The Annals of Applied Statistics
2023, Vol. 17, No. 1, 606–620
https://doi.org/10.1214/22-AOAS1644

DESIGNING EXPERIMENTS FOR ESTIMATING AN APPROPRIATE OUTLET SIZE FOR A SILO TYPE PROBLEM

BY JESUS LOPEZ-FIDALGO[1,a], CATERINA MAY[2,b] AND JOSE ANTONIO MOLER[3,c]

[1]*Instituto de Ciencia de los Datos e Inteligencia Artificial, Universidad de Navarra,* [a]*fidalgo@unav.es*
[2]*Dipartimento DiSEI, Università degli Studi del Piemonte Orientale,* [b]*caterina.may@uniupo.it*
[3]*Departamento de Estadística, Informática y Matemáticas, Universidad Pública de Navarra,* [c]*jmoler@unavarra.es*

Jam formation is a problem that may occur when granular material is discharged by gravity from a silo. The estimation of the minimum outlet size, which guarantees that the time to the next jamming event is long enough, can be crucial in the industry. The time is modeled by an exponential distribution with two unknown parameters, and this goal translates to precise estimation of a nonlinear transformation of the parameters. We obtain c-optimum experimental designs with that purpose, applying the graphic Elfving method. Because the optimal experimental designs depend on the nominal values of the parameters, we conduct a sensitivity analysis on our dataset. Finally, a simulation study checks the performance of the approximations, first with the Fisher Information matrix, then with the linearization of the function to be estimated. The results are useful for experimenting in a laboratory and then translating the results to a real scenario. From the application we develop a general methodology for estimating a one-dimensional transformation of the parameters of a nonlinear model.

1. Introduction. 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, for example, 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 Carson and 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 Carson (2008)), causing a blockage or jam. 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 shutdown 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.

Received July 2021; revised April 2022.
Key words and phrases. Bulk solid storage, jam formation, nonlinear heteroscedastic model, optimal design of experiments.

First, regarding the lack of knowledge of critical properties in the nature of the stored material, as biomass particles or granular materials, see Miccio, Barletta and Poletto (2013), Artoni, Santomaso and 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 et al. (2012), Chen and 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 and 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 et al. (2020), Li et al. (2004), Saleh, Golshan and 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 analysing 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 and 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 et al. (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, $i = 1, \ldots, r$ for r different outlet sizes. An optimal experimental design will provide specific values n_i and ϕ_i to obtain the estimates optimizing a criterion fixed beforehand (classical references on optimal experimental designs are, for instance, Atkinson, Donev and Tobias (2007) and Fedorov (1972)). 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 and López-Fidalgo

(2016) and Amo-Salas et al. (2016) obtained D-optimal experimental designs to estimate the parameters in several models for the time elapsed between jams. In particular, in Amo-Salas et al. (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, but López-Fidalgo and 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 2 explains in detail the motivating problem of the paper, while Section 3 illustrates the mathematical method adopted to solve it and provides the general results obtained. Section 4 presents some numerical examples as well as 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 5 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 and Moler (2023)).

2. 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. Consider, 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 et al. (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 one. Additionally, they proved empirically that the experiment with regular and identical spheres replicates well enough with other irregular and nonuniform shapes such as rice, lentils or stones. In particular, the increase in variability introduced by nonregular shapes in those cases is negligible. Figure 1 shows the 3D and 2D silos that they used in the laboratory to do the experiments. In this paper we refer to the experimental study presented in Janda et al. (2008) and further studied in Amo-Salas et al. (2016) and in Amo-Salas, Delgado-Márquez and 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

