Systems & Control Transactions

Research Article - Peer Reviewed Conference Proceeding
ESCAPE 35 - European Symposium on Computer Aided Process Engineering
Ghent, Belgium. 6-9 July 2025
Jan F.M. Van Impe, Grégoire Léonard, Satyajeet S. Bhonsale,
Monika E. Polańska, Filip Logist (Eds.)

PSE PRESS

From Experiment Design to Data-Driven Modeling of Powder Compaction Process

René Brands[a], Vikas Kumar Mishra[b*], Jens Bartsch[a], Mohammad Al Khatib[b], Markus Thommes[a] , and Naim Bajcinca[b]

[a] Laboratory of Solids Process Engineering, Department of Biochemical and Chemical Engineering, TU Dortmund University, 44227 Dortmund, Germany
[b] Department of Mechanical and Process Engineering, RPTU Kaiserslautern, 67663 Kaiserslautern, Germany
* Corresponding Author: vikas.mishra@rptu.de.

ABSTRACT

Tableting is a dry granulation process for compacting powder blends into tablets. In this process, a blend of active pharmaceutical ingredients (APIs) and excipients are fed into the hopper of a rotary tablet press via feeders. Inside the tablet press, rotating feed frame paddle wheels fill powder into dies, with tablet mass adjusted by the lower punch position during the die filling process. Pre-compression rolls press air out of the die, while main compression rolls apply the force necessary for compacting the powder into tablets. In this paper, process variables such as feeder screw speeds, feed frame impeller speed, lower punch position during die filling, and punch distance during main compression have been systematically varied. Corresponding responses, including pre-compression force, ejection force, and tablet porosity have been evaluated to optimize the tableting process. After implementing an open platform communications unified architecture (OPC UA) interface, process variables can be monitored in real-time. To enable in-line monitoring of tablet porosity, a novel UV/Vis fiber optic probe has been implemented into the rotary tablet press. To further analyze the overall process, a data-driven modeling approach is adopted. Data-driven modeling is a valuable alternative to modeling real-world processes where, for instance, first principles modeling is difficult or infeasible. Due to the complex nature of the powder compaction process, several model classes need to be explored. To begin with, linear autoregressive models with exogenous inputs (ARX) have been considered. Thereafter, nonlinear autoregressive models with exogenous inputs (NARX) have been considered. Notably, several experiments have been designed to gather the data required for the development of the model.

Keywords: Modeling, Process control, Process monitoring, Industry 4.0, Big data, powder compaction, Tableting, UV/Vis spectroscopy

INTRODUCTION

The pharmaceutical industry is increasingly striving for autonomous control of their processes [1]. In this context, process and product data are continuously generated, evaluated and interpreted in order to assess the quality of the drug product and to control the corresponding process based on this data. Thereby leading to improved product quality and reduced costs [2].

Tableting represents a vital process step in the production of tablets, the most common dosage form, and is therefore a particular focus of research [3]. Here, in a rotary tablet press, powder is filled into a die, pressed into a tablet by a two-stage compression process of the upper and lower punch and then ejected by an upward movement of the lower punch [4]. These tablet presses are characterized by a particularly high throughput and double-sided compression of the tablets.

Traditionally, individual samples are taken after tableting and subsequently evaluated off-line and used to assess the quality of the entire batch. In order to ensure real-time data, the implementation of an OPC UA connection and process analytical technology is crucial. This enables process and product data to be generated in real-time [5].

