in I$ the vector F_i is a feasible point of $P_i(F_{-i}, E)$, so $F \notin S(E)$ if there exists $i_0 \in I_{\mathrm{act}}$ such that $F_{i_0,0} > 0$. Let $(i_0, j) \in E_{i,\mathrm{act}}$ and let us consider the vector F_{i_0}' given by

$$F_{i_0,k}' = \begin{cases} F_{i_0,k} & \text{if } k \in I \setminus \{0, j\}, \\ 0 & \text{if } k = 0, \\ F_{i_0,j} + F_{i_0,0} & \text{if } k = j. \end{cases}$$

We have that $F_{i_0}' \geq 0$ and also

$$z_i(F_{-i_0}) + \sum_{(k,i_0)\in E} F_{k,i_0} = \sum_{(i_0,j)\in E} F_{i_0,j} = \sum_{(i_0,j)\in E} F_{i_0,j}'.$$

Thus, since F_{-i_0} remains the same, F'_{i_0} is a feasible point of $P_i(F_{-i_0}, E)$. Furthermore, we have that

$$\begin{aligned}\mathrm{Cost}_{i_0}(F'_{i_0}, F_{-i_0}, E) - \mathrm{Cost}_{i_0}(F_{i_0}, F_{-i_0}, E) &= \delta \sum_{(i_0,j)\in E, j\neq 0} F'_{i_0,j} + \beta F'_{i_0,0} - \left(\delta \sum_{(i_0,j)\in E} F_{i_0,j} + \beta F_{i_0,0}\right) \\ &= (\delta - \beta)\, F_{i_0,0} \\ &< 0.\end{aligned}$$

This yields that F_{i_0} doesn't solve $P_i(F_{-i_0}, E)$, which is a contradiction. Thus, $F \in S(E)$. Finishing the proof.

3.2 Mixed-Integer Formulation

Theorem 2 establishes the remarkable revelation that the authority's problem can be reformulated into a "classical" programming problem through the MPEC formula. Yet, due to the presence of exchange network topologies as variables in this programming context, a numerical implementation might pose a challenge. Consequently, within this section, we will illustrate the process of effectively dealing with a more conventional mixed-integer programming problem.

Let's start by introducing the crucial idea? what we refer to as arc classes? that we will employ to arrive at the final formulation: We define the arc classes exiting from each enterprise $i \in I$ as the sets:

$$C_{i,p} = \{(i,j) \in E_{\max} \ : \ j \in I\} \quad \text{and} \quad C_{i,0} = \{(i,0)\}.$$

The family of all arc classes that exit from i is denoted by the symbol $\mathcal{C}_i$, which is defined as $\mathcal{C}_i = \{C_{i,p}, C_{i,0}\}$.

Observe that, for each enterprise $i \in I_{\text{act}}$, the class $C_{i,p}$ is always satisfied

$$E_{i,\text{act}} \subseteq C_{i,p}. \tag{16}$$

This class is given by $C = C_{i,p}$ where (i,p) is any element of $E_{i,\text{act}}$.

Now, for each enterprise $i \in I$, we add two integer variables, $y_{i,p}, y_{i,0} \in \{0,1\}$ with the following interpretation:

- When $y_{i,p}$ is equal to 1, it signifies the inclusion of the connections within $C_{i,p}$ in the network structure.
- Similarly, when $y_{i,0}$ is set to 1, it denotes that the connection $(i,0)$ serves as the exclusive exit pathway for node i, while also denoting the involvement of node i in the EIP.

We establish the following constraints with this new boolean variable:

1. For each enterprise i within the set I, we assign the condition:

$$y_{i,p} + y_{i,0} = 1, \tag{17}$$

indicating that only one class is in an active state.

2. For each enterprise i in the set I, we establish the following condition to hold:

$$\sum_{(i,j)\in C_{i,p}} F_{i,j} \leq K \cdot y_{i,p}, \tag{18}$$

where $K > 0$ is a sufficiently large constant ensuring that all fluxes within the park remain below K. This constraint guarantees that when the connections $C_{i,p}$ are excluded from the network, the flow values $F_{i,j}$ become zero for any $j \in I$.
3. For each enterprise i in the set I, we establish the following condition to hold:

$$F_{i,0} \leq K \cdot y_{i,0}, \tag{19}$$

for some constant $K > 0$ large enough. This constraint ensures that, whenever $C_{i,0}$ are not included in the network, then $F_{i,0} = 0$.

Starting from the binary vector $y \in \{0,1\}^{2n}$, we proceed to construct the graph associated with y as depicted by the equation:

$$E(y) = \left(\bigcup (i,p) \ : \ y_{i,p} = 1\right) \cup (i,0) \ : \ i \in I. \tag{20}$$

We consider then the following Mixed-Integer optimization problem:

$$\min_{F\in\mathbb{R}^{|E_{\max}|}, y\in\{0,1\}^{2n}} Z(F)$$
$$s.t. \begin{cases} z_i + \sum_{(k,i)\in E_{\max}} F_{k,i} = \sum_{(i,j)\in E_{\max}} F_{i,j}, & \forall i \in I \\ \sum_{(k,i)\in E_{\max}} C_{k,\text{out}} F_{k,i} \leq C_{i,\text{in}} \left(z_i + \sum_{(k,i)\in E_{\max}} F_{k,i} \right), & \forall i \in I \\ y_{i,p} + y_{i,0} = 1, & \forall i \in I \\ \sum_{(i,j)\in C_{i,p}} F_{i,j} \leq K \cdot y_{i,p}, & \forall i \in I \\ F_{i,0} \leq K \cdot y_{i,0}, & \forall i \in I \\ z_i(F_{-i}) \geq 0, & \forall i \in I \\ F \geq 0, & \\ \text{Cost}_i(F_i, F_{-i}, E(y)) \leq \alpha \cdot \text{STC}_i, & \forall i \in I \end{cases} \tag{21}$$

