Vol. 42 Núm. 2
Estadística

Optimal Design of Experiments for Minimum Outlet Size Estimation

J. López-Fidalgo C. May J. Moler

Instituto de Ciencia de los Datos e Inteligencia Artificial, Universidad de Navarra, Spain. fidalgo@unav.es

Dipartimento DiSEI, Università degli Studi del Piemonte Orientale, Italy. caterina.may@uniupo.it

Departamento de Estadística, Informática y Matemáticas, Universidad Pública de Navarra, Spain. jmoler@unavarra.es

Cite as:
López-Fidalgo, J., May, C., Moler, J. (2026). Optimal Design of Experiments for Minimum Outlet Size Estimation. Boletín de Estadística e Investigación Operativa. 42(2), 7-19. https://doi.org/10.63552/beio.2026.42.2.02

Abstract

Jam formation can occur when granular material is discharged by gravity from a silo. Estimating the minimum outlet size that ensures a sufficiently long time before the next jamming event can be crucial in industrial applications. The time between two jams is modeled by an exponential distribution with two unknown parameters. The main objective is to estimate the outlet diameter of the silo for a user-selected threshold for the expected time of a jam. This quantity to be estimated is a nonlinear transformation of the parameters. Optimal design of experiments guarantee a desired precision in estimation reducing the cost of experimentation. We obtain c-optimum experimental designs with that purpose, applying the graphical Elfving method. Since the optimal experimental designs depend on the nominal values of the parameters, we conduct a sensitivity analysis to detect changes in the efficiency. Finally, a simulation study assesses the performance of two approximations used in the method: one based on the Fisher information matrix and the other on linearization of the target function. The results support laboratory experimentation and their transfer to real-world settings. Building on this application, we develop a general methodology for estimating a one-dimensional transformation of the parameters of a nonlinear model.

Keywords: bulk solid storage, jam formation, non-linear heteroscedastic model, optimal design of experiments

1 Introduction

The Spanish Society of Statistics, Operations Research and Data Science– BBVA Foundation awarded, in 2024, the paper "Designing experiments for estimating an appropriate outlet size for a silo type problem," by Jesús López-Fidalgo, José A. Moler and Caterina May, as the best applied contribution in the social area, innovation or knowledge transfer in the statistics field. The paper was originally published in 2023 in The Annals of Applied Statistics, López-Fidalgo, May, & Moler (2023a). The Editor of BEIO invited the authors to prepare a version of the awarded paper for a non-specialist audience. This work responds to that request. In this version, mathematical formulas and technical details have been reduced to improve accessibility while preserving rigor.

Silos and hoppers are usual containers in industry for storing a large variety of solids and liquids. In this work we focus on silos for granular bulk solids. Their main use is as a buffer between one transport activity or chemical process and another in many economic activities as power generation, steel making, quarrying plastics, food processing, mining, farming and agricultural industries, see e.g. Nedderman (1992). Therefore, materials stored are quite variable and the structural form of the silo depends greatly on several properties of the material, as size, shape, weight, cohesion, homogeneity, etc. Particle sizes range from fine powders of micron size to agricultural grains, pellets, minerals or crushed rocks.

Silos require careful structural construction to avoid failures. Many codes and standards have been published to help engineers in the construction process, but they have limitations, see J. Carson & Craig (2015). One of the main goals is to ensure reliable, steady and complete discharge of solid from the vessel, see Rotter (2001). Flow obstructions in the discharge under gravity are common problems in the operation of a silo because of the formation of a cylindrical pipe (rathole) around or a stable arch-shaped obstruction over the silo outlet opening (see Chapter 2 in J. W. Carson (2008)), causing a blockage or jam (see Fig. 1).

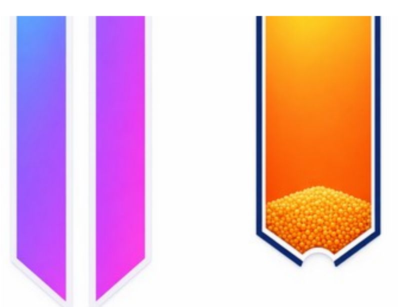

Fig. 1. Schemes of obstructions in a silo. Rathole (left), arch-shaped obstruction (right).

However, if the bulk solid properties are well known, reliable criteria for silo construction are established in the standards and obstruction problems are almost or completely eliminated with a right dimension of the outlet size called critical size, see Schulze (2008), Jenike (1961).

Silo failures carry important economic costs to the industry because of shut-down periods in production plants, damage or waste of the stored material or even the replacement of the silo. Reliable flow by standard silo construction is not always possible for several reasons. First, the lack of knowledge of critical properties in the nature of the stored material, as biomass particles or granular materials, see Miccio, Barletta, & Poletto (2013), Artoni, Santomaso, & Canu (2009). Second, the bulk solid nature can be modified by the storage period, moisture, temperature, aeration, silo degradation, climate conditions, etc, see, for instance, Mitra et al. (2017), Samuelsson, Larsson, Thyrel, & Lestander (2012), Chen & Roberts (2018) and references therein. Finally, silo outlet size can be smaller than the critical size because of improperly sized equipment, conservative estimations provided by theoretical models, etc, see Bell (2005), Fitzpatrick, Barringer, & Iqbal (2004). Under these situations, industry managers need information about the silo performance to minimize the chance of failures.

Direct experimentation can be very costly, risky and difficult to replicate because we are dealing with several tons of stored material, and waste of material or silo damage are too expensive for the companies. So that, computational simulation based on physical models of the dynamics of the stored material and statistical experimental designs are the two main solutions to deal with this problem, see Golsham, Esgandari, Zarghami, Blais, & Saleh (2020), Li, Langston, Webb, & Dyakowski (2004), Saleh, Golshan, & Zarghami (2018). Schulze (2008) remarks that the study of flow problems is similar when silos are studied on a smaller scale so that experimental design techniques are a practical way to provide information about silo reliability to industry managers. Indeed, the difficulties of modeling with granular materials make the statistical techniques to process experimental data as a relevant solution to deal with scale–up issues of granular bulk solids in industry, see Bell (2005).

As stated before, it is widely accepted that jam problems are closely related to the outlet size ϕ. Hence, the goal of this paper is to propose optimal experimental designs that provide precise estimation of the minimum outlet size necessary to guarantee that the expected time between two blockage events will exceed a fixed time of interest. This is valuable information to precisely schedule expert interventions for blockage interruption as silo cleaning, the use of vibrator systems of air cannons, etc.

The elapsed time between two jams is the starting point of this study. For analyzing the exit time distribution of a particle in a silo, a deterministic theoretical model based on the kinematic theory is developed and experimental studies are conducted to calibrate the model, see Nedderman (1992), Able, Othen, & Nedderman (1996), Golsham et al. (2020). This model depends on geometrical characteristics of the silo, density of the stored material in different silo zones and some kinematic constants, and the exit time decreases exponentially when the diameter of the outlet grows. However, unlike for the exit time, theoretical models for the time between jams do not exist, so ad-hoc models that fit experimental data are proposed instead, see To (2005), Janda, Zuriguel, Garcimartin, Pugnaloni, & Maza (2008). As expected, this function exponentially decreases as the diameter of the outlet grows and, also, several constants related to the characteristics of the silo and the granular bulk solid appear in the mathematical formulation. The exponential model chosen in this work agreed well with experimental data in To (2005) and also with the model proposed in Janda et al. (2008) whereas "avalanche" (amount of material dropped between jamming events) is observed.

