PROGRESSIVE RELAXATIONS FOR EFFICIENT DETERMINATION OF CONSERVATIVE DESIGN SPACES

D.J. Laky[1,*], M.L. Bynum[2], S. Vaidyaraman[3], S. García Muñoz[3], C.D. Laird[4]

[1]University of Wisconsin – Madison,
Madison, WI 53706, USA
[2]Discrete Mathematics and Optimization, Sandia National Laboratories
1515 Eubank SE, Albuquerque, NM 87185, USA
[3]Small Molecule Design and Development, Lilly Research Laboratories,
Eli Lilly & Company, Indianapolis, IN 46285, USA
[4]Carnegie Mellon University
Pittsburgh, PA 15289, USA

Abstract

In recent decades, quality-by-design and quality-by-control have sparked interest in identifying the design space of pharmaceutical processes. The design space of a pharmaceutical process is defined as the operating region in which quality assurance is guaranteed over the interaction of process inputs and model parameters. Traditional approaches to identify such regions are often expensive, which led to the reemergence of flexibility and feasibility analysis as a mathematical tool to substitute sampling-driven approaches. However, solving such nonlinear programs to global optimality can become a difficult or intractable task. In this work, progressive relaxations of the feasibility analysis problem are used to solve a series of mixed-integer, quadratically constrained programs to identify the probabilistic design space of a pharmaceutical process. The method is shown to approach the global solution and represents a practical approach to identify a conservative design space.

Keywords

Flexibility Test, Critical Quality Attributes, Global Optimization, Progressive Relaxation

Introduction

In 2004, the FDA launched the quality-by-design (QbD) initiative to prevent a shortage of critical medicines due to poor quality in pharmaceutical manufacturing processes (FDA 2004). Subsequently, in 2009, the so-called design space was defined to assist approval of operating procedures during pharmaceutical process design and in pharmaceutical manufacturing processes. The design space represents a region of quality assurance for a pharmaceutical process over the full interaction between process inputs and parameters (FDA 2009).

Quality in a pharmaceutical process is ensured by strictly adhering to product or process-specific critical quality attributes (CQAs). Naturally, the design space for a given pharmaceutical process is identified by the limits of the CQAs over the range of process inputs and system parameters. However, system parameters are often estimated from experimental data and are inherently uncertain. How well the CQAs hold over the entire random space of these uncertain parameters is not captured with a standard design space. Therefore, the so-called probabilistic design space can be identified as the region for which all CQAs hold with confidence α while considering model uncertainty.

Identification of such a region typically requires rigorous experimentation to understand the impact of input variables on the CQAs. These interactions are then used to fit mechanistic or data-driven models of a given process. These models, with or without their associated uncertainty,

are then used to generate a design space. Subsequent confirmatory experiments are run to substantiate the design space found during the modeling phase. Methods using this structure for design space identification through surface modeling (Kumar, Gokhale and Burgess, 2014), or PCA/PLS (Facco et. al. 2015) have been successful. However, time and cost of experiments as well as the requirement of high-quality data may become significant obstacles when employing such data-driven techniques.

Processes where mechanistic models are unavailable or intractable to use, but have significant data, have been explored. Boukouvala, Muzzio and Ierapetritou (2010) identify the probabilistic design space without explicit models. The work discusses identifying regions in black-box processes with various sampling techniques. Over recent years, progress on feasibility analysis using black-box and surrogate-based modeling in pharmaceuticals has been made. Rogers and Ierapetritou (2015a; 2015b) gave a good perspective on general feasibility analysis using surrogate and data-driven modeling. Wang and Ierapetritou (2017) used radial basis functions for the same task and later applied these techniques globally to pharmaceutical processes (Wang and Ierapetritou 2018). Most recently, Ding and Ierapetritou (2021) applied these strategies to continuous chromatography in biopharmaceutical manufacturing. García-Muñoz et al. (2015) identified the probabilistic design space by evaluating CQAs over process uncertainties through Monte Carlo simulation. With accurate distributions or ample data, sampling techniques provide excellent approximation of the probabilistic design space but carry significant computational cost.