Theorem 3. *For every feasible point (F, y) that satisfies (21), the corresponding pair $(F, E(y))$ is a feasible point for (15). Conversely, given any feasible point (F, E) that adheres to (15), the associated pair (F, y^E) is a feasible point for (21), where $y^E = (y^E_{i,p}, y^E_{i,0})_{i\in I} \in \{0,1\}^{2n}$ is defined as follows:*

$$y^E_{i,p} = \begin{cases} 1 & \textit{if the connections in } C_{i,p} \textit{ are part of the network } E, \\ 0 & \textit{otherwise,} \end{cases}$$

and

$$y^E_{i,0} = \begin{cases} 1 & \text{if the connection } (i,0) \text{ is the sole exit connection for } i, \\ 0 & \text{otherwise.} \end{cases}$$

In light of this, the implications are as follows:

1. *If* (F, E) *represents an optimal solution for* (15)*, then* (F, y^E) *stands as an optimal solution for* (21)*.*
2. *If* (F, y) *constitutes an optimal solution for* (21)*, then* $(F, E(y))$ *serves as an optimal solution for* (15)*.*

Proof. The proof follows a similar path as outlined in [15, Theorem 4.2]. Consider a feasible point (F, y) in the context of (21).

- Let's take an enterprise $i \in I_{\text{act}}$ and denote the unique class within $\mathcal{C}i$ for which $y_{i,p} = 1$ as $C_{i,p}$. By construction, we can ascertain that

  $$E(y)_{i,\text{act}} = C_{i,p} \quad \text{and} \quad \sum_{(i,j)\in E_{\max}\setminus C_{i,p}} F_{i,j} \leq K \cdot \sum_{C\in\mathcal{C}_i\setminus\{C_{i,p}\}} y_C = 0.$$

  Consequently, we derive that

  $$F_i\big|_{E(y)^c_{i,\text{act}}} = 0 \Longleftrightarrow F_{i,0} = 0, \quad \forall i \in I_{\text{act}}. \tag{22}$$

  Since this constraint holds for all active enterprises $i \in I_{\text{act}}$, we can reformulate the balance constraint within problem (21) as

  $$z(F_{-i}) + \sum_{(k,i)\in E(y)} F_{k,i} = \sum_{(i,j)\in E(y)_{i,\text{act}}} F_{i,j}, \quad \forall i \in I_{\text{act}}. \tag{23}$$

- Now, for an enterprise $i \in I_{\text{inact}}$, we can express the balance constraint in problem (21) as

  $$z_i(F_{-i}) + \sum_{(k,i)\in E(y)} F_{k,i} = F_{i,0}, \quad \forall i \in I_{\text{inact}}. \tag{24}$$

By combining (22), (23), and (24), we then infer that $(F, E(y))$ qualifies as a feasible point for problem (15).

Now, let (F, E) be a feasible point of problem (15). Let us define $y^E = (y^E_{i,p}, y^E_{i,0})_{i\in I} \in \{0,1\}^{2n}$ as in the statement of the theorem. Then, for every $i \in I$, $y_{i,p} + y_{i,0} = 1$.

- Now, let $i \in I_{\text{act}}$, we have that

$$\sum_{(i,j)\in C_{i,p}} F_{i,j} \leq \begin{cases} K = K \cdot y^E_{i,p} & \text{if the connections in } C_{i,p} \text{ are included in } E, \\ 0 = K \cdot y^E_{i,p} & \text{otherwise,} \end{cases}$$

For an enterprise $i \in I_{\text{act}}$, the fact that $E_{i,\text{act}} \subseteq E(y^E)$ lead us to the fact that

$$\text{Cost}_i(F_i, F_{-i}, E(y^E)) = \text{Cost}_i(F_i, F_{-i}, E),$$

and so, the constraint (12) is satisfied for any $i \in I_{\text{act}}$.

– For each enterprise $i \in I_{\text{inact}}$,

$$F_{i,0} \leq \begin{cases} K = K \cdot y^E_{i,0} & \text{if the connections in } C_{i,0} \text{ are included in } E, \\ 0 = K \cdot y^E_{i,0} & \text{otherwise,} \end{cases}$$

For an enterprise $i \in I_{\text{inact}}$, we have

$$\text{Cost}_i(F_i, F_{-i}, E(y^E)) = \text{Cost}_i(F_i, F_{-i}, E),$$

and so, the constraint (12) is satisfied for any $i \in I_{\text{inact}}$.

Consequently, we establish that (F, y^E) is a valid solution for (21), as all other constraints are automatically fulfilled when (F, E) proves to be feasible for (15).

The final two implications of the theorem are a direct result of the preceding developments.

3.3 Null Class as Exit Option

Perceiving that the network consistently possesses a viable configuration in the form of the standalone setup E_{st}, it's worth noting that the introduction of constraint (12) can potentially render problem (15) unfeasible.

The infeasibility of (15) implies that the authority faces a challenge in achieving a solution that abides by the constraint (12) across all enterprises. This prompts the need to consider the possibility of excluding certain enterprises from the network.

To address this, we introduce a boolean variable $y_{i,\text{null}} \in \{0, 1\}$ for each enterprise $i \in I$, defined as follows:

$$y_{i,\text{null}} = \begin{cases} 1 & \text{if } i \text{ violates the contract (12),} \\ 0 & \text{otherwise.} \end{cases}$$

We modify problem (21) with this extra variable, adding the following constraints:

1. For every enterprise i within the set I, we establish

$$y_{i,\text{null}} + y_{i,p} + y_{i,0} = 1, \tag{25}$$

which signifies that either a single arc class is operational, or the enterprise remains unlinked to the network.

