Data-Driven Chance-Constrained Optimization for Minimizing the Influence of Material Uncertainty on Product Quality

Qingbo Meng,a I. David L. Bogle,a Vassilis M. Charitopoulosa

aSargent Centre for Process Systems Engineering, Department of Chemical Engineering, University College London, UK

v.charitopoulos@ucl.ac.uk

Abstract

Minimizing the impact of process uncertainties, caused by estimation and measurement errors, unplanned disturbances, or environmental changes, is one of the crucial practical challenges in the pharmaceutical industry. In this work we propose an approach using data-driven chance constraints to eliminate the influence caused by physical property uncertainty in the raw materials via model-based optimization. A flowsheet for the pharmaceutical tableting manufacturing process of the Diamond Pilot Plant (DiPP) at the University of Sheffield is used to test the methodology. Firstly, the Kernel Density Estimation (KDE) technique is applied to generate the inverse cumulative density function for the uncertain variable using historical raw material quality data obtained from the supplier. Different uncertainty risk levels are then considered when obtaining the optimal operating conditions to explore the trade-off between product quality and economic performance. Process operating limitations are addressed for different risk levels to guarantee the desired product quality. Results indicate that the proposed approach can effectively reduce the uncertainty in the product quality caused by raw material physical property changes.

**Keywords**: Data-driven chance-constrained; Kernel Density Estimation (KDE); Model-Based Optimization; Pharmaceutical Tableting Manufacturing Process.

* 1. Introduction and Background

The manufacturing industry is currently experiencing a profound revolution as it undergoes a paradigmatic shift towards smarter manufacturing, encompassing the industry 4.0 concepts. The smart manufacturing revolution has been divided into three phases (Bogle, 2017): factory and enterprise integration and plant-wide optimization, exploiting manufacturing intelligence, and creating disruptive business models.

Over the past few decades, significant progress has been made in the first phase and has resulted in plenty of benefits. The process systems engineering community (PSE) is now focusing on the integration of the first two phases to achieve more self-adaptive manufacturing to market changes or demands (Bogle, 2017). In the pharmaceutical industry uncertainty with changes in raw material supply is a key challenge in pharmaceutical formulated product manufactory (Litster and Bogle, 2019).

Uncertainty sources exist as a result of a diverse array of factors such as variation of raw material quality, changing customer or market demand, environmental conditions, or faults in measurement devices. In recent years, mathematical programming techniques have been proposed and enabled successful applications for solving the uncertainty issues using techniques such as stochastic programming, robust optimization, and data-driven chance constrained optimization (Calfa et al, 2015).

Chance-constrained methods (CC) were first introduced by (Charnes and Cooper, 1959) and gained attention due to their ability to quantify profitability and reliability in a probabilistic formulation useable in optimization problems (Li et al, 2008). Applying chance constrained methods enables the model to be optimized with respect to an objective function while ensuring the solution satisfies the constraints with selected confidence level in the presence of uncertain parameters. The generic formulation of CC is described in Eq.1 (Calfa et al, 2015).

$P\left\{g\_{j}\left(x\right)\geq δ\_{j}\right\}\geq 1-α\_{j}, j=1, 2, …, m$ (1)

where P is the probability of an event, 1- $α$ stands for the confidence level, and $α$ the risk level within the range [0,1]. Eq.1 can be stated as “the probability of the random variable $δ\_{j}$ to achieve a value less than or equal to $g\_{j}\left(x\right)$ must be at least $1-α\_{j}$”. In the PSE community, CC is mostly applied for customer demand satisfaction, product quality specification, reliability level of chemical processes (Ning and You, 2019).

In this study, we apply data-driven chance constraint programming to reduce the uncertainty influence from raw materials on the product quality through optimization of the operation of a pharmaceutical formulation plant shown in Fig. 1. The intra-particle void fraction (IVF) of feeding material lactose is considered as the uncertain variable and granule moisture content (MC) exiting the fluid bed dryer is the controlled output which must be maintained at less than 8%. An assumption is made that the IVF information is obtainable from the material supplier and follows the normal distribution IVF ~ N (0.5, 0.05) (Fig. 2a). Based on the data, the Kernel Density Estimation technique is applied to generate the inverse cumulative distribution function for converting the probabilistic constraints into the equivalent deterministic algebraic constraints. Different confidence levels (or risk levels) are used to explain the trade-off between product quality and process reliability. The rest of this paper is organized as follows: in section 2, the proposed method is outlined, in section 3, we apply the data-driven CC on a segmented fluid bed dryer which is part of the DiPP, and present results and draw conclusion in section 4 and 5.