In general, flexibility in manufacturing may be viewed as the ability to mitigate or remove the impact of system disturbances caused by uncertainty within or around a process. Specifically, the flexibility of a process and the tradeoff of this flexibility with operational cost has been extensively studied in the literature through mathematical programming (Halemane and Grossmann, 1983; Grossmann, Halemene and Swaney, 1983; Grossmann and Floudas, 1987; Pistikopoulos and Grossmann, 1989a; Pistokopoulos and Grossmann 1989b; Ostrovsky et al. 1994; Pistikopoulos and Ierapetritou, 1995).

Since then, the field of flexibility analysis has reemerged in relevance for design space identification of pharmaceutical processes and other fields. Pulsipher and Zavala (2018) developed a framework considering gaussian uncertainty for flexibility analysis which was later used to construct a general tool (Pulsipher and Zavala 2019). Laky et al. (2019) developed a framework to identify probabilistic designs spaces in pharmaceutical processes with computationally efficient approaches when compared to sample-based techniques. Ochoa et al. (2019) developed a new framework in flexibility analysis specifically for design space identification and extended the formulation to address uncertain parameters that are unmodeled (Ochoa et al. 2020). Recent advances using black-box models for flexibility analysis (Zhao et al. 2021) and the so-called design centering problem (Ochoa et al. 2021; Zhao et al. 2022) have also been made.

For a pharmaceutical process, flexibility may be interpreted as the maximum allowable deviation from nominal operation that satisfies the CQAs. In general, satisfying a CQA in a pharmaceutical process may be represented mathematically by the following inequality

$$g(\theta_p, x, z, \theta_m) \leq 0 \tag{1}$$

Here, θ_p are the process parameters, $\theta_m \epsilon T_m$ are the uncertain parameters, x are the state variables of the system, and z are the control variables. With this idea in mind, the feasibility of a process as a function of the process parameters, what we refer to as the *flexibility test* (FT), may be evaluated with the following formulation:

$$\chi(\theta_p) = \max_{\theta_m \epsilon T_m} \min_{z \epsilon T_z} \max_{k \epsilon K} g_k(\theta_p, x, z, \theta_m) \tag{2}$$

$$\text{s.t. } h_l(\theta_p, x, z, \theta_m) = 0 \qquad \forall\, l \epsilon L \tag{3}$$

In the formulation above, the goal is to find the maximum violation of the constraints g_k, over the uncertainty space T_m, while model equations, h_l, hold and control actions, z, assist. If the maximum value is below 0, (i.e., Eq. (1) holds for all CQAs), then the conditions θ_p over all $\theta_m \epsilon T_m$ are feasible. Grossmann and Floudas (1987) provide a good overview of such feasibility and flexibility formulations. In previous work (Laky et al. 2019), we modified the flexibility test formulation to iteratively identify the largest design space over θ_p in which all CQAs hold for a pharmaceutical process. Because of the nature of the problem, global solution of formulation (FT) is required to ensure the feasible space is not overestimated. However, determining global solutions to such a formulation can be computationally challenging for non-convex problems.

In this work, a modification to previous work (Laky et al. 2019) on the flexibility test formulation is proposed. The problem is relaxed, and a design space is generated using this relaxed problem. If the region is sufficiently large, then we have confirmed flexibility. However, if the region is too small, progressively tighter relaxations are implemented until the region is satisfactory or the gap is closed. This method provides a low-cost screening approach to mitigate the difficulty of solving the unrelaxed program globally.

Problem Definition

The purpose of this section is to provide background on the flexibility test problem and present a progressive relaxation approach to the algorithm developed by Laky et al. (2019). In the following section, the new method will be analyzed on an industrial case study.

Here, θ_p are the process parameters, or the process and design inputs (i.e., feed concentrations, residence time, temperature profiles, etc.). θ_m, are model parameters with implicit uncertainty (i.e., reaction rate coefficients, Arrhenius coefficients, heat transfer coefficients, etc.). It is assumed that this uncertainty has been estimated from

experimental or simulated data. Variables x represent the state variables (i.e., species concentrations), and z represent the control variables.