For each silo geometry or bulk material, model constants change and must be estimated. An experimental design, in this context, consists of repeating n_i times a silo discharge fixing an outlet size ϕ_i, $i = 1, \cdots, r$, for r different outlet sizes. Optimal experimental designs are fundamental in this context because they can drastically lower the cost of experimentation by reducing the total number $n = \sum_{i=1}^{r} n_i$ of runs needed to guarantee a desired precision in parameter estimation. D-optimality is a popular criterion that can be useful when the target is the entire unknown parameter vector of a model. Amo-Salas, Delgado-Márquez, & López-Fidalgo (2016) and Amo-Salas, Delgado-Márquez, Filová, & López-Fidalgo (2016) obtained D-optimal experimental designs to estimate the parameters in several models for the time elapsed between jams. In particular, in Amo-Salas, Delgado-Márquez, Filová, & López-Fidalgo (2016), the D-optimal design to estimate the parameters of the model considered in this work was obtained.

When the target is estimating a linear combination of the unknown parameters of the model, a c-optimal design will provide the maximum likelihood estimator (MLE) of this combination with the minimal variance. Elfving (1952) showed an elegant graphical method to determine c-optimal experimental designs of a linear model on a compact experimental domain, based on the construction of a *convex hull*. This method is not easy to use for more than two parameters. López-Fidalgo & Rodríguez-Díaz (2004) provided an iterative procedure based on the graphical Elfving technique that successfully computes c-optimal experimental designs for more than two parameters.

The minimum outlet size which guarantees a minimum expected time between two blockages can be evaluated as a function of the model parameters. Because the model considers an exponential probability distribution, a local c-optimal design can be determined by using the Fisher Information Matrix (FIM) for nonlinear models. Then a first–order linear approximation of the function of the parameters around some nominal values of the parameters will be used.

Section 1.1 introduces the theory of optimal design of experiments, while Section 1.2 provides guidance on how to obtain a c-optimal design graphically. Section 1.3 explains in detail the motivating problem of the paper. Section 2 illustrates the mathematical method adopted to solve it, and provides the general results obtained. Section 3 presents some numerical examples, a sensitivity analysis, and the results are demonstrated on real data. We also provide insights into the two approximations by means of a simulation study. Section 4 concludes the paper with a discussion.

All computations have been done with Python 3.7. The dataset is real data coming from the Granular Means Group at the University of Navarre. The dataset and code are given in the supplementary material López-Fidalgo, May, & Moler (2023b).

1.1 Brief introduction to optimal experimental design theory

Optimal experimental design theory is very clean for linear models, but it can be extended to any non-linear model, including correlated observations with some limitations. Let $T = \theta^T f(\phi) + \varepsilon$ be a typical linear model with $\theta^T = (\theta_1, \ldots, \theta_m)$ the vector of unknown parameters to be estimated, $f^T(\phi) = (f_1(\phi), \ldots, f_m(\phi))$ the vector or linearly independent regressors and ϕ a non-random experimental condition chosen on a *design space*, χ, which is assumed compact, typically a cartesian product closed intervals or finite sets. The response T is assumed normal with mean $\theta^T f(\phi)$ and constant variance σ^2. A linear model assumes independent observations (experiments) as well.

Thus, let $Y = (T_1, \ldots, T_n)^T$ be uncorrelated measurements observed as results of some experimental conditions $\phi_1, \ldots, \phi_n$ (*exact design of size n*). The objective is to compute Maximum Likelihood Estimates (MLEs) of $\theta_1, \ldots, \theta_m, \sigma^2$ and make inferences with them. The matrix

$$X = \begin{pmatrix} f_1(\phi_1) & \cdots & f_m(\phi_1) \\ & \cdots & \\ f_1(\phi_n) & \cdots & f_m(\phi_n) \end{pmatrix},$$

with the typical notation $x_{ij} = f_j(\phi_i)$, $i = 1, \ldots, n$; $j = 1, \ldots, m$, is the so called *design matrix*.

Let $\mathscr{E} = (\varepsilon_1, \cdots, \varepsilon_n)^T$ be the actual errors. Thus, $Y = X\theta + \mathscr{E}$ is the matrix form of the model. The MLEs, which for a linear model are also the Least Squared Estimates (LSEs), are given by

$$\hat{\theta} = (X^T X)^{-1} X^T Y,\ \Sigma_{\hat{\theta}} = \sigma^2 (X^T X)^{-1}.$$

Some of the $\phi_1, \ldots, \phi_n$ may be repeated and a probability measure can be defined through the probability mass function (pmf): $\xi_n(\phi) = \frac{n_x}{n}$, where n_x is the number of times ϕ appears in the design. Then, the covariance matrix of the estimates can be written as

$$\Sigma_{\hat{\theta}} = \sigma^2 (X^T X)^{-1} = \frac{\sigma^2}{n} M^{-1}(\xi),$$

where $M(\xi) = \sum_{\phi \in \chi} f(\phi) f^T(\phi) \xi_n(\phi)$.

This suggests an extension of the concept of experimental design to any probability measure on χ, which is the called *Approximate design*. This idea is due to Kiefer & Wolfowitz (1959). Caratheodory's theorem jointly with a probability theory argument (see. e.g. López-Fidalgo, 2023, Appendix F.4 for details) states that the matrices associated to designs with less than $m(m+1)/2+1$ points are actually all possible information matrices. Thus, designs defined by their pmf are considered for computations and for real application,

$$\xi = \left\{ \begin{array}{cccc} \phi_1 & \phi_2 & \ldots & \phi_k \\ p_1 & p_2 & \ldots & p_k \end{array} \right\},$$

with $p_i > 0, \sum_i p_i = 1$. In practice, the practitioner should perform $n_i \approx np_i$ experiments at ϕ_i, where $n = \sum_i n_i$ is the total number of experiments. If a proportion p_i is not multiple of $1/n$, it must be approximated. An extension of the information matrix is defined as $M(\xi) = \sum_{\phi \in \chi} f(\phi) f^T(\phi) p_i$. It is straightforward to prove that it is symmetric, non-negative definite and it is singular if the number of different points in the design is less than m.

A *criterion function* is defined as $\Phi : \mathscr{M} \to \mathbb{R} \cup \{+\infty\}$ to be minimized. It is assumed that it is non-increasing in the Loewner sense (if $M \leq N$, that is $M - N$ is semi-definite positive, then $\Phi(M) \leq \Phi(N)$). It could be global or partial if the interest is on part of the parameters. Other properties are optional, but very convenient, such as convexity, differentiability or positive homogeneity: $\Phi(\delta M) = \frac{1}{\delta}\Phi(M), \quad \delta > 0$. A Φ*–optimal design* is a design ξ^* minimizing Φ.

There is a collection of possible criteria aiming different statistical properties. In this paper we will focus mainly in a linear combination of the parameters defined by a vector $c \in \mathbb{R}^m$. The criterion minimizing the variance of its estimator is called c-optimality,

$$\Phi_c[M(\xi)] \quad = \quad \left\{ \begin{array}{ll} c^T M^{-1}(\xi) c, & \text{if } c \text{ is in the image of } M(\xi), \\ \infty, & \text{otherwise.} \end{array} \right.$$

The theory is now extended to non-linear models in the widest sense. An general statistical parametric model can be defined as a family of distributions with pdf's (pmf's) $h(Y|\phi, \theta)$, where θ is in a parametric space $\Theta \subset \mathbb{R}^m$. This definition includes t-tests, ANOVA, any type of regression with correlated or uncorrelated observations and mixed models. There is not information matrix $X^T X$ now. Nevertheless, the inverse of the so called *Fisher Information Matrix (FIM)* approximates the covariance matrix of the MLEs. Let ξ be a general design and assume uncorrelated observations. The FIM is then

$$M(\xi, \theta) = \sum_{\phi \in \chi} \xi(\phi) I(\phi, \theta),$$

where

$$I(\phi, \theta) = \mathbf{E}_h \left[\left(\frac{\partial L(\theta)}{\partial \theta} \right) \left(\frac{\partial L(\theta)}{\partial \theta^T} \right) \right] = -\mathbf{E}_h \left[\left(\frac{\partial^2 L(\theta)}{\partial \theta^2} \right) \right],$$

and $L(\theta) = \log h(Y|\phi, \theta)$ the log-likelihood.