2. For every enterprise i within the set I, we put

$$F_{i,0} \leq K \cdot (y_{i,0} + y_{i,\text{null}}), \tag{26}$$

$$\sum_{(i,j)\in E_{\max}, j\neq 0} F_{i,j} \leq K \cdot (1 - y_{i,\text{null}}), \tag{27}$$

for some constant $K > 0$ large enough. This ensures that if the enterprise violates contract (12), he will employ the discharge arc $(i, 0)$.
3. For every enterprise i within the set I, we put

$$\sum_{(k,i)\in E_{\max}} F_{k,i} \leq K \cdot (1 - y_{i,\text{null}}), \tag{28}$$

for some large enough constant $K > 0$. This constraint ensures that in the event of an enterprise contravening the contract stipulated in (12), no other entity is allowed to transmit any flux to it.
4. For each enterprise $i \in I$, we establish

$$\text{Cost}_i(F_i, F_{-i}, E(y)) \leq \alpha_i \cdot \text{STC}_i \cdot (1 - y_{i,\text{null}}) + \text{STC}_i \cdot y_{i,\text{null}}. \tag{29}$$

In this context, the constraint of individual rationality holds true exclusively when $y_{i,\text{null}} = 0$. Conversely, when the enterprise is not integrated into the network, its cost aligns with STC_i.

Denoting

$$\text{STC}_i(y_{i,\text{null}}) := \alpha_i \cdot \text{STC}_i \cdot (1 - y_{i,\text{null}}) + \text{STC}_i \cdot y_{i,\text{null}},$$

the new optimization problem becomes

$$\begin{aligned} &\min_{F,y} \quad Z(F) \\ &s.t. \begin{cases} z_i + \sum_{(k,i)\in E_{\max}} F_{k,i} = \sum_{(i,j)\in E_{\max}} F_{i,j}, & \forall i \in I \\ \sum_{(k,i)\in E_{\max}} C_{k,\text{out}} F_{k,i} \leq C_{i,\text{in}} \left(z_i + \sum_{(k,i)\in E_{\max}} F_{k,i} \right), & \forall i \in I \\ y_{i,\text{null}} + y_{i,p} + y_{i,0} = 1, & \forall i \in I \\ \sum_{(i,j)\in C_{i,p}} F_{i,j} \leq K \cdot y_{i,p}, & \forall i \in I \\ F_{i,0} \leq K \cdot (y_{i,0} + y_{i,\text{null}}), & \forall i \in I \\ \sum_{(i,j)\in E_{\max}, j\neq 0} F_{i,j} \leq K \cdot (1 - y_{i,\text{null}}), & \forall i \in I \\ \sum_{(k,i)\in E_{\max}} F_{k,i} \leq K \cdot (1 - y_{i,\text{null}}), & \forall i \in I \\ z_i(F_{-i}) \geq 0, & \forall i \in I \\ F \geq 0, & \\ \text{Cost}_i(F_i, F_{-i}, E(y)) \leq \alpha \cdot \text{STC}_i(y_{i,\text{null}}), & \forall i \in I \end{cases} \end{aligned} \tag{30}$$

Given that the optimization problem (30) can yield multiple solutions, we incorporate an additional term into the objective function to guide the selection towards the solution with a greater number of participating enterprises. We replace $Z(F)$ with the following expression:

$$Z(F) + \text{Coef} \cdot \sum_{i \in I} y_{i,\text{null}}, \tag{31}$$

where $\text{Coef} \geq 0$ serves as a coefficient to penalize optimal solutions that exclude a larger number of enterprises from the park.

In the upcoming sections, we will tackle the MILP problem (30) through the utilization of the programming language `Julia v1.0.5`, employing the `Cplex` solver.

4 Design of Experiments

As mentioned in Sect. 1, the objective of this study is to establish a cause-and-effect relationship between a number of input factors and the economic feasibility, as well as between a number of input parameters and the overall satisfaction of operating enterprises in EIPs. To identify these input factors, we use the design of experimental (DoE) techniques. Thus, we conduct a series of experimental runs with the change of the inputs and test the results in order to collect the outputs to evaluate the corresponding change. Each run of an experiment involves a combination of the levels of the investigated factors. The DoE inputs are studied based on the original parameter values given in Table 1. In fact, the parameters in Table 1 include part of the hypothetical literature example originally developed by Olesen and Polley [13]. The DoE consists of 7 levels regarding factors $\text{C}_{i,\text{in}}, \text{C}_{i,\text{out}}$, and M_i of enterprises 5, 6, 7, 8, 9, and 10, being values in Table 1 as the base level. Other levels correspond to 0.5, 0.8, 0.9, 1.1, 1.2, 1.5 times the base value for $\text{C}_{i,\text{in}}, \text{C}_{i,\text{out}}$, and M_i. Additionally, we define $\rho = \dfrac{c_f}{\delta}$ as another factor considered in the DoE evaluated in five levels, i.e., 0.2, 0.5, 2.1, 3.1, and 12.4. Then, for evaluation purposes, it is assumed that $c_f = 6.2000$ [\$/T] and $\beta = 34.875$ [\$/T]. Furthermore, assume that the EIP operates for one hour, i.e., $\text{A} = 1\,\text{h}$.