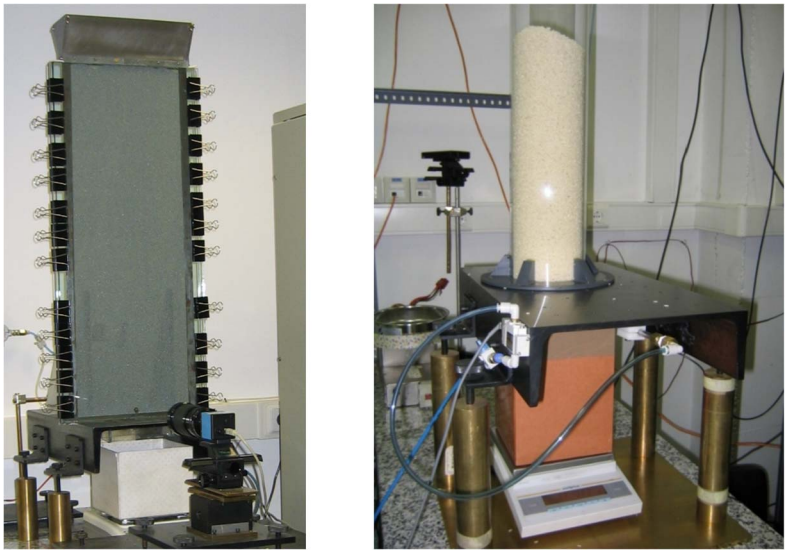

FIG. 1. *2D (left) and 3D (right) silos in the laboratory.*

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 shown in To (2005) (see also Section 4.3).

An experimental design ξ can be defined as a discrete probability distribution on a domain $\mathcal{X} = [a, b]$,

$$\xi = \begin{Bmatrix} \phi_1 & \cdots & \phi_r \\ p_1 & \cdots & p_r \end{Bmatrix}$$

with $r \leq n$. The experimenter takes $n \cdot p_i$ observations at each ϕ_i, $i = 1, \ldots, r$. In the practice, if the proportions $\{p_i\}$ are not multiples of $1/n$, they must be approximated. The extremes of the experimental domain $\mathcal{X} = [a, b]$ 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 et al. (2016) and Amo-Salas, Delgado-Márquez and López-Fidalgo (2016)).

Optimal experimental designs maximize some appropriate criterion function related to the covariance matrix of an estimator, chosen on the base of the particular aim of the experiment. 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,

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

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

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

for some inverse function $g(\boldsymbol{\theta}; T_0)$. Hence, our statistical goal is to obtain a proper optimal design ξ^* to minimize the variance of the MLE of this function of the parameter vector $\boldsymbol{\theta}$. This will provide a better precision of the resulting estimate, compared to any other nonoptimal design based on $n = \sum_{i=1}^{r} n_i$ experiments. An optimal design is then crucial to save cost of experimentation.

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).

REMARK 2.1. An alternative goal could be the estimation of the minimal diameter such that, for given values of α and T_1, $P(T > T_1) \geq 1 - \alpha$. However, for an exponential probability distribution model, this is equivalent to consider (4) with $T_0 = -T_1/\log(1-\alpha)$ because

$$1 - \alpha \leq P(T > T_1) = \exp\bigl(-T_1/\eta(\phi; \boldsymbol{\theta})\bigr)$$

and then

$$\eta(\phi; \boldsymbol{\theta}) \geq \frac{-T_1}{\log(1-\alpha)}.$$

For instance, for $\alpha = 0.05$, $T_0 = 19.5T_1$. In this particular case the threshold for the probability is about 20 times the one for the expectation. Thus, both approaches are equivalent in terms of estimation and designing.

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

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

where $\boldsymbol{\theta}^T = (C, L)$. This is also the model considered in equation (3) of Amo-Salas et al. (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 $\phi \in \mathcal{X}$ for which (1) holds. If the the expectation model is given by (3), this means

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

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

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

of T. These observations follow an exponential distribution; therefore, they have nonconstant variance,

$$\operatorname{Var}(\varepsilon_i) = \eta(\phi_i; \boldsymbol{\theta})^2.$$

When the inferential goal of an experiment is an efficient estimation of $\boldsymbol{\theta}$, an optimal design maximizes a suitable functional of the FIM. Let us denote the FIM of a model by $M(\xi, \boldsymbol{\theta})$. Note that $M(\xi, \boldsymbol{\theta})$ 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 $\boldsymbol{\theta}$ derived from guesses or previous knowledge of the possible likely values of the parameters.

When the target is to estimate a linear combination $\mathbf{c}^T\boldsymbol{\theta}$ 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 $\mathbf{c}^T\hat{\boldsymbol{\theta}}$,

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

The target of this paper is the precise estimation of a nonlinear function of the unknown parameters. Additionally, the model herein considered is nonlinear 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.