In a traditional extension of formulation (FT), the flexibility index, or the measure of the size of the design space, may be maximized with respect to the process variables as $\hat{\theta}_m$ shown in the *flexibility index* formulation (FI) below (Grossmann and Floudas 1987):

$$F(\theta_p) = \max_{\delta} \delta \quad (4)$$

$$\text{s.t. } \chi(\theta_p) = \left(\max_{\theta_m \epsilon T_m} \min_{z \epsilon T_z} \max_{k \epsilon K} g_k(\theta_p, x, z, \theta_m)\right) \leq 0 \quad (5)$$

$$\text{s.t. } h_l(\theta_p, x, z, \theta_m) = 0 \qquad \forall\, l \epsilon L \quad (6)$$

$$\delta \cdot \Delta\theta_m^- \leq (\theta_m - \hat{\theta}_m) \leq \delta \cdot \Delta\theta_m^+ \quad (7)$$

Of course, such a problem is cumbersome to solve as the constraint $\chi(\theta_p)$ is a complex, multi-level problem that must be solved globally. An important distinction for this problem is that δ is solved over the uncertainty space, and θ_p, the design variables, are inputs. In later formulations, the interest will be placed on δ in the θ_p space with formulation (FT*).

Formulation (FI) may be solved through a series of simplifications. The first is removing the inner-most maximization by introducing variable u. Second, a major simplification identified by Swaney and Grossmann (1985) by employing an active set strategy may be used. The simplification exploits the fact that it is guaranteed that a subset of constraints must be active at the solution to (FI).

As shown in Grossmann and Floudas (1987), when the number of control variables is zero (i.e., $n_z = 0$), a simplification to (FI) can be made for the resulting *active set flexibility index* formulation (FI*):

$$F(\theta_p) = \min_{\delta, x, s, y, \theta_m} \delta \quad (8)$$

$$\text{s.t. } h_l(\theta_p, x, \theta_m) = 0 \qquad \forall\, l \epsilon L \quad (9)$$

$$s_k + g_k(\theta_p, x, \theta_m) = 0 \qquad \forall\, k \epsilon K \quad (10)$$

$$\sum_k y_k = 1 \quad (11)$$

$$s_k - U(1 - y_k) \leq 0 \qquad \forall\, k \epsilon K \quad (12)$$

$$s_k \geq 0, y_k \epsilon \{0, 1\} \qquad \forall\, k \epsilon K \quad (13)$$

$$\delta \geq 0 \quad (14)$$

$$\delta \cdot \Delta\theta_m^- \leq (\theta_m - \hat{\theta}_m) \leq \delta \cdot \Delta\theta_m^+ \quad (15)$$

Formulation (FI*) may be solved globally to provide the maximum deviation, δ, over the hyperrectangle in Eq. (15) for which the CQAs of the process hold. The active set nature of the problem arises from eq. (11) where one of the CQAs has the worst constraint value over the entire uncertainty space. Here, $\hat{\theta}_m$ represents the nominal value of the model parameters θ_m.

Although the hyperrectangle in Eq. (15) represents a valid region, statistical information on the probability of lying within a hyperrectangle is not easily made explicit. For this reason, Rooney and Biegler (1999) suggested the use of an ellipsoid region. Using this terminology, a constraint to substitute Eq. (15) may be represented as:

$$(\theta_m - \hat{\theta}_m)^T \Sigma_{\theta_m}^{-1} (\theta_m - \hat{\theta}_m) \leq \delta \quad (16)$$

Here, the value of δ corresponds to $\chi^2_{n_m}(\alpha)$. This is the value of the chi-squared distribution with n_m degrees of freedom that corresponds to probability level α. One may use an LDL decomposition to reduce Eq. (16) to a set of linear constraints and sum of squares, as is used in this work, Pulsipher and Zavala (2018), and Laky et al. (2019).

