



**HAL**  
open science

## Robustness in optimal design of Eco-Industrial Parks under the lens of two-stage stochastic optimization

Gianfranco Liberona, Alessandro Di Pretoro, Stéphane Negny, Ludovic  
Montastruc, David Salas

► **To cite this version:**

Gianfranco Liberona, Alessandro Di Pretoro, Stéphane Negny, Ludovic Montastruc, David Salas. Robustness in optimal design of Eco-Industrial Parks under the lens of two-stage stochastic optimization. *Computers & Chemical Engineering*, 2023, 179, pp.108399. 10.1016/j.compchemeng.2023.108399 . hal-04312536

**HAL Id: hal-04312536**

**<https://hal.science/hal-04312536v1>**

Submitted on 12 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Robustness in optimal design of Eco-Industrial Parks under the lens of two-stage stochastic optimization

Gianfranco Liberona\*

Universidad de O'Higgins, Rancagua, Chile.

Alessandro Di Pretoro†

Laboratoire de Génie Chimique, Université de Toulouse, CNRS, INPT, UPS, Toulouse, France.

Stéphane Negny‡

Laboratoire de Génie Chimique, Université de Toulouse, CNRS, INPT, UPS, Toulouse, France.

Ludovic Montastruc§

Laboratoire de Génie Chimique, Université de Toulouse, CNRS, INPT, UPS, Toulouse, France.

David Salas¶

Instituto de Ciencias de la Ingeniería, Universidad de O'Higgins, Rancagua, Chile.

## Abstract

An eco-industrial park (EIP) is a community of businesses located together on a common property, that seek to reduce environmental and economical impact of their operation by collaborating and sharing materials and wastes. In practice, operations within an EIP have daily variations, and therefore are constantly facing uncertainty. In this work, a methodology to design efficient EIPs that are also *robust* to daily (uncertain) variations of the nominal operation of the enterprises is proposed. The attention is mainly focused in water exchange networks. Probability functions are used to measure robustness and propose a Sample Average Approximation technique to solve the associated optimization problem. Simulations based on literature examples are performed to illustrate the approach.

**Keywords:** Eco-Industrial park; Optimal Design; Robustness; Probability functions; Sample Average Approximation.

## Highlights:

- A robustness indicator is introduced to design eco-industrial parks.

---

\*e-mail: gianfranco.liberona@uoh.cl (corresponding author)

†e-mail: alessandro.dipretoro@ensiacet.fr

‡e-mail: stephane.negny@ensiacet.fr

§e-mail: ludovic.montastruc@ensiacet.fr

¶e-mail: david.salas@uoh.cl

- The EIP optimization aims at maximizing the amount of *good-enough* daily economical operation scenarios.
- Two-stage optimization addresses the EIP water consumption robust design problem.
- Sample Average Approximation technique is used to tackle daily operation uncertainty.
- Robustness, as chance operator, is dealt both in objective function and as constraint.

<b>Nomenclature</b>	
<b>Latin symbols:</b>	
$n$	number of participating agents
$N$	sample size for the SAA method
$I$	index set of participating agents
$0$	sink node
$V$	assembly of index set $I$ and sink node $0$
$\vec{E}$	network topology
$L$	vector of capacities for connections in $\vec{E}$
$L_{ij}$	capacity for the connection $(i, j) \in \vec{E}$
$z$	vector of total fresh water consumption of the network
$z_i$	fresh water consumption of agent $i$ [ton/h]
$\xi$	random vector of contaminant load production of the network
$M_i$	contaminant load of agent $i$ [kg/h]
$C_{i,in}$	maximum contaminant concentration allowed in inlet of processes [mg/l]
$C_{i,out}$	maximum contaminant concentration allowed in outlet of processes [mg/l]
$F$	complete vector of fluxes through the network
$F_{ij}$	flux through the connection $(i, j) \in E$ [ton/h]
$O$	vector of residual water sent to the sink node from all the agents
$O_i$	residual water sent to the sink node from agent $i$ [ton/h]
$SA_i$	stand-alone cost of agent $i$
$Cost_i$	operating cost of agent $i$
$Q(L, \xi)$	optimal daily operation, given $L$ and $\xi$
$X(L, \xi)$	feasible set of solution for the optimal daily operation problem
$G_\alpha$	$(1 - \alpha)$ level of <i>goodness</i> for an EIP
$Rob_\alpha(\cdot)$	EIP <i>robustness</i> , for a $(1 - \alpha)$ level of <i>goodness</i>
$\mathbb{E}$	expected value
$\mathbb{P}$	probability measure
$\mathbb{1}$	indicator function
$y$	boolean variable
$B$	maximum investment budget, when considered as a constraint
$c$	marginal cost for consuming fresh water
$d$	marginal cost for discharging water to the sink node
$e$	marginal cost for using the connections of the EIP
$w$	weight vector
$c_1, c_2$	budget and Robustness constraints
	boolean weights
<b>Acronyms:</b>	
EIP	Eco-Industrial Park
SA	total net cost if every agent operates on stand-alone conditions
SAA	Sample Average Approximation
KKT	Karush-Kuhn-Tucker
MILP	Mixed-integer linear programming
<b>Greek Symbols:</b>	
$\alpha$	cost reduction compared to the <i>stand-alone</i> global operation of the EIP (%)
$\beta$	minimal asked <i>Robustness</i> , when considered as a constraint (%)

## 1 Introduction

An Eco-Industrial Park (EIP), as defined in [17], is a community of manufacturing and service businesses located together on a common property. Member businesses seek enhanced environmental, economic, and social performance through collaboration in managing environmental and

resource issues.

A canonical example of EIP corresponds to the modeling of water exchange networks (see, e.g., [6, 29, 30] and the references therein). In this model, each participant of the EIP needs to consume fresh water for its industrial processes, and to send away partially contaminated water. In parallel, there is a central authority of the EIP, which is in charge of design the park and operate it afterwards, following some criteria that reflects environmental, economic and/or social benefits. The EIP design problem has already been studied under the setting of deterministic optimization, as reviewed in [5]. In this section, the main goal is to describe the model, in order to approach the design problem considering uncertainty. Most part of the water exchange model presented here is well-known in the literature. This work is mainly focused on the ones presented in [29, 30].

In fact, during the last decades, uncertainty has gained major attention in the vast majority of the engineering and scientific research fields [1, 9] due to the need to deal with both the market volatility and the variation related to the operating conditions. According to the works of Oberkampf et al. in [25], uncertainty can be classified into three main categories, namely aleatory uncertainty or variability, epistemic or reducible uncertainty and error. The concept aleatory refers to inherent variation of the physical environment under consideration while reducible is related to any lack of knowledge or information. Finally, error is defined as a recognizable deficiency during the modelling or simulation phases that is not due to the lack of knowledge. According to this definition, the fluctuation of input data or operating conditions, as in this research work, can be directly related to the so called aleatory or stochastic uncertainty. The approach to integrate uncertainty in modeling and simulation has been discussed in detail in [24] and several applications with different methodologies are available in literature. One of the possible applications in the process industry is design for flexibility that exploits both deterministic [33] and stochastic [28] flexibility indexes for various applications such as process units design [8] and optimal scheduling [7].

The problems of flexible and resilient EIP were already discussed in previous studies, carried out by Montastruc et al [22] and Valenzuela et al [35] respectively. In the latest, the definition of an index for the study of EIP performances and properties was developed: an indicator combining the number of connections among participants (Network Connectivity Index - NCI), and the capacity of the participants to compensate the flow demand when one participant interrupts its activity (Flows adaptability index -  $\phi$ ) was proposed. The concept of resilience, in general, accounts for daily operational fluctuations as well as system performance during significant disruptions, as can be thoroughly reviewed in [12, 26, 27].

The main goal of this work is to treat the problem of robust design of Eco-Industrial Parks considering no such indicator, but using directly the performance of the park under perturbations. That is, to design a park that can absorb fluctuations and keep a *good enough* performance in some sense. If directly estimate the impact of the random variations in the performance of the park were possible, one could approach the problem by studying the marginal variations of the performance function, as in [8]. However, this is not possible in the tackled problem, since the performance function is in fact the value function of another optimization problem (the optimal operation of a given EIP).

To deal with this obstruction, the robust design problem of the EIP is modeled by means of two-stage optimization (see, e.g., [4, 32]): in a first stage, the connections within the park are decided, and, in a second stage, the daily operation is treated, after the revealing of the uncertain

operation. The obtained problem is solved using the Sample Average Approximation (SAA) method, which allows us to tackle the problem as a single (large-scale) optimization problem, without the necessity of solving the second stage problem analytically. For the implementations of the SAA method and the assessment of the quality of the obtained solution, the main directions over viewed in [11] are applied.

The article is organized as follows: In Section 2 the EIP model about to be used is presented, as well as the general two-stage formulation. In Section 3, the notion of Robustness for an EIP is introduced, based on the level of *goodness* of an optimal operation. Section 4 is devoted to summarize the SAA method, and how it applies to the EIP model.

## 1.1 Our Contributions

A novel way of modeling the Eco-Industrial park design and daily operation problem is developed in this work, using two-stage optimization tools. This allows to pose the problem of optimal design without using any pre-fixed cost function nor any global indicator. Instead, the performance of the park in the second stage is directly considered. This model allows us to obtain robust designs for EIP on the long run.

The model is implemented numerically, following the Sample Average Approximation method, reducing it to a single (but larger) optimization problem. The methodology is tested numerically, showing that it is tractable for real-size EIPs, which validates the general approach of using the two-stage models.

