Analyzing the effects of control Strategies for Determining Process Feasible Space

Margherita Geremiaa, Fabrizio Bezzoa, Marianthi G. Ierapetritoub,\*

aCAPE-Lab – Computer-Aided Process Engineering Laboratory. Department of Industrial Engineering, University of Padova, via Marzolo 9, 35131 Padova (PD), Italy

bDepartment of Chemical and Biomolecular Engineering, University of Delaware, Newark, DE, USA

Corresponding author e-mail: mgi@udel.edu

Abstract

The identification of process Design Space (DS) is key to support the development of pharmaceutical processes, where strict requirements on manufacturability and product quality must be satisfied. If the process can be controlled by a set of manipulated variables, the DS can be enlarged with respect to an open-loop scenario, where there are no controls in place. Since pharmaceutical models are typically complex and computationally expensive, surrogate-based feasibility analysis can be suitably exploited to determine whether the process satisfies all constraints by adjusting the process control inputs, and mitigate the effect of uncertainty. The approach is successfully implemented on a pharmaceutical case study; results demonstrate that different control actions can be effectively exploited to mitigate uncertainty and operate the process in a wider range of inputs. The framework can conveniently be exploited to support decisions on control strategies for real industrial applications.

**Keywords**: feasibility analysis, design space, pharmaceutical manufacturing, surrogate models, process control.

* 1. Introduction

Process Design Space (DS) is defined as the subset of combinations of input parameters that have been demonstrated to provide assurance of product quality and satisfy all the relevant operating and production constraints (ICH, 2009); its description is particularly relevant in highly regulated sectors such as the pharmaceutical industry, where assurance of manufacturability and quality of the product is key for process development and optimization (Destro and Barolo, 2022). Feasibility analysis can be exploited to determine the range of conditions within which the process can be safely operated, i.e., the subset of combinations of input factors that satisfy all the relevant operating, quality, and production constraints (Grossmann et al., 2014). If the process can be controlled by a set of manipulated variables, the DS may be enlarged with respect to an open-loop scenario, where there are no controls in place. Namely, feasibility analysis can be used to determine whether the process satisfies all process constraints by adjusting the process control inputs to operate in a larger range of input factors and reduce the effect of uncertainty.

The use of surrogate-based approaches for feasibility analysis has been demonstrated to effectively identify the boundaries of the process DS, particularly when the available model is computationally expensive or consists of various black-box constraints – which is typically the case of pharmaceutical manufacturing models (Wang and Ierapetritou, 2017). However, the influence of control variables has not been included.

In this study we adopt a surrogate-based feasibility approach with the objective of investigating the effect of manipulated variables in the characterization of the process DS and in mitigating sources of uncertainty. More specifically, we aim at analyzing and quantifying the benefits of a proper control action on the process DS for pharmaceutical manufacturing development.

* 1. Methodology

Feasibility analysis is mathematically formulated as the maximum of the process constraint violation (Grossmann et al., 2014):

|  |  |
| --- | --- |
|  | (1) |

where is the process feasibility function; is the vector of control variables, and is the vector of critical input variables (e.g., raw material properties and critical process parameters); are the functions of the process constraints in the form 0 which must be satisfied during process operation. defines the range in which critical inputs can vary: with andrepresenting lower and upper bounds, respectively. Solving problem (1) determines whether for given critical input variables ,the control vector can be adjusted to satisfy all the problem constraints, and attain feasibility.

Since solution of Eq. (1) might be computational complex, we approximate the feasibility function using the surrogate-based approach proposed by Geremia et al. (2023), which is suitable to guide the selection of one best surrogate model for feasibility approximation and attain preset level of accuracy with the minimum requirement of additional training data. The procedure consists of three sequential steps (Figure 1), which are described in the following, and must be iteratively repeated until the surrogate approximates the real feasibility function with a given preset level of accuracy (stop criterion), or the total iterations exceeds the user-defined maximum number of iterations.

|  |
| --- |
|  |
| **Figure 1**. Schematic of the surrogate-based approach used in this study. |

Step 1. Analysis of the available dataset

Before any candidate surrogate is trained, statistical metrics (Sun and Braatz, 2021) and topological data analysis (Smith and Zavala, 2021) are combined to assess problem characteristics and complexity. A statistical analysis is suitable to detect nonlinearity between predictors and response and correlations, while algebraic topology aims at number disjointed feasible regions – which correspond to 0-dimensional Betti numbers (Edelsbrunner and Harer, 2009). This is key to acquire information regarding the original feasibility function, and to consequently narrow down a number of surrogate models to be further trained.