Process parameters that influence drug product

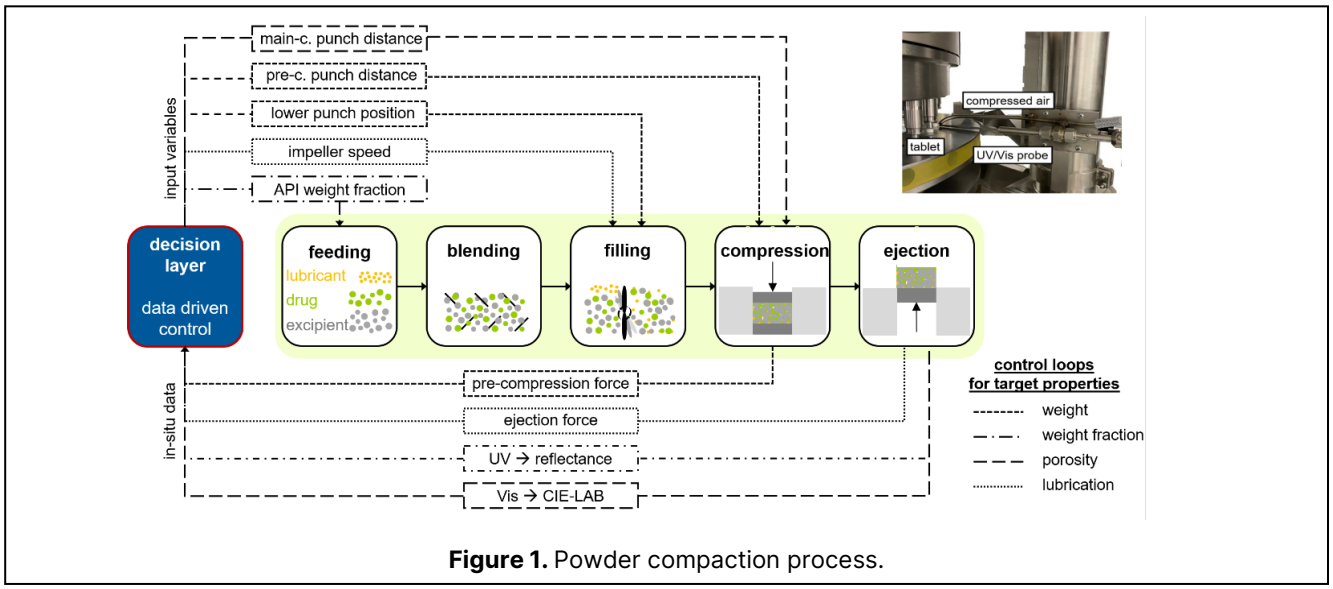

Figure 1. Powder compaction process.

quality, referred to as critical process parameters, are in particular the feed frame speed when filling the die with powder, the pre-compression force, the main-compression force, the compression roll distance and the ejection force. In order to evaluate these critical process parameters in real-time, an OPC UA interface was implemented.

Spectroscopic methods have proven particularly valuable for critical quality attribute (CQA) monitoring, as they are non-destructive, non-invasive and fast [6]. CQAs such as the active pharmaceutical ingredient (API) content and porosity are particularly important here, as these have a significant influence on the quality of the drug product. Near-infrared spectroscopy has proven to be useful for monitoring the active pharmaceutical ingredient content or porosity [7]. However, simultaneous determination was not possible and the evaluation was carried out by multivariate data analysis. In this work, a comparatively new method was used that is based on UV/Vis diffuse reflectance spectroscopy, measures API content and porosity simultaneously and can be evaluated univariately. For this purpose, the spectrum is separated into the ultraviolet and the visible range whereby the ultraviolet range is utilized to determine the active pharmaceutical ingredient content and the visible range to determine the porosity. Therefore, the visible range was transformed into the CIELAB color space.

To further analyze the process, we used the experimental data to develop a mathematical model. Recently, data-driven modeling has received a great deal of attention to model real-world applications. Parametric and nonparametric approaches are prevalent in the literature [8]. In this work, we are interested in the parametric modeling approach. In the parametric modeling approach, a model class is chosen a priori. Thereafter, the parameters are estimated, for instance, by using output-error methods [8, Chapter 7]. Here, we use nonlinear auto-regressive with exogeneous inputs (NARX) approach [9] to develop a parametric model for the process. Remarkably, NARX models have already found several applications in the literature [10-12]. As the powder compaction process is a complex process consisting of several subprocesses, it is challenging to determine a model class. In this regard, both linear and nonlinear autoregressive models with exogeneous inputs are considered to model the process. As is shown, the whole process is modeled as a combination of linear autoregressive and nonlinear autoregressive equations. We now summarize the contributions of this study. We collect experimental data for the powder compaction process on a rotary tablet press. We analyze full factorial experimental designs using multiple linear regression analysis to identify significant factors influencing product quality. We then combine process parameters with data obtained from a real-time spectroscopy method based on UV/Vis spectroscopy. To gain further insights about the relationships among various inputs and outputs of the process shown in Figure 1, we develop a parametric model based on the collected experimental data. This data-driven model of the process is developed using the NARX modeling approach.