The interpretation of the concept of *robustness* as the probability of maintaining certain level of production when facing variations, gives us the possibility of considering an alternate way of designing EIPs, which, as will be later shown, can provide “good-enough” results with important decreases on the investment cost for design. As this robustness operator is computed as a probability, *chance constraints* have to be dealt with, when solving the problem, aiming to a *desired* minimal robustness; and probability operators when trying to *maximize* the global robustness of the EIP, dealing with it at the objective function level. A reformulation of the main problem is developed, that allows us to use MIP tools, in order to solve an equivalent problem in both cases.

## 2 EIP Model and Uncertainty

### 2.1 Network model

A water exchange network can be modeled as a simple directed graph  $G = (V, \vec{E})$  with some specific conditions. First, the set  $V = \{0, 1, \dots, n\}$  represents the  $n$  agents/enterprises participating of the EIP, and 0 is an extra *sink node*.  $I = \{1, \dots, n\}$  is denoted to describe the set of agents of the park, and so  $V = I \cup \{0\}$ . The sink node 0 receives all the residual water exiting the park as final waste. Every agent is connected to the sink node, that is,  $(i, 0) \in \vec{E}$  for each  $i \in I$ , and 0 is not connected back to any participant, that is,  $(0, i) \notin \vec{E}$  for each  $i \in I$ .

In this model, capacities  $L = (L_{ij} : i, j \in I)$  [ton/h] for the connections between every pair of participants  $i, j \in I$  are considered. For a pair  $i, j \in I$ ,  $L_{ij}$  is non-negative and it determines how much water can be sent from  $i$  to  $j$ . The capacities of the connections  $(i, 0)$  are assumed to be unbounded. If  $L_{ij} = 0$ , then  $(i, j) \notin \vec{E}$ . Therefore, each EIP is uniquely defined by its capacity

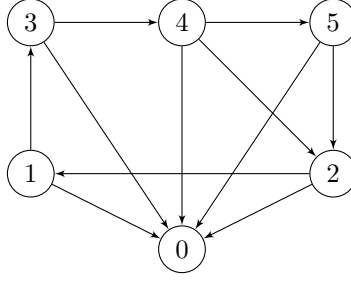


Figure 2.1: Example of  $G = (V, \vec{E})$  for 5 agents.

vector  $L$ .

At the design stage, the design problem is to decide the vector  $L$ . When  $L$  is decided, the the problem of deciding the connections in  $\vec{E}$  is implicitly solved.

## 2.2 Physical operation model

After the design stage, for a given EIP with capacities  $L = (L_{ij})$ , there is a set of operational constraints that every participant  $i \in I$  must follow. These constraints must hold every day of the park's operation, during the whole lifetime horizon.

Each day, each enterprise  $i \in I$  produces a mass load of contaminant  $M_i$  [kg/h] that needs to be diluted, considering a maximum outlet concentration  $C_{i,\text{out}}$  [mg/l]. To do so, the enterprise can buy fresh water  $z_i$  [ton/h] (tons per hour) and receive partially polluted water from other enterprises.  $F_{ki}$  [ton/h] will denote the water sent from participant  $k \in I$  to participant  $i$ .

A relevant assumption here, is that the operation of every participant is optimal, in the sense that their maximal outlet concentration is always attained and no excess of fresh water is consumed (see, e.g., [29, 30]).

Finally, after diluting  $M_i$ , participant  $i$  must send away the polluted water, by either sending part of their polluted water to another participant  $j \in I$ , through the flux  $F_{ij}$  [ton/h], or by discharging residual polluted water to the sink node through the flux  $O_i$  [ton/h].

Also, for an exchange to be valid, every participant can accept polluted water with a maximum inlet concentration  $C_{i,\text{in}}$  [mg/l]. This constraint is given by the process description of each participant. All inlet fluxes, including the fresh water, are assumed to get mixed before entering the process. Thus, the inlet concentration constraint is evaluated in the mixed inlet flux.

To sum up, the operation variables of the park are given by: 1) the fresh water consumption of each agent, given by the vector  $z = (z_i : i \in I)$ ; 2) the exchange water matrix  $F = (F_{ij} : i \neq j \in I)$ ; and 3) the discharge of each agent to the sink node, given by the vector  $O = (O_i : i \in I)$ . A valid operation is then given by values of  $(z, F, O)$  satisfying the following operation constraints:

1. **Water Mass Balance:** for every participant  $i \in I$ , the total inlet flux must coincide with

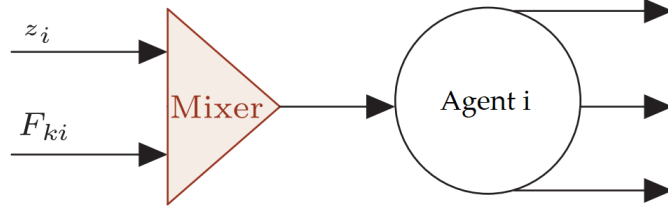


Figure 2.2: Water Mixture Description for Agent  $i$ . Original figure from [30].

the total outlet flux.

$$z_i + \sum_{k \neq i} F_{ki} = \sum_{j \neq i} F_{ij} + O_i. \quad (2.1)$$

At the sink node, there is no balance constraint.

2. **Contaminant Mass Balance:** For every participant  $i \in I$ , the total inlet contaminant mass must coincide with the total outlet contaminant mass, that is,

$$M_i + \sum_{k \neq i} C_{k,\text{out}} F_{ki} = C_{i,\text{out}} \left( O_i + \sum_{j \neq i} F_{ij} \right). \quad (2.2)$$

The mass is computed from the fluxes  $F$ , given the optimality assumption that the outlet concentration is always attained.

3. **Inlet/Outlet Concentration Constraints:** for every participant  $i \in I$ ,

$$\sum_{k \neq i} C_{k,\text{out}} F_{ki} \leq C_{i,\text{in}} \left( z_i + \sum_{k \neq i} F_{ki} \right). \quad (2.3)$$

The above inequality is the inlet concentration constraint expressed in terms of contaminant mass.

4. **Positivity of Fluxes:** all the fluxes in the EIP must be non-negative, that is,

$$F_{ij} \geq 0, \quad \forall i \neq j \in I \quad \text{and} \quad z_i, O_i \geq 0, \quad \forall i \in I. \quad (2.4)$$

5. **Boundedness of exchanges:** all the fluxes in the EIP must be within the capacities given by the vector  $L$ , that is,

$$F_{ij} \leq L_{ij}, \quad \forall i \neq j \in I. \quad (2.5)$$

### 2.3 Economical constraints for participation

The central authority's goal at the daily operation is to minimize the global fresh water consumption. Nevertheless, this is not necessarily aligned with the individual participants' interests,

which is to minimize their operational costs. Hence, jointly with the already detailed physical constraints, economical ones must be added for the daily operation of every participant, considering their *individual rationality*; a well-known principle : any enterprise will take part of the EIP only if this participation is economically convenient (see, e.g., [13]).

Consider the operation of an isolated agent  $i \in I$  (i.e. no participation in an EIP). This agent should operate by *stand-alone* conditions, where all the water needed to dilute the mass load of contaminant should be bought fresh, and all the water waste should then be discharged to the sink. This means that all the fresh water  $z_i$  and the discharged waste  $O_i$ , are given by the same following expression:

$$z_i = \frac{M_i}{C_{i,\text{out}}} \quad \text{and} \quad O_i = \frac{M_i}{C_{i,\text{out}}}. \quad (2.6)$$

Then, the daily stand-alone fresh water consumption is given by

$$\text{SA}_i(M_i) = \frac{M_i}{C_{i,\text{out}}}. \quad (2.7)$$

Now, let us consider the operation of an agent  $i \in I$  connected to other participants in an EIP. Now, this agent has two additional options for its operation: it can partially replace fresh water consumption by receiving partially polluted water from other participants, and it also can send some of its own water waste to other agents as partially polluted water. Hence, the daily operational cost in this new scenario is given by

$$\text{Cost}_i(F, z) = e \left( \sum_{j \neq i} F_{ij} + \sum_{k \neq i} F_{ki} \right) + dO_i + cz_i, \quad (2.8)$$

where  $c$  is the marginal cost for consuming fresh water,  $d$  is the marginal cost for the discharging water, and  $e$  models a marginal costs for using the connections of the park, which is co-paid for the sending and receiving participants.

After this discussion, an additional constraint for the model is considered, in order to tackle the *individual rationality* of every participant, which ensures that the operational costs for every agent are less than the *stand-alone* operation, that is,

$$\text{Cost}_i(F, z) \leq (c + d)\text{SA}_i(M_i). \quad (2.9)$$

This economic constraint was first introduced in [30] to validate models with a central operation under the assumption of selfish agents. It was then used in [3] with the same purposes. While here are not necessarily selfish agents, constraint (2.9) is still needed to model rational (cooperative) agents that follow the central operation.