For a linear model the FIM is just the information matrix. The main issues with the FIM for non-linear models are that

1. The convergence guarantying the approximation is valid just for uncorrelated observations, which is the case considered in the paper.
2. It depends on the parameters, which need to be guessed apriori through some nominal values, a Bayesian approach or using sequential design.
3. The approximation needs enough sample size to be good.
4. Sometimes, the expectations cannot be computed analytically.

If there is a function $f(\phi, \theta)$ such that $I(\phi, \theta) = f(\phi, \theta) f^T(\phi, \theta)$ then the theory for linear models applies straightforward, if nominal values of the parameters are available. It may be proved that this happens for the exponential family, including the typical generalized linear models.

1.2 Elfving graphical procedure for c-optimality

Elfving (1952) gave an elegant procedure to compute c-optimal designs graphically. The set $f(\chi)$ is a compact region of the same dimension than χ in $\mathbb{R}^m$. If $\chi = [a,b]$ is a compact interval, then the region is a curve. If $\chi = [a_1,b_1]\times[a_2,b_2]$ then it is a surface and so on. The convex hull of the set $f(\chi)\cup -f(\chi)$ determines all the c-optimal designs. In particular, if the interest is in estimating $c_1\theta_1 + c_2\theta_2$ in a two parameter model, then the straight line trough the origin defined by the vector $c^T = (c_1, c_2)$ cuts the boundary of the convex hull at a point that defines the c-optimal design. In particular, this point has to be, by definition of the convex hull, a convex combination of points of $f(\chi)\cup -f(\chi)$, e.g $(x_c, y_c) = p_1 f(\phi_1) + p_2 f(\phi_2)$. Then the c-optimal design is

$$\xi = \left\{ \begin{array}{cc} \phi_1 & \phi_2 \\ p_1 & p_2 \end{array} \right\}.$$

It may cut directly the curve, $(x_c, y_c) = f(\phi_c)$ leading to a one-point design concentrated at ϕ_c.

Although this applies exactly for linear combinations of the parameters, frequently there is interest in estimating the maximum of the model curve, the point where this maximum is reached or the area under the curve, to mention a few possibilities. Very often this function, say $g(\theta)$ is nonlinear on the parameters. If $g(\theta)$ is a differentiable function of the parameters and $\hat{\theta}$ are the MLEs, the estimator $g(\hat{\theta})$ has an asymptotic variance

$$\frac{\partial g(\theta)}{\partial \theta^T} M^{-1}(\xi) \frac{\partial g(\theta)}{\partial \theta}.$$

Then for some nominal values of the parameters, θ_0, the optimal design can be approximated by the c–optimal design for $c = \left(\frac{\partial g(\theta)}{\partial \theta}\right)_{\theta=\theta_0}$.

Simpler cases are $c^T = (1,0)$ or $c^T = (0,1)$ if there is just interest in estimating θ_1 or θ_2 while the efficiency of estimating θ_2 or θ_1 does not matter at all. Actually, computing all the optimal designs for estimating each of the parameters give interesting references of how good a design is for estimating each of the parameters.

1.3 Outline of the problem

Consider the situation, discussed in the introduction, of solid material stored in a silo and its discharge due to the force of gravity through an outlet at the bottom of the container, in particular, the problem of the formation of blockages.

The time T when a first jam happens, or between two jamming events, is a random variable that depends on the diameter ϕ of the outlet at the bottom of the silo. To estimate the probability distribution of T, an experimenter may collect n observations $t_1,\ldots,t_n$ of the times between two jams for $r \leq n$ experimental conditions $\phi_1,\ldots,\phi_r$ chosen according to an experimental design. Experiments can be replicated for a particular diameter.

Due to the difficulties of direct experimentation with real silos, collecting data in a smaller scale experiment in a laboratory may be needed to obtain the desired information. Then, the choice of the possible diameters of the outlet to perform the experiments is of critical importance for the physicists. Zuriguel, Garcimartín, Maza, Pugnaloni, & Pastor (2005) made a thoughtful study of how the laboratory experiment can be used to replicate the real potential experiment in a similar phenomenology. In particular, they proved how the spheres in a 3D silo can be emulated by spheres in a 2D. Additionally they proved empirically that the experiment with regular and identical spheres replicates good enough other irregular and non-uniform shapes such as rice, lentils or stones. In particular, the increase in variability introduced by non-regular shapes in those cases is negligible.

In this paper we refer to the experimental study presented in Janda et al. (2008), and further studied in Amo-Salas, Delgado-Márquez, Filová, & López-Fidalgo (2016) and in Amo-Salas, Delgado-Márquez, & López-Fidalgo (2016).

A two-dimensional silo is reproduced in laboratory, consisting of two vertical glass plates between which spherical beads are poured. The beads flow constantly through an aperture at the bottom of the silo due to gravity, until an arch is formed at this outlet. An arch is indeed a structure where particles are mutually stable and hence it generates a jam. If one of the particles that form an arch is removed, the arch collapses and the flow restarts until the formation of a new jam. In this experimental setup the *avalanche size*, that is, the number of balls fallen between two jamming events, can be directly measured. As a matter of fact they just weigh the set of balls, which is equivalent and much easier. The physicists have proved that, for a given orifice diameter, the flow rate is constant, and then the time between two jamming events is proportional to the avalanche size. Moreover, Janda et al. (2008) notice that the same distribution of the avalanche size is found in the discharge of a three-dimensional silo (other studies can be seen in the references contained therein). An exponential distribution for T shows good agreement with experimental data, as showed in To (2005) (see also Section 3.3).

The extremes of the experimental domain $\mathscr{X} = [a,b]$ have to satisfy $d < a < b < \phi_C$, where d is the diameter of the granular material and ϕ_C is the critical size, which is a practical limit above which jamming is practically impossible. For some models this is a parameter to be estimated (see Amo-Salas, Delgado-Márquez, Filová, & López-Fidalgo, 2016 and Amo-Salas, Delgado-Márquez, & López-Fidalgo, 2016).

As previously specified, the focus of the present work is the estimate of the minimum outlet diameter necessary to guarantee a minimum expected time, T_0, between two jamming events:

$$E[T|\phi] = \eta(\phi;\theta) \geq T_0. \tag{1}$$

If $\eta(\cdot,\cdot)$ is a suitable expectation model invertible, eq. (1) implies

$$\phi \geq g(\theta;T_0), \tag{2}$$

hence, our statistical goal is to obtain a proper optimal design ξ^* to minimize the variance of the MLE of $g(\theta;T_0)$. This will provide a better precision of the resulting estimate, compared to any other non-optimal design based on $n = \sum_{i=1}^{r} n_i$ experiments.

To find an optimal design the standard methods for linear models are here adapted: the covariance matrix of the estimators is approximated by the inverse of the FIM, then the function of the parameters to be estimated is linearized. In this paper we evaluate the impact of these two approximations via simulations; see Section 5 (Conclusion).

