

An Optimization Framework to Combine Operable Space Maximization with Design of Experiments

Qi Chen, Remigijus Paulavičius and Claire S. Adjiman*

Dept. of Chemical Engineering, Centre for Process Systems Engineering, Imperial College London, South Kensington Campus, London, SW7 2AZ, U.K.

Salvador García-Muñoz

Small Molecule Design and Development, Eli Lilly & Co., Indianapolis, IN, 46221

DOI 10.1002/aic.16214

Published online September 8, 2018 in Wiley Online Library (wileyonlinelibrary.com)

The introduction of Quality by Design in the pharmaceutical industry stimulates practitioners to better understand the relationship of materials, processes and products. One way to achieve this is through the use of targeted experimentation. In this study, an optimization framework to design experiments that effectively leverage parameterized process models is presented to maximize the space covered in the output variables while also obtaining an orthogonal bracketing study in the process input factors. The framework considers both multi-objective and bilevel optimization methods for relating the two maximization objectives. Results are presented for two case studies—a spray coating process and a continuously stirred reactor cascade—demonstrating the ability to generate and identify efficient designs with fit-for-purpose trade-offs between bracketed orthogonality in the input factors and volume explored in the process output space. The proposed approach allows a more complete understanding of the process to emerge from a small set of experiments. © 2018 The Authors. AIChE Journal published by Wiley Periodicals, Inc. on behalf of American Institute of Chemical Engineers. AIChE J, 64: 3944–3957, 2018
 Keywords: quality by design, pharmaceuticals, design of experiments, multiobjective optimization, bilevel optimization

Introduction

The acquisition of increased process understanding is an important step in drug product development. It is typically approached by carrying out extensive experimental studies. However, the benefits of such studies must be weighed against the time necessary to conduct them, which is a major cost driver^{1–3} and therefore impediment to developing better processes. The ability to gain process understanding more quickly through systematic model-based experimental design can thus be a competitive advantage. Placed in the context of recent regulatory changes in the pharmaceutical industry, such as Quality by Design (QbD), knowledge-driven time-saving techniques such as modeling and optimization technology open up new opportunities for a better understanding of the process and product relationships, and increase the benefits that can be derived from this newly acquired understanding.⁴ QbD allows the operation of the process in a flexible manner within an

approved design space, lessening the burden on both the manufacturer and the regulator. This change implies a shift away from the traditional approach of a single target for process conditions to operation within robust regions. To enable this shift, the relationships between materials, process and products need to be well-understood. This entails a need not only to develop adequate models, but also to use these models in the exploration of the process operational space.

Significant work has gone into the development of optimal experiment design to properly parameterize a model of a given system.^{5–11} Generally speaking these approaches are based on optimizing a metric of the Fisher information matrix to improve the statistical properties of the estimated parameters. Other researchers have focused on developing techniques to design optimal experiments to discriminate competing model structures.⁵ Although both avenues are variants of “experiment design,” the underlying objectives in model discrimination and model parametrization are completely different.

In our work, we introduce a third class of experiment design techniques where we seek to design experiments to gain additional process understanding and communicate this knowledge effectively to regulatory bodies in the pharmaceutical industry. This process understanding has to be derived in two spaces: that of inputs to the process and that of process outputs. In undertaking such an investigation, it is assumed that a suitable model with appropriately estimated parameters is already available. The use of traditional design of experiments (DoE) approaches for the exploration of the design space has so far

Additional Supporting Information may be found in the online version of this article.

Current address of [Qi Chen]: Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213

Correspondence concerning this article should be addressed to C. S. Adjiman at c.adjiman@imperial.ac.uk.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2018 The Authors. AIChE Journal published by Wiley Periodicals, Inc. on behalf of American Institute of Chemical Engineers.

been focused on bracketing the input factors to a given process using linear experiment design methods. As a result, regulators are accustomed to reviewing and approving process operation (submissions) based on orthogonal bracketing-based studies where the process *inputs* are varied at selected corners of a design, with popular approaches such as full factorial design, fractional or composite design in common use.^{6,7,12,13}

However, there has been less work on using parameterized models to gain additional understanding of the process *outputs* and communicate this knowledge effectively to regulatory bodies in the pharmaceutical industry. To save time and resources, it is desirable to maximize the amount of process information obtained from a fixed number of experiments by carefully and systematically selecting experimental measurements, simultaneously obtaining information on both the process input and output spaces. The application of DoE methods to the input space has so far not explicitly considered the implications of the chosen design on the exploration and understanding of the space of process outputs. For example, in the selection of atomization parameters in spray coater design, the typical industrial approach to identifying the output space is “a combination of trial-and-error and previous experience.”¹⁴ Thus, Ref. 7 chose design points for scale-up on the basis of their match to an output region that had already been determined, perhaps heuristically. However, the critical quality attributes of the process are usually found in the output space.¹⁵ It is therefore advantageous for manufacturers to properly explore the space of process outputs while also seeking to maintain orthogonality in the input space, by exploiting the strong interplay between the input and output spaces.

In our work, we present a new class of model-based design of experiments and propose two strategies to solve the resulting optimization problem which is bilevel/multiobjective in nature and thus requires orthogonal exploration in the input space (also referred to as *bracketing*) as well as the identification of points that are widely distributed in the output space. First, multiobjective optimization (MOO) provides a way to systematically consider both objectives. MOO has been applied by several authors to the design of experiments.⁸⁻¹⁰ These examples demonstrate the potential for MOO to help process engineers design experiments that balance multiple criteria, in this case improving coverage in the output space with orthogonal input factor values. A second way to represent the relationship between the input and output space objectives is using bilevel optimization (BLO), which represents the interaction as a hierarchical system with two decision-makers, the leader and the follower. Such an approach has been used in the context of min-max economic experiment design.¹¹ Here, the objectives of the two decision-makers correspond to the previously mentioned output and input space objectives, respectively.

We develop an optimization framework considering objectives in both the process input space and the output operating space. To relate the input and output spaces, we rely on a deterministic process model, which is assumed to have been previously derived. We investigate the suitability of MOO and BLO to provide valuable insights and experimental designs. Both approaches have different strengths: the multiobjective approach allows several good designs to be visualized on a trade-off curve and the bilevel optimization approach identifies a suitable compromise point without the need to locate multiple trade-off points. By maximizing the information value of each experiment, we find *efficient* designs that reduce the total number of experiments necessary to extract maximum understanding of the input and output spaces and the time required

for conducting them. The effectiveness of this framework in relation to the traditional design approach is demonstrated on two case studies: a spray coating process and a system of two continuous reactors.

The rest of this article is organized as follows. First, a motivating example based on tablet spray coating is introduced. This is followed by the presentation of the mathematical formulations that underpin our approach. Next, the results of the framework when applied to two process development case studies, including the motivating example, are discussed. Finally, conclusions and further thoughts are given in the Conclusions section.

Motivating Example

As a motivating example, we examine a spray coating process. Tablet coating is typically one of the final steps in the manufacture of a pharmaceutical product, applied for both functional and/or aesthetic considerations.^{7,16} Systematic DoE for this processing step would give process engineers better insight into the regions of safe operation while helping them communicate their process knowledge to regulators. Thus, it is advantageous to design experiments during model verification that simultaneously maximize information on the input factors and maximize understanding of the output space.

The tablet film coating process involves the spraying of an atomized solution (or suspension) of polymers onto a tablet bed, diagrammed in Figure 1. Energy for the evaporation of the spray is provided by a warm air stream. Ebey¹⁷ developed a thermodynamic model for tablet coating with aqueous solutions based on material and energy balances around the coater unit. This model was later extended to represent the use of organic solvents and to capture energy requirements better with the inclusion of a heat loss adjustment factor.¹⁸ The process inputs include the inlet temperature and volumetric flow rate of the drying air stream and the mass flow rate of the coating solution. The relevant output conditions are the outlet temperature of the drying air stream and its outlet humidity.

To motivate the need for new ways to design experimental investigations, we consider a film coating process for which the input factors are defined in the following ranges: $T_{\text{air,in}} (^{\circ}\text{C}) \in [20, 85]$, $M_{\text{coat}} (\text{g}/\text{min}) \in [10, 80]$ and $Q_{\text{air}} (\text{ft}^3/\text{min}) \in [150, 450]$. Two four-point experimental designs that are based

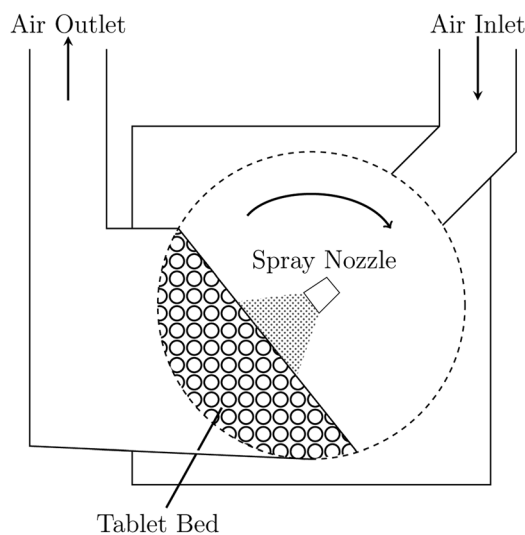


Figure 1. Schematic of a tablet coater, adapted from Ref. 18.