## 2.4 Uncertainty and Two-stage model

In order to design the EIP, a new element has to be considered: the mass load production of contaminant of every participant,  $\xi = (M_1, \dots, M_n)$  is uncertain at the design stage, because each process has daily unpredictable variations. These variations are only revealed during the daily operation of the EIP, and of course, they can be different every time. Thus, after  $\xi$  is revealed, for a given capacity vector  $L$ , the daily operation problem  $Q(L, \xi)$  can be defined as



$$Q(L, \xi) = \begin{cases} \min_{(z, F, O)} & Z := \sum_{i=1}^n z_i \\ \text{s.t.} & z_i + \sum_{k \neq i} F_{ki} = O_i + \sum_{j \neq i} F_{ij}, \forall i \in I \\ & M_i + \sum_{k \neq i} C_{k, \text{out}} F_{ki} = C_{i, \text{out}} \left( O_i + \sum_{j \neq i} F_{ij} \right), \forall i \in I \\ & \sum_{k \neq i} C_{k, \text{out}} F_{ki} \leq C_{i, \text{in}} \left( z_i + \sum_{k \neq i} F_{ki} \right), \forall i \in I \\ & \text{Cost}_i(F, z) \leq (c + d) \text{SA}_i(M_i), \forall i \in I \\ & 0 \leq F \leq L \\ & z \geq 0. \end{cases} \quad (2.10)$$

For  $L$  and  $\xi$  fixed,  $X(L, \xi)$  is defined as the feasible set of Problem (2.10), which is given by linear constraints only. That is,

$$X(L, \xi) := \left\{ (z, F, O) : \begin{cases} z_i + \sum_{k \neq i} F_{ki} = O_i + \sum_{j \neq i} F_{ij} \\ M_i + \sum_{k \neq i} C_{k, \text{out}} F_{ki} = C_{i, \text{out}} \left( O_i + \sum_{j \neq i} F_{ij} \right) \\ \sum_{k \neq i} C_{k, \text{out}} F_{ki} \leq C_{i, \text{in}} \left( z_i + \sum_{k \neq i} F_{ki} \right) \\ \text{Cost}_i(F, z) \leq (c + d) \text{SA}_i(M_i) \\ 0 \leq F \leq L \\ z \geq 0. \end{cases} \right\} \quad (2.11)$$

This is the general setting of a two-stage problem: minimize a cost function, which depends on the design variables  $L$  (decided here-and-now), taking into account the optimal operation of every day, which is given by the parametric problem (2.10).

An option to solve this two-stage problem, is to obtain or define first an analytical expression for  $Q(L, \xi)$ . See, e.g. [8] for an application. However, this is not always possible. Instead, this problem will be solved using the implicit expression of  $Q(L, \xi)$  as the optimal value of the problem (2.10). This general approach is known as *stochastic optimization with recourse* (see, e.g. [4, 16, 32]). Using this tools, the determination of the analytical expression for  $Q$  can be avoided, and working directly with the optimal value obtained in the daily operation stage is plausible.

Therefore, for the first stage, the minimization of an objective function of the form

$$\langle c, L \rangle + \mathcal{R}[Q(L, \xi)], \quad (2.12)$$

can be considered, where  $c$  models investment costs and  $\mathcal{R}$  is a operator that takes into account the stochasticity of the model, and computes what is called a *risk measure*. These kind of measure aim to characterize the uncertain value of  $Q(L, \xi)$  by two relevant characteristics, its mean (to measure the expected outcome) and its risk or dispersion (to measure its uncertainty).

This kind of functions allow us to go beyond classic computations (such as the expected value) and consider risk-averse design (see, e.g. [32, Chapter 6]), in order to get, for example, an EIP that is *robust* in some sense to difficult or undesired conditions during its lifetime. Other works (see, e.g. [10]) discuss these ideas in different contexts.

### 3 Optimization criteria: Performance and Robustness

Once defined  $Q(L, \xi)$ , the optimal design problem can be tackled: to decide the best capacity vector  $L$ , taking into account the uncertain operational cost-to-go  $Q(L, \xi)$ .

On a first approach, it is natural to consider the averaged fresh water consumption costs of the EIP as the objective function to minimize. This correspond to expected value  $A\mathbb{E}[Q(L, \xi)]$ , where  $A$  is the lifetime factor, and it allows us to control some kind of “average day” on its long-term operation. However, this does not necessarily give us an optimal EIP considering other indicators, such as its operational costs for every participant.

Another possible approach is to design an EIP that admits *good-enough* operations in most of the uncertain scenarios, for example by fixing an admissible level of improvement on the total expendings for the daily operation, and look for a configuration that allows us to get the greatest possible number of scenarios where this improvement is achieved.

In order to quantify this goal, the  $(1 - \alpha)$ -level *goodness* for an EIP is defined as

$$G_\alpha(L, \xi) \doteq Q(L, \xi) - \alpha SA(\xi), \quad (3.1)$$

where  $SA(\xi) = \sum_{i \in I} SA_i(M_i)$  is the total net cost, if all the agents worked on *stand-alone* operation. If  $G_\alpha \leq 0$ , it means that the EIP operation is better than the stand-alone one. Hence, the *robustness* of the EIP can be defined as

$$\text{Rob}_\alpha(L) = \mathbb{P}[G_\alpha(L, \xi) \leq 0] \quad (3.2)$$

In order to maximize the number of *good* scenarios, the value of  $\text{Rob}_\alpha(L)$  (correspondingly minimizing  $-\text{Rob}_\alpha(L)$ ) can also be maximized, as a part of the objective function. Using this idea, more *robust* designs can be obtained, ensuring that most of the scenarios work well for the imposed improvement level.

Hence, an unifying formulation that considers the average fresh water consumption, the *robust* design and the investment costs, is given by:

$$\begin{cases} \min_L & w_1 \langle c, L \rangle + w_2 \mathbb{E}[Q(L, \xi)] - w_3 \text{Rob}_\alpha(L) \\ \text{s.t.} & L \geq 0 \end{cases} \quad (3.3)$$

where  $w = (w_1, w_2, w_3) \geq 0$  is a specific weight vector, which allows us to prioritize the criteria. Therefore, the general two-stage problem is given by the design problem (3.6) in the first stage, and the daily operation (2.10) in the second one.

Finally, at the constraints level in the design stage, similarly to the objective function construction, criteria for the design of the EIP can be added. Consider for this, the following models:

- **Budget constraints:** For a budget  $B \geq 0$  [\$], a budget constraint is given by

$$\langle c, L \rangle \leq B \quad (3.4)$$

The total budget for the design of the EIP can be controlled, combined with a minimization of its robustness and/or its fresh water consumption.

- **Minimal Desired Robustness:** considering the definition (3.2), this can be also considered as a constraint, writing for example

$$\text{Rob}_\alpha(L) \geq \beta. \quad (3.5)$$

There's a lot of work in this area, known as *chance constraints optimization* (see, e.g., [32, Section 5.7]), where the value of a probability function is fixed over a desired quantity.

Therefore, the general problem can be seen as

$$\mathcal{P} = \begin{cases} \min_L & w_1 \langle c, L \rangle + w_2 \mathbb{E}[Q(L, \xi)] - w_3 \text{Rob}_\alpha(L) \\ \text{s.t.} & L \geq 0 \\ & c_1(\langle c, L \rangle - B) \leq 0 \\ & c_2(\text{Rob}_\alpha(L) - \beta) \geq 0 \end{cases}, \quad (3.6)$$

where  $c_1, c_2 \in \{0, 1\}$  will be used in order to eventually consider budget or robustness constraints, as recently defined.

## 4 The Sample Average Approximation Method

The Sample Average Approximation (SAA), firstly shown in [15], allows one to tackle two-stage problems through the use of sampling and optimization Monte-Carlo based methods, in order to be able to use deterministic programming tools to solve them. In this section, the Sample Average Approximation Method will be briefly visited. For recent surveys, please refer to [11, 32].

Let us consider a minimization problem to be the expected value of a function, that is,

$$\min_{x \in X} q(x) = \mathbb{E}[Q(x, \xi)], \quad (4.1)$$

where  $x \in \mathbb{R}^n$  and  $\xi \in \mathbb{R}^m$ ,  $X \subseteq \mathbb{R}^n$  is a fixed set and  $\xi = \xi(\omega)$  is a random vector.  $\Xi \in \mathbb{R}^m$  denotes the support of the probability distribution of  $\xi$ , that is, the smallest closed set of  $\mathbb{R}^m$  such that  $\mathbb{P}(\xi \in \mathbb{R}^m \setminus \Xi) = 0$ .

If a set of i.i.d. random vectors with the same distribution as  $\xi$  is considered, say  $\{\xi^1, \xi^2, \dots, \xi^N\}$  (normally called a *sample*) of size  $N$ , the estimation of  $q(x)$  can be obtained by averaging the values of  $Q(x, \xi^j)$ , with  $j = 1, \dots, N$ . This leads to the main idea of the Sample Average Approximation (SAA) method, where  $q(x)$  is replaced with its average approximation

$$q_N(x) \doteq \frac{1}{N} \sum_{j=1}^N Q(x, \xi^j). \quad (4.2)$$

From a statistics approach, since each  $\xi^i$  has the same distribution as  $\xi$ , it is clear that

$$\mathbb{E}[q_N(x)] = q(x), \quad (4.3)$$

and then,  $q_N(x)$  is an unbiased estimator for  $q(x)$ . Then, given a realization  $\{\hat{\xi}^1, \hat{\xi}^2, \dots, \hat{\xi}^N\}$  of the sample,

$$\hat{q}_N(x, \hat{\xi}^1, \dots, \hat{\xi}^N) \doteq \frac{1}{N} \sum_{j=1}^N Q(x, \hat{\xi}^j), \quad (4.4)$$

is now a deterministic expression, so it can be tackled with standard optimization tools, in order to get an approximated solution of (4.1) by computing the solution of

$$\min_{x \in X} \hat{q}_N(x, \hat{\xi}^1, \dots, \hat{\xi}^N) = \frac{1}{N} \sum_{j=1}^N Q(x, \hat{\xi}^j), \quad (4.5)$$

## 4.1 Statistical Properties of SAA Estimators

The optimal value  $\hat{\nu}_N$  and an optimal solution  $\hat{x}_N$  of the problem (4.5) will be considered statistical estimators of their counterparts of the original problem (4.1). Naturally, when using this technique, some questions arise

1. Is this method consistent? That is, do the solution of the SAA version of the problem converge in some way to the original problem solution?
2. Can any guarantees about a solution obtained by this method be given? Are the obtained solutions any good?