2 Methodology

Assume, as in Janda et al. (2008), that the expected value of T given ϕ is

$$\eta(\phi;\theta) = \frac{1}{C}\exp(L\phi^2) - 1,\ \phi \in \mathscr{X} = [a,b], \tag{3}$$

where $\theta^T = (C,L)$. This is also the model considered in eq. (3) of Amo-Salas, Delgado-Márquez, Filová, & López-Fidalgo (2016). The time between jams, and therefore the mean η, is increasing with respect to ϕ. It has to be positive, so $L > 0$ and $0 < C < \exp(L\phi^2)$ for any ϕ.

Our main goal is the efficient estimation of the minimal diameter $\phi \in \mathscr{X}$ for which (1) holds. If the expectation model is given by (3), this means

$$\phi \geq g(\theta;T_0) = \sqrt{\frac{\log(C(T_0+1))}{L}}. \tag{4}$$

Our aim is then to obtain an optimal design to minimize the variance of the maximum-likelihood estimator of the bound $g(\theta;T_0)$ given in (4) based on n uncorrelated observations $\{t_1, t_2, \cdots, t_n\}$ of T:

$$t_i = \eta(\phi_i;\theta) + \varepsilon_i, \quad i = 1,...,n. \tag{5}$$

These observations follow an exponential distribution, therefore they have non-constant variance,

$$Var(\varepsilon_i) = \eta(\phi_i;\theta)^2.$$

When the inferential goal of an experiment is an efficient estimation of θ, an optimal design maximizes a suitable functional of the FIM. Let us denote the FIM of a model by $M(\xi,\theta)$. Note that $M(\xi,\theta)$ depends on the value of the unknown parameters except in the case of linear models; to overcome this problem, an optimal design can be computed on some nominal values of θ derived from guesses or previous knowledge of the possible likely values of the parameters.

When the target is to estimate a linear combination $\mathbf{c}^T\theta$ of the unknown parameters of the model, for some vector $\mathbf{c}$, a c-optimal design will provide the maximum likelihood estimator with the minimal variance for this combination. A c-optimal design ξ_c^*, minimizes the asymptotic variance of $\mathbf{c}^T\hat{\theta}$:

$$\xi_c^* = \arg\min_{\xi} \mathbf{c}^T M^{-1}(\xi;\theta)\mathbf{c}. \tag{6}$$

The target of this paper is the precise estimation of a non-linear function of the unknown parameters. Additionally, the model herein considered is non-linear and heteroscedastic. To solve this problem, we apply two linear approximations and we follow the Elfving procedure to determine a c-optimal design, as detailed in the next subsections.

2.1 Double approximation

Let $\mathscr{L}(\theta;t,\phi)$ be the log-likelihood function of T. For an exponential distribution model with mean $\eta(\phi,\theta)$, we have

$$\mathscr{L}(\theta;t,\phi) = \log\left(\frac{1}{\eta(\phi,\theta)}\exp - \frac{t}{\eta(\phi,\theta)}\right).$$

As $\theta^T = (C,L)$, the FIM is a 2×2 matrix that depends on C and L. There exists a linear model with the same FIM as (5):

$$t_i = \theta^T f(\phi_i;\theta_t) + \varepsilon_i, \tag{7}$$

with

$$f(\phi;\theta) = \frac{1}{\eta(\phi,\theta)}\nabla\eta(\phi,\theta), \tag{8}$$

where ∇ stands for the gradient. Eq. (8) is evaluated in the true value θ_t of the parameter vector. In this way, we have approximated model (5) in a neighborhood of the true value.

On the other hand, the MLE estimator of $g(\theta;T_0)$ is $g(\hat{\theta};T_0)$, so we approximate the non-linear function $g(\cdot;T_0)$ using a Taylor expansion around the true value θ_t, i.e., $g(\hat{\theta};T_0) \approx g(\theta_t;T_0) + \nabla g(\theta_t;T_0)(\hat{\theta} - \theta_t)$. The variance of $g(\hat{\theta};T_0)$ can then be approximated by

$$\nabla g(\theta_t;T_0)^T M(\xi,\theta_t)^{-1} \nabla g(\theta_t;T_0) \tag{9}$$

and a c-optimal design for model (5) is a design satisfying (6) with $\mathbf{c} = \mathbf{c}(\theta)$ given by

$$\mathbf{c}(\theta) = \nabla g(\theta;T_0). \tag{10}$$

Notice that two procedures of approximation have been used, and that the c-optimum design satisfying (6) depends on the unknown parameters both through the vector $\mathbf{c}(\theta)$ and the FIM, $M(\xi,\theta)$. Hence we choose an initial guess θ_0 for the true value θ_t and the design obtained will be *locally optimum*, that is, it will depend on this choice.

2.2 Graphical Elfving procedure

A convenient way to compute c-optimal experimental designs, especially in two dimensions, is the graphical Elfving procedure (first proposed by Elfving, 1952). This procedure allows us to estimate a linear transformation $\mathbf{c}^T\theta$ of the unknown parameters of a linear homoscedastic model $t = \theta^T f(\phi) + \varepsilon$, like the model (7), where $f(\phi) = (f_1(\phi), f_2(\phi))^T$.

To apply the Elfving's graphical method, the first step is to obtain and plot in the Cartesian plane the *Elfving locus*, that is, the convex hull of the union of the curve defined by the regressors $f(\phi) = (f_1(\phi), f_2(\phi))$ and its reflection through the origin. Figure 2 shows the case considered in this paper. The boundary of the Elfving locus in this case will be denoted by $\overline{A_1A_2A_3A_4}$. Then, the second step is to plot the line defined by the vector $\mathbf{c}$ through the origin until the boundary $\overline{A_1A_2A_3A_4}$ is reached. The c-optimal design is then defined by the crossing point. The extremes of the segment of the boundary of the Elfving locus crossed by the line defined by c are the support points. The crossing point is then a convex combination of the extreme points giving the design weights for each point. For more details on the Elfving method in the continuous case here considered, see Section 1.1 of Rivas-López, López-Fidalgo, & del Campo (2014) or the appendix in Flournoy, Moler, & Sada (2025).

The information matrix at a point ϕ for model (3) is

$$I(\phi,\theta) = \frac{e^{2\phi^2 L}}{C(e^{\phi^2 L} - C)^2} \begin{pmatrix} \frac{1}{C} & -\phi^2 \\ -\phi^2 & C\phi^4 \end{pmatrix}. \tag{11}$$

Hence,

$$f(\phi,\theta) = (f_1(\phi), f_2(\phi))^T = G(\phi,\theta)\,(1/C, -\phi^2)^T,$$

with

$$G(\phi,\theta) = \frac{e^{\phi^2 L}}{e^{\phi^2 L} - C}. \tag{12}$$

Finally, the parametric equations defining the curve $f([a,b])$ are

$$\begin{cases} x(\phi) = G(\phi,\theta)/C, \\ y(\phi) = -G(\phi,\theta)\,\phi^2, \\ \phi \in [a,b]. \end{cases} \tag{13}$$

For the second step, let $\mathbf{c}(\theta)$, as defined in (10), be the gradient of $g(\theta;T_0)$, that is,

$$\mathbf{c}(\theta) = \frac{1}{2\sqrt{L}} \left(\frac{1}{C\sqrt{\log(C(T_0+1))}}, -\frac{\sqrt{\log(C(T_0+1))}}{L} \right)^T; \tag{14}$$

hence, $\mathbf{c}$ will be given by (14) evaluated at some nominal values (C_0, L_0). Depending on the value of T_0, $\mathbf{c}$ will have a different angle and the line defined by $\mathbf{c}$ will cross the convex hull in $\overline{A_1A_2}$, $\overline{A_2A_3}$ or $\overline{A_3A_4}$ (see Figure 2).

The next propositions provide some general results to obtain c-optimal designs for estimating the bound (4). The proofs are detailed in Section 1 of the Supplementary Material López-Fidalgo et al. (2023b).