Table 1. Comparison of Input Values for Two Highly Orthogonal Four-Experiment Designs Derived from DoE

Exp	Design 1			Design 2		
	$T_{\text{air,in}}$ [°C]	M_{coat} [g/min]	Q_{air} [ft ³ /min]	$T_{\text{air,in}}$ [°C]	M_{coat} [g/min]	Q_{air} [ft ³ /min]
1	20	10	450	20	10	150
2	85	10	150	85	10	150
3	85	10	450	85	10	450
4	85	80	450	85	80	158

Differences between the two designs are highlighted in bold.

on achieving a high D-criterion in the space of input factors, without regard to the space of output variables, are given in Table 1. Both designs are based on similar input choices and are highly orthogonal in the input space: the D-criterion for Design 1 has the maximum value that can be achieved, and that of Design 2 is 1% smaller. However, the two designs result in very different values of the output variables, as seen in Figure 2. Whereas Design 1 leads to the exploration of only a small sliver in the output space, the measurements of Design 2 map out a much larger range of possible outputs. Even with Design 2, a large fraction of the feasible output space (as approximated by the convex region delimited by dotted lines in Figure 2) remains unexplored. Furthermore, one of the experimental points in Design 2 does not contribute to increasing the explored range. This shows that an experiment design derived by examining the input space only does not always correspond to a good design in terms of coverage of the output space. There is therefore a need to consider both criteria in the course of experiment design, to try and maximize the value that can be extracted from the experiments.

Methodology

Design criteria and optimization formulations

In this section, we describe a new framework based on multiobjective and bilevel optimization that considers simultaneously the spaces of process inputs and outputs in generating experiment designs that are optimal from both perspectives. To determine an optimal design, we first define quantitative criteria for the input and output space objectives. Then, we present details of the mathematical formulations of the relevant multiobjective and bilevel optimization problems and describe the solution strategies utilized for obtaining results.

Criteria and Formulation for the Input Space. In the input space, the primary objective is to explore each input factor as independently as possible, leading to an orthogonal design. One way to quantify this objective is by adopting a modified D-optimality criterion and applying it to the input factors. D-optimality is one of several design criteria commonly used for model parameterization.¹⁹ Geometrically, this criterion can be interpreted as minimizing the volume of the confidence ellipsoid of the parameter estimates.²⁰ The use of D-optimality confers several advantages, such as the ability to scale the model to avoid ill-conditioned information matrices²¹ and the ability to accommodate feasible spaces of varying shapes, allowing for operating constraints to be imposed on the process model.⁵ Crucially, applying D-optimality to a linear response model yields an orthogonal design in the parameter space. In this work, we apply D-optimality to the input factors rather than to the parameters, obtaining an orthogonal design in the space of input factors. To avoid confusion between parameter-based criteria and input factor-based criteria, we will refer to D-optimality applied to the input factors as input

space optimality or D_I -optimality, and to the D-criterion applied to the input factors as the D_I -criterion.

Finding a design that is D_I -optimal involves maximizing the D_I -criterion, which is defined as the determinant of the Fisher information matrix M . To represent the calculation of this determinant within an algebraic optimization problem, we exploit the fact that the Fisher information matrix is symmetric and positive semidefinite¹⁹ to express the determinant calculation as a set of equality constraints using LDL decomposition,²² where $M = LDL^T$, L is a lower-triangular matrix and D is a diagonal matrix. L and D are recursively defined by Eqs. 1 and 2

$$D_{pq} = \begin{cases} M_{pq} - \sum_{r=1}^{p-1} L_{pr}^2 D_{rr}, & p = q \\ 0, & p \neq q \end{cases} \quad \text{for } p = 1, \dots, \phi, q = 1, \dots, \phi \quad (1)$$

$$L_{pq} = \begin{cases} \frac{1}{D_{qq}} \left[M_{pq} - \sum_{r=1}^{q-1} L_{pr} L_{qr} D_{rr} \right], & p > q \\ 0, & p < q \\ 1, & p = q \end{cases} \quad \text{for } p = 1, \dots, \phi, q = 1, \dots, \phi \quad (2)$$

Here, p , q , and r are indices on the ordered set of ϕ process input factors and X_{pq} defines the element in row p , column q of matrix X . Using this notation, the input space objective function $f_{\text{in}}(u)$ to be maximized can therefore be expressed as the D_I -criterion value

$$f_{\text{in}}(u) = \prod_{p=1}^{\phi} D_{pp}, \quad (3)$$

where u are the input variables and D_{pp} is given by the LDL decomposition of Eqs. 1 and 2.

In some cases, the use of such an objective function can lead to degenerate designs in which all experiments have the same value for one or more factors, without an adverse effect on the D_I -criterion. To exclude such undesirable solutions from the feasible space of designs, a constraint is introduced to ensure a minimum variance for each input factor. Given a set of input factors $\hat{u}_{i,p}$, $i = 1, \dots, n$, $p = 1, \dots, \phi$, where i denotes the experiment number, p the input variable number, n the total number of experiments, and ϕ the total number of input factors, it is possible to define a centroid of measurements \bar{u}_p

$$\bar{u}_p = \frac{1}{n} \sum_{i=1}^n \hat{u}_{i,p}, \quad p = 1, \dots, \phi. \quad (4)$$

From this, the variance of input factor p , $\sigma_{\hat{u}_p}^2$, can be calculated as

$$\sigma_{\hat{u}_p}^2 = \frac{1}{n-1} \sum_{i=1}^n (\hat{u}_{i,p} - \bar{u}_p)^2, \quad p = 1, \dots, \phi. \quad (5)$$

By constraining the variance of an input factor across all experiments to be greater than some small value ($\sigma_{\hat{u}_p}^2 \geq \epsilon_{\sigma^2}$), degenerate solutions can be avoided. The choice of a value for ϵ_{σ^2} depends on the range of the variables. In our work, the input factors are scaled and a default value of $\epsilon_{\sigma^2} = 1 \times 10^{-3}$ is used.

As experimental trials are indistinguishable by ordering, the same solution may have multiple symmetric representations. To reduce the number of redundant solutions, linear

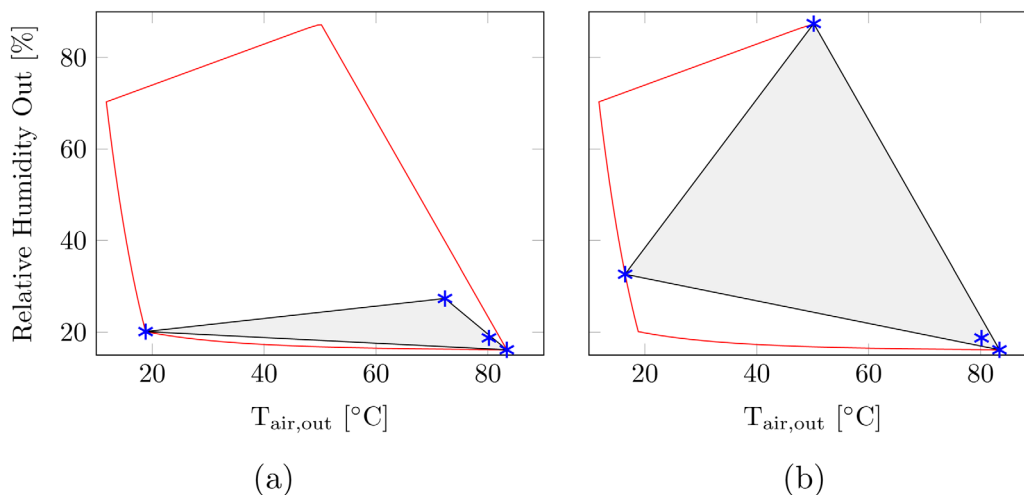


Figure 2. Comparison of output space coverage (in gray) for two similar input designs. Measurements are given by stars, with the outer boundary indicating the convex feasible region approximation generated by input domain and operating constraints. (a) Design 1 output space with 10% coverage. (b) Design 2 output space with 63% coverage.

[Color figure can be viewed at wileyonlinelibrary.com]

symmetry-breaking constraints are introduced. Two types of symmetry-breaking constraints are used: Eq. 6, which enforces an ordering for one of the input factors p (e.g., $p = 1$), and a lexicographic ordering in Eq. 7:

$$u_{p,i} \leq u_{p,i+1}, \quad i = 1, \dots, n-1 \quad (6)$$

$$\sum_{p=1}^{\phi} 10^p (\hat{u}_{i,p} - \hat{u}_{i+1,p}) \leq 0, \quad i = 1, \dots, n-1 \quad (7)$$

Collectively, Eqs. 6 and 7 are denoted by $A_{in}u \leq 0$ in subsequent optimization formulations.