In previous work (Laky et al. 2019), (FI*) is used to identify a probabilistic design space of a pharmaceutical synthesis process. However, an iterative approach can be used on a modification of formulation (FT) to identify a region that corresponds to a flat confidence α over the entire design space. The modification of (FT) can be solved by introducing another shape factor in the process parameter space, δ_p^r, which is defined below as formulation (FT*):

$$\chi(\delta_p^r, \delta_m^\alpha) = \max_{u, \theta_p, x, s, y, \theta_m} u \quad (17)$$

$$\text{s.t. } h_l(\theta_p, x, \theta_m) = 0 \qquad \forall\, l \epsilon L \quad (18)$$

$$s_k + g_k(\theta_p, x, \theta_m) - u = 0 \qquad \forall\, k \epsilon K \quad (19)$$

$$\sum_k y_k = 1 \quad (20)$$

$$s_k - U(1 - y_k) \leq 0 \qquad \forall\, k \epsilon K \quad (21)$$

$$s_k \geq 0, y_k \epsilon \{0, 1\} \qquad \forall\, k \epsilon K \quad (22)$$

$$\Sigma_{\theta_m}^{-1} = LDL^T \quad (23)$$

$$q^T = (\theta_m - \hat{\theta}_m)^T LD^{1/2} \quad (24)$$

$$q^T q \leq \delta_m^\alpha \quad (25)$$

$$\delta_p^r \cdot \Delta\theta_p^- \leq (\theta_p - \hat{\theta}_p) \leq \delta_p^r \cdot \Delta\theta_p^+ \quad (26)$$

Formulation (FT*) takes two inputs and returns the maximum violation of the CQAs. In the standard flexibility index formulation (FI), δ is a decision variable over $\theta_m \epsilon T_m$, however in (FT*), δ_m^α is the value corresponding to the chi-squared distribution with fixed confidence level α. Because the confidence level for $\theta_m \epsilon T_m$ is fixed, an *operating region* for θ_p may be found instead of an *operating point* where δ is maximized. Thus, the operating region in θ_p is defined by a new shape factor, δ_p^r, which in this case is a hyperrectangle region. Using bounds from 0 to 1 for δ_p^r, a bisection search can be used to find the maximum value of δ_p^r, or the largest operating region in θ_p where process CQAs are satisfied with confidence level α. A more detailed description of the algorithm to determine δ_p^r is shown in Laky et al. (2019).

Unfortunately, (FT*) must still be solved to global optimality at each iteration. Often, knowing the full extent of the design space is not necessary. Rather, we only require that a pre-specified design space be confirmed which may be possible even with a conservative estimate of the design space. Exploiting this property of the probabilistic design space allows the proposal of a progressively refined relaxation of the nonlinear program in (FT*). The method for doing this is as follows.

An upper bound on the maximum constraint violation, u, can be obtained by solving a relaxation of (FT*) for a given δ_p. Because the optimal value of u is a non-decreasing function of δ_p, any δ_p for which the maximum u of the relaxed problem is less than or equal to 0 is guaranteed to

satisfy the CQAs for the original problem. Thus, the final value of δ_p^r, δ_p^*, obtained from the bisection method using the relaxed problem will necessarily be less than or equal to the true value of δ_p^* for (FT*).

Many methods exist for constructing relaxations of general MINLPs, including αBB, generalized McCormick relaxations, and factorable programming (Adjiman et al. 1998; Tsoukalas and Mitsos, 2014; Misener and Floudas, 2014). Here, we construct convex MIQCP relaxations (i.e., the continuous relaxation is a convex quadratically constrained program) using the factorable programming method in Coramin (Bynum et al. 2019). We use Coramin to construct piecewise linear relaxations of the nonlinear constraints h and g. The remaining constraints are either linear or convex quadratic constraints. Initially, the relaxation is constructed with only one piecewise segment for each relaxed term, and bounds tightening is performed. If the solution to this problem does not produce a flexible region that is sufficiently large, then the segments are further refined (i.e., split in half) and the bisection algorithm is run again. Using these new segments is guaranteed to produce a tighter relaxation because the vertices of the piecewise linear segments of the tighter relaxation are a superset of those in the previous iteration. This process is repeated until an adequate region is accepted or the computational cost of solving the problem with many piecewise segments is too high. The resulting solution is a conservative region in θ_p which is guaranteed to abide by the CQAs with at least probability α.