Proposition 2.1 *Consider the curve (13) and its reflection through the origin. Let A_1, A_2, A_3, A_4 be the outermost points: $A_1 = (-x(b), -y(b))$, $A_2 = (x(a), y(a))$, $A_3 = (x(b), y(b))$, $A_4 = (-x(a), -y(a))$. Then, for any value of (C_0, L_0) such that $0 < C_0 < \exp(L_0\phi^2)$ and $L_0 > 0$, the boundary of the convex hull of these curves is given by the polygonal $\overline{A_1A_2A_3A_4}$.*

The next proposition gives the c-optimal experimental designs obtained by the crossing point between the line defined by $\mathbf{c}$ and the boundary of the convex hull, according to the Elfving method; this depends on the fixed value of T_0 and on the choice of (C_0, L_0). There are actually three possible cases, but it is always a two-point design in the extremes a and b.

Denote by (x_i, y_i), $i = 1, \ldots, 4$, the coordinates of the extremes A_i of the convex hull (2) and by ϕ_i the corresponding values of ϕ in curve (13) or in its symmetric, see Figure 2. Then the following result can be proved.

Proposition 2.2 *When the crossing point P_0 is in $\overline{A_iA_{i+1}}$, $i = 1, 2, 3$, the c-optimal design is*

$$\left\{ \begin{array}{cc} \phi_i & \phi_{i+1} \\ 1-p_i & p_i \end{array} \right\}, \qquad \text{with } p_i = \sqrt{\frac{(Kx_0 - y_i)^2 + (x_0 - x_i)^2}{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}}, \tag{15}$$

where (x_0, y_0) are the coordinates of P_0 given by

$$x_0 = \frac{y_i - \dfrac{y_{i+1} - y_i}{x_{i+1} - x_i} x_i}{K - \dfrac{y_{i+1} - y_i}{x_{i+1} - x_i}} \quad \text{and} \quad y_0 = Kx_0,$$

and

$$K = \frac{\partial g(\theta; T_0)/\partial L|_{\theta_1 = C_0, \theta_2 = L_0}}{\partial g(\theta; T_0)/\partial C|_{\theta_1 = C_0, \theta_2 = L_0}}.$$

3 Application

The theoretical results and techniques provided in previous sections are now applied to particular cases. Because the choice of the nominal values is crucial in the method considered, we conduct a sensitivity analysis. Finally, we show an application to a real dataset provided by the Granular Media Laboratory of the University of Navarra (Gella, Zuriguel, & Maza, 2018).

3.1 Illustrative examples

To illustrate the method and the theoretical results, the experimental case in Janda et al. (2008) is considered, and the estimates obtained in Amo-Salas, Delgado-Márquez, Filová, & López-Fidalgo (2016) from data are used as nominal values of the parameters.

Example 1 *Let $\phi \in \mathscr{X} = [1.53, 5.63]$, $C_0 = 0.671741$ and $L_0 = 0.373098$. Figure 2 represents the parametric curve (13), its reflection, and the Elfving locus $\overline{A_1A_2A_3A_4}$ obtained in this case. It is worth observing that the vertexes of the convex hull in Figure 2 are not tangential points of the curve but outermost points of the curve.*

The crossing point P_0 is in a different side of $\overline{A_1A_2A_3A_4}$ depending on whether T_0 is either in the interval $(0.49, 2.57]$, in $(2.57, 203603.03]$ or in $(203603.03, \infty)$ (see Lemma 1.2 of the Supplementary Material López-Fidalgo et al. (2023b) for details). Figure 2 illustrates the three cases $T_0 = 1.5$, $T_0 = 1200$ and $T_0 = 900{,}000$; P_0 falls, respectively, in $\overline{A_1A_2}$, $\overline{A_2A_3}$ and $\overline{A_3A_4}$. Observe that the vectors from the origin to P_0 are divided into two parts. In each case, the standard deviation of $g(\hat{\theta}; T_0)$ is equal to the ratio of the dashed vector to the whole vector.

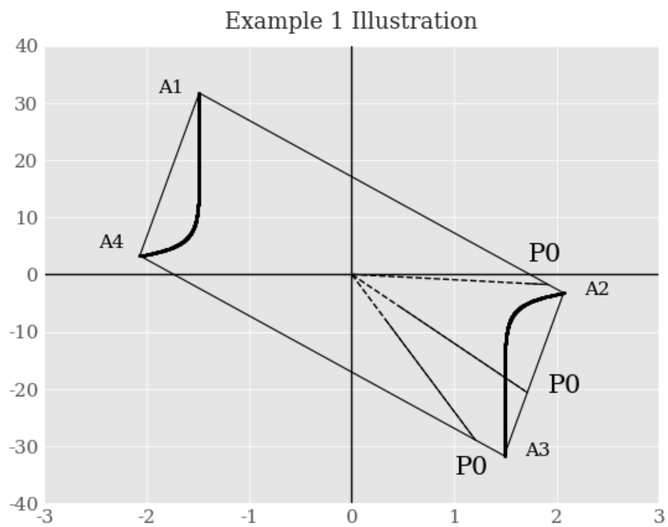

Fig. 2. Graphical representation for the framework in Example 1.

According to Proposition 2.2, the optimal design in each case is

$$\left\{ \begin{array}{cc} 1.53 & 5.63 \\ 0.9582 & 0.0418 \end{array} \right\}_{T_0=1.5}; \quad \left\{ \begin{array}{cc} 1.53 & 5.63 \\ 0.6118 & 0.3882 \end{array} \right\}_{T_0=1200};$$

$$\left\{ \begin{array}{cc} 1.53 & 5.63 \\ 0.0792 & 0.9208 \end{array} \right\}_{T_0=900{,}000}.$$

Note that the closer the crossing point is to a vertex, the more unbalanced is the optimal design.

Example 2 *As showed in the proof of Proposition 2.1 (point 5), there are two possible cases for $\overline{A_1A_2A_3A_4}$, that is, when the point A_2 is a maximum of the curve and when it is not. The first one has been illustrated in Example 1 and the second one is illustrated in this example. If $L_0 = 0.373098$ and $\phi \in [1.53, 5.63]$, a value of C_0 in the interval $(1.2784, 2.395)$ must be chosen, say $C_0 = 2$. Assume $T_0 = 1.5$.*
Figure 3 represents the convex hull and **c** *in this example, when*

A_2 is not a maximum. Now, Proposition 2.2 holds, and then

$$\xi_c^* = \left\{ \begin{array}{cc} 1.53 & 5.63 \\ 0.3040 & 0.6960 \end{array} \right\}.$$

It is interesting to stress that this design puts more weight in the right extreme, and therefore longer experimentation times are required, although the limit T_0 is much smaller than in the previous example.

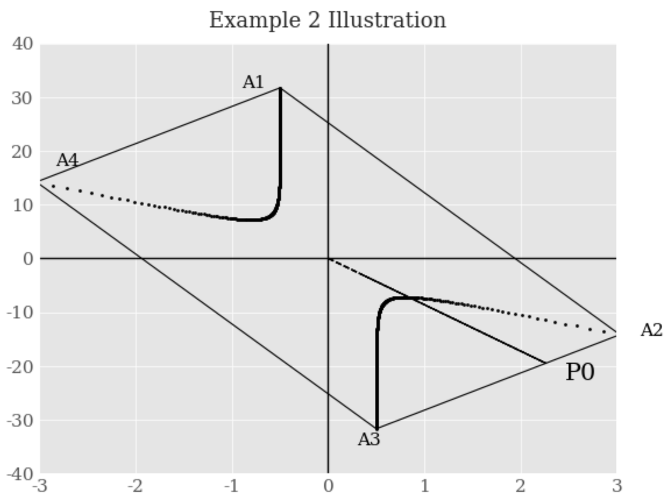

Fig. 3. Graphical representation for the framework in Example 2.