The overall input space optimization formulation is thus given by

$$\max_{u,y} f_{in}(u) = \prod_{p=1}^{\phi} D_{pp}$$

$$\text{s.t. } M = LDL^T$$

$$D_{pq} = \begin{cases} M_{pq} - \sum_{r=1}^{p-1} L_{pr}^2 D_{rr}, & p = q \\ 0, & p \neq q \end{cases} \quad \text{for } p = 1, \dots, \phi, q = 1, \dots, \phi$$

$$L_{pq} = \begin{cases} \frac{1}{D_{qq}} \left[M_{pq} - \sum_{r=1}^{q-1} L_{pr} L_{qr} D_{rr} \right], & p > q \\ 0, & p < q \\ 1, & p = q \end{cases} \quad \text{for } p = 1, \dots, \phi, q = 1, \dots, \phi$$

$$\bar{u}_p = \frac{1}{n} \sum_{i=1}^n \hat{u}_{i,p}, \quad p = 1, \dots, \phi$$

$$\sigma_{\hat{u}_p}^2 = \frac{1}{n-1} \sum_{i=1}^n (\hat{u}_{i,p} - \bar{u}_p)^2, \quad p = 1, \dots, \phi$$

$$\sigma_{\hat{u}_p}^2 \geq \epsilon_{\sigma^2}, \quad p = 1, \dots, \phi$$

$$g(u,y) \leq 0$$

$$A_{in}u \leq 0$$

$$u^{lb} \leq u \leq u^{ub}, \quad y^{lb} \leq y \leq y^{ub} \quad (8)$$

where u^{lb} and u^{ub} are lower and upper bounds, respectively, on the values of the input factors u , and y^{lb} and y^{ub} are the bounds on the values of the output variables. $g(u,y) \leq 0$ represents the process model, ensuring that selected input factor values are feasible. For simplicity, the first six constraints in Problem (8) are denoted by $g_{in}(u) \leq 0$ in the remainder of the paper so that the input space problem is written more concisely as

$$\begin{aligned} \max_{u,y} & f_{in}(u) \\ \text{s.t.} & g_{in}(u) \leq 0 \\ & g(u,y) \leq 0 \\ & A_{in}u \leq 0 \\ & u^{lb} \leq u \leq u^{ub} \\ & y^{lb} \leq y \leq y^{ub} \end{aligned} \quad (9)$$

For a given number of input factors, the maximum value of the D_I -criterion that can be attained depends on the number of experiments designed. Given an optimal value of the D_I -criterion, $f_{in}(u^*)$, at the solution u^* of Problem (8), it is thus desirable to calculate the D_I -optimality of design u^* , which makes it possible to compare the suitability of two designs that are based on different numbers of experiments. Let M_n^* correspond to the Fisher information matrix with the largest possible D_I -criterion value for a model with ϕ factors and a given number of experiments n . Let $M_n(u)$ be the Fisher information matrix for another design u with n experiments. The D_I -optimality of design u is given by

$$D_{I,opt} = \left(\frac{\det(M_n(u))}{\det(M_n^*)} \right)^{\frac{1}{\phi}}. \quad (10)$$

This value then indicates the fraction of input space information obtained relative to the best possible design, for the given number of experiments: two repetitions of a design with D_I -optimality of 50% are equivalent to one at 100%. An example calculation is included in Supporting Information Section 1.

Criteria and Formulation for the Output Space. In the space of process outputs, the objective of the experimental investigation is to explore the largest volume possible within the feasible limits of operation of the process. Determining the

size of the region explored with a finite set of points is a nontrivial problem and several approaches can be deployed; here, we investigate a convex hull interpretation and an ellipsoid approximation. If the feasible output region is convex or nearly so, then the volume of the convex hull of the image of the set of input points in the output space provides a good approximation of the size of the region explored. However, it is difficult to derive the convex hull of a set of points analytically^{23,24} and indeed all existing polynomial-time approximations of the convex hull are nonanalytical.²⁵⁻²⁷ This makes it challenging to embed the calculation of the volume of the convex hull within an optimization framework.²⁸ Therefore, the convex hull representation is reserved for the interpretation of the results of the optimization, where its intuitiveness is beneficial.

In its place, an ellipsoid approximation is used within the optimization formulation. The ellipsoid volume is given by the determinant of the covariance matrix of the output points. Given a matrix of process output points \hat{y}_{ic} , $i = 1, \dots, n$, $c = 1, \dots, \psi$ with i the index of each experimental point, c the index of each output variable, n the total number of experiments in the design, and ψ the total number of output variables, an output centroid \bar{y}_c can be computed as

$$\bar{y}_c = \frac{1}{n} \sum_{i=1}^n \hat{y}_{ic}, \quad c = 1, \dots, \psi. \quad (11)$$

This is used to calculate the covariance matrix \mathcal{M} with elements \mathcal{M}_{cd} where c and d denote two output variables

$$\mathcal{M}_{cd} = \frac{1}{n-1} \sum_{i=1}^n (\hat{y}_{ic} - \bar{y}_c)(\hat{y}_{id} - \bar{y}_d), \quad c = 1, \dots, \psi, \quad d = 1, \dots, \psi. \quad (12)$$

The volume of the ellipsoid is then

$$V = \frac{4}{3} \pi \sqrt{\det(\mathcal{M})}. \quad (13)$$

Due to numerical issues that may arise from the square root operator, the square of the volume and a scaling factor β are used to define an alternative metric, which serves as objective function, V_{obj}

$$V_{\text{obj}} = \beta \det(\mathcal{M}) \quad (14)$$

where β is chosen to be equal to 1000.

As with the input space optimization, a determinant calculation is required. The covariance matrix is also symmetric and positive semidefinite, so the LDL decomposition is used again here. The relevant equations are

$$\mathcal{M} = \mathcal{L} \mathcal{D} \mathcal{L}^T \quad (15)$$

$$\mathcal{D}_{cd} = \begin{cases} \mathcal{M}_{cd} - \sum_{r=1}^{c-1} \mathcal{L}_{cr}^2 \mathcal{D}_{rr}, & c = d \\ 0, & c \neq d \end{cases} \quad \text{for } c = 1, \dots, \psi, d = 1, \dots, \psi \quad (16)$$

$$\mathcal{L}_{cd} = \begin{cases} \frac{1}{\mathcal{D}_{dd}} \left[\mathcal{M}_{cd} - \sum_{r=1}^{d-1} \mathcal{L}_{cr} \mathcal{L}_{dr} \mathcal{D}_{rr} \right], & c > d \\ 0, & c < d \\ 1, & c = d \end{cases} \quad \text{for } c = 1, \dots, \psi, d = 1, \dots, \psi \quad (17)$$

A key assumption in using the ellipsoid approximation is that the volume of the ellipsoid varies monotonically with the equivalent convex hull volume. However, this is not always the case. Due to degeneracies in certain geometries, the ellipsoid approximation sometimes leads to duplicate points in the output space, which are not beneficial under the convex hull interpretation. Therefore, a penalty function based on the Euclidean norm between two output vectors is introduced to discourage clustering. A minimum squared distance, δ_{\min} , between any two vectors of output variables \hat{y}_i and $\hat{y}_{i'}$, where i, i' are indices corresponding to two experimental designs (i.e., two different sets of input factors), is given by

$$\delta_{\min} = \min_{i=1, \dots, n-1, i' > i} \delta_{i, i'} = \min_{i=1, \dots, n-1, i' > i} \sum_{c=1}^{\psi} (\hat{y}_{i,c} - \hat{y}_{i',c})^2. \quad (18)$$

Thus, δ_{\min} is bounded by a set of $\frac{n-1}{2}$ constraints

$$\sum_{c=1}^{\psi} (\hat{y}_{i,c} - \hat{y}_{i',c})^2 \geq \delta_{\min} \quad \forall i' > i. \quad (19)$$

A logarithmic penalty function, similar to barrier functions,²⁹ is adopted to favor larger values of δ_{\min}

$$f_{\text{out}} = V_{\text{obj}} + \mu \ln(\delta_{\min} + \epsilon_{\log}) - \mu \ln(\epsilon_{\log}) \quad (20)$$

The logarithmic form of function ensures that the penalty is most severe when the minimum squared distance approaches zero, and moderate elsewhere. The adjustable penalty parameter μ allows for tuning of the penalty magnitude according to the situation, with values typically ranging between 1 and 10, and a default value of $\mu = 3$ used here. A small value of ϵ_{\log} helps to avoid numerical issues near the logarithm of zero; a default value of 10^{-10} is used. The approximate penalized volume f_{out} is then used as the output space objective to be maximized.

Finally, as the aim is to design a set of experiments in the input space, a process model relating input and output variables is needed (e.g., mass and energy balances), as well as constraints on the feasibility of process (e.g., a maximum outlet air temperature). This is given by a constraint vector, $g(u, y) \leq 0$, represented here as a set of inequalities without loss of generality. The final formulation of the output space maximization problem is thus

$$\begin{aligned} \max_{u, y} \quad & V_{\text{obj}} + \mu \ln(\delta_{\min} + \epsilon_{\log}) - \mu \ln(\epsilon_{\log}) \\ \text{s.t.} \quad & V_{\text{obj}} = \beta d_{\mathcal{M}} \\ & d_{\mathcal{M}} = \prod_{c=1}^{\psi} \mathcal{D}_{cc} \\ & \mathcal{D}_{cd} = \begin{cases} \mathcal{M}_{cd} - \sum_{r=1}^{c-1} \mathcal{L}_{cr}^2 \mathcal{D}_{rr}, & c = d \\ 0, & c \neq d \end{cases} \quad \text{for } c = 1, \dots, \psi, d = 1, \dots, \psi \\ & \mathcal{L}_{cd} = \begin{cases} \frac{1}{\mathcal{D}_{dd}} \left[\mathcal{M}_{cd} - \sum_{r=1}^{d-1} \mathcal{L}_{cr} \mathcal{L}_{dr} \mathcal{D}_{rr} \right], & c > d \\ 0, & c < d \\ 1, & c = d \end{cases} \quad \text{for } c = 1, \dots, \psi, d = 1, \dots, \psi \\ & \sum_{c=1}^{\psi} (\hat{y}_{i,c} - \hat{y}_{i',c})^2 \geq \delta_{\min} \quad \forall i' > i \\ & g(u, y) \leq 0 \\ & u^{lb} \leq u \leq u^{ub}, \quad y^{lb} \leq y \leq y^{ub} \end{aligned} \quad (21)$$

where y^{lb} and y^{ub} denote lower and upper bounds on y , respectively, and $d_{\mathcal{M}}$ denotes the determinant of \mathcal{M} . For simplicity, the set of constraints in problem (21), excluding the vector g and the bound constraints on the inputs, is denoted by G_{out} in the remainder of the paper, i.e., the output space problem is

$$\begin{aligned} \max_{u,y} \quad & f_{out}(y) \\ \text{s.t.} \quad & G_{out}(y) \leq 0 \\ & g(u,y) \leq 0 \\ & u^{lb} \leq u \leq u^{ub} \\ & y^{lb} \leq y \leq y^{ub} \end{aligned} \quad (22)$$