Since the DoE consists of 7 levels regarding factors $\text{C}_{i,\text{in}}, \text{C}_{i,\text{out}}, M_i$ and 5 levels regarding factor ρ, thus the DoE will require 1715 experimental runs. For each experimental run, the MILP problem (30) was solved in proper sequence. Then the economic satisfaction of each enterprise was checked for each combination of levels. For such a purpose, the level of satisfaction of each enterprise is defined as:

$$L_i = \begin{cases} 1 & \text{if } \text{Cost}_\text{i}{}^* \leq \text{Cost}_\text{i}{}^{\text{LOW}}, \\ \dfrac{\text{Cost}_\text{i}{}^* - \text{Cost}_\text{i}{}^{\text{UP}}}{\text{Cost}_\text{i}{}^{\text{LOW}} - \text{Cost}_\text{i}{}^{\text{UP}}} & \text{if } \text{Cost}_\text{i}{}^{\text{LOW}} < \text{Cost}_\text{i}{}^* < \text{Cost}_\text{i}{}^{\text{UP}}, \\ 0 & \text{if } \text{Cost}_\text{i}{}^* \geq \text{Cost}_\text{i}{}^{\text{UP}}, \end{cases} \tag{32}$$

Table 1. Parameters of the Network.

Enterprise i	$C_{i,\text{in}}$(ppm)	$C_{i,\text{out}}$(ppm)	M_i(g/h)
1	0	100	2000
2	50	80	2000
3	50	100	5000
4	80	800	30000
5	400	800	4000
6	10	100	2000
7	60	80	2000
8	80	400	5000
9	100	800	30000
10	400	1000	4000

where $\text{Cost}_i{}^*$ is the optimal operating cost of enterprise i when operating inside the EIP, $\text{Cost}_i{}^{\text{UP}}$ is the operating cost of enterprise i when operating stand-alone, while $\text{Cost}_i{}^{\text{LOW}}$ is the desired cost of enterprise i. Note that, for each combination of levels of the factors, the MILP problem (30) was solved to obtain a generalized Nash equilibrium point (F_i, F_{-i}, y) then the value of $\text{Cost}_i{}^*$ is defined by $\text{Cost}_i(F_i, F_{-i}, E(y))$, on the other hand the value of $\text{Cost}_i{}^{\text{LOW}}$ is determined by 80% of the operating stand-alone cost $\text{Cost}_i{}^{\text{UP}}$.

Now, the overall satisfaction of operating enterprises in EIPs is measured by

$$L_{tot} = \sum_{i \in I} L_i. \tag{33}$$

For the EIP economic feasibility study, all 1715 simulations were included on the DoE built with the 4 variable mentioned above ($C_{i,\text{in}}$, $C_{i,\text{out}}$, M_i, ρ), as well as their interactions and the squared terms. When enterprises operate in an EIP, there are two possibilities either the EIP is economically feasible or the EIP is not economically feasible. So, we introduce a categorical variable Y defined as

$$Y = \begin{cases} 1 & \text{if } L_{tot} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{34}$$

where $Y = 1$ means that the EIP is economically feasible while $Y = 0$ means that the EIP is not economically feasible.

4.1 Binary Logistic Regression Model

In this part, we present a logistic regression model to evaluate whether the predictors (i.e., $C_{i,\text{in}}$, $C_{i,\text{out}}$, M_i, and ρ), their interactions, and squared terms affect the economic feasibility of the operating enterprises in EIPs or not. The model built was of the following form:

$$\ln\left[\frac{P(Y=1)}{P(Y=0)}\right] = \beta_0 + \beta_1 \mathrm{C}_{i,\mathrm{in}} + \beta_2 \mathrm{C}_{i,\mathrm{out}} + \beta_3 M_i + \beta_4 \rho + \beta_5 \mathrm{C}_{i,\mathrm{in}}^2 + \beta_6 \mathrm{C}_{i,\mathrm{out}}^2 + \beta_7 M_i^2 + \beta_8 \rho^2 + \beta_9 \mathrm{C}_{i,\mathrm{in}} \times \mathrm{C}_{i,\mathrm{out}} + \beta_{10} \mathrm{C}_{i,\mathrm{in}} \times M_i + \beta_{11} \mathrm{C}_{i,\mathrm{in}} \times \rho + \beta_{12} \mathrm{C}_{i,\mathrm{out}} \times M_i + \beta_{13} \mathrm{C}_{i,\mathrm{out}} \times \rho + \beta_{14} M_i \times \rho, \quad (35)$$

where P is the probability, Y is the categorical variable defined as in equation (34), $\mathrm{C}_{i,\mathrm{in}}$, $\mathrm{C}_{i,\mathrm{out}}$, M_i and ρ are the predictors, β_0 is the Y intercept, and $\beta_1, \beta_2, \ldots, \beta_{14}$ are the regression coefficients of the independent variables to the dependent variable Y. In literature, the coefficients $\beta_0, \beta_1, \ldots, \beta_{14}$ are typically estimated by maximum likelihood method. For the survey of binary logistic regression, we prefer the reader to [7].

For each combination of levels of factors, the MILP problem (30) is solved and the categorical variable Y is determined. Thus, we have a set of 1715 response variables Y_i corresponding to 1715 different sets of 4 input factors ($\mathrm{C}_{i,\mathrm{in}}$, $\mathrm{C}_{i,\mathrm{out}}$, M_i and ρ). Based on this collected data set, the logistic regression analysis was carried out by the Logistic procedure in SPSS software version 22 [5]. The result showed that

$$\ln\left[\frac{P(Y=1)}{P(Y=0)}\right] = 18.447 + 3.502\mathrm{C}_{i,\mathrm{in}} - 50.245\mathrm{C}_{i,\mathrm{out}} - 10.949 M_i + 3.053\rho - 2.985\mathrm{C}_{i,\mathrm{in}}^2 + 14.684\mathrm{C}_{i,\mathrm{out}}^2 - 4.097 M_i^2 - 0.210\rho^2 + 4.794\mathrm{C}_{i,\mathrm{in}} \times \mathrm{C}_{i,\mathrm{out}} - 0.826\mathrm{C}_{i,\mathrm{in}} \times M_i - 0.104\mathrm{C}_{i,\mathrm{in}} \times \rho + 14.207\mathrm{C}_{i,\mathrm{out}} \times M_i + 0.120\mathrm{C}_{i,\mathrm{out}} \times \rho + 2.057 M_i \times \rho. \quad (36)$$