3.2 Sensitivity analysis

Since the optimal experimental designs depend on the nominal values of the parameters, we conduct a sensitivity analysis to asses how the performance of the c-optimal design (e.g., its efficiency) changes when the nominal values of the unknown parameters are varied. Assume that T_0 is a given value; the efficiency reduces as the true parameters (C^*, L^*) are far from the nominal values (C_0, L_0) chosen as a guess. The theoretical details of how the sensitivity analysis is actually performed can be found in Section 2 of the Supplementary Material López-Fidalgo et al. (2023b).

Example 3 *Consider the framework of Example 1, where T_0 can take values $2; 200$ or $300,000$. The first one is too small to have a practical interest, while the last one needs a diameter longer than those ones in the design space. Thus, they are extreme cases that will not be considered in this paper. Instead, we focus on the case $T_0 = 1200$ which covers more usual situations.*

The following steps describe the procedure to perform a sensitivity analysis for the choice of the nominal values of the parameters.

***Step 1**: Consider $\phi \in [a,b]$ and the nominal values (C_0, L_0). The c-optimal design is obtained from Proposition 3.2. The weights and diameters are identified with the superscript (0).*

***Step 2**: We consider a grid where the parameters C and L take potential actual values (C^*, L^*) in a neighborhood of the nominal values (C_0, L_0). Thus, we obtain the c-optimal design, as in Step 1, but when the true values of the parameters is a pair (C^*, L^*) in the grid. The weights and diameters are identified with the superscript $*$*

***Step 3**: For each (C^*, L^*) in the grid, we assume them as the real values and calculate the variance, $Var_1(g)$, with the expression (9):*

$$M_1 = (1-p_i^*) I(\phi_i^*, C^*, L^*) + p_i^* I(\phi_{i+1}^*, C^*, L^*),$$
$$Var_1(g) = \nabla g(C^*, L^*; T_0)^T M_1^{-1} \nabla g(C^*, L^*; T_0).$$

• *Consider the nominal values (C_0, L_0), where $p_i^{(0)}$ and $\phi_i^{(0)}$ were obtained in Step 1, then, we calculate the expression $Var_0(g)$, using the actual values for computing its FIM,*

$$M_0 = (1-p_i^{(0)}) I(\phi_i^{(0)}, C^*, L^*) + p_i^{(0)} I(\phi_{i+1}^{(0)}, C^*, L^*),$$
$$Var_0(g) = \nabla g(C^*, L^*; T_0)^T M_0^{-1} \nabla g(C^*, L^*; T_0).$$

• *Compute, for each (C^*, L^*), the relative efficiency given by $Var_1(g)/Var_0(g)$, which indicates the rate of change in variance when using the diameters and weights from Step 1 instead of the optimal diameters and weights obtained in Step 2.*

Observe that the relative efficiency measures the inaccuracy resulting from using nominal values in a practical problem when the right values are (C^*, L^*) .

In Table 1, the efficiencies for $T_0 = 1200$ are shown for a grid of points (C^*, L^*), where C^* varies in the interval $(C_0 - 0.3, C_0 + 0.3)$ while L^* varies in the interval $(L_0 - 0.15, L_0 + 0.1)$. Because the value of T_0 is fixed, the crossing point can be in a different segment $\overline{A_i A_{i+1}}$ for (C_0, L_0) and the point of the grid (C^*, L^*). For instance, in Table 1, the largest reduction of the efficiency happens for large values of C^* combined with small values of L^*. We could say that, when both L^* and C^* grow or decrease, the efficiency is more stable; instead, changes of C^* and L^* in opposite directions make the efficiency to reduce faster.

L^* vs C^*	0.37	0.47	0.57	0.67	0.77	0.87	0.97
0.22	0.767	0.799	0.826	0.851	0.874	0.895	0.914
0.27	0.829	0.863	0.891	0.916	0.936	0.954	0.969
0.32	0.898	0.930	0.955	0.973	0.986	0.995	0.999
0.37	0.965	0.986	0.997	1.000	0.997	0.989	0.977
0.42	0.997	0.995	0.981	0.961	0.939	0.914	0.888
0.47	0.947	0.909	0.872	0.837	0.805	0.777	0.751

Tab. 1. Efficiency values in each point of the grid for $T_0 = 1200$.

3.3 Real-data application

As stated in the introduction, time between jams has been first studied experimentally and, then, a statistical model was proposed to fit the data. The basis of the experimentation in this topic was established in the seminal paper by To (2005). The experimentation procedure requires a warm-up period after each jam to reach a steady flow of the particles. This introduces a nuisance parameter in the fitting model. In Janda et al. (2008), the experimentation procedure was modified fitting the description in Section 1.3. The warming up period is avoided by introducing a large amount of balls in the silo at the beginning of the experimentation and the device is refilled from time to time to maintain the pressure at the bottom of the hopper. When a jam happens, the time since the previous jam is recorded. Moreover, the jam is broken with a controlled jet of pressurized air to avoid relevant changes in the internal relationships among particles which makes realistic the assumption of independence between time observations.

Let X be the random variable representing the size of an "avalanche", that is, the number of disks that pass through the outlet between two jams. In To (2005) the system is modelled by a Markov Chain and each ball that passes through the outlet is a transition. The warming-up period requires n_0 balls, after this, the system reaches the steady flow state in which each ball can remain with probability q or generate a jam, which is an absorbing state, with probability $1-q$. Based on this model, they confirm with the experimentation a good agreement of an exponential decay for the tail distribution $G(n) = P(X \geq n)$, for several outlet diameter values ϕ. Consistent results are obtained in Janda et al. (2008) avoiding the nuisance parameter n_0, so that X follows a Geometric distribution with success probability $1-q$. From Brown (1990), an exponential distribution with the same mean, $E[X] = q/(1-q)$, can be approximated to the distribution of the time between jams when $1-q$ is not large. Following Janda et al. (2008), $1-q = Ce^{-L\phi^2}$ and the model (3) is obtained.

Members of the University of Navarra's research team, who also authored Janda et al. (2008), kindly provided us with the micro-data of a similar experiment to the one just described (Gella et al., 2018). In this case, the silo is filled with spherical stainless steel beads with diameter 4.00 mm. The experiment is repeated for 12 different values of ϕ ($\{9.3, 9.7, 10.5, 11.2, 11.8, 12.4, 13, 13.6, 14.2, 14.9, 15.3, 15.4\}$); for each value, a different number n of observations of the time T between jams are obtained and the average time $\overline{T}$ is calculated, see Table 2. To find the best probability distribution function that fits the experimental data, the procedure in Rigby, Stasinopoulos, Heller, & De Bastiani (2019) is followed here. It is based on the GAIC (the Generalized Akaike Information Criterion) for generalized additive models and it has been implemented by *gamlss* R package. For each diameter ϕ considered in the experimentation, the exponential distribution results to be the best fit for the data.

ϕ	9.3	9.7	10.5	$(\cdots)$	14.9	15.3	15.4
n_i	1510	2005	2139	$(\cdots)$	2026	1009	1001
$\overline{T}_j$	19.04	18.38	33.37	$(\cdots)$	1212.22	3109.30	4611.61

Tab. 2. Setting for the most extremal design points of the real experiment .