Multiobjective formulation

Multiobjective optimization (MOO) permits the simultaneous consideration of both input and output space objectives. MOO enables the generation of a Pareto-optimal front, a set of trade-off points where improvement in one objective is not possible without a worsening in the value of other objectives. The Pareto-points are thus efficient designs, among which the decision-maker can choose based on business needs. Many alternative formulations are available for multiobjective optimization,³⁰ including the widely used L_p -norm formulation, Eq. 23, originally proposed by Lightner and Director³¹

$$\begin{aligned} \max_x \quad & \left(\sum_{i=1}^K (w_i f_i(x))^q \right)^{1/q} \\ \text{s.t.} \quad & x \in S \end{aligned} \quad (23)$$

where the f_i 's are the K objective functions, x is the vector of all the variables used to evaluate the objective functions, and defined over a feasible set S , w_i is a weight for function f_i , and q is a user-defined parameter that defines the norm to be used. The variables x are thus adjusted to maximize a norm of the weighted sum of the objective functions.

In our work, we utilize both the L_1 - and L_∞ -norm variants, which have complementary advantages and are easy to implement. For $q = 1$, the weighted sum formulation is obtained

$$\begin{aligned} \max_{u,y} \quad & w_{in} f_{in}(u) + w_{out} f_{out}(y) \\ \text{s.t.} \quad & g(u,y) \leq 0 \\ & g_{in}(u) \leq 0 \\ & A_{in} u \leq 0 \\ & G_{out}(y) \leq 0 \\ & u \in [u^{lb}, u^{ub}] \\ & y \in [y^{lb}, y^{ub}] \end{aligned} \quad (24)$$

The maximization of the input and output space objectives $f_{in}(u)$ and $f_{out}(y)$ with weights w_{in} and w_{out} takes place with respect to the process input and state variables u and the output space variables y . The process model and operation constraints, $g(u,y) \leq 0$, link these two variable vectors together. Global solutions to Problem (24) for the case of positive weights are guaranteed to be Pareto-optimal³²; however, this formulation cannot be used to generate points along nonconvex portions of the Pareto-front.^{33,34} Therefore, the infinity norm, the limit of Eq. 23 where $q \rightarrow \infty$, is also used to generate trade-off solutions. The algebraic interpretation of the L_∞ -norm is the weighted Tchebycheff formulation, Eq. (25)

$$\max_{x \in S} \min_{i=1, \dots, K} [\alpha_i f_i(x)] \quad (25)$$

which is commonly expressed as the following equivalent problem

$$\begin{aligned} \max_{x \in S, \gamma} \quad & \gamma \\ \text{s.t.} \quad & \alpha_i f_i(x) \geq \gamma \quad \forall i = 1, \dots, K \end{aligned} \quad (26)$$

Though this approach only guarantees weak Pareto-optimality, the formulation allows the identification of all Pareto-optimal points by varying the objective weights. The weights in this formulation, α_i , can also be set as the reciprocal of the desired values for each objective.³⁴ We use a slight modification of this formulation, where the weights are expressed as α_{in} and α_{out} , with $\alpha_i = \frac{1}{w_i}$, $i \in \{in, out\}$

$$\begin{aligned} \max_{u,y,\gamma} \quad & \gamma \\ \text{s.t.} \quad & f_{in}(u) \geq \gamma w_{in} \\ & f_{out}(y) \geq \gamma w_{out} \\ & g(u,y) \leq 0 \\ & g_{in}(u) \leq 0 \\ & A_{in} u \leq 0 \\ & G_{out}(y) \leq 0 \\ & u \in [u^{lb}, u^{ub}] \\ & y \in [y^{lb}, y^{ub}] \\ & \gamma \in \mathbb{R} \end{aligned} \quad (27)$$

Bilevel formulation

Bilevel optimization (BLO) provides a different approach to relate the input and output space objectives. Whereas with MOO a range of possible trade-offs between the two objectives is sought, BLO delivers a single solution that features the best possible value of the objective function for one of the optimization problems, with the other optimization problem serving as a constraint. With BLO, two separate decision makers — the leader and the follower — each optimize one of the objectives, with the leader selecting a decision first, then the follower. Two formulations of the two-space design of experiments problem are possible with BLO. The first is given in Problem (28), where the leader is focused on the maximization of the output space objective

$$\begin{aligned} \max_{u,y} \quad & f_{out}(y) && \text{output space objective} \\ \text{s.t.} \quad & G_{out}(y) \leq 0 && \text{output space objective calculation} \\ & \max_u f_{in}(u) && \text{input space objective} \\ \text{s.t.} \quad & g(u,y) \leq 0 && \text{process model} \\ & g_{in}(u) \leq 0 && \text{input space objective calculation} \\ & A_{in} u \leq 0 && \text{2D symmetry-breaking constraints} \\ & u \in [u^{lb}, u^{ub}] \\ & y \in [y^{lb}, y^{ub}] \end{aligned} \quad (28)$$

In this formulation, the process model, operating constraints and symmetry-breaking constraints are embedded within the inner optimization problem. In the case of the process model and operating constraints, this is because the constraints must be satisfied for both the leader (outer problem) and the follower (inner problem) to identify a physically realizable solution. As

previously, the intent of the symmetry-breaking constraints is to limit the number of redundant solutions for both problems.

An optimistic BLO formulation³⁵ is assumed, which implies that, when the follower is indifferent to several solutions among a set, the leader has the ability to choose the best solution from its perspective. In this formulation, the manufacturer, as leader, selects process output variables that maximize coverage of the process output space, and then identifies the set of input factors that match the outputs, while maximizing input space orthogonality, which can be seen as the regulator's objective.

The other possible BLO formulation is based on optimizing the input space (regulator's) objective as the outer problem, with output space coverage as the inner objective. This is well-suited to the case of a regulator setting policy and the manufacturer responding to it, or when the manufacturer wants to emphasize regulatory approval over output space exploration. Formulation (29) can be generated by interchanging the input and output space objectives and some of the constraints from (28)

$$\begin{aligned}
 & \max_{u,y} f_{in}(u) && \text{input space objective} \\
 \text{s.t. } & g_{in}(u) \leq 0 && \text{inputs objective calculation} \\
 & A_{in}u \leq 0 && \text{symmetry breaking cuts} \\
 & \max_y f_{out}(y) && \text{output space objective} \\
 \text{s.t. } & g(u,y) \leq 0 && \text{process model} \\
 & G_{out}(y) \leq 0 && \text{outputs objective calculation} \\
 & y \in [y^{lb}, y^{ub}] \\
 & u \in [u^{lb}, u^{ub}]
 \end{aligned} \tag{29}$$

Due to the special structure of Problems (28) and (29), in which the inner variables do not appear in the outer constraints or in the outer objective function, each bilevel problem can be reformulated as a pair of connected optimization problems as described in Ref. 36. In the case of Problem (28), in which the input space maximization is the inner (follower) problem, this is given by

$$\begin{aligned}
 & \max_{u,y} f_{in}(u) \\
 \text{s.t. } & f_{out}(y) \geq \xi \\
 & G_{out}(y) \leq 0 \\
 & g(u,y) \leq 0 \\
 & g_{in}(u) \leq 0 \\
 & A_{in}u \leq 0 \\
 & u \in [u^{lb}, u^{ub}] \\
 & y \in [y^{lb}, y^{ub}]
 \end{aligned} \tag{30}$$

where ξ is given by the solution of

$$\begin{aligned}
 \xi &= \max_{u',y'} f_{out}(y') \\
 \text{s.t. } & G_{out}(y') \leq 0 \\
 & g(u',y') \leq 0 \\
 & g_{in}(u') \leq 0 \\
 & A_{in}u' \leq 0 \\
 & u' \in [u'^{lb}, u'^{ub}] \\
 & y' \in [y'^{lb}, y'^{ub}]
 \end{aligned} \tag{31}$$

The solution of Problem (31), followed by that of Problem (30), gives the same result as solving Problem (28). Observe that Problem (31) is a relaxation of Problem (28), where maximization of the inner objective is no longer enforced as a constraint. Therefore, its optimal value ξ is an upper bound on Problem (28). Also note that the set of all feasible solutions of Problem (30) corresponds to the set of global optima for Problem (31) if ξ is set to the optimal value of Problem (31). Thus, the solution of Problem (30) corresponds to the point with the best possible value of the input space objective among the optimal solutions of the output space maximization. This implies that the constraint $f_{out}(y) \geq \xi$ is active at the solution of Problem (30). A similar strategy can be adopted for Problem (29).