MATERIAL AND METHODS

Materials

The formulation consists of 10 wt% theophylline monohydrate (Theophylline Monohydrate, Thermo Fisher Scientific, Waltham, USA) used as the active pharmaceutical ingredient, 0.5 wt% magnesium stearate (Ligamed MF-2-V, Peter Greven, Bad Münstereifel, Germany) used as the lubricant and lactose monohydrate (Foremost NF Lactose 310, Foremost Farms USA, Baraboo, USA) used as the filler and binder. Accordingly, the lactose content was changed when the active pharmaceutical ingredient

content was adjusted to 7 and 13 wt%.

Tableting

The tableting process is carried out on a rotary tablet press (102i, Fette Compacting, Schwarzenbek, Germany). The throughput was set to 20,000 tablets per hour. In a first step, the material is fed into the hopper of the tablet press via loss-in-weight feeders (KT-20, Cooperion, Niederlenz, Switzerland). The powder is then filled into the dies via the feed frame equipped with round spokes paddle wheels (Fill-O-Matic (FOM), Fette Compacting, Schwarzenbek, Germany). Each die is mounted on the die table and rotates in a circular motion with the pairs of punches. The die table then continues to rotate and pre-compression rolls press air out of the die, followed by the actual material compression step via the main-compression rolls. Finally, the finished tablet is ejected by an upward movement of the lower punch. A full-factorial experimental design containing three stages was developed for the investigations. The active pharmaceutical ingredient content was varied from 7 to 13 wt\%, the main-compression force from 5 to 15 kN and the speed of the FOM from 40 to 120 rpm. Here, the responses were the ejection force, the reflectance value and the CIELAB color space. Furthermore, another experimental design was performed to investigate the influence of the powder mass on the pre-compression force as a possible surrogate parameter for real-time monitoring of the tablet weight. Tablet masses from 280 to 360 mg were set by varying the filling depth. Also, the influence of the pre-compression punch distance was investigated.

Critical process parameter monitoring

Real-time monitoring of critical process parameters requires the implementation of an OPC UA interface. This guarantees real-time data exchange between the tablet press and the corresponding server. In this study, the ejection, pre- and main-compression force were recorded and evaluated using the OPC UA interface.

Critical quality attribute monitoring

In this study, the weight and porosity of the tablets were determined off-line and additionally supported by real-time UV/Vis spectroscopy data to capture the dynamics of the process.

Off-line determination

Tablet porosity (ε) (Eq. (1)) was determined by correlating the tablet density (ρ_t) derived from geometry and weight measurements (ST50, Sotax, Aesch, Switzerland) to the solid density (ρ_s) measured by helium pycnometry according to European Pharmacopoeia 11 (Chapter 2.9.23):

$$\varepsilon = \left(1 - \frac{\rho_t}{\rho_s}\right) 100\%. \quad (1)$$

Real-time UV/Vis spectroscopy

Real-time monitoring of the active pharmaceutical ingredient content and porosity was carried out using a UV/Vis diffuse reflectance probe (Inspectro X, ColVisTec, Berlin, Germany). For this purpose, the probe was implemented in the ejection position of the tablets. Here, the probe is oriented orthogonally to the tablet side wall meaning measurements are taken on the tablet side wall. The distance between the probe and the tablet is set to 4 mm. In this way, radiation from 224 to 820 nm is guided to the tablet via the six outer glass fibers, where the radiation is absorbed. The reflected light is guided back via the middle glass fiber and evaluated. In order to keep the 6 in 1 array of glass fiber optics free of particles, an additional air flow and an extraction system have been implemented. For the measurement, 35 light flashes are emitted per measurement, each lasting 2 ms leading to a measurement frequency of 1.5 Hz.

For data analysis and evaluation with regard to the API content, the reflectance (R) was formed on the basis of the intensity of the reflected light (I) and the emitted light (I_0). In addition, Kubelka-Munk was applied to establish a correlation between absorption and reflection. Here, the transformed reflectance (f(R)) is the quotient of absorption and scattering coefficients (k and s) (Eq. (3)):

$$R = \frac{I}{I_0} \quad (2)$$

$$f(R) = \frac{k}{s} = \frac{(1-R)^2}{2R}. \quad (3)$$

For the assessment, the reflectance intensity in the UV range, specifically in the wavelength range from 320 to 330 nm, was correlated with the active pharmaceutical ingredient content.