The model (36) is not selected because the terms of $\mathrm{C}_{i,\mathrm{in}}$, $\mathrm{C}_{i,\mathrm{in}} \times M_i$, $\mathrm{C}_{i,\mathrm{in}} \times \rho$, $\mathrm{C}_{i,\mathrm{out}} \times \rho$ are not statistically significant because there are large sig numbers, respectively 0.311, 0.573, 0.872, and 0.870. Now, we remove the terms of $\mathrm{C}_{i,\mathrm{in}}$, $\mathrm{C}_{i,\mathrm{in}} \times M_i$, $\mathrm{C}_{i,\mathrm{in}} \times \rho$, $\mathrm{C}_{i,\mathrm{out}} \times \rho$ from the model (35), then continue to run SPSS with the remaining terms, we get the following result

$$\ln\left[\frac{P(Y=1)}{P(Y=0)}\right] = 20.873 - 50.475\mathrm{C}_{i,\mathrm{out}} - 11.947 M_i + 3.011\rho - 2.104\mathrm{C}_{i,\mathrm{in}}^2 + 14.695\mathrm{C}_{i,\mathrm{out}}^2 - 4.054 M_i^2 - 0.205\rho^2 + 5.332\mathrm{C}_{i,\mathrm{in}} \times \mathrm{C}_{i,\mathrm{out}} + 14.193\mathrm{C}_{i,\mathrm{out}} \times M_i + 2.081 M_i \times \rho \quad (37)$$

The model (37) was chosen because all terms are statistically significant at the standard error of regression of 5%. Thus, the model (37) is the optimal logistic regression model in this study.

After estimating the coefficients of the binary linear regression model as shown in model (37), there are several steps involved in assessing the appropriateness, adequacy and usefulness of the model (37). Therefore, to evaluate the logistic regression model, one must attend to (a) overall model evaluation,

(b) goodness-of-fit statistics, (c) validations of predicted probabilities, and (d) statistical tests of individual predictors.

Overall model evaluation. A logistic regression model is considered to exhibit an enhanced fit to the data when it shows advancement over the intercept-only model, which lacks our explanatory variables and is commonly referred to as the null model.

Table 2. Omnibus Tests of Model Coefficients.

		Chi-square	df	Sig.
Step 1	Step	1743.079	10	0.000
	Block	1743.079	10	0.000
	Model	1743.079	10	0.000

The Omnibus Tests of Model Coefficients as shown in Table 2 will be used to check whether model (37) (with explanatory variable included) is an improvement over the intercept-only model. Here the chi-square is highly significant (chi-square = 1743.079, df=10, $p < 0.000$) so the logistic regression model (37) is significantly better the intercept-only model.

Goodness-of-fit statistics. The prevalent evaluation of the comprehensive model fit in logistic regression involves the likelihood ratio test, essentially indicating the chi-square disparity between the model with just the intercept and the one encompassing the predictors. In the context of the Model Summary depicted in Table 3, the -2 Log Likelihood statistic registers a value of 485.903. This metric gauges the model's efficacy in forecasting decisions, signifying that a lower value signifies a more proficient model prediction.

Table 3. Model Summary.

Step	−2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	485.903	0.638	0.877

Most statistical software packages offer supplementary statistics akin to the coefficient of determination (R^2) in linear regression, though not an exact analogy [3]. Among these, the Cox & Snell R^2 and the Nagelkerke R^2 stand out. The Cox and Snell R^2 outcome, indicating that 63.8% of the variance in the dependent variable is accounted for by the predictor variable, is considered satisfactory. It's worth noting that a limitation of the Cox-Snell R^2 is its maximum value being less than 1.

In contrast, the Nagelkerke R^2 represents a modified rendition of the Cox & Snell R^2 and encompasses a full range from 0 to 1, which makes it generally more preferred. The R^2 statistic doesn't measure the model's goodness of fit, rather it assesses the utility of explanatory variables in predicting the response variable, serving as a measure of effect size. With a value of 0.877, it indicates that the model effectively predicts the economic viability of operating enterprises within EIPs.

Table 4. Goodness-of-Fit Statistics

Step	Chi-square	df	Sig
1	3.388	8	0.908

In Table 4, the Hosmer-Lemeshow test, an inferential measure of goodness-of-fit, is also provided. The Hosmer-Lemeshow test statistic of 3.388 yielded insignificance ($p = 0.908 > 0.05$), implying that the logistic regression model (37) exhibited a satisfactory fit to the dataset.

Table 5. Classification Table.

			Predicted		
			Economic feasibility		Percentage Correct
	Observed		0.0	1.0	
Step 1	Economic feasibility	0.0	1029	79	92.9
		1.0	57	550	90.6
	Overall Percentage				92.1

Validations of predicted probabilities. In Table 5, the classification table showcases the alignment between anticipated probabilities and real outcomes. The overall accuracy of predictions, at 92.1%, signifies an enhancement beyond the 50% baseline probability level. This classification table facilitates the assessment of sensitivity, specificity, false positive, and false negative rates. Sensitivity gauges the fraction of accurately identified events, while specificity quantifies the ratio of accurately identified nonevents. False positive denotes the proportion of observations wrongly classified as events among those classified as such. On the other hand, false negative quantifies the portion of observations inaccurately labeled as nonevents within the nonevent cate

Statistical tests of individual predictors. However the most important of all output is the Variables in the Equation as shown in Table 6. Within the confines of Table 6, crucial information is presented, encompassing regression coefficients (denoted as βs), the Wald statistic for assessing statistical significance, and the pivotal Odds Ratio (Exp (β)) pertaining to each of the input variables.