Step 2. Surrogate-based feasibility approximation

After the training, quality of fitting and predictive capability of the different surrogates are compared, in such a way that it is possible to identify which of them best approximates the feasibility function. The Bayesian Information Criterion (BIC) is used to account for both model complexity and quality of fitting (Schwarz, 1978):

|  |  |
| --- | --- |
|  | (2) |

where is the likelihood corresponding to the model , is the number of model parameters, and is the number of training points. The first term,, acts as a measure of inaccuracy, with smaller values to be preferred, while is a penalty factor that increases with model complexity (higher ) and the number of training points (higher ). Therefore, the current best model is the one associated with the lowest value of BIC. Evaluation of accuracy is treated as a binary classification problem, relying on the metrics proposed by Wang and Ierapetritou (2017), namely the percentage of Correct Feasible region (), the percentage of Correct InFeasible region (), and the percentage of Not Conservative feasible region (). If none of the trained surrogates guarantees the preset level of accuracy (stop criterion), new sampling points are needed.

Step 3. Adaptive sampling

To improve accuracy, new points are iteratively sampled and included within the initial dataset. To compute new samples, we use an adaptive strategy based on the minimization of the BIC of the current best surrogate. Under the assumption of independent errors, each new adaptive point is computed by solving the following minimization problem:

|  |  |
| --- | --- |
|  | (3) |

where is the standard error of the predictor, and is the value of the feasibility function that is predicted by the current best surrogate. Note that the term in Eq. (3) corresponds to in Eq. (2) (Bard, 1974).

* + 1. Effect of uncertainty

Some critical input factors in vector may be affected by uncertainty, e.g., material properties entering the process, affecting the characterization of the DS. We describe uncertain inputs using a uniform distribution around an expected value and analyze the level of uncertainty that propagates to the process DS by randomly selecting values of uncertain inputs from those distributions. We, then, quantify the reliability provided by the process DS through a probability figure of merit (Peterson, 2008):

|  |  |
| --- | --- |
|  [0,1]**.** | (4) |

 is the reliability factor representing the probability that process constraints are satisfied for a given set . If 0, there is always at least one constraint that is being violated, namely no combination of inputs allows the process to be feasibly operated; if 1, no constraint is being violated, namely the process is feasible for any combinations of inputs; if 01, some combinations of inputs satisfy all process contraints, while others don’t. The closer is to 1, the higher the probability that process constraints are satisfied.

* 1. Case Study

A pharmaceutical roller compaction process is used as case study. We rely on the model proposed by Hsu et al. (2010), which combines the Johanson’s rolling theory (Johanson, 1965) with a dynamic material balance:

|  |  |
| --- | --- |
|  | (5) |
|  | (6) |
|  | (7) |

where is the roll radius, is the roll width, and are compression parameters, is the compact surface area, is the effective angle of friction, is the nip angle, is the inlet angle, is the inlet powder density, is the roll speed, is the roll pressure, and is the powder feed speed. The model is used to predict the critical quality attributes of the product, i.e., the ribbon density at outlet, , and the ribbon thickness, , which must satisfy the requirements on product quality [850–950] kg/m3, and [1.7 10-3 – 1.9 10-3] m.

We account for the combined effect of the critical operating factors and on the identification of the process DS (i.e., **)** and the benefits of a proper control action on (i.e., **z**) while hedging against variability in raw material properties. We assume that can be manipulated in between 10% its nominal value in order to enlarge the process DS,and compare results with an open-loop scenario where the DS is only determined by the operating conditions **.** We account for the effect of variability in raw material properties assuming that fluctuates as powder enters the process around its expected (nominal) value.

* 1. Results

We start with an initial dataset of 25 points, which are obtained using a Sobol’s sampling strategy in order to uniformly map the input domain (Saltelli et al., 2010).

According to Step 1 of the presented procedure, statistical analysis of the available dataset indicates nonlinearity between predictors and response, and suggests the training of nonlinear candidate surrogates at Step 2. Algebraic topology does not detect the presence of disconnected components (1), i.e., the process DS does not consist of disconnected feasible regions.