The porosity was evaluated on the basis of the Vis range (380 to 780 nm) and the corresponding spectral information were converted into the CIELAB color space. This describes the range of visible light for humans via three values (L*, a* and b*) ordered in a Cartesian coordinate system. From this, the color saturation chroma (C*) can be determined, which correlates with the surface roughness and in consequence with the porosity.

Multiple linear regression analysis

Multiple linear regression analyses were carried out using Modde software (MODDE 10, Sartorius, Göttingen, Germany) at a significance level of α=0.05 to visualize the effects of the individual factors on the responses. Therefore, models were developed for each set of factors (β) and response (Y_R)(Eq. 4).

$$Y_R = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{1,2} x_1 x_2. \quad (4)$$

The results of the analyses are graphically represented by contour plots using the developed models. Here, the factors are on the axes while the response is represented by the color scale. The centre points were

performed in triplicate.

Data-driven modeling approach

We use nonlinear auto-regressive with exogeneous inputs (NARX) models to develop a mathematical model for the considered tabletting process. The NARX formalism contains both model identification and model validation. Model identification step involves two steps: selecting model class, for example, polynomial NARX and parameter estimation. The identified model is then validated by using a different dataset. As our process contains multiple-input, multiple-output (MIMO), we model it using a MIMO NARX model. Assuming that we have m input variables and p output variables, the jth output of the NARX model is given by

$$y_j(k) = F_j\left(U_1(k), \ldots, U_m(k),\ Y_1(k-1), \ldots, Y_p(k-1)\right) + e_j(k), \quad (5)$$

where

$$U_i(k) = \left[u_j(k), u_j(k-1), \ldots, u_j(k-n_u)\right], \qquad i = 1,2,..,m,$$

$$Y_j(k-1) = \left[y_j(k-1), \ldots, y_j\left(k-n_y\right)\right], \qquad j = 1,2,..,p.$$

Here $u_i(k), y_i(k), e_i(k)$ denote, respectively, control input, process output, and noise terms. Function F_j is unknown. However, as we consider the model class of polynomial NARX to model the process, F_j is an unknown polynomial function. Finally, the parameters n_u, n_y represent the maximum lags for the input and output, respectively.

RESULTS

Experimental results

The experimental results are illustrated as contour plots in Figure 2. Here, increasing the feed frame speed during die filling leads to an increase in the amount of applied shear and the hydrophobic lubricant spreads over the binder particle surface. Friction between the die, punch and powder is reduced, resulting in a lower ejection force [13]. Likewise, an influence of the main-compression force is noted, whereby an increase leads to a stronger consolidation of the powder and thus to a higher ejection force. Based on the powder component properties, a correlation to the ejection force is also evident here. Furthermore, combinatorial effects can be identified. Thus, the ejection force can be utilized, taking into account the associated additional effects, to obtain information about the lubrication.

The color saturation chroma (C*), however, only indicates a significant influence of the main-compression force. Increasing the main-compression force results in decreased roughness on the surface of the tablet, which intensifies specular reflection and thus leads to a higher chroma (C*) value. Moreover, this decreases the volume of the tablet and causes an increase in porosity. This makes the chroma value (C*) particularly suitable for monitoring porosity.

The reflectance in the range from 320 to 330 nm serves to monitor the active pharmaceutical ingredient content. Here, the reflectance value increases with higher active pharmaceutical ingredient content. The experimental results provide an additional effect of the main-compression force as well as combinatorial effects. Thus, further models need to be developed to fully describe the process and to enable autonomous control.