Industrial Case Study Results

In this section, we analyze the results of using the described relaxation algorithm on an industrial case study provided by Eli Lilly and Company. The Michael addition reaction (Chen, Biegler and García-Muñoz, 2016) may be described by the following reaction kinetics:

$$AH + B \xrightarrow{k_1} A^- + BH^+ \quad (27)$$

$$A^- + C \xrightarrow{k_2} AC^- \quad (28)$$

$$AC^- \xrightarrow{k_3} A^- + B \quad (29)$$

$$AC^- + AH \xrightarrow{k_4} A^- + P \quad (30)$$

$$AC^- + BH^+ \xrightarrow{k_5} P + B \quad (31)$$

Here the starting materials AH and C are the Michael donor and the Michael acceptor, respectively. B is a base, BH^+, A^-, and AC^- are reaction intermediates, and P is the product. The reactions rates for eqs. (32) through (36) are given below:

$$r_1 = k_1 C_{AH} C_B \quad (32)$$

$$r_2 = k_2 C_{A^-} C_C \quad (33)$$

$$r_3 = k_3 C_{AC^-} \quad (34)$$

$$r_4 = k_4 C_{AC^-} C_{AH} \quad (35)$$

$$r_5 = k_5 C_{AC^-} C_{BH^+} \quad (36)$$

In this case, uncertainty is represented in the reaction coefficients k_i (i.e., $\theta_m = \{k_1, k_2, k_3, k_4, k_5\}$). Thus, the nominal point $\hat{\theta}_m$ is defined as $\hat{k}$ with the estimated reaction coefficients:

$$\hat{k} = [49.7796,\ 8.9316,\ 1.3177,\ 0.3109,\ 3.8781] \quad (37)$$

The units for $\hat{k}_i$ is L/(mol·min) for all entries except $\hat{k}_3$ which is 1/min. Also, the mean values of the estimated parameters in $\hat{k}$ correspond to a normal variance-covariance matrix Σ_{θ_m}. This matrix Σ_{θ_m} is provided in Laky et al. (2019).

For this case study, it is assumed that the reaction is carried out in a constant-volume CSTR. The process parameters θ_p are the residence time, τ, and feed ratio of species AH to B, $R_{A|B}$ (i.e., $\theta_p = \{\tau, R_{A|B}\}$). For simplicity, it is assumed the CSTR is operating at steady state. Using a steady state CSTR mass balance, and the rates of reaction r_i, the equations $h(\theta_p, x, \theta_m)$ can be defined for each species below:

$$(C_j^0 - C_j) + \tau \sum_i \nu_i r_i \quad (38)$$

Initial concentrations $[C_{AH}^0, C_B^0, C_C^0, C_{A^-}^0, C_{AC^-}^0, C_{BH^+}^0, C_P^0]$ are given by $[0.3955, 0.3955/R_{A|B}, 0.25, 0, 0, 0, 0]$ mol/L. Finally, the CQAs may be defined mathematically as follows:

$$\frac{C_C^0 - C_C - C_{AC^-}}{C_C^0} \geq 0.9 \Rightarrow C_C + C_{AC^-} - 0.1 C_C^0 \leq 0 \quad (39)$$

$$C_{AC^-} \leq 0.002 \quad (40)$$

Eq. (39) states that the conversion of the Michael donor, C, must be greater than or equal to 90 percent. Eq. (40) states that the concentration of intermediate AC^- may not exceed 0.002 mol/L.

Using parameters from Laky et al. (2019), tolerance for CQA violation is $\epsilon_{tol} = 10^{-5}$. The range of process parameters θ_p are given by 400 to 1400 minutes for residence time τ, and 10 to 30 for feed ratio $R_{A|B}$. The dimensions $\Delta\theta_p^-$ and $\Delta\theta_p^+$ were chosen for each process parameter θ_p as to scale δ_p within the domain $[0, 1]$ (i.e., initialize $\delta_p^L = 0$, $\delta_p^U = 1$). For uncertain parameters θ_m, a value of δ_m^α corresponding to confidence level $\alpha = 0.85$ was chosen using the chi-squared cumulative density function for five degrees of freedom.