Interestingly, as a result of the nature of this formulation, the solution of the bilevel problem is also Pareto-optimal. This can be understood based on the fact that at a Pareto point, an improvement in one objective must result in a worsening of the other objective. Consider the global solution (u^*, y^*) of Problem (28). As discussed, $f_{out}(u^*, y^*) = \xi$, where ξ is the solution of Problem (31), must hold. Therefore, there can be no $(u, y) \in [u^{lb}, u^{ub}] \times [y^{lb}, y^{ub}]$ such that $f_{out}(u, y) > f_{out}(u^*, y^*)$ and an improvement in the outer problem is not possible. The question remains of whether an improvement in $f_{in}(u)$ can result in an improvement in f_{out} . To explore this, consider any point (u^\dagger, y^\dagger) that satisfies $G_{out}(y^\dagger) \leq 0$, $g(u^\dagger, y^\dagger) \leq 0$, $g_{in}(u^\dagger) \leq 0$ and $A_{in}u^\dagger \leq 0$ and that provides an improvement in f_{in} : $f_{in}(u^\dagger) > f_{in}(u^*)$. To obtain a simultaneous improvement in f_{out} , we need $f_{out}(u^\dagger, y^\dagger) > f_{out}(u^*, y^*) = \xi$. However, this is impossible as ξ is the global optimal function value for Problem (31) and (u^\dagger, y^\dagger) is a feasible point in Problem (31). Therefore, the solution of bilevel problem (28) is Pareto-optimal. Again, a similar approach can be used to show that the solution of bilevel problem (29) is another Pareto-optimal point.

Solution strategy

The multiobjective and bilevel formulations are solved using the BARON global optimization software (version 12.7.7)³⁷ via the GAMS 24.2.3 modeling environment,³⁸ with the default optimality tolerance. To generate the Pareto front, the weights w_{in} and w_{out} are first varied in the interval $[0, 1]$ in increments of 0.1, with $w_{out} = 1 - w_{in}$ and Problems (24) and (27) are solved for each combination of weights. Intermediate weights are then added to fill any significant gaps along the Pareto-front. All GAMS files have been made available on zenodo.org (see Data Statement at the end of this article).

In certain runs, convergence to the desired tolerance was not attained with BARON, so some of the Pareto-optimal points presented in this work may be local optima. However, all obtained points are feasible, so they provide at least a pessimistic bound for the true Pareto-optimal value. In many of the problems for which convergence is not reached, increasing the available CPU resource does not lead to better local solutions. In view of the inherent approximation introduced by the use of the ellipsoid surrogate objective, the failure to guarantee the Pareto-optimality of all points does not in practice hinder the ability to provide valuable information to the decision maker.

Case Studies

The framework is applied to two case studies from pharmaceutical process development — one from drug product manufacturing and one from drug substance manufacturing.

Development of a tablet film coating process

Analysis Based on Two Output Variables. The proposed approach is first applied to the tablet coater presented as a motivating example, to illustrate how the framework can support better decision making by process engineers. The same ranges are used for the three input factors ($\phi=3$): $T_{\text{air,in}}$ ($^{\circ}\text{C}$) $\in [20, 85]$, M_{coat} (g/min) $\in [10, 80]$ and Q_{air} (ft³/min) $\in [150, 450]$. The process model with two output variables ($\psi = 2$), the outlet air temperature, $T_{\text{air,out}}$, and the outlet relative humidity, %RH_{out}, is given by $g(u, y) \leq 0$, representing the equations in Supporting Information Section 2.1.

The results of the Pareto and bilevel optimizations are shown in Figure 3. The two axes correspond to the two different objectives: the regulator's objective, D_T -optimality for a design with four experiments ($n = 4$), appears on the x -axis; the manufacturer's objective, the output space area, lies on the y -axis. It is shown as a scaled value relative to the maximum area that can be achieved for a four-experiment design, where the maximum area corresponds to the optimal value of V_{obj} in Problem (22) with $w_{\text{out}} = 1$ and $\mu = 0$. Both the ellipsoid approximation (EA) and the convex hull (CH) approximation of the output space objective are presented in the figure. A comparison of the two Pareto fronts indicates that the EA area results in a small overestimation of the CH area, but is nevertheless a good approximation overall in this case.

The results in Figure 3 highlight the benefits of using two objectives. It is clear that the input space and output space criteria are competing objectives and that focusing on one objective only, to the exclusion of the other, can in fact lead to a poor experiment design. To illustrate this, consider the two trade-off points highlighted in Figure 3 and their single-objective counterparts are listed in Table 2. The trade-off points offer much better options for process development than the single-objective ones.

This can be seen by comparing trade-off point 1 to the design obtained when maximizing the volume of the output space only ("Output Only" in the table). Both designs perform equally well in the output space, but the single-objective

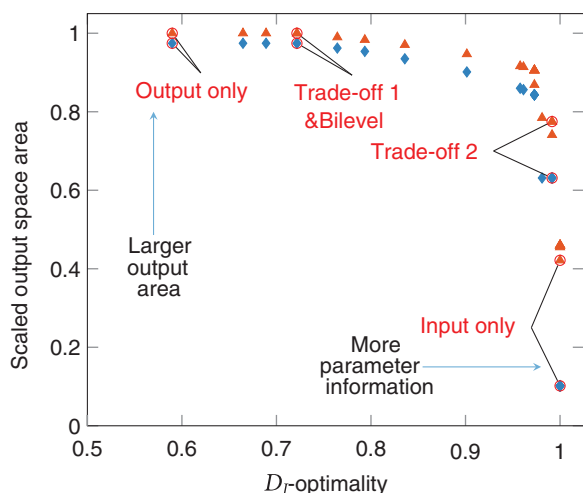


Figure 3. Pareto-front approximation for four-experiment coater design with two outputs, expressed as the scaled output space area vs. the D_T -optimality criterion.

Triangles denote the value of the scaled output space area in terms of the ellipsoid approximation, and the diamonds in terms of the convex hull approximation. [Color figure can be viewed at wileyonlinelibrary.com]

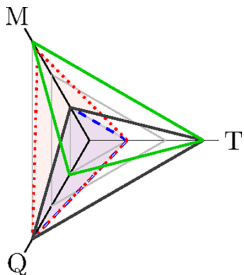
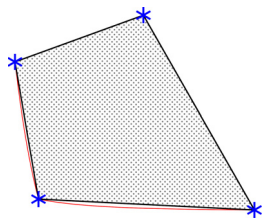
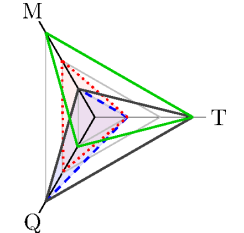
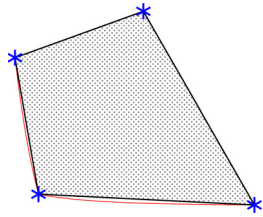
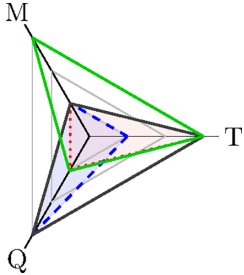
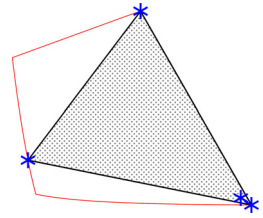
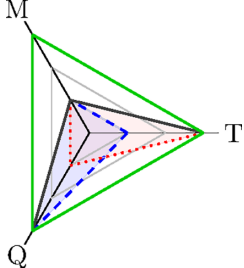
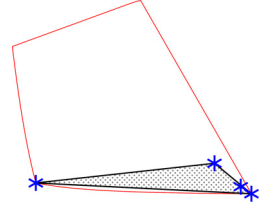
solution has a D_T -optimality of only 8% whereas the trade-off solution achieves a 72% D_T -optimality. This stark difference is due to the fact that, in the single-objective case, the solution is indifferent to the outcome in the input space; this is as likely to lead to a poor (input space) design from the perspective of the regulator as it is to lead to a good design. In this case study, a very poor design is achieved. Conversely, the trade-off point combines the desire for a large output area with the need to find a design with the best possible D_T -optimality value. The solution found, which corresponds to the bilevel solution, is much better as the proposed approach ensures that both objectives are taken into account.

Comparison of trade-off point 2 with the single-objective input space problem also shows the benefits of using two objectives to ensure that as much information as possible is derived from a single set of experiments. The trade-off design provides a nearly-maximal value of the D_T -criterion (over 99%) while maintaining a reasonable coverage of the output space, as seen in Table 2 (63% of the potential feasible area, as calculated with the convex hull approximation). In contrast, the single-objective case ("Input Only") is focused exclusively on optimizing the input-space information content. Such an approach is similar to the factorial design techniques that are commonly applied in industrial settings today, but it is found to lead to a poor exploration of the process output space. Indeed, while the single-objective design has a D_T -optimality of 100%, it covers only 10% of the entire feasible output region, as calculated with the convex hull approximation. Consequently, little process understanding is gained as the output space remains largely unexplored and this design may incur more expense and experiments in the future. Trade-off point 2, which covers over six times more output area at a cost of only a 1% decrease in D_T -optimality, is clearly superior to the single-objective design. This result further illustrates that failure to consider both spaces can lead to a missed opportunity in terms of the neglected objective.

Beyond these useful trade-off points, the proposed framework also allows the decision maker to visualize the impact of increasing or decreasing the number of experimental trials to perform. By increasing the number of experiments and hence the degrees of freedom for the experimental design, it may be possible to cover more of the output operating space while achieving a high D_T -optimality. However, this comes with increased costs for the time, equipment, and materials to perform additional experiments; these costs are not considered explicitly here. The process engineer must then decide whether each successive experiment and its corresponding gains in output space coverage and input space information are worth the additional expense.