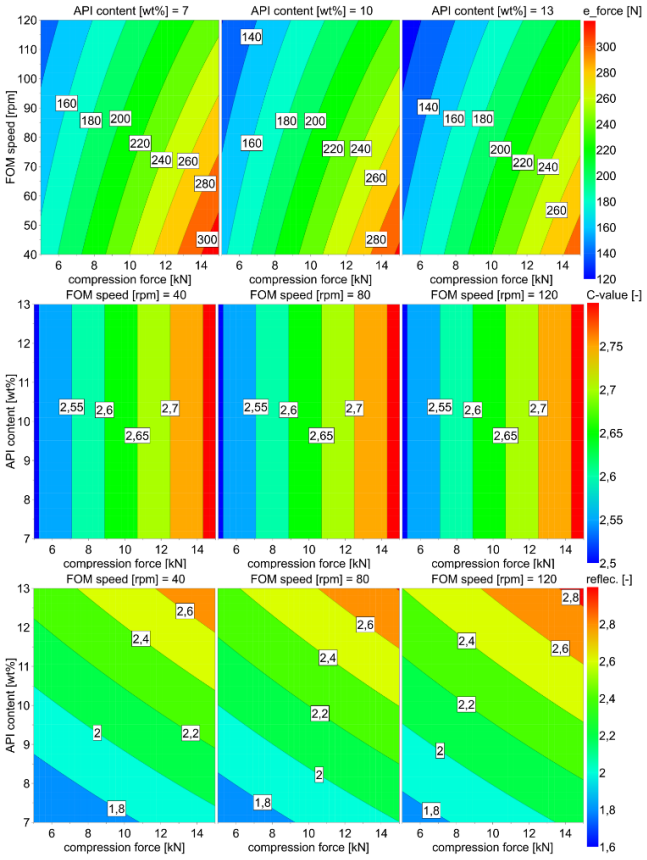

Figure 2. MLR results illustrated as contour plots for ejection force (e_force, top), chroma value (C-value, middle), and reflectance value (reflec., bottom).

In the second sub-experiment, the pre-compression force and the mass of the tablets were examined in relation to the filling depth and punch distance (Figure 3). The pre-compression force increases with decreasing the pre-compression roll distance. Here, more force is required to compress the same amount of powder into a smaller volume. In addition, an increased quantity of powder in the die leads to an increased force required to compact the powder to the defined height. Combinatorial effects can also be recognized here. However, the mass of the tablet only depends on the filling depth as this defines the volume filled with material prior to compression,

if the time for material filling in the die is sufficient.

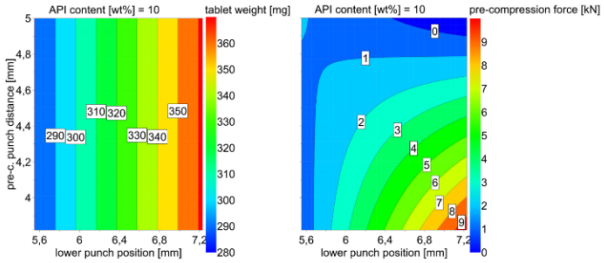

Figure 3. MLR results illustrated as countour plots for tablet weight (left) and pre-compression force (right).

Simulation results

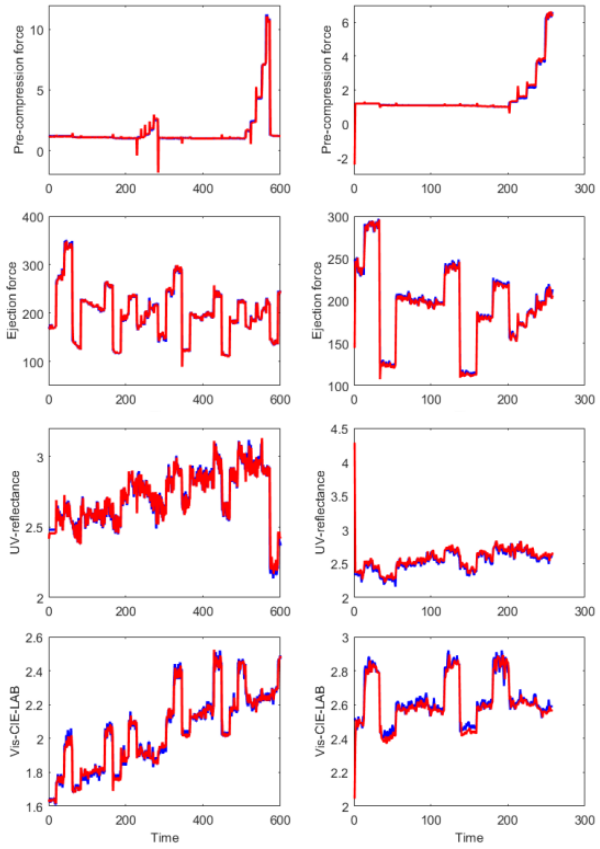

Figure 4. Identification (left) and validation (right) plots for all the outputs. The blue solid lines indicate the measured outputs, whereas red solid lines indicate computed outputs.