All multilinear terms will be relaxed with linear McCormick envelopes (McCormick 1976). Initially, the relaxation will be over the entire variable bound, and will be progressively tightened to have 2^n ($n \in \mathbb{Z}_0^+$) piecewise segments representing each bilinear relaxation variable w_i. An example of how nonlinear terms are linearized is applied to the quadrilinear term, $\tau(k_1 C_{AB} C_B)$, is shown below:

$$w_1 = k_1 C_{AH} \quad (42)$$
$$w_2 = k_1 C_{AH} C_B = w_1 C_B \quad (43)$$
$$w_3 = \tau(k_1 C_{AB} C_B) = \tau w_2 \quad (44)$$

Piecewise McCormick envelopes are then applied to these terms. The process is repeated for all nonlinear terms in the model. In this case, 5 iterations of bound tightening were performed on each variable before the piecewise segmentation. With the constraints relaxed and variable bounds tightened, a 2^n piecewise segment linear relaxation is applied to variables w_i. The problem is now a mixed-integer quadratically constrained program (MIQCP) due to the quadratic nature of the variance-covariance ellipsoid constraint.

The MIQCPs solved in this work utilized the Gurobi solver under version 7.5.2 (Gurobi). All models were written with Pyomo (Bynum et al. 2021). To find the global solution for comparison with the progressively relaxed solutions, BARON (Sahinidis 1996) was used to solve the NLP subproblems of (FI*).

Table 1. Maximum constraint violation at δ_p=0.5, optimal value δ_p^*, and solution time for progressively refined relaxations.

# of PW McCormick Relaxation Segments	Maximum violation of CQAs at $\delta_p = 0.5$	Flexibility Test Value (FT*) δ_p^*	Solve Time (s)
1	$1.45 \cdot 10^{-1}$	0	5.1
2	$5.61 \cdot 10^{-2}$	0	6.3
4	$1.96 \cdot 10^{-2}$	0.0352	6.9
8	$9.09 \cdot 10^{-3}$	0.2656	12.2
16	$4.07 \cdot 10^{-3}$	0.3125	42.5
32	$1.97 \cdot 10^{-3}$	0.3438	714.5
Global Solution	$5.72 \cdot 10^{-4}$	0.3789	245.3

Starting with n = 0, 1 piecewise linear segment was used to relax (FT*) for the Michael addition case study. With this relaxation, the entire operating region was infeasible, as shown in Table 1. However, as the relaxation was progressively refined, small operating region was feasible at n = 4. Eventually, using 32 segments for the relaxations yielded a solution within 10% of the extent of the global feasible region while also approaching the true global maximum of the CQA violations for a given value of δ_p. In Figure 1, the maximum constraint violation over all CQAs for a δ_p value of 0.5 is shown. Importantly, the constraint violation for finer relaxations conservatively represents the feasibility of the process and approaches the global value. The solution, δ_p^*, found by solving the flexibility test formulation with progressively finer relaxations is also shown. As expected, finer relaxations approach the global optimum, remaining conservative for all relaxations. However, timing considerations show that solution time for 32 segments requires more solution time than the global optimizer. It should be noted that all relaxed bilinear terms utilize 32 segments, and improvement can be made by using a quantitative approach to choose which variables are progressively refined. Also, the bounds tightening procedure accounts for approximately 5 seconds of the total solution time to find δ_p^* for a given number of segments.

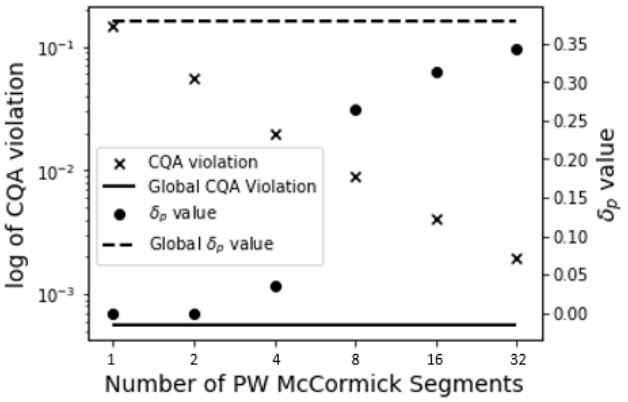

Figure 1: Visualization of results reported in Table 1. Size δ_p approaches the global solution, and CQA violation at δ_p=0.5 approaches the global value as well.