### 4.1.1 Consistency of Solutions

Let us write  $\nu^*$  and  $S$  to be the optimal value and the set of optimal solutions of the original problem (4.1), and  $\hat{\nu}_N$ ,  $\hat{x}_N$ ,  $\hat{S}_N$  the optimal value, an optimal solution and the set of optimal solutions of (4.5).  $\hat{\nu}_N$  is said to be a *consistent* estimator of  $\nu^*$  if  $\hat{\nu}_N \rightarrow \nu^*$  with probability one, as  $N \rightarrow \infty$ . That is, if

$$\mathbb{P} \left[ \lim_{N \rightarrow \infty} \hat{\nu}_N = \nu^* \right] = 1.$$

Similarly,  $\hat{x}_N$  is said to be *consistent* if  $\text{dist}(\hat{x}_N, S) \rightarrow 0$  with probability one, as  $N \rightarrow \infty$ . That is,

$$\mathbb{P} \left[ \lim_{N \rightarrow \infty} \text{dist}(\hat{x}_N, S) = 0 \right] = 1.$$

If no chance constraint is considered, consistency of the SAA method is well-studied (see, e.g., [32]). The version of the problem that considers chance constraints (that is, Problem 3.6 with constraints of the form (3.5) fits the setting of [18, 19]. There, consistency of SAA methods is studied for problems with chance constraints.

### 4.1.2 Quality of Solutions

Let us denote a candidate solution of the original problem as  $\hat{x} \in X$ . One of the approaches for assessing solution quality is to bound the candidate solution's optimality gap, for which a method called the *Multiple Replications Procedure* (MRP) can be used, as presented in [20].

The optimality gap for  $\hat{x}$  can be computed as  $f(\hat{x}) - \nu^*$ . The value of  $\nu^*$  is not known, but can be bound as in [23] by the bias result

$$\mathbb{E}[\nu_N] \leq \nu^*.$$

Therefore, an *upper bound* on the optimality gap of  $\hat{x}$  (from now on, *point estimator of the optimality gap* of  $\hat{x}$ ), can be estimated via

$$\mathcal{G}_N(\hat{x}) = f_N(\hat{x}) - \nu_N. \quad (4.6)$$

When viewed as an estimator of the optimality gap,  $\mathcal{G}_N(\hat{x})$  is biased,  $\mathbb{E}[\mathcal{G}_N(\hat{x})] \geq f(\hat{x}) - \nu^*$ .

In order to compute this estimator, the same i.i.d. random variables  $\xi^1, \dots, \xi^N$  from the distribution of  $\xi$  are used, for both terms in (4.6). That is, given a realization  $\{\hat{\xi}^1, \hat{\xi}^2, \dots\}$  of the random vector  $\xi$ ,

$$\hat{\mathcal{G}}_N(\hat{x}) = \hat{f}_N(\hat{x}, \hat{\xi}^1, \dots, \hat{\xi}^N) - \hat{\nu}_N(\hat{\xi}^1, \dots, \hat{\xi}^N),$$

where the notation  $\hat{\nu}_N(\hat{\xi}^1, \dots, \hat{\xi}^N)$  emphasizes that this quantity corresponds to the optimal value of the approximated problem, for the particular realization  $\{\hat{\xi}^1, \dots, \hat{\xi}^N\}$ . The use of the same observations in both terms of  $\hat{\mathcal{G}}_N$  results in variance reduction.

As shown in [31], statistical inference for  $\mathcal{G}_N(\hat{x})$  is usually hard to compute. This difficulty can be avoided employing a “batch-means” approach, commonly used in the simulation literature. Thinking of this approach as an algorithm, the process goes as follows:

1. First off, a number of “batches”  $N_G$  is decided, in order to compute a confidence interval for the point estimator of the optimality gap. As mentioned in [11], typical values for  $N_G$  move between 20 or 30.
2. Observations  $\{\xi^{k1}, \xi^{k2}, \dots, \xi^{kN}\}$ , for  $k = 1, 2, \dots, N_G$  are generated; and these are averaged to obtain a point estimator of the optimality gap

$$\bar{\mathcal{G}}(\hat{x}) = \frac{1}{N_G} \sum_{k=1}^{N_G} \mathcal{G}_N^k(\hat{x}),$$

where  $\mathcal{G}_N^k$  is computed following (4.6), using the  $k$ -th batch of observations.

3. The sample variance is computed as

$$S_{\mathcal{G}}^2 = \frac{1}{N_G - 1} \sum_{k=1}^{N_G} (\mathcal{G}_N^k(\hat{x}) - \bar{\mathcal{G}}(\hat{x}))^2.$$

4. An approximate  $(1 - \gamma)$ -level confidence interval estimator on the optimality gap of  $\hat{x}$  is given by

$$\left[ 0, \bar{\mathcal{G}}(\hat{x}) + \frac{z_\gamma S_{\mathcal{G}}}{\sqrt{N_G}} \right],$$

where  $z_\gamma$  denotes a  $(1 - \gamma)$ -quantile from a standard Normal distribution.

## 4.2 Reformulation under the SAA Method

Considering the elements detailed in Section 4, reformulation of the two-stage problem is feasible, using the Sample Average Approximation Method.

Given a sample  $\{\xi^1, \xi^2, \dots, \xi^N\}$  for the mass load production of contaminant, where each  $\xi^m = (M_1^m, M_2^m, \dots, M_n^m)$  for  $m = 1, \dots, N$ , the estimator for the expected value of  $Q$  can be computed as

$$q_N(L) = \frac{1}{N} \sum_{m=1}^N Q(L, \xi^m), \quad (4.7)$$

and given a realization  $\{\hat{\xi}^1, \hat{\xi}^2, \dots, \hat{\xi}^N\}$  of this sample, define

$$\hat{q}_N(L) = \frac{1}{N} \sum_{m=1}^N Q(L, \hat{\xi}^m), \quad (4.8)$$

which, as discussed before, is now a deterministic expression.

For  $\text{Rob}_\alpha(L)$ , as it is defined in (3.2) by a probability, it can also be written as an expected value, considering the corresponding indicator function for  $G_\alpha(L, \xi)$ . Therefore, for the same previously considered realization of the sample, define

$$\widehat{\text{Rob}}_\alpha(L) = \frac{1}{N} \sum_{m=1}^N \mathbb{1}_{(-\infty, 0]}[G_\alpha(L, \hat{\xi}^m)]. \quad (4.9)$$

With all of this in mind, the initial reformulation of the Problem (3.6) is given by

$$\hat{\mathcal{P}}_0 = \begin{cases} \min_L & w_1 \langle c, L \rangle + w_2 \hat{q}_N(L) - w_3 \widehat{\text{Rob}}_\alpha(L) \\ \text{s.t.} & L \geq 0 \\ & c_1 \langle c, L \rangle - B \leq 0 \\ & c_2 (\widehat{\text{Rob}}_\alpha(L) - \beta) \geq 0 \end{cases}. \quad (4.10)$$

Now, the following consistency proposition is stated, based on results that can be found in [32, Corollary 4.41, Theorem 5.30]. Recall that a non-negative function  $f(x)$  defined on a convex set  $\Omega \subseteq \mathbb{R}^n$  is said to be log-concave, if for all  $x, y \in \Omega$  and every  $t \in [0, 1]$ , the following inequality holds true

$$f(tx + (1-t)y) \geq f(x)^t f(y)^{1-t}.$$

**Theorem 4.1** *When  $\xi^m$  is a random vector that has a log-concave probability distribution for all  $m \in [N]$ , then the solutions of the reformulated Problem (4.10) are consistent estimators of the solutions of Problem (3.6).*

*Proof.* The proof of this result can be found in the Appendix. □

In order to compute the objective function for this problem, two main difficulties arise: an explicit expression for  $Q(L, \hat{\xi}^m)$  at any realization  $\hat{\xi}^m$  is not feasible to be obtained, and the indicator functions that define  $\widehat{\text{Rob}}_\alpha(L)$  do not have good properties for optimization in general (they are not convex nor continuous).

To avoid computation of  $Q(L, \hat{\xi}^m)$ , variables  $(z^m, F^m)$  are created that, a posteriori, will represent the optimal operation of the ecopark at the second-stage, given the realization  $\hat{\xi}^m$ . Similarly, to avoid computation of  $\mathbb{1}_{(-\infty, 0]}(G_\alpha(L, \hat{\xi}^m))$ , a binary variable  $y^m$  is included, that represents the value of the indicator function. The idea is to formulate a mixed linear problem that, a posteriori, will verify

$$Q(L, \hat{\xi}^m) = \sum_i z_i^m \quad \text{and} \quad \mathbb{1}_{(-\infty, 0]}[G_\alpha(L, \hat{\xi}^m)] = y_m.$$

In order to deal with the computation of  $\hat{q}_N(L)$ , for every index  $m \in [N] = \{1, \dots, N\}$ , the tuple  $(z^m, F^m, O^m) \in X(L, \hat{\xi}^m)$  is considered, and

$$Z^m = \sum_{i=1}^n z_i^m.$$

The sum of every  $Z^m$  will be used as a replacement of  $\hat{q}_N(L)$  in the final reformulations of Problem (4.10). Now, in order to compute the indicator functions needed at the robustness term, an additional binary variable  $y \in \{0, 1\}^N$  is defined, that works as follows: the objective function is reformulated as

$$f(L, y) := w_1 \langle c, L \rangle + \frac{w_2}{N} \sum_{m=1}^N Z^m - \frac{w_3}{N} \sum_{m=1}^N y_m, \quad (4.11)$$

where  $y_m$  verifies, for every realization  $\hat{\xi}^m$ , the additional constraint

$$G_\alpha(L, Z, \hat{\xi}^m) \leq SA(\hat{\xi}^m)(1 - y_m), \quad (4.12)$$

where  $G_\alpha(L, Z, \hat{\xi}^m) = Z^m - SA(\hat{\xi}^m)$ . This works as follows: if  $y_m = 1$ , then (4.12) implies that  $G_\alpha(L, Z, \hat{\xi}^m) \leq 0$ , or in other words,

$$y_m = 1 \implies \mathbb{1}_{(-\infty, 0]}[G_\alpha(L, Z, \hat{\xi}^m)] = 1.$$

If  $y_m = 0$ , then (4.12) implies that

$$G_\alpha(L, Z, \hat{\xi}^m) \leq SA(\hat{\xi}^m),$$

which becomes a non-active constraint,

$$\begin{aligned} G_\alpha(L, Z, \hat{\xi}^m) &= Z^m - \alpha SA(\hat{\xi}^m) \\ &\leq SA(\hat{\xi}^m) - \alpha SA(\hat{\xi}^m) \\ &= (1 - \alpha) SA(\hat{\xi}^m) \\ &\leq SA(\hat{\xi}^m). \end{aligned}$$