We aim to develop a parsimonious model for the process on the basis of gathered experimental data. Remarkably, several sets of experiments are performed to gather the necessary data. These data are then used for both identification and validation of the developed model. Total 859 samples of data are collected for various inputs and outputs mentioned in Figure 1. We split the dataset as follows: 70% for identification (training) and 30% for validation (generalization or testing).

We denote all the five input variables by u and all the four output variables by y. Furthermore, we denote the four outputs, namely, pre-compression force, ejection force, UV-reflectance, Vis-CIE-LAB, respectively, by y_1, y_2, y_3, y_4. With this notation in place, using the NARX approach, we obtain, based on the gathered dataset, the following model:

$$y_1(k) = \alpha_1 u(k) + \beta_1 u(k-1) + 0.9425 y_1(k-1)$$

$$y_2(k) = \alpha_2 u(k) + \beta_2 u(k-1) + \gamma_2 u^2(k) + \delta_2 u^2(k-1) + 0.9 y_2(k-1)$$

$$y_3(k) = \alpha_3 u(k) + \beta_3 u(k-1) + 0.9181 y_3(k-1)$$

$$y_4(k) = \alpha_4 u(k) + \beta_4 u(k-1) + \gamma_4 u^2(k) + \delta_4 u^2(k-1) + 0.6516 y_4(k-1) - 0.0756 y_4(k-2) + 0.0012 y_4^2(k-1) + 0.0510 y_4^2(k-2), \quad (6)$$

where

$$\alpha_1 = [0.0135 \ -1.7538 \ -0.4880 \ -0.7282 \ 2.5461]$$

$$\beta_1 = [-0.0129 \ 1.5106 \ 0.4653 \ 0.7258 \ -2.3407]$$

$$\alpha_2 = [2.6 \ 96.9 \ 1092.7 \ -74.6 \ -425.5]$$

$$\beta_2 = [2.5 \ -103.7 \ -1010.1 \ 730 \ 396]$$

$$\gamma_2 = [0 \ -12.5 \ -190.1 \ 2.4 \ 43.7]$$

$$\delta_2 = [0 \ 13.1 \ 176 \ -2.4 \ -40.7]$$

$$\alpha_3 = [0.0100 \ -0.1067 \ -0.7568 \ 0.2428 \ 0.5787]$$

$$\beta_3 = [-0.0096 \ 0.1160 \ 0.6993 \ -0.2406 \ -0.5288]$$

$$\alpha_4 = [-0.0007 \ -1.0966 \ -6.3087 \ 0.0093 \ 4.5086]$$

$$\beta_4 = [0.0011 \ 0.7916 \ 5.1898 \ 0.0104 \ -3.5830]$$

$$\gamma_4 = [0 \ 0.1243 \ 0.8290 \ 0.0010 \ -0.3157]$$

$$\delta_4 = [0 \ -0.0904 \ -0.6881 \ -0.0010 \ 0.2515].$$

For the purposes of illustration, both the identification and validation plots for all the four outputs are shown in Figure 4. It can be seen that the computed ones almost agree with the measured ones both in identification and validation plots. Thus, the model is generalizable and can be used for further analysis and control tasks. Moreover, the performance of the model is evaluated by computing the relative percentage error between measured and computed outputs defined as

$$\frac{\|y_{measured} - y_{computed}\|}{\|y_{measured}\|} \ 100\%.$$

The relative percentage error for both identification and validation of all the four outputs are given in Table 1. It can be seen that the errors are in the acceptable range. Hence, we conclude that the identified NARX model represents the process. In the future, we will use this identified model in real-time control of the process for desired product properties.

Table 1: Relative percentage error for id. (identification) and val. (validation) of various outputs.