Conclusions

Identification of the design space is a costly yet beneficial process in the pharmaceutical industry. Any advancement that can alleviate costs, either computationally or by reducing the number of required experiments is of great value. Here, we proposed an algorithm that is guaranteed to produce a conservative design space for non-linear pharmaceutical processes. This conservativeness of this approach is confirmed by overestimating infeasibility of the CQAs, as seen in Figure 1.

Also, it was seen that progressively finer relaxations produce solutions more representative of the global solution. Often, the full extent of the design space is not required so long as the conservative design space is large enough for implementation.

The form of the relaxations removes the difficult non-linearity of the problem reducing the flexibility test formulation from a MINLP to a sequence of MIQCPs, resulting in improved computational performance.

Acknowledgments

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the

paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

References

Adjiman, C.S, Dallwig, S., Floudas, C.A., and Neumaier, A. (1998) A global optimization method, αBB, for general twice-differentiable constrained NLPs – I. Theoretical advances. *Comp. & Chem. Eng.* 22 (9), 1137-1158.

Boukouvala, F., Muzzio, F.J., and Ierapetritou, M.G. (2010). Design space of pharmaceutical processes using data-driven-based methods. *J. Pharm. Innov.*, 5, 119-137.

Bynum, M.L., Hackebiel, G.A., Hart, W.E., Laird, C.D., Nicholson, B.L., Siirola J.D., Watson, J.-P. and Woodruff, D.L. (2021). *Pyomo-optimization modeling in Python*. Springer, 3rd edition.

Bynum, M.L., Castillo, A., Laird, C., and Watson, J.-P. (2019). Coramin v. 0.1 https://doi.org/10.11578/dc.20190311.1

Ding, C. and Ierapetritou, M. (2021). A novel framework of surrogate-based feasibility analysis for establishing design space of twin-column continuous chromatography. *Int. J. Pharm.*, 609.

Facco, P., Pastro, F.D., Meneghetti, N., Bezzo, F., and Barolo, M. (2015). Bracketing the design space within the knowledge space in pharmaceutical product development. *Ind. & Eng. Chem. Res.* 54, 5128-5138.

FDA (2004) Pharmaceutical cgmps for the 21st century-a risk-based approach final report.

FDA (2009) Guidance for industry q8(r2) pharmaceutical development.

García Muñoz, S., Luciani, C., Vaidyaraman, S., and Seibert, K.D. (2015). Definition of design spaces using mechanistic models and geometric projections of probability maps. *Org. Process Res. Dev.* 19, 1012-1023.

Grossmann, I.E., Halemane, K.P, and Swaney, R.E. (1983). Optimization strategies for flexible chemical processes. *Comp. & Chem. Eng.* 7 (4), 439-462.

Grossmann, I.E. and Floudas, C.A. (1987) Active constraint strategy for flexibility analysis in chemical processes. *Comput. Chem. Engng*, 11, 675–693.

Gurobi Optimization (2021). *Gurobi*. http://www.gurobi.com (accessed 2018-10-25)

Halemane, K.P. and Grossmann, I.E. (1983) Optimal process design under uncertainty. *AIChE Journal* 29, 425-433

Kumar, S., Gokhale, R. and Burgess, D.J. (2014). Quality by design approach to spray drying processing of crystalline nanosuspensions. *International Journal of Pharmaceutics*. 464, 234-242.

Laky, D., Xu, S., Rodriguez, J.S., Vaidyaraman, S., García Muñoz, S., Laird, C. (2019). An optimization-based framework to define the probabilistic design space of pharmaceutical processes with model uncertainty. *Processes*, 7.

McCormick, G.P. (1976) Computability of global solutions to factorable nonconvex programs part I – convex underestimating problems. *Mathematical Programming*, 10, 147-175.