According to Table 6, the factors ($C_{i,out}, M_i, \rho$), their interactions ($C_{i,in} \times C_{i,out}, C_{i,out} \times M_i, M_i \times \rho$) and squared terms ($C^2_{i,in}, C^2_{i,out}, M_i^2, \rho^2$) were significant and they influence the economic feasibility of operating enterprises in EIPs.

Table 6. Statistical Tests of Individual Predictors

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I for EXP(B)	
								Lower	Upper
Step 1	C_{out}	−50.475	4.757	112.561	1	0.000	0.000	0.000	0.000
	M	−11.947	3.178	14.131	1	0.000	0.000	0.000	0.003
	ρ	3.011	0.769	15.352	1	0.000	20.315	4.504	91.624
	C^2_{in}	−2.104	0.613	11.767	1	0.001	0.122	0.037	0.406
	C^2_{out}	14.695	1.599	84.494	1	0.000	2408641.387	104960.835	55273506.03
	M^2	−4.054	1.156	12.297	1	0.000	0.017	0.002	0.167
	ρ^2	−0.205	0.048	18.520	1	0.000	0.815	0.742	0.894
	$C_{in} \times C_{out}$	5.332	1.234	18.683	1	0.000	206.886	18.436	2321.600
	$C_{out} \times M$	14.193	1.748	65.961	1	0.000	1458833.066	4745.852	44826871.12
	$M \times \rho$	2.081	0.693	9.011	1	0.003	8.015	2.059	31.193
	Constant	20.873	3.041	47.123	1	0.000	1161411149		

Moreover, from Table 6, a cause-and-effect relationship between a number of input factors and the economic feasibility of operating enterprises in EIPs is given by

$$\ln\left[\frac{P(Y=1)}{P(Y=0)}\right] = 20.873 - 50.475C_{i,out} - 11.947M_i + 3.011\rho - 2.104C^2_{i,in} + 14.695C^2_{i,out} - 4.054M_i^2 - 0.205\rho^2 + 5.332C_{i,in} \times C_{i,out} + 14.193C_{i,out} \times M_i + 2.081M_i \times \rho \quad (38)$$

As observed from (38), the effect of each input factor ($C_{i,in}$, $C_{i,out}$, M_i and ρ respectively) on the economic feasibility of operating enterprises in EIPs depends on the value(s) of one or more other input factors.

A very powerful application of binary logistic regression is predictability. From the regression equation (38), we have the probability function to evaluate the economic feasibility of operating enterprises in EIPs as follows:

$$P(Y=1) = \frac{e^{20.873-50.475C_{i,out}-11.947M_i+3.011\rho-2.104C^2_{i,in}+14.695C^2_{i,out}-4.054M_i^2-0.205\rho^2+5.332C_{i,in}\times C_{i,out}+14.193C_{i,out}\times M_i+2.081M_i\times\rho}}{1+e^{20.873-50.475C_{i,out}-11.947M_i+3.011\rho-2.104C^2_{i,in}+14.695C^2_{i,out}-4.054M_i^2-0.205\rho^2+5.332C_{i,in}\times C_{i,out}+14.193C_{i,out}\times M_i+2.081M_i\times\rho}}.$$

To demonstrate how the economic feasibility of enterprises in EIPs is sensitive when changing the levels of input parameters ($C_{i,in}, C_{i,out}, M_i, \rho$), we consider the following two scenarios:

– First, the levels of $\mathrm{C}_{i,\mathrm{in}}, \mathrm{C}_{i,\mathrm{out}}, M_i, \rho$ are 0.5, 0.5, 0.5, 12.4, respectively. Then the probability of the economic feasibility of enterprises in EIPs for this scenario is given by

$$P(Y = 1) = 0.99$$

– Second, the levels of $\mathrm{C}_{i,\mathrm{in}}, \mathrm{C}_{i,\mathrm{out}}, M_i, \rho$ are 0.5, 0.5, 0.5, 0.2, respectively. Then the probability of the economic feasibility of enterprises in EIPs for this scenario is given by

$$P(Y = 1) = 0.07$$

As we can observe, the economic feasibility of enterprises in EIPs is very sensitive when changing the levels of input parameters. More precisely, when the levels of $\mathrm{C}_{i,\mathrm{in}}, \mathrm{C}_{i,\mathrm{out}}, M_i, \rho$ are 0.5, 0.5, 0.5, 12.4 then the probability of the economic feasibility of enterprises in EIPs is 0.99. However, when the levels of $\mathrm{C}_{i,\mathrm{in}}, \mathrm{C}_{i,\mathrm{out}}, M_i, \rho$ are 0.5, 0.5, 0.5, 0.2 then the probability of the economic feasibility of enterprises in EIPs is 0.07.

4.2 Multiple Linear Regression Model

In this part, we present a multiple linear regression model to predict the value of overall satisfaction L_{tot} based on the variables $\mathrm{C}_{i,\mathrm{in}}, \mathrm{C}_{i,\mathrm{out}}, M_i$, and ρ. The data consists of only 607 simulations (economically feasible configurations) built with the same 4 variables, their interactions, and the square terms.