In what follows we compare the performance of the design used in practice to collect these data and the optimal design obtained in Proposition 2.2, in terms of precise estimation. Let λ_j be the parameter of the exponential distribution associated with each outlet size ϕ_j; applying the maximum likelihood principle to our data, the MLE $\hat{\lambda}_j$ is

$$\hat{\lambda}_j = \frac{1}{\overline{T}_j}, \qquad j = 1, \cdots, r, \tag{16}$$

where r is the number of outlet sizes used in the experimentation; $r = 12$ in this example. Recalling that we assume model (3), when $r = 2$ explicit expressions for the MLEs of C and L can be then obtained. When $r > 2$ numerical optimization procedures must be applied to maximize the likelihood function. Observe that the case $r = 2$ is exactly the case of a D-optimal design, see Amo-Salas, Delgado-Márquez, Filová, & López-Fidalgo (2016), and also of the c-optimal design obtained in Proposition 2.2. In both cases, the two design points are the two bounds of the design space but the D-optimal is balanced whereas the c-optimal weights depend on the initial conditions.

Because MLEs are asymptotically unbiased, variance is a good measure of the precision when there is a large number of experiments. In small samples, the MLE may incur some non-negligible bias, which could be estimated via the jackknife or bootstrap and be adjusted accordingly. To guarantee the application of the c-optimality, a large sample size of 1,000 was chosen because this is the number of observations in the upper bound 15.4 of the dataset considered.

First we calculate the MLE when $r = 12$, where the entire set of experimental conditions is required. For each diameter ϕ_j, $j = 1, \cdots, 12$, 84 values are randomly chosen so that, $n \sim 1,000$. By maximizing the ML function, we obtain $\tilde{C} = 1.1362$, $\tilde{L} = 0.0329$, from which the estimate $\tilde{g}$ of the bound $g = g(\theta; T_0)$ in (4) is calculated. We take T_0 equal to the round values of $\overline{T}_j$, $j = 1, \cdots, 12$ that appear in Table 2, so that, the corresponding g value is close to ϕ_j. This procedure is repeated $1,000$ times and then an estimation of the standard deviation $\hat{\sigma}_{\tilde{g}}$ of $\tilde{g}$ is also obtained.

To do a comparative study with $r = 2$ according to the c-optimality, let $C_0 = 1.1362$ and $L_0 = 0.0329$ and we compute from Proposition 2.2 the proportions of observations to take in the two bounds (9.3 and 15.4). Hence we obtain the corresponding estimate $\hat{g}$ and the exact variance $\sigma_{\hat{g}}$ of the MLE. The whole set of results are in Table 3; the proportion $p_{9.3}$ of experiments in 9.3 in the c-optimal design is also indicated.

ϕ	9.3	9.7	10.5	$(\cdots)$	14.9	15.3	15.4
T_0	19	18	33	$(\cdots)$	1212	3109	4611
$p_{9.3}$	0.999	0.991	0.89	$(\cdots)$	0.237	0.070	0.001
$\tilde{g}$	9.35	9.3	10.1	$(\cdots)$	14.2	15.1	15.4
$\hat{\sigma}_{\tilde{g}}$	0.39	0.40	0.28	$(\cdots)$	0.18	0.26	0.29
$\hat{g}$	9.4	9.2	10.0	$(\cdots)$	14.2	15.1	15.4
$\sigma_{\hat{g}}$	0.06	0.06	0.05	$(\cdots)$	0.04	0.05	0.05

Tab. 3. Results of the comparative study between the c-optimal and the MLE with the most extremal design points

It is worth highlighting several points from the results presented in Table 3. For c-optimality two–point designs are obtained with closed–form expressions for the MLEs. Moreover, the variability decreases dramatically. Finally, from the sensibility study presented in Section 3.2, the c-optimal design could benefit from better guesses for the initial values C_0 and L_0 reducing bias and variability.

4 Conclusions

In this work we consider the problem of estimating the parameters of a non-linear model for the time elapsed between two jams in the emptying of a silo. This may be applied to a number of phenomena such as delivering some material on a mine through a vertical tunnel. In most of the cases a jam might be rather dramatic involving some expensive procedures to break the jam. In the case of the mine some explosive has to be used including costs, risks and delays. Then, determining the diameter of the outlet, say ϕ, to guarantee a period of time long enough is of great interest. This could be considered as a specific expected time, say T_0, or else a specific probability of reaching a target time without jams. Either "expected time" (T_0) or probability of "success" (reach target time jam-free) entails the estimation of a lower bound expressed as a non-linear function that depends on the unknown parameters. To obtain an analytical solution of the problem, first we use the classical Fisher Information approximation for the covariance matrix of the estimates of the parameters. Then the non-linear lower bound, which is the target for estimation, is linearized in such a way its gradient will play the role of the c-vector for c-optimality. A model with two parameters is chosen, and the Elfving graphic procedure to find the c-optimal design is used.

Propositions 2.1 and 2.2 establish, respectively, the main characteristics of the convex hull depending on the parameter values and then an explicit expression for the c-optimal design can be provided in all cases. Actually, the latter indicates that the c-vector may intersect the convex hull in three possible sides of the convex hull depending on three intervals where T_0 can lie. The vertices produce one–point c-optimal experimental designs, otherwise two points are needed. Thus, the optimal experiment involves only two outlet diameters, which is very convenient in the laboratory.

The vertices of the convex hull are critical points in the sensitivity analysis because they indicate a change in the type of design. For this analysis a uniform grid with values for the parameters around the nominal values was considered to detect sensitive changes in the efficiency. In particular, a dramatic loss of efficiency happens when the parameter values considered in the grid produce a edge change for the crossing point of the c-vector. A smaller decrease is observed for changes in the crossing point without changing the edge of the convex hull. Both facts imply a very important change in weights of the c-optimal design in Proposition 2.2. Also, for very large values of T_0, the sensitivity of the design with respect to the selection of the nominal values is large. For instance, a small change in one of the parameters produces a dramatic decrease in the efficiency, so sensitivity analysis requires a reduced scale on this parameter. This analysis and interpretation is important for a safe choice of the nominal values of the parameters.

Finally, a simulation study was conducted to verify the adequacy of the approximations (7) and (9) (cf. Section 2.1). The a priori approximated variances and covariances of the estimates are compared with the empirical variances and covariances of the estimates obtained by simulation. Given the original non-linear model the MLEs are obtained by simulating a large number n of observations allocated in the c-optimal design. A value of T_0 is chosen jointly with some nominal values from the literature. Results are generally very close. Nevertheless, the approximation procedure produces slightly higher variances of the lower bound for the silo outlet size than the simulated one. When T_0 is close to its lower bound, the convergence is slower and n must be enlarged. Details are provided in Section 3 of the Supplementary Material López-Fidalgo et al. (2023b).

Acknowledgements

The authors would like to thank the Granular Media Laboratory of the University of Navarra, in particular professors Gella, Zuriguel, Maza, Janda and Kiwing To for providing real data from their experiments.

The first and third author were sponsored by the Spanish State Research Agency (PID2024-156302OB-I00).

About the authors

Jesús López-Fidalgo is Director of the Institute of Data Science and AI, and Full Professor of Statistics and Operations Research at the University of Navarra. He has held positions at the universities of Salamanca and Castilla-La Mancha, where he served as Director of the School of Industrial Engineering for 8 years. He was in charge of the National Mathematics Plan at the Ministry of Science. He has served as President of the Spanish Society of Statistics and Operations Research and as Editor-in-Chief of the journal Test. He has carried out research stays at the universities of Manchester (United Kingdom), Glasgow, California, Los Angeles (UCLA), and Alberta (Canada). He taught for one term at the University of California, Riverside. He has international collaborators, reflected in the co-authorship of many of his more than 120 publications. He has authored or edited 6 books, two of them with popular science titles, “El azar no existe” (“Chance Does Not Exist”) and “IA: Desafíos y Oportunidades” (“AI: Challenges and Opportunities”). He has supervised fifteen doctoral theses. He was co-founder and coordinator of a node of the National Biostatistics Network.