These trade-offs are illustrated in Figure 4, which shows how the Pareto-front approximation changes as the number of experiments per design increases from three to eight. Here, the x -axis denotes the number of equivalent fully D_T -optimal experiments, a measure of total input space information content. Thus, a five-experiment design that has a D_T -optimality of 0.6 would be equivalent to a three-experiment fully D_T -optimal design (see the designs indicated by label A in the figure). Displaying the data in this way allows the decision maker to easily see what types of trade-offs are possible with differing numbers of experiments. For example, in deciding whether to do four or five experiments, one can see that with five trials, it is possible, using design B in the figure, to cover more than 90% of the output space, as indicated by the convex hull coverage fraction, without much loss of information in the input

Table 2. Results for Selected Four-Trial Spray Coater Experimental Designs

Case	D-Opt.	Input Space	Output Coverage	Output Space
Trade-off Point 1 (Bilevel)	72%		97%	
Output Only	8%		97%	
Trade-off Point 2	99%		63%	
Input Only	100%		10%	

Lines in the input space Kiviat diagrams illustrate scaled inlet temperature (T), coating solution mass flow rate (M), and inlet air volumetric flow rate (Q) values for each trial of a design. The output space diagrams illustrate the convex hull coverage (shaded region) of the total feasible space in the output variables, denoted by the outer thin red line. Each triangle in the input space Kiviat diagram corresponds to a proposed experimental trial and thus also to one of the blue stars in the output space diagrams. [Color figure can be viewed at wileyonlinelibrary.com]

space ($D_{I, \text{opt}} = 5$). Conversely, achieving a coverage of 90% or more with a four-experiment design requires losing about half of an experiment's worth of information in the input space. This is indicated by design C in the figure, where $D_{I, \text{opt}} \approx 3.5$.

Moreover, there is not much room for improvement in the output objective beyond five trials. Indeed, for five trials and more, one can achieve essentially the maximum D_I -optimality while exploring a reasonable fraction of the output space. The proposed approach gives the decision-maker additional insights on the number of experiments to perform in addition to the best measurement points to use for each experiment.

A promising five-experiment design (the “compromise” design) is compared to the four-experiment trade-off designs previously presented, as an illustration of how the Pareto-curve approximations in Figure 4 can be used to suggest alternatives not just between the two objectives, but also between carrying out more or fewer experiments. The details for this new five-trial design are given in Table 3, with an

accompanying visualization of the measurements in terms of the output space in Figure 5. This new design yields a D_I -optimality of 98%, with an output space area coverage of 96%. Compared to trade-off point 1 from the four-experiment case, this new compromise solution results in a small 1% sacrifice in the coverage (which remains high) and an increase in D_I -optimality by 26 percentage points. This makes the new D_I -optimality just 1% less than in the four-experiment trade-off design 2. Therefore, the Pareto-curve shows how the addition of one more experiment can yield large gains in either information efficiency or coverage fraction, helping the decision-maker to choose how many experiments to conduct.

It is also clear from Figure 4 why using the proposed framework provides better results than using a combination of two single-objective designs. In the single-objective four-experiment cases, the maximization of the input space objective yielded a perfect D_I -optimality but with only 10% output coverage and the output space maximization resulted in a

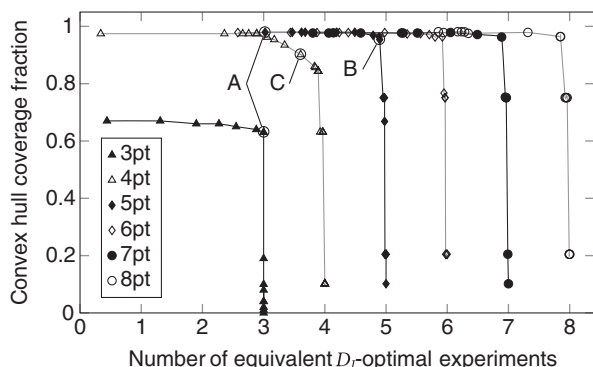


Figure 4. Pareto-front approximation for the tablet coating process with two outputs, demonstrating the compromise between output space coverage, as expressed by the convex hull coverage fraction, and input space optimality, expressed in terms of the number of equivalent D_T -optimal experiments.

Each symbol corresponds to a different number of designed experiments as indicated in the legend (e.g., “5pt” refers to a design with five experiments).

coverage of 99% of the output space with a very low D_T -optimality. As both designs had two points in common, it might reasonably be concluded that those two experiments could be shared between the designs, leading to a total of six experiments to achieve high performance in both objectives. In this composite design, the six points would cover 99% of the design space, providing a total information content slightly above four D_T -optimal experiment equivalents. Using the proposed framework, however, it can be seen that on the six-point Pareto-curve, designs exist with 99% coverage and an information content above five D_T -optimal experiment equivalents. Therefore, this example shows how the proposed framework provides value in decision making that would not be easily accessible by considering only single-space optimal designs.

Extension to Include Atomization Phenomena. Having demonstrated the utility of the proposed algorithm with a coater model that accounts for two output variables and three input factors, additional equations describing atomization properties of the coating spray are now appended to form an extended coater model. The Sauter Mean Diameter (SMD)³⁹ is added as a model output to characterize droplet size distribution. In the input space, the atomizing gas flow rate is added as the key metric.⁶ The model now has four input and three output dimensions and includes the equations given in Supporting Information Section 2.2.

Because the information matrix would be rank-deficient for three-experiment designs, the model is investigated only for four- to eight-experiment designs.

Table 3. A Five-Experiment Design for the Simple Tablet Coater Example with Two Output Variables

Exp	$T_{air,in}$ [°C]	M_{coat} [g/min]	Q_{air} [ft ³ /min]	$T_{air,out}$ [°C]	%RH _{out} [%]
1	20	10.5	150	16.3	33.7
2	20	10	450	18.8	20.1
3	23.7	80	450	14.2	71.0
4	85	10	450	83.4	16.2
5	85	80	157.6	50.2	87.4

D_T -optimality (% relative to maximum) = 98%. Output space area (% relative to maximum) = 96%.

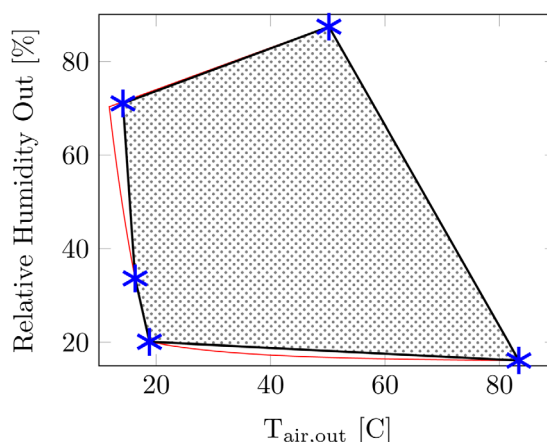


Figure 5. Output space covered by the five-experiment compromise design in Table 3 for the coater example.

The stars denote the two outputs for the five experiments, the shaded area is the corresponding convex hull and the outer lines denote the largest possible output space. [Color figure can be viewed at wileyonlinelibrary.com]

Despite the increased computational complexity arising from the increased dimensionality of the model, the results provide actionable insights. Examining the Pareto-front approximations in Figure 6 shows that the trade-off between the two objectives still stands. It has become harder to achieve high performance in both objectives. As with the simplified coater model, it is possible to examine the benefits of deploying differing numbers of experiments using the framework. Increasing the number of experiments in the extended model appears to lend additional flexibility to the trade-off with the output space objective, although the extent of the improvement diminishes as the number of experiments becomes larger. With one additional variable in each problem, at least seven experiments are now needed to achieve a coverage of the output space volume of over 80%. Under the proposed framework, the value of additional and expensive physical experiments can be assessed computationally before undertaking such an investment.

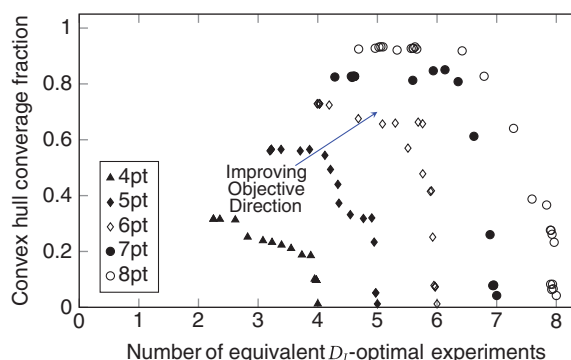


Figure 6. Pareto-front approximation demonstrating the compromise between output space coverage, input space optimality, and experimental effort for the extended coater model.

Each symbol corresponds to a different number of designed experiments as indicated in the legend (e.g., “5pt” refers to a design with five experiments). [Color figure can be viewed at wileyonlinelibrary.com]

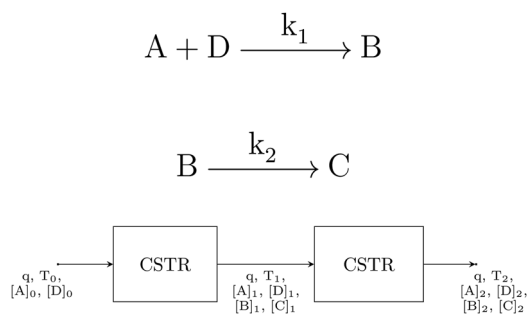


Figure 7. Schematic of the two reactors in series considered in the second case study, showing the reactions and flows involved.