To ensure the credibility and validity of our analysis when performing a multiple linear regression, it is imperative to assess our data against various assumptions. These prerequisites encompass:

1. The dependent variable must be measured on a continuous scale.
2. There should be a presence of two or more independent variables, which could either be continuous or categorical in nature.
3. The residuals' values need to exhibit independence.
4. The connection between the dependent and independent variables ought to be linear.
5. The residuals' variance should remain constant.
6. The data should be devoid of multicollinearity issues.
7. The model should not be unduly influenced by outlier cases.
8. The residuals, or errors, should approximate a normal distribution.

Considering none of the eight previously mentioned assumptions were violated, multiple regression outputs for the given data were generated through the following steps performed in SPSS statistical software.

Determining how well the model fits. The first table of interest is the *model summary* (Table 7). Within this table, essential metrics such as R, R^2, and adjusted R^2 are presented, aiding in the assessment of the regression model's goodness of fit to the data.

Table 7. Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin Watson
1	0.905	0.819	0.817	0.5582011102	1.518

In the "R" column, you will find the R value, denoting the *multiple correlation coefficient* between the actual and projected values of the dependent variable. R can be considered one gauge of how accurately the dependent variable, in this instance, *overall satisfaction*, is predicted. A value of $R = 0.905$ suggests a commendable level of prediction.

The "R Square" column represents the value of R^2, the *coefficient of determination*. R Square is the proportion of variance in the dependent variable (overall satisfaction) that can be explained by the independent variables $(\mathrm{C}_{i,\mathrm{in}}, \mathrm{C}_{i,\mathrm{out}}, M_i, \rho)$. This value indicates that 81.9% of the variance in overall satisfaction can be predicted from the variables $\mathrm{C}_{i,\mathrm{in}}, \mathrm{C}_{i,\mathrm{out}}, M_i$, and ρ. It also means that 18.1% of the variation is caused by factors other than the predictors included in this model.

A greater R^2 value indicates a more optimal fit. Nonetheless, a higher R^2 doesn't necessarily signify a robust fit or a superior regression model, as including additional variables consistently elevates the R^2 value, leading to less effective predictions. To address this, the adjusted R^2 is introduced, which isn't swayed by mere variable additions and provides a more reliable gauge.

Furthermore, a notable disparity between the R-squared and Adjusted R Square values implies a subpar model fit. In the current instance, the value stands at 0.817, not significantly distant from 0.819, signifying a reasonably good fit.

Statistical significance of the model. The F-ratio within ANOVA evaluates the suitability of the entire regression model concerning the dataset. As illustrated in Table 8, the independent variables exhibit a statistically significant predictive influence on the dependent variable, with $F(4, 602) = 679.108$, $p < 0.0005$ (indicating a strong alignment between the regression model and the data).

Table 8. ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	846.408	4	211.602	679.108	0.000
	Residual	187.576	602	0.312		
	Total	1033.985	606			

Estimated model coefficients. The "B" column displays the coefficients corresponding to the regression equation, used to predict the dependent variable based on the independent variables (Table 9).

Table 9. Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% C.I for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	Constant	1.433	0.134		10.708	0.000	1.170	1.695		
	C_{in}	0.524	0.080	0.115	6.593	0.000	0.368	0.681	0.995	1.005
	C_{out}	−0.449	0.071	−0.112	−6.356	0.000	−0.588	−0.310	0.972	1.028
	M	−0.659	0.081	−0.143	−8.155	0.000	−0.818	−0.501	0.980	1.020
	ρ	0.246	0.005	0.892	51.142	0.000	0.237	0.256	0.992	1.008

The "Sig." column contains the p-values for each of the independent variables ($\mathrm{C}_{i,\text{in}}$, $\mathrm{C}_{i,\text{out}}, M_i, \rho$). A p-value < 0.05, provides evidence that the coefficient is different to 0. Thus $\mathrm{C}_{i,\text{in}}$, $\mathrm{C}_{i,\text{out}}, M_i$, and ρ are all significant predictors of *overall satisfaction*.

Therefore, the general form of the equation to predict *overall satisfaction* from $\mathrm{C}_{i,\text{in}}, \mathrm{C}_{i,\text{out}}, M_i, \rho$ is:

$$L_{tot} = 1.433 + 0.524\mathrm{C}_{i,\text{in}} - 0.449\mathrm{C}_{i,\text{out}} - 0.659M_i + 0.246\rho \qquad (39)$$

These regression coefficients tell us about the relationship between the independent variables ($\mathrm{C}_{i,\text{in}}$, $\mathrm{C}_{i,\text{out}}, M_i, \rho$) and the dependent variable (overall satisfaction L_{tot}). These regression coefficients provide the expected change in the dependent variable for a one-unit increase in the independent variable. The coefficient for $\mathrm{C}_{i,\text{in}}$ is 0.524. So, for every unit increase in $\mathrm{C}_{i,\text{in}}$, there is 0.524 unit increase in overall satisfaction, holding all other variables constant. But each unit increase in $\mathrm{C}_{i,\text{out}}$ (resp. M_i) causes reduction (the negative sign of the coefficient) in overall satisfaction by 0.449 (resp. 0.659) unit. the coefficient for ρ is 0.246. Hence, for every unit increase in ρ we expect a 0.246 unit increase in overall satisfaction, holding all other variables constant. Thus, we can use equation (39) to find the estimated *overall satisfaction*, based on the level of $\mathrm{C}_{i,\text{in}}$, $\mathrm{C}_{i,\text{out}}, M_i$, and ρ.