3.1. *Double approximation.* Let $\mathcal{L}(\boldsymbol{\theta};t,\phi)$ be the log-likelihood function of T. The FIM is defined by

$$M(\xi,\boldsymbol{\theta}) = \int_{\mathcal{X}} I(\phi,\boldsymbol{\theta})\,d\xi(\phi),$$

where

$$I(\phi,\boldsymbol{\theta}) = -\mathrm{E}_T\left[\frac{\partial^2}{\partial\boldsymbol{\theta}^2}\mathcal{L}(\boldsymbol{\theta};t,\phi)\right]$$

is a 2×2 matrix representing the FIM at one point ϕ. For an exponential distribution model with mean $\eta(\phi,\boldsymbol{\theta})$, we have

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

Then, for model (5),

$$I(\phi,\boldsymbol{\theta}) = \frac{1}{\eta^2(\phi,\boldsymbol{\theta})}\nabla\eta(\phi,\boldsymbol{\theta})\nabla\eta(\phi,\boldsymbol{\theta})^T,$$

and, for a design ξ,

$$M(\xi,\boldsymbol{\theta}) = \int_{\mathcal{X}} \frac{1}{\eta^2(\phi,\boldsymbol{\theta})}\nabla\eta(\phi,\boldsymbol{\theta})\nabla\eta(\phi,\boldsymbol{\theta})^T\,d\xi(\phi), \tag{7}$$

where the transpose is indicated with the superscript T and ∇ stands for the gradient.

Equation (7) is also the FIM of the following linear model:

$$t_i = \boldsymbol{\theta}^T f(\phi_i;\boldsymbol{\theta}_t) + \epsilon_i, \tag{8}$$

where

$$f(\phi;\boldsymbol{\theta}) = \frac{1}{\eta(\phi,\boldsymbol{\theta})}\nabla\eta(\phi,\boldsymbol{\theta}) \tag{9}$$

is evaluated in a the true value $\boldsymbol{\theta}_t$ of the parameter vector. In this way we have approximated model (5) in a neighborhood of the true value.

The MLE estimator of $g(\boldsymbol{\theta};T_0)$ is $g(\hat{\boldsymbol{\theta}};T_0)$, so we approximate the nonlinear function $g(\cdot;T_0)$ using Taylor expansion around the true value $\boldsymbol{\theta}_t$, that is, $g(\hat{\boldsymbol{\theta}};T_0)\approx g(\boldsymbol{\theta}_t;T_0)+\nabla g(\boldsymbol{\theta}_t;T_0)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_t)$. The variance of $g(\hat{\boldsymbol{\theta}};T_0)$ can then be approximated by

$$\nabla g(\boldsymbol{\theta}_t;T_0)^T M(\xi,\boldsymbol{\theta}_t)^{-1}\nabla g(\boldsymbol{\theta}_t;T_0)$$

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

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

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 $\mathbf{c}(\boldsymbol{\theta})$ and the FIM, $M(\xi,\boldsymbol{\theta})$. Hence, we choose an initial guess θ_0 for the true value θ_t, and the design obtained will be locally optimum.

3.2. *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 $\mathbf{c}^T\boldsymbol{\theta}$ of the unknown parameters of a linear homoscedastic model $t = \boldsymbol{\theta}^T f(\phi) + \epsilon$, like the model (8), where $f(\phi) = (f_1(\phi), f_2(\phi))^T$.

To apply the Elfving's graphical method, the first step is to obtain and to 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 and del Campo (2014).

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

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

Hence,

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

with

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

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

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

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

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

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, May and Moler (2023)).

PROPOSITION 3.1. *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 3.2. *When the crossing point P_0 is in $\overline{A_i A_{i+1}}$, $i = 1, 2, 3$, the c-optimal design is*

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

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

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

and

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

4. 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 and Maza (2018)).

4.1. *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 et al. (2016) from data are used as nominal values of the parameters.

EXAMPLE 1. Let $\phi \in \mathcal{X} = [1.53, 5.63]$, $C_0 = 0.671741$ 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, ∞) (see Lemma 1.2 of the Supplementary Material (López-Fidalgo, May and Moler (2023)) for details). Figure 2 illustrates the three cases $T_0 = 2$, $T_0 = 200$ and $T_0 = 300{,}000$; 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 $g(\hat{\boldsymbol{\theta}}; T_0)$ is equal to the ratio of the dashed vector to the whole vector.