The inclusion of 186 new adaptive points to the initial dataset (Step 3) allows to accurately approximate the feasibility boundaries through a Gaussian Process () with exponential kernel ( 98%, 98%, 2%). The surrogate-based prediction of the feasible boundaries is shown in Figure 2a, and compared to both the original contour, and the one obtained for an open-loop scenario. It is evident that a proper control action can extensively enlarge the feasible DS, allowing the process to be safely operated in a wider range of . For clarity purpose, values of control and effect on defining the feasible DS for different levels of compaction pressure are visualized in Figure 2b.

|  |  |
| --- | --- |
| Immagine che contiene testo, Carattere, linea, numero  Descrizione generata automaticamenteImmagine che contiene testo, schermata, linea, Parallelo  Descrizione generata automaticamente |  |
| *(a)* | *(b)* |
| **Figure 2**. (a) Comparison between real contour plot of enlarged DS (continue black line) and surrogate-based approximation (dotted orange line) through a with exponential kernel after the inclusion of 186 adaptive samples. Blue dotted lines represent an open-loop scenario. (b) Values of control and effect on for different levels *.* Nominal value of (no control action) is shown as continuous red line. |

* + 1. Accounting for uncertainty in raw material

We assume that the inlet powder density, , fluctuates as powder enters the process with a deviation of 25 kg/m3 with respect to the nominal value of 300 kg/m3. The stochastic DS is identified by evaluating the probability that all process constraints are satisfied for different values of . 102 scenarios are simulated for any combination of operating conditions (, and manipulated ) by randomly selecting values of in the predefined range of variability. 3600 evaluation points are used in order to heavenly map the input space (Kucherenko et al., 2015). Results are shown in Figure 3, from which it can be seen that a robust DS (i.e., = 1) has shrink in width if compared to Figure 2a, where no uncertainty was considered. However, a proper control action on is suitable to operate in a larger range if compared to the case of an open-loop scenario.

|  |
| --- |
|  |
| **Figure 3.** Stochastic DS considering a deviation of 25 kg/m3 from the expected value of *.* |

* 1. Conclusions

In this study, we presented a surrogate-based approach, which combines different mathematical tools to: (*i*) identify the process DS relying on the available training dataset, and (*ii*) investigate the effect of manipulated variables in order to operate the process in a wider range of operating conditions. The presented workflow combines different mathematical tools which are proved to effectively identify the process DS while illustrating the effect of control variables, and can be suitably exploited to guide decisions on convenient control strategies for industrial processes.

Acknowledgments

M. G. acknowledges “Fondazione Cariparo” for her scholarship, and “Fondazione Ing. Aldo Gini” for financial support during her stay at University of Delaware.

References

Y. Bard, 1974. Nonlinear Parameter Estimation. Academic Press, New York, NY.

F. Destro, M. Barolo, 2022, A review on the modernization of pharmaceutical development and manufacturing – Trends and perspectives, and the role of mathemtical modeling. Int. J. Pharm., 620, 121715.

H. Edelsbrunner, J. Harer, 2009. Computational Topology: An Introduction. American Mathematical Society.

M. Geremia, F. Bezzo, M.G. Ierapetritou, 2023. A novel framework for the identification of complex feasible space. Comput. Chem. Eng*.,* 79, 108427.

I.E. Grossmann, B.A. Calfa, P. Garcia-Herreros, 2014. Evolution of concepts and models for quantifying resiliency and flexibility of chemical processes. Comput. Chem. Eng., 70, 22–34.

S.H. Hsu, G.V. Rklaitis,V. Venkatasubramanian, 2010. Modeling and control of roller compaction for a pharmaceutical manufacturing. Part I: Process dynamics and control framework. J. Pharm. Innov*.,* 5, 14–23.

ICH, 2009. ICH harmonised tripartite guideline, guidance for industry, pharmaceutical development Q8(R2). In: Proceedings of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. Silver Spring MD: ICH, 2009.

J.R. Johanson, 1965. A rolling theory for granular solids. J. Appl. Mech. B., 32, 842–848.

S. Kucherenko, D. Albrecht, A. Saltelli, 2015. Exploring multi-dimensional spaces: a Comparison of Latin Hypercube and Quasi Monte Carlo Sampling Techniques. ArXiv150502350 Stat.

J. Peterson, 2008. A Bayesian approach to the ICH Q8 definition of Design Space. J. Biopharm. Stat., 18, 959–975.

A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comput. Phys. Commun., 181, 259-270.

G. Schwarz, 1978. Estimating the dimension of a model. Ann. Stats., 6, 461–464.

A. Smith, V.M. Zavala, 2021. The Euler characteristic: A general topological descriptor for complex data. Comput. Chem. Eng*.*, 154, 107463.

W. Sun, R.D. Braatz, 2021. Smart process analytics for predictive modeling. Comput. Chem. Eng., 144, 107134.

Z. Wang, M.S. Escotet-Espinoza, M. Ierapetritou, 2017. Process analysis and optimization of continuous pharmaceutical manufacturing using flowsheet models. Comput. Chem. Eng*.*, 107, 77–91.