Before stating the main result of this work, the following reformulations of problem (3.6) are defined, considering the additional constraint (4.12).

$$\widehat{\mathcal{P}}_1 = \begin{cases} \min_{L,z,F,y} & w_1 \langle c, L \rangle + \frac{w_2}{N} \sum_{m=1}^N Z^m - \frac{w_3}{N} \sum_{m=1}^N y_m \\ \text{s.t.} & L \geq 0 \\ & c_1 (\langle c, L \rangle - B) \leq 0 \\ & c_2 \left( \frac{1}{N} \sum_{m=1}^N y_m - \beta \right) \geq 0 \\ & (z^m, F^m) \in X(L, \hat{\xi}^m), \forall m \in [N] \\ & G_\alpha(L, Z, \hat{\xi}^m) \leq \text{SA}(\hat{\xi}^m)(1 - y_m) \end{cases} \quad (4.13)$$

**Theorem 4.2** *If  $L^*$  is an optimal solution of Problem (4.10), then there exists  $(z^*, F^*, O^*, y^*)$  such that  $(L^*, z^*, F^*, O^*, y^*)$  is an optimal solution of Problem (4.13). Conversely, if  $(L^*, z^*, F^*, O^*, y^*)$  is an optimal solution of Problem (4.13), then  $L^*$  is an optimal solution for Problem (4.10). In both cases, one has that*

$$w_1 \langle c, L^* \rangle + w_2 \widehat{q}_N(L^*) - w_3 \widehat{\text{Rob}}_\alpha(L^*) = w_1 \langle c, L^* \rangle + \frac{w_2}{N} \sum_{m=1}^N z^{*m} - \frac{w_3}{N} \sum_{m=1}^N y_m^*. \quad (4.14)$$

*Proof.* The proof of this result can be found in the Appendix. □

## 5 Numerical Experiments

In this section, the main numerical results are shown, using the SAA approach in order to solve the two-stage problem of the EIP design and operation.

For the simulations, water exchange networks between 4, 10 and 15 agents are considered. For each network, the inlet and outlet concentrations for every agent have to be defined, with a *nominal* value of contaminant mass loaded by each participant,  $\bar{M}_i$ . Using this nominal values as means, three different distributions (two of them symmetrical) are considered.

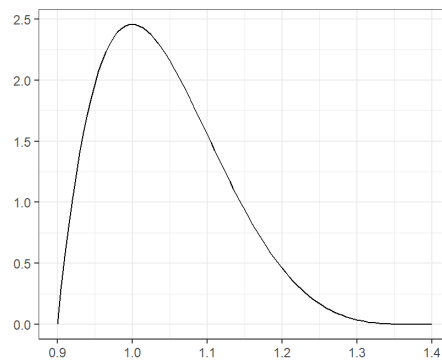


Figure 5.1: Beta Distribution,  $a = 2$ ,  $b = 5$ .

Specifically:



1. Uniform Distribution, with range  $[a, b] = [0.9\bar{M}_i, 1.1\bar{M}_i]$ .
2. Normal Distribution, with  $\mu = \bar{M}_i$  and  $\sigma = 0.1\bar{M}_i$ .
3. Beta Distribution, centered in  $\bar{M}_i$  with  $a = 2$ ,  $b = 5$ .

The Beta Distribution with these chosen parameters, has the shape seen in Figure 5.1.

Agent	$C_{in}$ [mg/L]	$C_{out}$ [mg/L]	Nominal Mass Load (kg/h)
1	0	200	7.00
2	100	500	22.40
3	200	650	62.55
4	0	200	2.00

Table 1: First Study Case,  $n = 4$  participants.

Agent	$C_{in}$ [mg/L]	$C_{out}$ [mg/L]	Nominal Mass Load (kg/h)
1	25	80	2.00
2	25	90	2.88
3	25	200	4.00
4	50	100	3.00
5	50	800	30.00
6	400	800	5.00
7	400	600	2.00
8	0	100	1.00
9	50	300	20.00
10	150	300	6.50

Table 2: Second Study Case,  $n = 10$  participants.

Agent	$C_{in}$ [mg/L]	$C_{out}$ [mg/L]	Nominal Mass Load (kg/h)
1	30	100	7.50
2	0	200	6.00
3	50	100	5.00
4	80	800	30.00
5	400	800	4.00
6	20	100	2.50
7	50	100	2.20
8	80	400	5.00
9	100	800	30.00
10	400	1000	4.00
11	30	60	2.00
12	25	50	2.00
13	25	75	5.00
14	50	800	30.00
15	100	900	13.00

Table 3: Third Study Case,  $n = 15$  participants.

The smaller network is considered in order to compare the results for this approach, with the

one proposed by [34] on their previous work. The bigger academic examples considered are the following, based on EIPs suggested by [30] and [3] respectively.

The marginal costs  $c = 0.13$  [USD/ton] for consuming fresh water,  $d = 0.22$  [USD/ton] for the discharging water and  $e = 0.01$  [USD/ton] for using the connections of the park, are considered for every simulation.

All the simulations were implemented in a computer with an Intel Core(TM) i7-10700F processor, running at 2.90GHz, with 16 GB of RAM, running Windows 10 Pro; with Julia v1.6.1 programming language, using Gurobi v9 as solver.

## 5.1 Sensitivity Analysis for the Sample Size

First, a sensitivity analysis of the sample size  $N$  is performed, in order to choose a value that ensures a good optimality gap, without increasing too highly the CPU time. For this, different values of  $N$  are considered, with a Beta distribution for every simulation, in order to get a comparable set of data.

Sample Size $N$	CPU time (s)	Opt.Gap (%)
500	3.64	0.09
1000	14.07	0.08
2000	29.85	0.05
3000	45.49	0.04
4000	66.68	0.04
5000	80.34	0.03
10000	178.75	0.02
20000	419.22	0.02
30000	662.95	0.02

Table 4:  $n = 4$  agents, increasing sample size, batch size = 30.

Sample Size $N$	CPU time (s)	Opt.Gap (%)
500	954.97	0.07
1000	5424.30	0.05
2000	10735.23	0.05
3000	16145.14	0.04
4000	21295.41	0.03
5000	26741.33	0.03
10000	135487.75	0.02
20000	248945.71	0.02
30000	389463.57	0.01

Table 5:  $n = 10$  agents, increasing sample size, batch size = 30.

As can be seen in the two smaller EIPs, fixing  $N = 5000$  gives us a good compromise between solution quality and CPU time.

## 5.2 Sensitivity Analysis for $\alpha$

In this part, a sensitivity analysis for  $\alpha$  will be presented, which measures the desired costs reduction comparing to the *stand-alone* global operation of the EIP (see equation (3.1)).

It is clear that, if a higher costs reduction is needed, the obtained robustness will decrease, because less tested scenarios will achieve the desired level. In order to show this, the problem of maximizing robustness is studied, considering cost penalization, for different values of  $\alpha$ .

$\alpha$	Rob $_{\alpha}$ (%)
0.78	0.00
0.79	0.06
0.80	0.98
0.81	7.13
0.82	26.94
0.83	56.99
0.84	89.21
0.85	98.86
0.86	99.99
0.87	100.00

Table 6:  $n = 4$  agents, increasing  $\alpha$ , batch size = 30.

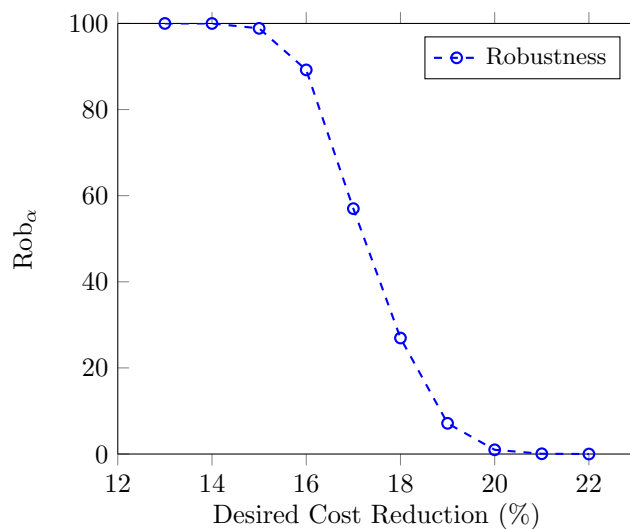


Figure 5.2: Obtained Robustness for EIP,  $n = 4$ , increasing desired Cost Reduction

$\alpha$	Rob $_{\alpha}$ (%)
0.64	0.12
0.65	3.00
0.66	15.36
0.67	47.86
0.68	77.86
0.69	94.93
0.70	97.99
0.71	99.21
0.72	99.86
0.73	100.00

Table 7:  $n = 10$  agents, increasing  $\alpha$ , batch size = 30.