* 1. Problem Formulation
		1. Raw Material Uncertainty – Intra-particle Void Fraction

Intra-particle void fraction (IVF) refers to the fraction of the total volume that is composed of void spaces or pores within the particle structure. A larger IVF value generally provides more void spaces which makes the moisture evaporating or escaping from the particles easier during the drying process and results in lower granule MC under the same operating condition. In the optimization problem, we obtain and implement the minimum corresponding IVF value via KDE as a boundary at different risk levels (Fig. 2b) to guarantee that the probability of 90% ($α=0.1$), 95% ($α=0.05$) or 99% ($α=0.01$) random selected IVF values are larger than the corresponding IVF value.

* + 1. Data-driven Chance Constraints

The data-driven chance constraint approach employed in this study was introduced by Calfa et al. (2015). It integrates observed data or historical information to approximate the level of uncertainty, rather than solely relying on a predetermined probability distribution. This allows more flexibility when the underlying probability distribution of the uncertainty is unknown or difficult to determine. It has been shown by (Jiang and Guan, 2013) that reformulation of classic CC and data-driven CC are equivalent. Therefore, according to Eq. 1, the reformulation of data-driven CC can be represented as follows (Calfa et al, 2015):

$g\_{j}\left(x\right)\geq F\_{δ\_{j}}^{-1}(1-α\_{j})$, *j* = 1, 2, …, m (2)

Where $F\_{δ\_{j}}^{-1}$ is the inverse cumulative density function that is estimated depending on the uncertain variable distribution/observed data information. Based on Eq.2, the uncertain variable value $g\_{j}\left(x\right)$ can be calculated to satisfy the desired level of probability.



Figure 1. Flowsheet of continuous pharmaceutical tableting process (Jiang et al, 2022)

* + 1. Kernel Density Estimation (KDE)

KDE is a probability density function (pdf) estimation technique. Compared to a traditional histogram, KDE has a number of advantages, for example (Weglarczyk, 2018):

* A smooth curve shows the details of the probability density function better,
* All data points are included to reveal comprehensive information and more convincing multimodality.

The KDE formulation is given by Bowman and Azzalini (1997):

$\hat{f\_{h}}\left(x\right)=\frac{1}{nh}\sum\_{i=1}^{n}K(\frac{x-x\_{i}}{h})$ (3)

where x1, x2, …, xn are random samples from distribution, *n* and *h* are the sample size and bandwidth, respectively, the normal kernel smoothing function is applied and represented using *K*. The formulation is given below:

$K\_{i}\left(x\right)=\frac{1}{\sqrt{2π}}e^{\frac{-(\frac{x-a\_{i}}{h})^{2}}{2}}$ (4)

where $a\_{i}$ is the $a$ value at point $i$. In this study, the KDE technique is applied for finding the inverse cumulative distribution (quantile function) of the uncertain variable IVF to determine its corresponding values under different confidence levels (see Fig.2).

* + 1. Dynamic Model Description

The dynamic model involved in this study includes a Twin-screw Granulator (TSG) and a segmented fluidized bed dryer (FBD), which are upstream process of the tableting manufacturing in a pharma plant (see Fig.1). gPROMS Formulated Products is the platform for the model implementation. Parameters have been validated using DiPP at the University of Sheffield. The solid and liquid binder source fed into TSG are lactose and water to mix and produce granules for FBD. In FBD, the granules are suspended in a hot air stream and moisture is evaporated to the desired target.



Figure 2. (a) Histogram of IVF, (b) Corresponding IVF values at different risk levels.

* + 1. Optimization

The optimization strategy is required to ensure the best design options and optimal operating performance of the drying process. In this study the three decision variables are lactose (*Fl*) and water feed flowrate (*Fw*) to the TSG, and feed vapour temperature (*Tv*) to the FBD. The objective is to minimize the feed vapour flowrate ($F\_{v}$) to the FBD. The uncertainty variable IVF has been converted from a probabilistic to a deterministic constraint via the KDE approach subject to different risk levels. To restrict the granule MC out from FBD less than 8% is the control objective. The optimization problem is expressed as follows:

$\genfrac{}{}{0pt}{}{min}{F\_{w}, F\_{l},T\_{v}}$$F\_{v}$

s.t. C1: 10kg/h ≤ *Fl* ≤ 20kg/h

 C2: 2kg/h ≤ *Fw* ≤ 6kg/h

 C3: 40°C ≤ *Tv* ≤ 80°C

 C4: MC ≤ 8%

 C5: 800µm ≤ Pas ≤ 1000µm

 C6: $\left\{\begin{array}{c}risk level \left(α\right)=0.1, P\geq 90\%, IVF\geq F\_{δ\_{j}}^{-1}\left(0.1\right)=0.426\\risk level \left(α\right)=0.05, P\geq 95\%, IVF\geq F\_{δ\_{j}}^{-1}\left(0.05\right)=0.407\\risk level \left(α\right)=0.01, P\geq 99\%, IVF\geq F\_{δ\_{j}}^{-1}\left(0.01\right)=0.354\end{array}\right.$

where the inequality constraints C1 to C3 indicate the operating range constraints of the three decision variables. MC and Pas in C4 and C5 are abbreviations for moisture content and average particle size, respectively. C6 describes the corresponding IVF values under different risk level ($α$)/confidence level (***P***). $F\_{δ\_{j}}^{-1}$ represents the quantile function (Eq.2).