Misener, R., and Floudas, C.A. (2014). ANTIGONE: Algorithms for continuous/Integer Global Optimization of Nonlinear Equations. *J. Glob. Optim.* 59, 503-526.

Ochoa, M.P., Deshpande, A., García Muñoz, S., Stamatis, S., and Grossmann, I.E. (2019) Flexibility analysis for design space definition. *Computer Aided Chemical Engineering*. 47, 323-328.

Ochoa, M.P. and Grossmann, I.E. (2020). Novel MINLP formulations for flexibility analysis for measure and unmeasured uncertain parameters. *Comp. & Chem. Eng.* 135.

Ochoa, M.P., García Muñoz, S., Stamatis, S., and Grossmann, I.E. (2021) Novel flexibility index formulations for the selection of the operating range within a design space. *Comp. & Chem. Eng.* 149.

Ostrovsky, G.M., Volin, Y.M., Barit, E.I., and Senyavin, M.M. (1994). Flexibility analysis and optimization of chemical plants with uncertain parameters. *Comp. & Chem. Eng.* 18 (8), 755-767.

Pistikopoulos, E.N. and Grossmann, I.E. (1989a) Optimal retrofit design for improving process flexibility in nonlinear systems—I. fixed degree of flexibility. *Comp. & Chem. Eng.* 13 (9), 1003-1016.

Pistikopoulos, E.N. and Grossmann, I.E. (1989a) Optimal retrofit design for improving process flexibility in nonlinear systems—II. optimal level of flexibility. *Comp. & Chem. Eng.* 13 (9), 1087-1096.

Pistikopoulos, E.N. and Ierapetritou, M.G. (1995). Novel approach for optimal process design under uncertainty. *Comp. & Chem. Eng.* 19 (10), 1089-1110.

Pulsipher, J. and Zavala, V.M. (2018). A mixed-integer conic programming formulation for computation of the flexibility index under multivariate gaussian uncertainty. *Comp. & Chem. Eng.* 119 (2), 302-308.

Pulsipher, J.L., Rios, D. and Zavala, V.M. (2019) A computational framework for quantifying and analysis system flexibility. *Comp. & Chem. Eng.* 126, 342-355.

Rogers, A. and Ierapetritou, M.G. (2015a) Feasibility and flexibility analysis of black-box processes part 1: surrogate-based feasibility analysis. *Chem. Eng. Sci.* 137, 986-1004.

Rogers, A. and Ierapetritou, M.G. (2015b) Feasibility and flexibility analysis of black-box processes part 1: surrogate-based flexibility analysis. *Chem. Eng. Sci.* 137, 1005-1013.

Rooney, W.C. and Biegler, L.T. (1999) Incorporating joint confidence regions into design under uncertainty. *Comp. & Chem. Eng.* 23, 1563-1575

Sahinidis, N.V. (1996) BARON: A general purpose global optimization software package. *J. Glob. Optim.* 8, 201-205.

Swaney, R.E. and Grossmann, I.E. (1985) An Index for Operational Flexibility in Chemical Process Design. *AIChE Journal* 31 (4), 621-630

Tsoukalas, A., and Mitsos, A. (2014). Multivariate McCormick relaxations. *J. Glob. Optim.* 59, 633-662.

Wang, Z. and Ierapetritou, M. (2017). A novel feasibility analysis method for black-box processes using a radial basis function adaptive sampling approach. *AIChE Journal*, 63 (2), 532-550.

Wang, Z. and Ierapetritou, M.G. (2018). Global sensitivity, feasibility, and flexibility analysis of continuous pharmaceutical manufacturing processes. *Comp. Aided Chem. Eng.* 41, 189-213.

Zhao, F., Grossmann, I.E., García-Muñoz, S., and Stamatis, S.D. (2021). Flexibility index of black-box models with parameter uncertainty through derivative-free optimization. *AIChE Journal*, 67 (5).

Zhao, F., Ochoa, M.P., Grossmann, I.E., García-Muñoz, S., and Stamatis, S.D. (2022). Novel formulations of flexibility index and design centering for design space definition. *Comp. & Chem. Eng.* 166.