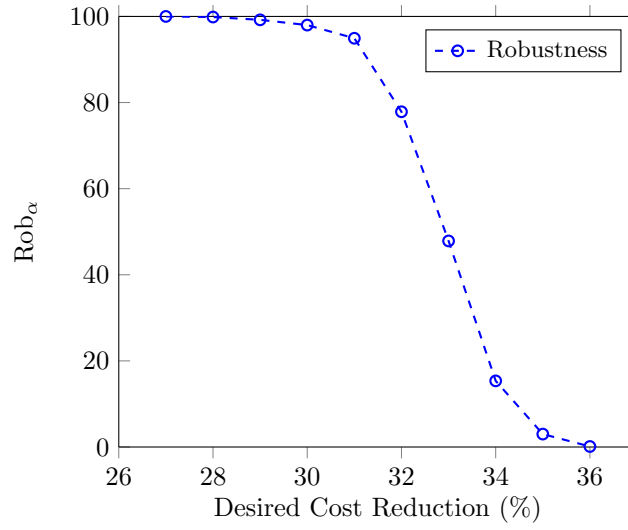


Figure 5.3: Obtained Robustness for EIP,  $n = 10$ , increasing desired Cost Reduction

$\alpha$	Rob $_{\alpha}$ (%)
0.60	0.06
0.61	0.85
0.62	1.03
0.63	6.29
0.64	28.71
0.65	62.86
0.66	84.99
0.67	92.87
0.68	98.13
0.69	98.45

Table 8:  $n = 15$  agents, increasing  $\alpha$ , batch size = 30.

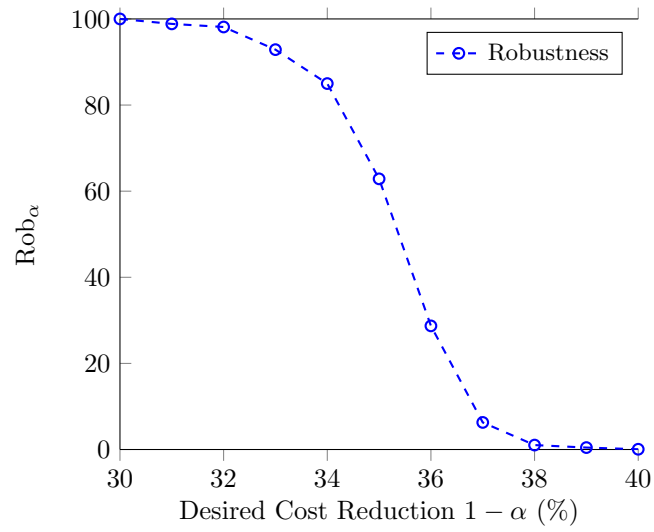


Figure 5.4: Obtained Robustness for EIP,  $n = 15$ , increasing desired Cost Reduction

Near the value  $\alpha = 0.83$  for the smaller EIP, a very sensible change for the obtained robustness is obtained. This value differs, according to the studied EIP, as  $\alpha = 0.68$  is obtained for the  $n = 10$  EIP critical value and  $\alpha = 0.65$  for the bigger one.

### 5.3 Efficiency vs. Robustness

In this section, results of solving the following two different versions of (3.6) are compared.

$$\mathcal{P}_1 = \begin{cases} \min_L & w_1 \langle \kappa, L \rangle + w_2 \mathbb{E}[Q(L, \xi)] \\ \text{s.t.} & L \geq 0 \end{cases} \quad (5.1)$$

$$\mathcal{P}_2 = \begin{cases} \min_L & w_1 \langle \kappa, L \rangle - w_3 \text{Rob}_\alpha(L) \\ \text{s.t.} & L \geq 0 \end{cases} \quad (5.2)$$

Problem (5.1) corresponds to (3.6) considering  $w_3 = c_1 = c_2 = 0$  and (5.2) corresponds to (3.6) considering  $w_2 = c_1 = c_2 = 0$ . This allows us to compare the efficiency and the robustness as optimization criteria.

Both experiments consider penalization over the investment costs term. By doing this, the investment costs are not too high, and actually constructing connections is feasible.

The results are given in Tables 9 to 12, for the  $n = 4$  and  $n = 10$  EIPs.

$M_i$ distrib.	$\alpha$	$Q(L, \xi)$ [ton/h]	$\text{Rob}_\alpha$ (%)	EIP size	Opt.Gap (%)	CPU time (s)
Beta	0.83	160.61	56.90	84.09	0.03	80.24
Uniform	0.83	154.03	59.26	75.35	0.02	79.78
Normal	0.83	154.01	55.71	82.63	0.03	80.78

Table 9:  $n = 4$  agents, penalized costs and efficiency, sample = 5000, batch size = 30.

$M_i$ distrib.	$\alpha$	$Q(L, \xi)$ [ton/h]	$\text{Rob}_\alpha$ (%)	EIP size	Opt.Gap (%)	CPU time (s)
Beta	0.83	162.44	61.11	63.15	0.04	289.08
Uniform	0.83	155.26	64.36	53.53	0.03	275.64
Normal	0.83	156.05	58.95	62.90	0.03	267.95

Table 10:  $n = 4$  agents, penalized costs and robustness, sample = 5000, batch size = 30.

$M_i$ distrib.	$\alpha$	$Q(L, \xi)$ [ton/h]	$\text{Rob}_\alpha$ (%)	EIP size	Opt.Gap (%)	CPU time (s)
Beta	0.83	175.31	100.00	164.27	0.02	26741.33
Uniform	0.83	175.34	100.00	164.36	0.03	25465.91
Normal	0.83	168.56	100.00	165.33	0.02	24712.66

Table 11:  $n = 10$  agents, penalized costs and efficiency, sample = 5000, batch size = 30.

### 5.4 Budget Constraints

In this part, the goal is to compare the EIPs designs, when budget constraints are considered, instead of having the investment costs on the objective function. In particular, analyzing the

$M_i$ distrib.	$\alpha$	$Q(L, \xi)$ [ton/h]	Rob $_{\alpha}$ (%)	EIP size	Opt.Gap (%)	CPU time (s)
Beta	0.83	217.37	100.00	138.75	0.02	90549.33
Uniform	0.83	208.92	100.00	139.36	0.02	89447.31
Normal	0.83	208.44	100.00	141.33	0.03	91290.13

Table 12:  $n = 10$  agents, penalized costs and robustness, sample = 5000, batch size = 30.

obtained robustness in two cases is the main goal: when maximizing efficiency, or maximizing robustness.

In Tables 13 to 15, the optimal budget obtained in the previous simulations is altered by a percentage of itself, varying from 80% to 120%.

Budget	Rob $_{\alpha}$ (%)
80%	45.48
90%	54.70
95%	56.82
100%	57.06
105%	57.06
110%	57.10
120%	57.10

Table 13:  $n = 4$  agents,  $\alpha = 0.83$ , increasing budget.

Budget	Rob $_{\alpha}$ (%)
80%	66.59
90%	74.61
95%	75.32
100%	77.17
105%	77.19
110%	77.22
120%	77.23

Table 14:  $n = 10$  agents,  $\alpha = 0.68$ , increasing budget.

Budget	Rob $_{\alpha}$ (%)
80%	53.52
90%	59.16
95%	60.33
100%	62.77
105%	63.19
110%	63.24
120%	63.31

Table 15:  $n = 15$  agents,  $\alpha = 0.65$ , increasing budget.

## 5.5 Robustness Constraints

In this last analysis part, the goal is to compare the EIPs designs, when robustness constraints are considered. Particularly, in this case the interest is focused in minimizing the investment costs, without maximizing the efficiency of the EIP.

For the maximum efficiency EIP in the  $n = 4$  case, the obtained robustness for  $\alpha = 0.83$  is  $\text{Rob}_{0.83} = 61.6\%$ , with an investment of 89.1. Considering only the minimization of investment, with increasing asked robustness, the following results are obtained

Rob $_{\alpha}$ (%)	Budget
53	53.44
54	54.96
55	56.47
56	57.78
57	58.65
58	59.29
59	59.84
60	60.35

Table 16:  $n = 4$  agents,  $\alpha = 0.83$ , increasing  $\beta$ .

## 6 Discussion and Future Work

The results of Section 5.2 shows us that, when dealing with robustness, there exists some kind of *critical value* for the parameter  $\alpha$ , at which is not possible to build connections between participants that gets us the desired cost reduction.

Looking at the results obtained in Section 5.3 it can be observed that designing a robust EIP, instead of designing a more efficient one, is cheaper in terms of investment costs. This makes sense, as a minimum of *good enough* scenarios are asked for, instead of trying to maximize the efficiency of the whole EIP when dealing with robustness instead. Nevertheless, the CPU time is higher (in between 3 or 4 times) when dealing with this indicator. The increase in CPU time is expected, since Problem 5.2 is more difficult due to the integer variables  $y \in \{0, 1\}^N$  used to model the risk measure  $\widehat{\text{Rob}}_{\alpha}(L)$ .

According to the experiments developed in Section 5.4, past from the total reference budget, increasing significantly the obtained robustness is not feasible. For every value of  $\alpha$ , a *limit robustness* is reached, not depending on the considered budget to do so. At the network configuration, adding more budget to the simulations, gives us bigger, but not more robust EIPs.

The results of the last set of experiments are, from the authors point of view, the most important ones. Here the conclusion is that, minimizing investment but asking for a desired level of robustness, investment deductions up to a 25% can be obtained, only losing 2% of robustness.

Considering bigger EIPs, the results are still promising, considering that the investment cost is still reduced by at least 25% in both examples, only losing 2% of robustness. At the  $n = 10$  and  $n = 15$  EIPs, less connections between participants are needed (downgrade from 40 to 32 at the  $n = 10$  EIP, and from 66 to 57 at the  $n = 15$  EIP).