* 1. Results and Discussions
		1. Optimization Results

The overall drying process, comprising loading, drying, and discharging stages, has a total duration of 400 seconds, with 200 seconds allocated for loading and 50 seconds for discharging. The optimization process is divided into 4 time intervals. Decision variables are optimized (see Fig. 3) in each time interval within the constraints to minimize the vapour feed flowrate to the FBD to reach the desired MC level. The exact optimized vapour temperature values in bottom Fig.3 are 51.35°C ($α$**=**0.1), 52.52°C ($α$**=**0.05) and 53.1°C ($α$**=**0.01).



Figure 3. Optimized decision variables in each time interval

Fig.3 reveals that although a lower risk level represents a relatively smaller chance that the controlled granule MC exceeds the desired boundary (8%), it requires higher vapour temperature to achieve this objective which results in a higher energy consumption. This is a compromise between manufacturing reliability and cost-saving.

* + 1. Validation Results

To validate the optimization performance, we randomly generated 18 different IVF values within the distribution range to run the drying process with optimized decision variables at each risk level. The aim is to observe how many IVF values lead to granules exceeding the desired MC level of 8% when discharged from the FBD. To avoid redundancy, a representative case where the IVF equals 0.551 is used to encompass other IVF values that result in output granule MC less than 8% (see Fig. 4).

It is noted (Fig. 4d, e, f) that when the risk level is reduced, the number of controlled MC higher than 8% is reduced. This observation highlights the inherent trade-off between the MC and temperature requirements across different risk levels.

* 1. Conclusions

A data-driven chance constrained approach is implemented in this study to minimize the impact of raw material uncertainty. Different risk levels were tested to explore the trade-off between the desired moisture content and energy consumption. Results demonstrate that the proposed approach can eliminate the uncertainty impact from raw material variation and guarantee the output MC at a desired value with various risk levels. Lower risk levels are associated with lower MC violation rate but necessitate a comparatively higher vapour temperature resulting in a trade-off between process reliability and energy consumption. This approach presents a structured framework to mitigate the impact of uncertainty through process optimization.

**Acknowledgements:** This work was funded by the EPSRC (EP/V034723/1) RiFTMaP



**(b)**

**(c)**

**(d)**

**(e)**

**(f)**

**(a)**

Figure 4. Validation results

Table 1 Moisture content violation rate at different risk level

|  |  |  |
| --- | --- | --- |
| Risk Level | Violations (MC ≥ 8%) | Rate |
| $α$**=0.1** | 2/18 | 11.11% |
| $α$**=0.05** | 1/18 | 5.56% |
| $α$**=0.01** | 0/18 | 0 |

**References:**

A. Charnes, W.W. Cooper, 1959, Chance-constrained Programming, Management Science, 6 (1), 73-79.

A.W. Bowman, A. Azzalini, 1997, Applied Smoothing Techniques for Data Analysis, New York: Oxford University Press Inc.

B.A. Calfa, I.E. Grossmann, A. Agarwal, S.J. Bury, J.M. Wassick, 2015, Data-driven individual and joint chance-constrained optimization via kernel smoothing, Computers and Chemical Engineering, 78, 51-69.

C. Ning, F. You, 2019, Optimization under Uncertainty in the Era of Big Data and Deep Learning: When Machine Learning Meets Mathematical Programming, Comput. Chem. Eng., 125 (9), 434-448.

I.D.L. Bogle, 2017, A Perspective on Smart Process Manufacturing Research Challenges for Process Systems Engineers, Engineering, 3, 161-165.

J. Litster, I.D.L. Bogle, 2019, Smart Process Manufacturing for Formulated Products, Engineering, 5, 1003-1009.

P. Li, H. Arellano-Garcia, G. Wozny, 2008, Chance constrained programming approach to process optimization under uncertainty, Computers & Chemical Engineering, 32, 25-45.

S. Weglarczyk, 2018, Kernel density estimation and its application, ITM Web of Conferences 23, 00037.

S.L. Jiang, L.G. Papageorgiou, I.D.L. Bogle, V.M. Charitopoulos, 2022, Investigating the Trade-Off between Design and Operational Flexibility in Continuous Manufacturing of Pharmaceutical Tablets: A Cast Study of the Fluid Bed Dryer, Processes, 10, 454.

R. Jiang, Y. Guan, 2013, Data-driven chance constrained stochastic program, Optimization Online, <http://www.optimization-online.org/DBFILE/2013/09/4044.pdf>.