5 Conclusions and Future Work

This study evaluated the influence of different input parameters on the economic feasibility and overall satisfaction of operating enterprises in EIPs. The proposed method integrated the design of experiments (DoE) method and the SLMF game model to identify and quantify the impact of different parameters on the economic feasibility and overall satisfaction of operating enterprises in EIPs. As the results show the most critical parameters in a potential EIP environment are those related to process constraints and those related to the inherent production of each enterprise. Moreover, the economic feasibility of enterprises operating in eco-industrial parks is quite sensitive to changing input parameters, which can produce a variety of scenarios that can be a potential reason to reject EIP cooperation between different enterprises. However, by combining the DoE method and the SLMF model, we can adjust the input parameters so that enterprises operating in the eco-industrial park become economically feasible. Thus, it's an advantage in the design of EIPs.

For future research, the results of this study can be extended to the assessment of the influence of input parameters on the economic feasibility and overall satisfaction of other resource exchanges in the eco-industrial park, such as the exchange of steam, electricity, energy, ...

Acknowledgments. This research is funded by Nguyen Tat Thanh University, Ho Chi Minh city, Vietnam.

References

1. Aussel, D., Van Cao, K., Salas, D.: Optimal design of exchange water networks with control inputs in eco-industrial parks. Energy Econ. **120**, 106480 (2023). https://doi.org/10.1016/j.eneco.2022.106480, https://www.sciencedirect.com/science/article/pii/S0140988322006090
2. Baumrucker, B.T., Renfro, J., Biegler, L.: MPEC problem formulations and solution strategies with chemical engineering applications. Comput. Chem. Eng. **32**(12), 2903–2913 (2008)
3. Bewick, V., Cheek, L., Ball, J.: Statistics review 7: Correlation and regression. Critical Care (London, England) **7**, 451–459 (2004). https://doi.org/10.1186/cc2401
4. Boix, M., Montastruc, L., Azzaro-Pantel, C., Domenech, S.: Optimization methods applied to the design of eco-industrial parks: a literature review. J. Clean. Prod. **87**, 303–317 (2015)
5. Field, A.: Discovering Statistics Using IBM SPSS Statistics. SAGE Publications (2018). https://books.google.com.vn/books?id=JIrutAEACAAJ
6. Fisher, R.: The Design of Experiments, 9th edn. MacMillan, New York (1971)
7. Hosmer, D., Lemeshow, S., Sturdivant, R.: Applied Logistic Regression. Wiley Series in Probability and Statistics, Wiley (2013). https://books.google.com.vn/books?id=64JYAwAAQBAJ
8. Jankovic, A., Chaudhary, G., Goia, F.: Designing the design of experiments (doe) - an investigation on the influence of different factorial designs on the characterization of complex systems. Energy Build. **250**, 111298 (2021). https://doi.org/10.1016/j.enbuild.2021.111298, https://www.sciencedirect.com/science/article/pii/S037877882100582X
9. Laoun, B., Kasat, H.A., Ahmad, R., Kannan, A.M.: Gas diffusion layer development using design of experiments for the optimization of a proton exchange membrane fuel cell performance. Energy **151**, 689–695 (2018). https://doi.org/10.1016/j.energy.2018.03.096, https://www.sciencedirect.com/science/article/pii/S036054421830505X
10. Lowe, E.A.: Creating by-product resource exchanges: Strategies for eco-industrial parks. J. Clean. Prod. **5**(1), 57–65 (1997)
11. Luo, Z.Q., Pang, J.S., Ralph, D.: Mathematical Programs with Equilibrium Constraints. Cambridge University Press, Cambridge (1996)
12. Montgomery, D.: Design and Analysis of Experiments. Design and Analysis of Experiments. John Wiley & Sons, Hoboken (2008). https://books.google.de/books?id=kMMJAm5bD34C
13. Olesen, S., Polley, G.: Dealing with plant geography and piping constraints in water network design. Process Saf. Environ. Prot. **74**(4), 273–276 (1996). https://doi.org/10.1205/095758296528626

14. Ramos, M., Boix, M., Aussel, D., Montastruc, L., Domenech, S.: Water integration in eco-industrial parks using a multi-leader-follower approach. Comput. Chem. Eng. **87**, 190–207 (2016)
15. Salas, D., Van Cao, K., Aussel, D., Montastruc, L.: Optimal design of exchange networks with blind inputs and its application to eco-industrial parks. Comput. Chem. Eng. **143**, 107053 (2020)
16. Sieira, P., Mendes, P., Castro, A., Pradelle, F.: Impact of spinning conditions on the diameter and tensile properties of mesophase petroleum pitch carbon fibers using design of experiments. Mater. Lett. **285**, 129110 (2021). https://doi.org/10.1016/j.matlet.2020.129110, https://www.sciencedirect.com/science/article/pii/S0167577X20318176
17. Suchana, S., Passeport, E.: Optimization of a solid-phase microextraction technique for chloro- and nitro- substituted aromatic compounds using design of experiments. J. Chromatogr. A **1621**, 461083 (2020). https://doi.org/10.1016/j.chroma.2020.461083, https://www.sciencedirect.com/science/article/pii/S0021967320303046
18. Tseveendorj, I.: Mathematical programs with equilibrium constraints: a brief survey of methods and optimality conditions. In: Chinchuluun, A., Pardalos, P., Enkhbat, R., Pistikopoulos, E. (eds.) Optimization, Simulation, and Control. Springer Optimization and its Applications, vol. 76, pp. 49–61. Springer, New York (2013). https://doi.org/10.1007/978-1-4614-5131-0_4
19. Zaibi, M., Cherif, H., Champenois, G., Sareni, B., Roboam, X., Belhadj, J.: Sizing methodology based on design of experiments for freshwater and electricity production from multi-source renewable energy systems. Desalination **446**, 94–103 (2018). https://doi.org/10.1016/j.desal.2018.08.008, https://www.sciencedirect.com/science/article/pii/S0011916418305812