The extended coater model offers a test of the framework in higher dimensional spaces. It is found that the surrogate output objective function has some limitations due to the maximum possible value of the determinant in Eq. 10 scaling exponentially with the number of experiments. Nevertheless, insights on the trade-offs inherent in experiment design are achieved. The visualization tools presented in this work are also shown to scale to higher dimensions, and to be of even greater relevance as the additional flexibility derived from a greater number of input variables is found to elicit a greater impact when changing the number of experiments in a design.

CSTR case study

Having demonstrated the applicability and limitations of the proposed framework to pharmaceutical spray coater development, it is applied to investigate the design of experiments for a system of two continuously stirred tank reactors (CSTRs) in series. This case study makes it possible to examine how the proposed framework performs with even greater dimensionality in the input space. It also provides an opportunity to investigate the benefits of the proposed experiment design methodology on the development of new continuous pharmaceutical processes, a context where the CSTR assumption has been used previously to investigate process development options. A model system of two reactions is examined: the formation of product B through the reaction of A and D and the degradation of B into waste product C. The reactions take place across two well-insulated CSTRs, as shown in Figure 7.

In this system, pure reactants A and D are fed into the first CSTR and allowed to react. The effluent stream is then fed into the second reactor to react further before being removed for more processing downstream. The system is assumed to operate at steady state with perfect mixing. The desired reaction (formation of B) is kinetically more favorable than the degradation of B and is exothermic.

Six inputs are included among the important factors. These include the inlet concentration of reactant A, $[A]_0$, the inlet feed ratio of the concentration of A to that of D, $[A]_0/[D]_0$, the feed flow rate, q , the reactor volumes V_1 and V_2 and the inlet feed temperature T_0 . The choice of feed flow rate and reactor volumes effectively sets the residence times for the lower temperature (first) and higher temperature (second) reactors. These factors are explored with the D_F -optimality objective.

In the output space, there are three critical quality attributes: the exit mole fraction of impurity C, x_{C2} , the exit mole fraction of reactant A, x_{A2} , and the exit temperature, T_2 . The impurity mole fraction in the product stream must be kept under 0.002 to remain within specifications. As a measure of reaction completeness, the exit mole fraction of reactant A must remain below 0.02 because there is no recycle in the system. Finally, the exit temperature must be kept under 85°C to prevent degradation of reactant D. Within these constraints, the feasible output space should be explored to as great an extent as possible.

The penalty parameter (μ in Eq. 20) also needs to be adjusted based on the number of experiments to avoid degenerate results in which clustering of the outputs occurs as a result of the convex hull approximation used. This issue is illustrated in Table 4 for an eight-experiment design. While the outputs of the first two experiments have reasonable separation, the next six points are arranged in three clusters, denoted by groups C through E in the table, for which the values of x_{C2} , x_{A2} , and T_2 are very similar to each other.

This clustering can be better visualized in Figure 8. The use of the surrogate ellipsoid objective function with the default value of $\mu = 3$ results in the placement of all output points at extremities of the space to increase the variance of the distribution, but regions of the feasible space are missed in doing so.

Although the default penalty multiplier value of 3 prevents the cluster points from overlapping, it is not sufficient to correct the behavior for all the instances of this case study, in particular for the seven- and eight-experiment problems. A number of tests are therefore conducted by varying the penalty multiplier for the seven- and eight-experiment reactor model designs, as listed in Supporting Information Section 3. The lowest penalty multiplier values that mitigate the clustering behavior are selected for further investigation: $\mu = 5$ in the seven-experiment case and $\mu = 8$ for the eight-experiment case.

Using these penalty multiplier values, we obtain the approximate Pareto curves in Figure 9. The relative positions of the trade-off curves indicate that there is limited value in making small increments to the number of experiments for the purpose of mapping out more of the output space. The main benefits of such increments are felt in the input space. Conversely, trying to achieve a coverage of the output space very close to

Table 4. Inputs and Outputs for an Eight-Experiment Design for the Reactor Case Study

Exp #	$[A]_0$ [M]	$[A]_0/[D]_0$ [-]	q_0 [L/h]	V_1 [L]	V_2 [L]	T_0 [°C]	x_{C2} [-]	x_{A2} [-]	T_2 [°C]	[Group]
1	0.800	0. (-)800	0.025	0.729	2.000	22.0	0.0020	0.020	67.9	A
2	0.800	0.800	0.036	1.743	1.826	22.0	0.0018	0.020	67.9	B
3	1.079	0.800	0.029	0.880	0.901	22.0	0.0020	0.015	84.3	C
4	1.085	0.800	0.040	1.944	0.500	22.0	0.0020	0.020	84.2	D
5	1.092	0.800	0.019	0.556	0.569	22.0	0.0020	0.015	85.0	C
6	1.099	0.800	0.020	0.500	0.501	22.0	0.0016	0.020	85.0	E
7	1.099	0.806	0.037	1.698	0.500	22.0	0.0020	0.020	85.0	D
8	1.099	0.800	0.021	0.500	0.501	22.0	0.0016	0.020	85.0	E

The "Group" column indicates clusters in the outputs.

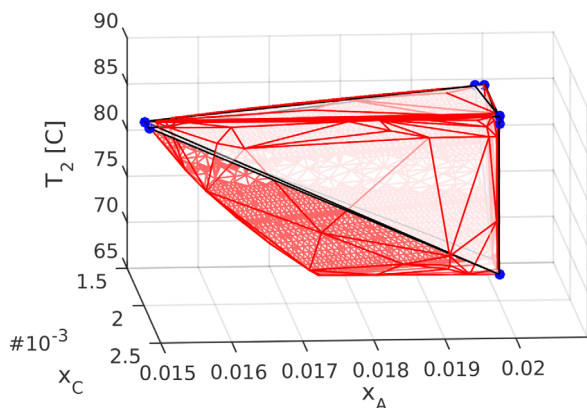


Figure 8. Convex hull (thick lines connecting the circles, lighter shading) formed by eight measurement points (circles) in the output space compared to the total feasible region (all lines, all shading).

[Color figure can be viewed at wileyonlinelibrary.com]

100% may require significant experimental effort. It can be seen that achieving a high coverage percentage (above 80%) involves the loss of approximately one D_T -optimal experiment equivalent of information.

From a computational perspective, the increased dimensionality of this case study presents some challenges. The D_T -optimality calculation involves the computation of the determinant of the information matrix, as explained in the Methodology section. When all experimental measurements are scaled to $[-1, 1]$, the information matrix values are bounded by $[-n, n]$, where n is the number of experimental trials in the design. The information matrix determinant is then bounded by $[0, n^\phi]$, where ϕ is the number of input factors. This leads to numerical difficulties: as ϕ increases, the D_T -criterion scaling becomes progressively worse. This may contribute to the gaps observed in the Pareto-front approximation at higher D_T -optimality weightings, such as for the seven-experiment case seen in Figure 9, as the increased value of the determinant may cause scaling issues.

In fact, scaling issues also make it much more difficult for the D_T -criterion optimization problem to converge, even when

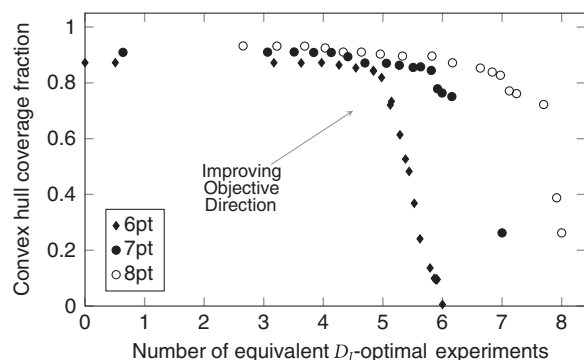


Figure 9. Pareto-front approximation for the reactor case study, illustrating the trade-offs between output space coverage, input space optimality, and experimental effort for the reactor model.

Each symbol corresponds to a different number of designed experiments as indicated in the legend (e.g., “7pt” refers to a design with seven experiments).

Table 5. Input Space D_T -Optimality Runs for the Reactor Model

# Exps	D_T -Criterion	Upper Bound	Theoretical Max	Run Time
6	683	46,635	46,656	3 days
7	2182	117,619	117,649	3 days
8	4331	262,124	262,144	3 days

considering the input space only. The theoretical maximum D_T -criterion values that can be achieved for systems of six input factors and six through eight experiments are shown Table 5. They are given by n^6 , where n is the number of experiments in the design and 6 is the number of input variables in the reactor case study, based on assuming the best possible realization of the information matrix, which may or may not be achievable in reality given problem constraints. Also reported in the table are the results of runs to find the global optima of the D_T -criterion value. Due to the difficulty of bounding the multilinear determinant calculation, the calculated upper bounds obtained with BARON were not improved significantly from their theoretical maximum values by the solver, even after three days of CPU time (see Table 5). Nevertheless, feasible points were found for all the cases, providing a lower bound on the achievable D_T -optimality.

The inability to prove convergence to the global solution does not deter from the usefulness of the proposed approach. For a given point in Figure 9, failure to prove global optimality implies there may exist another point with the same D_T -optimality and a better coverage of the output space. However, even in such a case, the solution obtained is a valid design — the point represents a feasible set of experiments and the corresponding values of D_T -optimality and output coverage are evaluated correctly regardless of whether full convergence has been achieved. As an example of the usefulness of the points obtained, consider the eight-experiment designs in Figure 9. It is clear that the point with the third highest number of equivalent D_T -optimal experiments (slightly below 8) is a very good compromise point, which enables a significant increase in the output coverage relative to the case where only D_T -optimality is taken into account. The set of valid designs identified thus allows the decision maker to choose between prioritization of input space information or output space exploration, while maintaining good performance from both perspectives.