Caterina May has been an Associate Professor of Statistics at at the Department of Economics and Business Studies (DISEI) of the University of Eastern Piedmont since 2018, where she previously held positions as a Postdoctoral Research Fellow and Assistant Professor. She obtained her Ph.D. in Mathematics from the University of Milan, specializing in Probability and Statistics, under the supervision of Piercesare Secchi. Her research has focused on stochastic processes, functional data analysis, and optimal design of experiments. She has been a visiting scholar at the University of Missouri–Columbia under the supervision of Nancy Flournoy, and at the Public University of Navarra, where she collaborated with José A. Moler and Jesús López-Fidalgo. In Italy, she has also had the opportunity to collaborate extensively with Giacomo Aletti and Chiara Tommasi. In addition, she has served as a Research Associate at King’s College London.

José A. Moler Cuiral is a PhD in Mathematics from the Public University of Navarra (UPNA). He is an associate professor in the Department of Statistics, Computer Science and Mathematics at the UPNA since the 1992-93 academic year. He was head of the department and Director of the Faculty Area, in the Vice- Rectorate for Faculty.
His research career has always been linked to the study of stochastic processes and has evolved towards issues of statistical inference and design of experiments with applications in the fields of computing and medicine. She has furthered the methodological study of response-adaptive designs for clinical trials with a research stay at Columbia University- Missouri. He is member of several interdisciplinary research teams where probabilistic and statistical models and data analysis tools are preeminent.

References

Able, R., Othen, S., & Nedderman, R. (1996). The exit time distribution during the batch discharge of a cylindrical bunker. *Chemical Engineering Science*, *51*(20), 4605-4610.

Amo-Salas, M., Delgado-Márquez, E., Filová, L., & López-Fidalgo, J. (2016). Optimal designs for model discrimination and fitting for the flow of particles. *Statist. Papers*, *57*(4), 875–891.

Amo-Salas, M., Delgado-Márquez, E., & López-Fidalgo, J. (2016). Optimal experimental designs in the flow rate of particles. *Technometrics*, *58*(2), 269–276.

Artoni, R., Santomaso, A., & Canu, P. (2009). Simulation of dense granular flows: dynamics of wall stress silos. *Chemical Engineering Science*, *64*, 4040–4050.

Bell, T. (2005). Challenges in the scale-up of particulate processes - an industrial perspective. *Powder Technology*, *150*, 60-71.

Brown, M. (1990). Error bounds for exponential approximations of geometric convolutions. *The Annals of Probability*, *18, 3*, 1388-1402.

Carson, J., & Craig, D. (2015). Silo design codes: their limits and inconsistencies. *Procedia Engineering*, *102*, 647-656.

Carson, J. W. (2008). *In bulk solids handling. equipment selection and operation, mcglinchey, d. (ed).* Oxford, U.K.: Blackwell Publishing.

Chen, W., & Roberts, A. (2018). A modified flowability classification model for moist and cohesive bulk solids. *Powder Technology*, *325*, 639-650.

Elfving, G. (1952). Optimum allocation in linear regression theory. *Ann. Math. Stat.*, *23*, 255-262.

Fitzpatrick, J., Barringer, S., & Iqbal, T. (2004). Flow property measurement of food powders and sensitivity of jenike's hopper design methodology to the measured values. *Journal of Food Engineering*, *61*, 399-405.

Flournoy, N., Moler, J., & Sada, M. (2025). Influence of the slope in percentile estimation through binary regression for dose-finding experiments. *Statistics*, *59*(2), 305–329.

Gella, D., Zuriguel, I., & Maza, D. (2018). Decoupling geometrical and kinematic contributions to the silo clogging process. *Physical Review Letters*, *121*, 138001.

Golsham, S., Esgandari, B., Zarghami, R., Blais, B., & Saleh, K. (2020). Experimental and dem studies of velocitiy profiles and residence time distribution of non-spherical particles in silos. *Powder Technology*, *373*, 510-521.

Janda, A., Zuriguel, I., Garcimartin, A., Pugnaloni, L. A., & Maza, D. (2008). Jamming and critical outlet size in the discharge of a two-dimensional silo. *Europhysics Letters*, *23*, 255-262.

Jenike, A. W. (1961). *Gravity flow of bulk solids*. Bulletin of the University of Utah.

Kiefer, J., & Wolfowitz, J. (1959). Optimum designs in regression problems. *The Annals of Mathematical Statistics*, *30*, 271-294.

Li, J., Langston, P., Webb, C., & Dyakowski, T. (2004). Flow of sphero-disc particles in rectangular hoppers-a dem and experimental comparison in 3d). *Chemical Engineering Science*, *59*, 5917-5929.

López-Fidalgo, J., May, C., & Moler, J. A. (2023a). Designing experiments for estimating an appropriate outlet size for a silo type problem. *Annals of Applied Statistics*, *17*(1), 606-620. doi: 10.1214/22-AOAS1644

López-Fidalgo, J., May, C., & Moler, J. A. (2023b). Supplement to "designing experiments for estimating an appropriate outlet size for a silo type problem".

López-Fidalgo, J., & Rodríguez-Díaz, J. M. (2004). Elfving's method for*m*-dimensional models. *Metrika*, *59*(3), 235–244.

López-Fidalgo, J. (2023). *Optimal experimental design: A concise introduction for researchers*. Heidelberg: Springer.

Miccio, F., Barletta, D., & Poletto, M. (2013). Flow properties and arching behavior of biomass particulate solids. *Powder Technology*, *235*, 312-321.

Mitra, H., Pushpadass, H. A., Eljeeva, M., Franklin, E., Kingsby, R. P., Ghoroi, C., … Battula, S. (2017). Influence of moisture content on the flow properties of basundi mix. *Powder Technology*, *312*, 133-143.

Nedderman, R. (1992). *Statics and kinematics of granular materials*. Cambridge: Cambridge University Press.

Rigby, R., Stasinopoulos, M., Heller, G., & De Bastiani, F. (2019). *Distributions for modeling location, scale, and shape: Using gamlss in r*. Chapman and Hall/CRC. ((1st ed.)) doi: 10.1201/9780429298547

Rivas-López, M. J., López-Fidalgo, J., & del Campo, R. (2014). Optimal experimental design for accelerated failure time with TypeI and random censoring. *Biometrical Journal*, *56*(5), 819–837. Retrieved from https://doi.org/10.1002/bimj.201300209 doi: 10.1002/bimj.201300209

Rotter, J. (2001). *Guide for the economic design of circular metal silos.* London, New York: Taylor & Francis Routledge.

Saleh, K., Golshan, S., & Zarghami, R. (2018). A review on gravity floe of free-flowing granular solids in silos - basics and practical aspects. *Chemical Engineering Science*, *192*, 1011-1035.

Samuelsson, R., Larsson, S., Thyrel, M., & Lestander, T. A. (2012). Moisture content and storage time influence the binding mechanisms in biofuel wood pellets. *Applied Energy*, *99*, 109-115.

Schulze, D. (2008). *Powders and bulk solids.* Berlin: Springer-Verlag.

To, K. (2005). Jamming transition in two-dimensional hoppers and silos. *Physical Review E*, *71*, 060301.

Zuriguel, I., Garcimartín, A., Maza, D., Pugnaloni, L., & Pastor, J. (2005). Jamming during the discharge of granular matter from a silo. *Physical Review E*, *71 (5)*, 051303.