According to Proposition 3.2, the optimal design in each case is

$$\begin{Bmatrix} 1.53 & 5.63 \\ 0.9789 & 0.0211 \end{Bmatrix}_{T_0=2}; \quad \begin{Bmatrix} 1.53 & 5.63 \\ 0.5526 & 0.4474 \end{Bmatrix}_{T_0=200}; \quad \begin{Bmatrix} 1.53 & 5.63 \\ 0.0240 & 0.9760 \end{Bmatrix}_{T_0=300,000}.$$

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

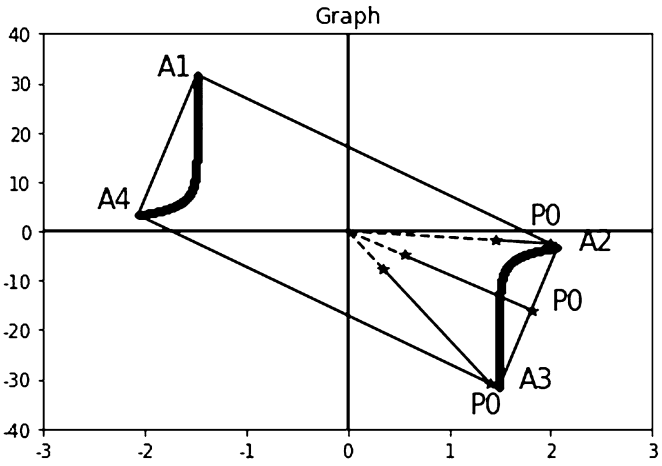

FIG. 2. *Graphical representation for the framework in Example 1.*

EXAMPLE 2. As shown in the proof of Proposition 3.1 (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 $C_0 = 2.3$. Assume $T_0 = 2$.

Figure 3 represents the convex hull and $\mathbf{c}$ in this example, when A_2 is not a maximum. Now, Proposition 3.2 holds, and then

$$\xi_c^* = \begin{Bmatrix} 1.53 & 5.63 \\ 0.2706 & 0.7294 \end{Bmatrix}.$$

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.

4.2. *Sensitivity analysis.* 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, May and Moler (2023)).

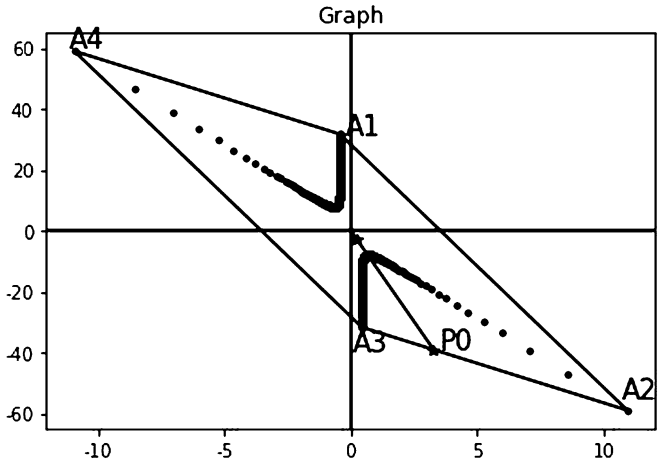

FIG. 3. *Graphical representation for the framework in Example 2.*

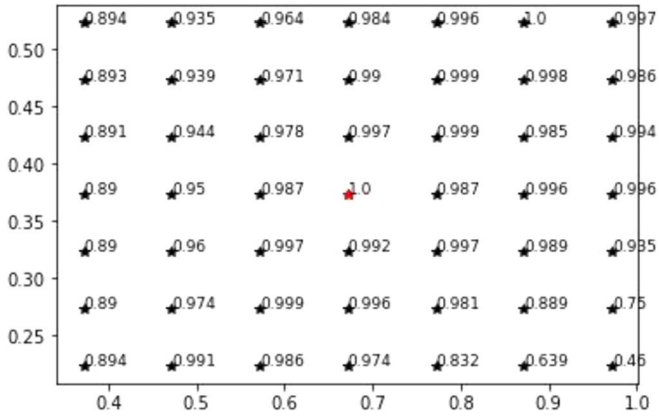

FIG. 4. *Efficiency values in each point of the grid for* $T_0 = 2$.

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 but interesting to be considered in this study.

Examine a grid of points (C^*, L^*) appropriate to detect sensitive changes in the efficiencies. In Figures 4, 5 and 6 the efficiencies for the three cases considered are shown. In all the three cases, C^* varies in the interval $(C_0 - 0.3, C_0 + 0.3)$, while L^* varies in the interval $(L_0 - 0.15, L_0 + 0.15)$ in cases 1 and 2 and in the interval $(L_0 - 0.05, L_0 + 0.05)$ in case 3.