Conclusions

Model-based experiment design enables research and development organizations to leverage the information already encoded into a mathematical model to direct experimental efforts. Enhancing the extent of process understanding obtained from each experiment allows for a reduction in the number of experiments necessary during process development, saving resources and potentially accelerating time-to-market. In a resource-constrained environment, it is desirable to coordinate the objectives of an experimental campaign. In this article, we have demonstrated the value of designing experiments that will result in an optimal bracketing study for the process parameters (input space objective), while also maximizing the understanding or exploration of the output space (output space objective). We have put forward a framework for optimal design of experiments based on multiobjective and bilevel optimization. It has been found to be a useful tool to identify an experiment design that achieves an optimal trade-off between the two objectives. The presented method has been

illustrated with two examples, one from drug product manufacturing (tablet coating) and one from drug substance manufacturing (use of two continuous reactors in series). An investigation of the space of Pareto-optimal designs in both cases has shown that focusing on only one of the two objectives results in poor performance in the other and that much better trade-off or compromise solutions exist on the Pareto curve. While global optimization software has been used to find Pareto points, it is found that for larger dimensions, it is not always possible to guarantee global optimality. This can be partly attributed to the deterioration of problem scaling as the number of input parameters increases. Despite this challenge, the experiment designs obtained provide a significant improvement over a single-objective approach. Further extensions of the proposed framework could include the use of different optimization algorithms to increase the likelihood of finding global solutions and the quantification of the impact of uncertainty in the model parameters on the output space objective. This latter consideration could enable the robust design of optimal experiments. Even in the deterministic case considered here, the outcome of the experiments as designed yields significant value without increasing the experimental cost: it expands the knowledge of the process output space and simultaneously provides a bracketing study to complement the process development section of a common technical document.

Acknowledgments

This contribution was identified by nominator Zhigang Sun (US Food Drug Administration) as the Best Presentation in the session “Application of Quality by Design in Drug Product Formulation Design & Process Development” of the 2015 AIChE Annual Meeting in Salt Lake City. Financial support for this work from the US-UK Fulbright Commission and from the Engineering and Physical Sciences Research Council of the UK via a Leadership Fellowship (EP/J003840/1) is gratefully acknowledged. We would also like to thank Carla Luciani for her assistance with the reactor model.

Data Statement

Data underlying this article can be accessed on Zenodo at <https://doi.org/10.5281/zenodo.805943>, and used under the Creative Commons Attribution license.

References

- Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, Schacht AL. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov*. 2010;9(3):203-214.
- DiMasi JA. The value of improving the productivity of the drug development process: faster times and better decisions. *Pharmacoconomics*. 2002;20(Suppl. 3):1-10.
- DiMasi JA, Hansen RW, Grabowski HG, Lasagna L. Cost of innovation in the pharmaceutical industry. *J Health Econ*. 1991;10(2):107-142.
- Food and Drug Administration. Pharmaceutical CGMPs for the 21st Century - A risk-based approach. Tech. Rep. U.S. Department of Health and Human Services, September 2004.
- Franceschini G, Macchietto S. Model-based design of experiments for parameter precision: state of the art. *Chem Eng Sci*. 2008;63(19):4846-4872.
- Müller R, Kleinebudde P. Comparison study of laboratory and production spray guns in film coating: effect of pattern air and nozzle diameter. *Pharm Dev Technol*. 2006;11(4):425-433.
- Prpich A, Am Ende MT, Katschner T, Lubczyk V, Weyhers H, Bernhard G. Drug product modeling predictions for scale-up of tablet film coating: a quality by design approach. *Comput Chem Eng*. 2010;34(7):1092-1097.
- Maheshwari V, Rangaiah GP, Samavedham L. Multiobjective framework for model-based design of experiments to improve parameter precision and minimize parameter correlation. *Ind Eng Chem Res*. 2013;52(24):8289-8304.
- Maheshwari V, Rangaiah GP, Samavedham L. A Novel multi-objective optimization based experimental design and its application for physiological model of type 1 diabetes. In: *IFAC Proceedings Volumes*, 2012;45:638-643. Elsevier, <https://org.doi/10.3182/20120710-4-SG-2026.00117>
- Kao MH, Mandal A, Lazar N, Stufken J. Multi-objective optimal experimental designs for event-related fMRI studies. *Neuroimage*. 2009;44(3):849-856.
- Telen D, Houska B, Logist F, Impe JV. Multi-purpose economic optimal experiment design applied to model based optimal control. *Comput Chem Eng*. 2016;94:212-220.
- Mueller R, Kleinebudde P. Comparison of a laboratory and a production coating spray gun with respect to scale-up. *AAPS PharmSciTech*. 2007;8(1):3.
- Chatterjee S. Design space considerations. In: *AAPS Annual Meeting*. Chicago, 2012:1-29.
- Aulton ME, Twitchell AM. Solution properties and atomization in film coating. In: Michael A, Graham C, John H, editors. *Pharmaceutical Coating Technology*. London: CRC Press, 1995:64-117.
- Kamalid A. OOS Results in FDA warning letters: Meeting of the Southern California Pharmaceutical Discussion Group (SCPDG) of AAPS. Nov 8, 2012. http://www2.aaps.org/uploadedFiles/Content/Sections_and_Groups/Regional_Discussion_Groups/Southern_California_Pharmaceutical/Announcements/SCPDG%20OOS%20Presentation-%20Nov.%208%20event.pdf
- Am Ende MT, Herbig SM, Korsmeyer RW, Chidlaw MB. Osmotic drug delivery from asymmetric membrane film-coated dosage forms. In: Wise DL, editor. *Handbook of Pharmaceutical Controlled Release Technology, chap. 36*. New York: CRC Press, 2000:751-786.
- Ebey GCA. Thermodynamic model for aqueous film-coating. *Pharm Technol*. 1987;11(4):40-50.
- Am Ende MT, Berchielli A. A thermodynamic model for organic and aqueous tablet film coating. *Pharm Dev Technol*. 2005;10(1):47-58.
- Pukelsheim F. Matrix Means. In: *Optimal Design of Experiments, Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, 2006:135-157.
- Atkinson AC, Donev AN. *Optimum Experimental Designs*. Oxford: Oxford Science Publications, Clarendon Press, 1992.
- Fedorov VV. *Theory of Optimal Experiments*. New York: Academic Press, 1972.
- Golub GH, Van Loan CF. *Matrix Computations*. 4th ed. The Johns Hopkins University Press, Baltimore, Maryland, 2013.
- Bárány I, Füredi Z. Computing the volume is difficult. *Discrete Comput Geom*. 1987;2(4):319-326.
- Bringmann K, Friedrich T. Approximating the volume of unions and intersections of high-dimensional geometric objects. *Comput Geom*. 2010;43(6-7):601-610.
- Dyer M, Frieze A, Kannan R. A random polynomial-time algorithm for approximating the volume of convex bodies. *J ACM*. 1991;38(1):1-17.
- Lovász L, Vempala S. Simulated annealing in convex bodies and an volume algorithm. *J Comput Syst Sci*. 2006;72(2):392-417.
- Bueler B, Enge A, Fukuda K. Exact volume computation for polytopes: a practical study. In: Kalai G, Ziegler GM, editors. *Polytopes: Combinatorics and Computation*. Basel: Springer Science & Business Media, 2000:131-154.
- Floudas CA, Gounaris CE. A review of recent advances in global optimization. *J Glob Optim*. 2009;45(1):3-38.
- Bazaraa MS, Sherali HD, Shetty CM. *Nonlinear Programming: Theory and Algorithms*. New York: John Wiley & Sons, 1993.
- Miettinen K. *Nonlinear Multiobjective Optimization*. Boston, MA: Springer US, 1998.
- Lightner M, Director S. Multiple criteria optimization and statistical design for electronic circuits. *Tech. Rep. DRC-13-4-79*, Carnegie Mellon University, 1979.
- Zadeh L. Optimality and non-scalar-valued performance criteria. *IEEE Trans Automat Contr*. 1963;8(1):59

33. Das I, Dennis JE. A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems. *Struct Optim.* 1997; 14(1):63-69.
34. Clark P, Westerberg A. Optimization for design problems having more than one objective. *Comput Chem Eng.* 1983;7(4): 259-278.
35. Dempe S. Foundations of bilevel programming. In: *Nonconvex Optimization and its Applications*, vol. 61. Dordrecht: Springer US, 2002.
36. Waltz F. An engineering approach: hierarchical optimization criteria. *IEEE Trans Automat Contr.* 1967;12(2):179-180.
37. Tawarmalani M, Sahinidis NV. A polyhedral branch-and-cut approach to global optimization. *Math Program.* 2005;103(2): 225-249.
38. Brook A, Kendrick D, Meeraus A. GAMS, a user's guide. https://www.gams.com/latest/docs/UG_MAIN.html#UG_INSTALL
39. Aliseda A, Hopfinger E, Lasheras J, Kremer D, Berchielli A, Connolly E. Atomization of viscous and non-Newtonian liquids by a coaxial, high-speed gas jet. Experiments and droplet size modeling. *Int J Multiphase Flow.* 2008;34(2):161-175.

Manuscript received Jan. 19, 2018, and revision received Apr. 7, 2018.