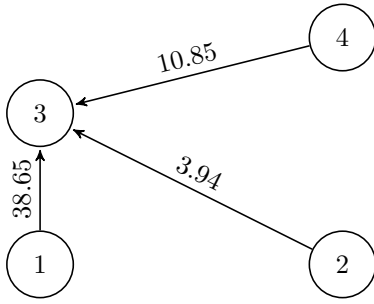


Figure 6.1: EIP at  $\text{Rob}_\alpha = 53\%$

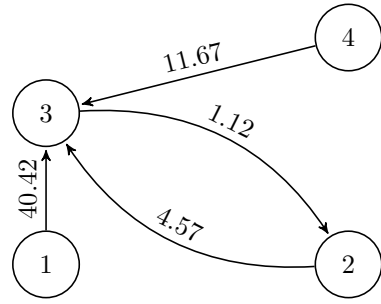


Figure 6.2: EIP at  $\text{Rob}_\alpha = 56\%$

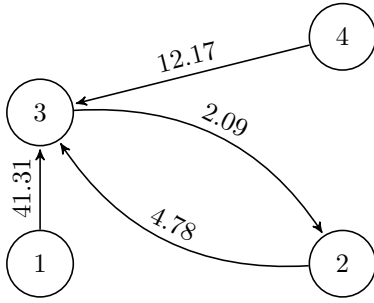


Figure 6.3: EIP at  $\text{Rob}_\alpha = 59\%$

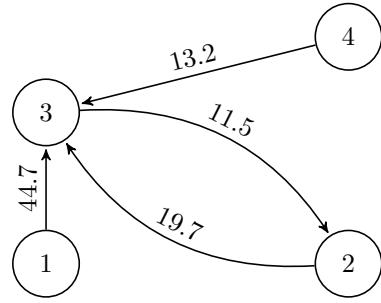


Figure 6.4: Eff. EIP,  $\text{Rob}_\alpha = 61.6\%$

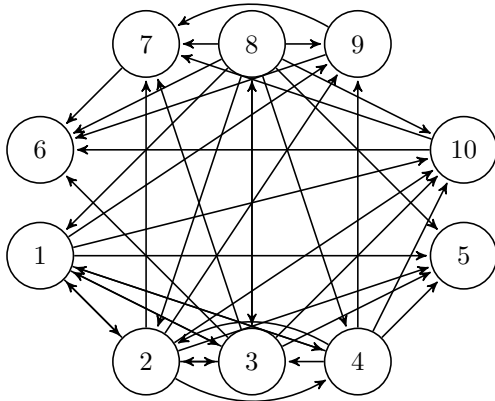


Figure 6.5: Most Efficient  $n = 10$  EIP,  $\text{Rob}_\alpha = 77.17\%$

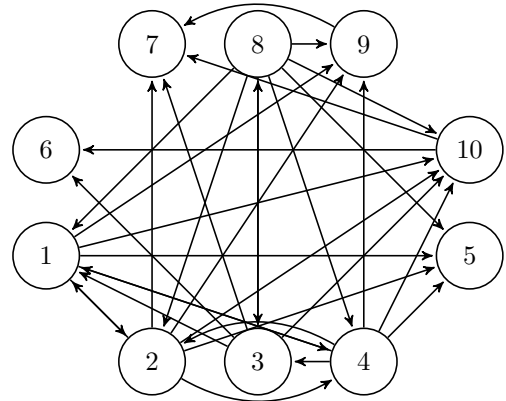


Figure 6.6:  $n = 10$  EIP at  $\text{Rob}_\alpha = 75\%$ , minimizing investment.

From this starting point, a lot of work can be done at designing eco-industrial parks that can be able to *endure* different uncertain scenarios, such as the developed here, where *robustness* is defined as an indicator that measures the capability of an EIP to have a minimum level of *good* operating scenarios, looked from an economical point of view.

As a natural perspective of this work, more uncertainty can be included to the model, adding to the operational time horizon, a probability that, at the end of any given period (for example,

yearly), a number of participants could *leave* the EIP. It would be desirable that, given this *disruptive* event, the ecoindustrial park could still maintain its operations going, studying it from a *system resilience* point of view, as can be seen in previous works like [35].

Another extension, considering the works on EIP design by [3, 30], non-cooperative models can be added to the model, e.g., blind-input, control-input or de-regulated exchange markets. In this last scenario, adding *decision uncertainty* between the participants adds a challenging, but also interesting component.

**Acknowledgements:** This work was partially supported by Center of Mathematical Modelling, FB210005, BASAL funds for centers of excellence from ANID-Chile, and by Fondecyt project 11220586 ANID-Chile.

## References

- [1] Erdem Acar, Gamze Bayrak, Yongsu Jung, Ikjin Lee, Palaniappan Ramu, and Suja Shree Ravichandran. Modeling, analysis, and optimization under uncertainties: a review. *Structural and Multidisciplinary Optimization*, 64, 08 2021.
- [2] Zvi Artstein and Roger J.-B. Wets. Consistency of minimizers and the SLLN for stochastic programs. *Journal of Convex Analysis*, 2(1-2):1–17, 1995.
- [3] Didier Aussel, Kien Cao Van, and David Salas. Optimal design of exchange water networks with control inputs in eco-industrial parks. *Energy Economics*, 120:106480, 2023.
- [4] John R. Birge and François Louveaux. *Introduction to Stochastic Programming*. Springer Series in Operations Research and Financial Engineering. Springer, July 1997.
- [5] Marianne Boix, Ludovic Montastruc, Catherine Azzaro-Pantel, and Serge Domenech. Optimization methods applied to the design of eco-industrial parks: a literature review. *Journal of Cleaner Production*, 87:303–317, 2015.
- [6] Marianne Boix, Ludovic Montastruc, Luc Pibouleau, Catherine Azzaro-Pantel, and Serge Domenech. Industrial water management by multiobjective optimization: from individual to collective solution through eco-industrial parks. *Journal of Cleaner Production*, 22(1):85–97, 2012.
- [7] Alessandro Di Pretoro, Francesco D’Iglio, and Flavio Manenti. Optimal cleaning cycle scheduling under uncertain conditions: A flexibility analysis on heat exchanger fouling. *Processes*, 9(1), 2021.
- [8] Alessandro Di Pretoro, Ludovic Montastruc, Flavio Manenti, and Xavier Joulia. Flexibility analysis of a distillation column: Indexes comparison and economic assessment. *Computers & Chemical Engineering*, 124:93–108, 2019.
- [9] Urmila M. Diwekar. *Optimization Under Uncertainty*, pages 145–208. Springer US, Boston, MA, 2003.
- [10] Ramin Giahi, Cameron A. MacKenzie, and Chao Hu. Design optimization for resilience for risk-averse firms. *Computers & Industrial Engineering*, 139:106122, 2020.
- [11] Tito Homem-de Mello and Güzin Bayraksan. Monte Carlo sampling-based methods for stochastic optimization. *Surveys in Operations Research and Management Science*, 19(1):56–85, 2014.
- [12] Seyedmohsen Hosseini, Kash Barker, and Jose E. Ramirez-Marquez. A review of definitions and measures of system resilience. *Reliability Engineering & System Safety*, 145:47–61, 2016.
- [13] M. O. Jackson. *Mechanism Theory*. 2014. Available at SSRN: <https://ssrn.com/abstract=2542983> or <http://dx.doi.org/10.2139/ssrn.2542983>.
- [14] Rajeshwar S. Kadadevaramath, Jason C. H. Chen, B. Latha Shankar, and K. Rameshkumar. Application of particle swarm intelligence algorithms in supply chain network architecture optimization. *Expert Systems with Applications*, 39(11):10160–10176, 2012.

- [15] Anton J. Kleywegt, Alexander Shapiro, and Tito Homem-de Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502, 2002.
- [16] Francois Louveaux and John R. Birge. *Two-stage stochastic programs with recourse*, pages 3959–3961. Springer US, Boston, MA, 2009.
- [17] Ernest Lowe. *Eco-industrial Park Handbook for Asian Developing Countries Report to Asian Development Bank*. Environment Department, Indigo Development, Oakland, CA, 2001.
- [18] James Luedtke and Shabbir Ahmed. A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization*, 19:674–699, 06 2008.
- [19] James Luedtke, Shabbir Ahmed, and George Nemhauser. An integer programming approach for linear programs with probabilistic constraints. volume 122, pages 247–272, 01 2010.
- [20] Wai-Kei Mak, David P. Morton, and R. Kevin Wood. Monte carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24(1):47–56, 1999.
- [21] L. Mellouk, M. Boulmalf, A. Aaroud, K. Zine-Dine, and D. Benhaddou. Genetic algorithm to solve demand side management and economic dispatch problem. *Procedia Computer Science*, 130:611–618, 2018. The 9th International Conference on Ambient Systems, Networks and Technologies (ANT 2018) / The 8th International Conference on Sustainable Energy Information Technology (SEIT-2018) / Affiliated Workshops.
- [22] Ludovic Montastruc, Marianne Boix, Luc Pibouleau, Catherine Azzaro-Pantel, and Serge Domenech. On the flexibility of an eco-industrial park (eip) for managing industrial water. *Journal of Cleaner Production*, 43:1–11, 2013.
- [23] Vladimir Norkin, Georg Pflug, and Andrzej Ruszczyński. A branch and bound method for stochastic global optimization. *Mathematical Programming, Series B*, 83:425–450, 11 1998.
- [24] William L. Oberkampf, Sharon M. DeLand, Brian M. Rutherford, Kathleen V. Diegert, and Kenneth F. Alvin. Error and uncertainty in modeling and simulation. *Reliability Engineering & System Safety*, 75(3):333–357, 2002.
- [25] William L. Oberkampf, J. Helton, and K. Sentz. Mathematical representation of uncertainty. *Proceedings of the 19th AIAA Applied Aerodynamics Conference*, 2012.
- [26] Bhushan Pawar, Sunhwa Park, Pingfan Hu, and Qingsheng Wang. Applications of resilience engineering principles in different fields with a focus on industrial systems: A literature review. *Journal of Loss Prevention in the Process Industries*, 69:104366, 2021.
- [27] João Pires Ribeiro and Ana Barbosa-Povoa. Supply chain resilience: Definitions and quantitative modelling approaches – a literature review. *Computers & Industrial Engineering*, 115:109–122, 2018.
- [28] E.N. Pistikopoulos and T.A. Mazzuchi. A novel flexibility analysis approach for processes with stochastic parameters. *Computers & Chemical Engineering*, 14(9):991–1000, 1990.