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 Figures 4 and 5 the largest reduction of the efficiency happens for large values of C^*, combined with small values of L^*. In Figure 6 a smaller interval is chosen to vary L^* because dramatic changes of the efficiency are observed for further values of L^*. Now, the efficiency decreases when L^* grows and C^* decreases (top left on Figure 6) and when L^* decreases and C^* grows (bottom right on the table). This is because the cross points of the gradient with the convex hull are far away from the cross point of (C_0, L_0) or even in another segment.

In the three cases 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.

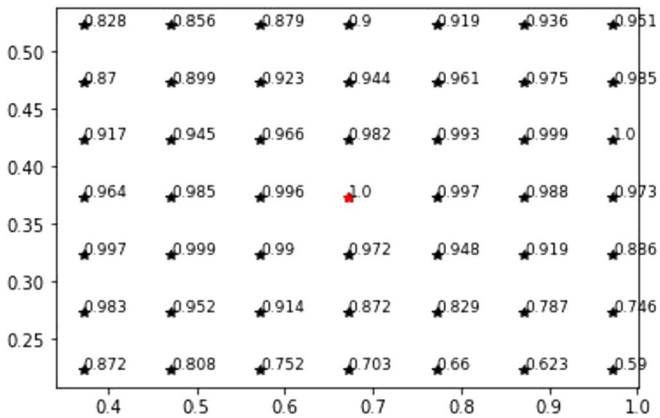

FIG. 5. *Efficiency values in each point of the grid for* $T_0 = 200$.

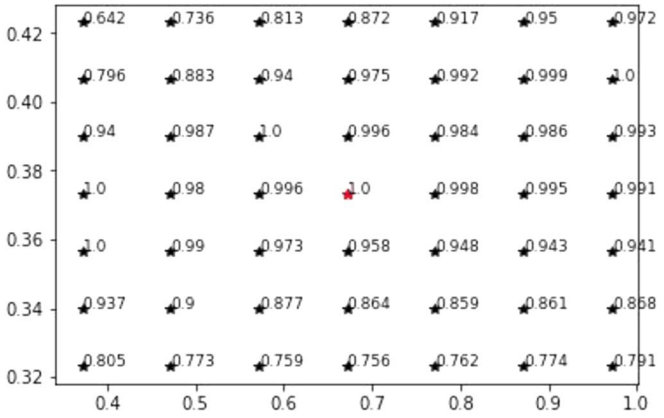

FIG. 6. *Efficiency values in each point of the grid for* $T_0 = 300{,}000$.

4.3. *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 2. 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 microdata of a similar experiment to the one just described (Gella, Zuriguel and Maza (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; 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 1. To find the best probability distribution function that fits the experimental data, the procedure in Rigby et al. (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.

TABLE 1
Setting for the real experiment

ϕ	9.3	9.7	10.5	11.2	11.8	12.4	13	13.6	14.2	14.9	15.3	15.4
n_i	1510	2005	2139	2124	2164	2168	2113	2037	2085	2026	1009	1001
$\overline{T}_j$	19.04	18.38	33.37	62.11	86.85	163.43	219.51	581.13	873.55	1212.22	3109.30	4611.61

In what follows, we compare the performance of the design, used in practice to collect these data, and the optimal design, obtained in Proposition 3.2, 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

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

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 et al. (2016) and also of the c-optimal design obtained in Proposition 3.2. 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 nonnegligible 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 1000 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, $j = 1, \ldots, 12$, 84 values are randomly chosen so that, $n \sim 1000$. 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 $\overline{T}_j$, $j = 1, \ldots, 12$ that appear in Table 1 so that, the corresponding g value is close to ϕ_j. This procedure is repeated 1000 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 3.2 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 2; the proportion $p_{9.3}$ of experiments in 9.3 in the c-optimal design is also indicated.