Outputs	Id.	Val.
Pre-compression force	12.23	13.90
Ejection force	3.02	4.39
UV-reflectance	2.32	5.42
Vis-CIE-LAB	1.56	1.86

CONCLUSION

In this study, experimental data were collected for the powder compaction process on a tablet press. Full factorial experimental designs were analyzed in terms of multiple linear regression analysis and significant factors associated with product quality were identified. Thereby, process parameters were combined with data from a real-time spectroscopy method based on UV/Vis spectroscopy. These collected data and gained insights from the process were further utilized in developing a parametric data-driven model of the process based on the NARX modeling approach. An avenue of future research work would be the application of this developed data-driven model in real-time control of the process.

ACKNOWLEDGEMENTS

This work was funded by the German Research Foundation (DFG) – GZ: BA 5511/9-1 and TH 1817/8-1.

REFERENCES

1. Hole G, Hole AS, McFalone-Shaw I. Digitalization in pharmaceutical industry: What to focus on under the digital implementation process?. *Int J Pharm* X 3:100095 (2021) https://doi.org/10.1016/j.ijpx.2021.100095
2. Su Q, Ganesh S, Moreno M, Bommireddy Y, Gonzalez M, Reklaitis GV, Nagy ZK. A perspective on Quality-by-Control (QbC) in pharmaceutical continuous manufacturing. *Comput Chem Eng* 125:216–231 (2019) https://doi.org/10.1016/j.compchemeng.2019.03.001
3. Ervasti T, Simonaho S P, Ketolainen J, Forsberg, P, Fransson M, Wikström H, Folestad S, Lakio S, Tajarobi P, Abrahmsén-Alami S. Continuous manufacturing of extended release tablets via powder mixing and direct compression. *Int J Pharm* 495:290–301 (2015) https://doi.org/10.1016/j.ijpharm.2015.08.077
4. Siegmann E, Forgber T, Toson P, Martinetz MC, Kureck H, Brinz T, Manz S, Grass T, Khinast J. Powder flow and mixing in different tablet press feed frames. *Adv Powder Technol* 31:770–781 (2020) https://doi.org/10.1016/j.apt.2019.11.031
5. Markl D, Warman M, Dumarey M, Bergman E-L, Folestad S, Shi Z, Manley LF, Goodwin DJ, Zeitler JA. Review of real-time release testing of pharmaceutical tablets: State-of-the-art, challenges and future perspective. *Int J Pharm* 582:119353 (2020) https://doi.org/10.1016/j.ijpharm.2020.119353
6. Zhong L, Gao L, Li L, Zang H. Trends-process analytical technology in solid oral dosage manufacturing. *Eur J Pharm Biopharm* 153:187–199 (2020) https://doi.org/10.1016/j.ejpb.2020.06.008
7. Pauli V, Roggo Y, Pellegatti L, Nguyen Trung NQ, Elbaz F, Ensslin S, Kleinebudde P, Krumme M. Process analytical technology for continuous manufacturing tableting processing: A case study. *J Pharm Biomed Anal* 162:101–111 (2019) https://doi.org/10.1016/j.jpba.2018.09.016
8. Verhaegen M, Verdult V. Filtering and System Identification: A Least Squares Approach. Cambridge University Press (2007).
9. Billings SA. Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains. John Wiley & Sons (2013).
10. Pisoni E, Farina M, Carnevale C, Piroddi L. Forecasting peak air pollution levels using NARX models. *Eng Appl Artif Intell* 22:593-602 (2009) https://doi.org/10.1016/j.engappai.2009.04.002
11. Ruiz LGB, Cuéllar MP, Calvo-Flores MD, Jiménez MDCP. An application of non-linear autoregressive neural networks to predict energy consumption in public buildings. *Energies* 9:1-21 (2016) https://doi.org/10.3390/en9090684
12. Mohanty S, Patra PK, Sahoo SS. Prediction of global solar radiation using nonlinear auto regressive network with exogenous inputs (NARX). In *2015 39th National Systems Conference (NSC)* (pp. 1-6). IEEE. (2015) https://doi.org/10.1109/NATSYS.2015.7489103
13. Brands R, Mathias C, Bartsch J, Thommes M. Advancing tablet lubrication: A systematic comparison of feed frame lubrication and internal lubrication. *Powder Technol* 434:119369 (2024) https://doi.org/10.1016/j.powtec.2024.119369