- [29] M. Ramos, M. Boix, D. Aussel, L. Montastruc, and S. Domenech. Water integration in eco-industrial parks using a multi-leader-follower approach. *Computers and Chemical Engineering*, 87:190–207, 2016.
- [30] David Salas, Kien Cao Van, Didier Aussel, and Ludovic Montastruc. Optimal design of exchange networks with blind inputs and its application to eco-industrial parks. *Computers & Chemical Engineering*, 143:107053, 2020.
- [31] Alexander Shapiro. Monte carlo sampling methods. *Handbooks in Operations Research and Management Science*, 10, 12 2003.
- [32] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on stochastic programming*, volume 9 of *MOS-SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, second edition, 2014. Modeling and theory.
- [33] Ross E. Swaney and Ignacio E. Grossmann. An index for operational flexibility in chemical process design. Part I: Formulation and theory. *Aiche Journal*, 31:621–630, 1983.
- [34] Raymond R. Tan and Dennis E. Cruz. Synthesis of robust water reuse networks for single-component retrofit problems using symmetric fuzzy linear programming. *Computers & Chemical Engineering*, 28(12):2547–2551, 2004.
- [35] Guillermo Valenzuela-Venegas, Francisco Henríquez-Henríquez, Marianne Boix, Ludovic Montastruc, Fernando Arenas-Araya, Jenny Miranda-Pérez, and Felipe A. Díaz-Alvarado. A resilience indicator for eco-industrial parks. *Journal of Cleaner Production*, 174:807–820, 2018.

## A Omitted Proofs

This appendix is devoted to prove the main results of Section 4.

*Proof of Theorem 4.1.* Considering results that can be found in [32, Corollary 4.41, Theorem 5.30], the only need is to prove that  $Q$  is a jointly convex function, and therefore,

$$-G_\alpha(L, \xi) = \alpha \text{SA}(\xi) - Q(L, \xi)$$

is a jointly concave function.

For two given pairs  $(L_1, \xi_1)$  and  $(L_2, \xi_2)$ , let us consider  $x_1 \in X(L_1, \xi_1)$ ,  $x_2 \in X(L_2, \xi_2)$  and  $t \in [0, 1]$ . Defining

$$x_t = tx_1 + (1-t)x_2, \quad L_t = tL_1 + (1-t)L_2, \quad \xi_t = t\xi_1 + (1-t)\xi_2.$$

As every physical operation constraint (2.1)-(2.5) and the economical constraint (2.9) can be written as one of the following forms

$$A_1x \leq B_1\xi, \quad A_2x \leq B_2L, \quad A_3x \leq b_3,$$

It can be seen that, for example,

$$\text{Cost}_i(x_t) = t\text{Cost}_i(x_1) + (1-t)\text{Cost}_i(x_2) \leq \frac{(c+d)}{C_{i,\text{out}}}(t\xi_1 + (1-t)\xi_2) = \text{SA}_i(\xi_t)$$

By repeating this with every constraint, it can be concluded that  $x_t \in X(L_t, \xi_t)$ . Considering the optimal value of  $Q(L_t, \xi_t)$ , computations show that

$$Q(L_t, \xi_t) \leq Z(x_t) = tQ_1(L_1, \xi_1) + (1-t)Q_2(L_2, \xi_2),$$

that is,  $Q$  is jointly convex. □

*Proof of Theorem 4.2.* If  $L^*$  is an optimal solution of Problem (4.10), define for each  $m \in [N]$ ,  $(z^{*m}, F^{*m}, O^{*m})$  as the optimal solution of the second-stage problem (2.10), for  $L^*$  and the realization  $\hat{\xi}^m$ . Then, one has that

$$Z^{*m} = \sum_{i=1}^n z_i^{*m} = Q(L^*, \hat{\xi}^m).$$

Let also define  $y^* = (y_m^*)$  as

$$y_m^* = \mathbb{1}_{(-\infty, 0]}[G_\alpha(L^*, Z^{*m}, \hat{\xi}^m)].$$

Then, it follows that

$$\hat{q}_N(L) = \sum_{m=1}^N Z^{*m}, \quad \widehat{\text{Rob}}_\alpha(L) = \frac{1}{N} \sum_{m=1}^N y_m^*,$$

and that constraint (4.12) is verified for each  $m \in [N]$ .

Let us suppose that Problem (4.13) had a different optimal solution, let us say  $(\bar{L}, \bar{z}, \bar{F}, \bar{O}, \bar{y})$ . As discussed before, if  $\bar{y}_m = 1$ , then  $G_\alpha(\bar{L}, \bar{Z}, \hat{\xi}^m) \leq 0$ , implying that

$$\frac{1}{N} \sum_{m=1}^N \bar{y}_m \leq \widehat{\text{Rob}}_\alpha(\bar{L}).$$

Moreover,  $\frac{1}{N} \sum_{m=1}^N \bar{Z}_m \geq \hat{q}_N(\bar{L})$ . Thus, it can be written:

$$\begin{aligned} w_1 \langle c, L^* \rangle + w_2 \hat{q}_N(L^*) - w_3 \widehat{\text{Rob}}_\alpha(L^*) &= w_1 \langle c, L^* \rangle + \frac{w_2}{N} \sum_{m=1}^N Z^{*m} - \frac{w_3}{N} \sum_{m=1}^N y_m^* \\ &> w_1 \langle c, \bar{L} \rangle + \frac{w_2}{N} \sum_{m=1}^N \bar{Z}_m - \frac{w_3}{N} \sum_{m=1}^N \bar{y}_m \\ &\geq w_1 \langle c, \bar{L} \rangle + w_2 \hat{q}_N(\bar{L}) - w_3 \widehat{\text{Rob}}_\alpha(\bar{L}), \end{aligned}$$

which is a contradiction, given that  $L^*$  was an optimal solution for Problem (4.10). Therefore,  $(L^*, z^*, F^*, O^*, y^*)$  is optimal for Problem (4.13), and moreover, (4.14) holds.

Conversely, if  $(L^*, z^*, F^*, O^*, y^*)$  is an optimal solution of Problem (4.13), the claim is that  $L^*$  is feasible for (4.10). Clearly,  $L^*$  is nonnegative and it verifies the budget constraint whenever  $c_1 = 1$ . It only remains to show that, whenever  $c_2 = 1$ ,  $L^*$  also verifies the robustness constraint  $\widehat{\text{Res}}_\alpha(L^*) \geq \beta$ . Note that, always

$$G_\alpha(L^*, Z^{*m}, \hat{\xi}^m) = Z^{*m} - \text{SA}(\hat{\xi}^m) \geq Q(L^*, \hat{\xi}^m) - \text{SA}(\hat{\xi}^m) = G_\alpha(L^*, \hat{\xi}^m).$$

Then, constraint (4.12) yields that

$$\frac{1}{N} \sum_{m=1}^N y_m^* \leq \frac{1}{N} \sum_{m=1}^N \mathbb{1}_{(-\infty, 0]} G_\alpha(L^*, Z^{*m}, \hat{\xi}^m) \leq \frac{1}{N} \sum_{m=1}^N \mathbb{1}_{(-\infty, 0]} G_\alpha(L^*, \hat{\xi}^m) = \widehat{\text{Res}}_\alpha(L^*).$$

Thus, since  $\frac{1}{N} \sum_{m=1}^N y_m^* \geq \beta$ , the conclusion follows. The claim is then proved.

Now, optimality of  $(L^*, z^*, F^*, O^*, y^*)$  entails that

$$\mathbb{1}_{(-\infty, 0]} [G_\alpha(L^*, Z^*, \hat{\xi}^m)] = y_m^*, \quad \forall m \in \{1, \dots, N\}.$$

Indeed, if this were not the case, one would have at least a  $j \in \{1, \dots, N\}$  such that  $\mathbb{1}_{(-\infty, 0]} [G_\alpha(L^*, Z^*, \hat{\xi}^j)] \neq y_j^*$ . That situation only happens if  $G_\alpha(L^*, Z^*, \hat{\xi}^j) \leq 0$  and  $y_j^* = 0$ . However, that is not possible, because in that case, one could define  $\tilde{y}$  such that

$$\tilde{y}_m = \begin{cases} y_m^* & \text{if } m \neq j \\ 1 & \text{if } m = j \end{cases},$$

which is a contradiction, because  $(L^*, z^*, F^*, O^*, \tilde{y})$  would be a better point than  $(L^*, z^*, F^*, O^*, y^*)$ . Noting that for each  $m \in [N]$ , optimality of  $(L^*, z^*, F^*, O^*, y^*)$  yields that  $Z^{*m} = Q(L^*, \hat{\xi}^m)$ , conclusion is, (4.14) holds.

To finish, let us suppose that there is another optimal solution of Problem (4.10),  $\tilde{L}$ . In this case, using the previous development, there would exist  $(\tilde{z}, \tilde{F}, \tilde{O}, \tilde{y})$  such that  $(\tilde{L}, \tilde{z}, \tilde{F}, \tilde{O}, \tilde{y})$  would be feasible and, from (4.14), it would be a better solution of Problem (4.13). This contradicts the optimality of  $(L^*, z^*, F^*, O^*, y^*)$ , finishing the proof.

□