TABLE 2
Results of the comparative study between the c-optimal and the MLE with the 12 *design points*

ϕ	9.3	9.7	10.5	11.2	11.8	12.4	13	13.6	14.2	14.9	15.3	15.4
T_0	19	18	33	62	87	163	220	581	874	1212	3109	4611
$p_{9.3}$	0.999	0.991	0.89	0.7817	0.719	0.602	0.547	0.370	0.296	0.237	0.070	0.001
$\tilde{g}$	9.35	9.3	10.1	10.9	11.3	12.1	12.4	13.4	13.8	14.2	15.1	15.4
$\hat{\sigma}_{\tilde{g}}$	0.39	0.40	0.28	0.18	0.13	0.06	0.04	0.11	0.15	0.18	0.26	0.29
$\hat{g}$	9.4	9.2	10.0	11.0	11.3	12.0	12.4	13.4	13.9	14.2	15.1	15.4
$\sigma_{\hat{g}}$	0.06	0.06	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.05	0.05

It is worth highlighting several points from the results presented in Table 2. 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 4.2, the c-optimal design could benefit from better guesses for the initial values C_0 and L_0 reducing bias and variability.

5. Conclusions. In this paper we consider the problem of estimating the parameters of a nonlinear 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) entail the estimation of a lower bound expressed as a nonlinear 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 nonlinear 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 3.1 and 3.2 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 an 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 3.2. 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 (8) and (3.1) (cf. Section 3.1). 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 nonlinear 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, May and Moler (2023)).

Acknowledgments. The authors would like to thank the Granular Media Laboratory of the University of Navarra (https://en.unav.edu/web/laboratorio-medios-granulares), in particular, Professors Gella, Zuriguel, Maza, Janda and Kiwing To for providing real data from their experiments. They also wish to acknowledge the anonymous reviewers, an Associate Editor as well as the Editor-in-Chief whose insights and efforts led to this favorable outcome.

Funding. The first author was sponsored by Ministerio de Ciencia e Innovación PID2020-113443RB-C21 and the third one by Ministerio de Ciencia e Innovación PID2020-116873GB-I00 and PID2020-114031RB-I00. The second author thanks the Departamento de Estadística, Informática y Matemáticas of the Universidad Pública de Navarra for enduring her scientific visit to the department.

SUPPLEMENTARY MATERIAL

Suplementary material (DOI: 10.1214/22-AOAS1644SUPP; .pdf). A .zip folder includes a .pdf file divided into three sections: Section 1 provides the proofs of the theoretical results in Section 3; Section 2 gives details for the sensitivity analysis and, finally, in Section 3 a simulation study to check the goodness of the approximations is presented. In addition, Python 3.7. code is provided in two separate files to replicate the computations. The file dataset.txt includes the real dataset used in the paper. There are two columns, first column provides the observed times between jams and the second displays the different diameters.

REFERENCES

ABLE, R., OTHEN, S. and NEDDERMAN, R. (1996). The exit time distribution during the batch discharge of a cylindrical bunker. *Chem. Eng. Sci.* **51** 4605–4610.

AMO-SALAS, M., DELGADO-MÁRQUEZ, E. and LÓPEZ-FIDALGO, J. (2016). Optimal experimental designs in the flow rate of particles. *Technometrics* **58** 269–276. MR3488305 https://doi.org/10.1080/00401706.2015.1042169

AMO-SALAS, M., DELGADO-MÁRQUEZ, E., FILOVÁ, L. and LÓPEZ-FIDALGO, J. (2016). Optimal designs for model discrimination and fitting for the flow of particles. *Statist. Papers* **57** 875–891. MR3571178 https://doi.org/10.1007/s00362-016-0792-5

ARTONI, R., SANTOMASO, A. and CANU, P. (2009). Simulation of dense granular flows: Dynamics of wall stress silos. *Chem. Eng. Sci.* **64** 4040–4050.

ATKINSON, A. C., DONEV, A. N. and TOBIAS, R. D. (2007). *Optimum Experimental Designs, with SAS. Oxford Statistical Science Series* **34**. Oxford Univ. Press, Oxford. MR2323647

BELL, T. (2005). Challenges in the scale-up of particulate processes—an industrial perspective. *Powder Technol.* **150** 60–71.

BROWN, M. (1990). Error bounds for exponential approximations of geometric convolutions. *Ann. Probab.* **18** 1388–1402. MR1062073

CARSON, J. and CRAIG, D. (2015). Silo design codes: Their limits and inconsistencies. *Proc. Eng.* **102** 647–656.

CARSON, J. W. (2008). *Bulk Solids Handling. Equipment Selection and Operation* (D. McGlinchey, ed.) Blackwell Publishing, Oxford.

CHEN, W. and ROBERTS, A. W. (2018). A modified flowability classification model for moist and cohesive bulk solids. *Powder Technol.* **325** 639–650.

ELFVING, G. (1952). Optimum allocation in linear regression theory. *Ann. Math. Stat.* **84** 44002–1–44002–6.

FEDOROV, V. V. (1972). *Theory of Optimal Experiments*. Academic Press, New York. MR0403103

FITZPATRICK, J. J., BARRINGER, S. A. and IQBAL, T. (2004). Flow property measurement of food powders and sensitivity of Jenike's Hopper design methodology to the measured values. *J. Food Eng.* **61** 399–405.

GELLA, D., ZURIGUEL, I. and MAZA, D. (2018). Decoupling geometrical and kinematic contributions to the silo clogging process. *Phys. Rev. Lett.* **121** 138001.

GOLSHAM, S., ESGANDARI, B., ZARGHAMI, R., BLAIS, B. and SALEH, K. (2020). Experimental and DEM studies of velocitiy profiles and residence time distribution of non-spherical particles in silos. *Powder Technol.* **373** 510–521.

JANDA, A., ZURIGUEL, I., GARCIMARTIN, A., PUGNALONI, L. A. and MAZA, D. (2008). Jamming and critical outlet size in the discharge of a two-dimensional silo. *Europhys. Lett.* **23** 255–262.

JENIKE, A. W. (1961). *Gravity Flow of Bulk Solids*. Bulletin of the Univ. Utah.

LI, J., LANGSTON, P. A., WEBB, C. and DYAKOWSKI, T. (2004). Flow of sphero-disc particles in rectangular hoppers-a DEM and experimental comparison in 3D). *Chem. Eng. Sci.* **59** 5917–5929.

LOPEZ-FIDALGO, J., MAY, C. and MOLER, J. A (2023). Supplement to "Designing experiments for estimating an appropriate outlet size for a silo type problem." https://doi.org/10.1214/22-AOAS1644SUPP

LÓPEZ-FIDALGO, J. and RODRÍGUEZ-DÍAZ, J. M. (2004). Elfving's method for m-dimensional models. *Metrika* **59** 235–244. MR2066727 https://doi.org/10.1007/s001840300281

MICCIO, F., BARLETTA, D. and POLETTO, M. (2013). Flow properties and arching behavior of biomass particulate solids. *Powder Technol.* **235** 312–321.

MITRA, H., PUSHPADASS, H. A., ELJEEVA, M., FRANKLIN, E., KINGSBY, R. P., GHOROI, C., NATH, S. and BATTULA, S. (2017). Influence of moisture content on the flow properties of basundi mix. *Powder Technol.* **312** 133–143.

NEDDERMAN, R. M. (1992). *Statics and Kinematics of Granular Materials*. Cambridge Univ. Press, Cambridge.

RIGBY, R. A., STASINOPOULOS, M. D., HELLER, G. Z. and DE BASTIANI, F. (2019). *Distributions for Modeling Location, Scale, and Shape: Using GAMLSS in R*, 1st ed. CRC Press, Boca Raton, FL.

RIVAS-LÓPEZ, M. J., LÓPEZ-FIDALGO, J. and DEL CAMPO, R. (2014). Optimal experimental design for accelerated failure time with Type I and random censoring. *Biom. J.* **56** 819–837. MR3258098 https://doi.org/10.1002/bimj.201300209

ROTTER, J. M. (2001). *Guide for the Economic Design of Circular Metal Silos*. Taylor & Francis Routledge, London.

SALEH, K., GOLSHAN, S. and ZARGHAMI, R. (2018). A review on gravity floe of free-flowing granular solids in silos—basics and practical aspects. *Chem. Eng. Sci.* **192** 1011–1035.

SAMUELSSON, R., LARSSON, S. H., THYREL, M. and LESTANDER, T. A. (2012). Moisture content and storage time influence the binding mechanisms in biofuel wood pellets. *Appl. Energy* **99** 109–115.

SCHULZE, D. (2008). *Powders and Bulk Solids*. Springer, Berlin.

TO, K. (2005). Jamming transition in two-dimensional hoppers and silos. *Phys. Rev. E* **71** 060301.

ZURIGUEL, I., GARCIMARTÍN, A., MAZA, D., PUGNALONI, L. A. and PASTOR, J. M. (2005). Jamming during the discharge of granular matter from a silo. *Phys. Rev. E* **71** 051303.