Developing the final product attribute prediction model in a continuous direct compression process

Yuki Kobayashia, Sanghong Kimb, Takuya Nagatoc, Takuya Oishib,c, Manabu Kanoa,\*

aDepartment of Systems Science, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto, 6068501, Japan

bDepartment of Applied Chemistry, Tokyo University of Agriculture and Technology, 2-24-16 Naka-cho, Koganei, 1840012, Japan

cResearch and Development Division, Powrex Corporation, 5-5-5 Kitagawara, Itami, 6640837, Japan

manabu@human.sys.i.kyoto-u.ac.jp

Abstract

To analyze and design pharmaceutical manufacturing processes, models that can predict critical quality attributes (CQAs) of final products from process parameters (PPs) and material attributes (MAs) have been investigated. Such a model depends on a type of equipment; thus, the entire model needs to be rebuilt when equipment is changed. To reduce the effort of model rebuilding, we propose a series model, in which an equipment-dependent model predicts inputs of the equipment-independent model that predicts CQAs. This study focuses on the equipment-independent model of a continuous direct compression (CDC) process. In experiments by changing nine PPs, MAs of intermediate products and CQAs of final products were measured. CQA prediction models were constructed by using different sets of input variables. As a result, equipment-independent models accurately predicted the disintegration time of tablets. Moreover, ten MAs and the main compression force were identified as important variables for accurate prediction.

**Keywords**: Continuous manufacturing, continuous direct compression, variable selection, process model, intermediate product material attribute

* 1. Introduction

Quality assurance methods in the pharmaceutical industry are shifting from Quality by Testing (QbT), in which the quality is ensured by product testing, to Quality by Design (QbD), in which the quality is ensured by the good design of the products and manufacturing processes. To realize QbD, mathematical models that express the relationship between CQAs of final products and PPs and MAs play a significant role. This study focuses on a continuous direct compression (CDC) process. CDC is one of the methods of continuous manufacturing of tablets and its process consists of feeders, mixers, and a tablet press.

For integrated CDC processes, Kreiser et al. (2022) developed multiple linear regression models that predict final product CQAs from the PPs of the mixer. Bekaert et al. (2022) developed a partial least squares (PLS) model to predict final product CQAs from the PPs and the blend properties. These models use the PPs of the mixers, however, the types of mixers used are different and the mixers have different PPs. As shown in Figure 1 (top), previous studies have predicted CQAs from equipment-dependent PPs, equipment-independent PPs, and MAs. Because such models depend on equipment, the entire CQA prediction model needs to be rebuilt when equipment is changed. On the other hand, as shown in Figure 1 (bottom), this study aims at a series model that predicts MAs independent of equipment from variables including equipment-dependent PPs and then predicts CQAs from equipment-independent PPs and MAs. Such a model has the advantage of requiring less effort when equipment is changed since only the equipment-dependent model needs to be rebuilt.

In this study, aiming for the concurrent use of equipment-independent models that predict CQAs and equipment-dependent models that predict MAs in a CDC process, an attempt was made to construct equipment-independent models. Moreover, important variables for accurate prediction were identified.

|  |
| --- |
| 図形, 四角形  自動的に生成された説明 |
| Figure 1: Comparison of prediction models in the previous (top) and this study (bottom). |

* 1. Materials and Methods
		1. Process

Figure 2 shows the process flow diagram investigated in this study. Acetaminophen (APAP) (Spera Nexus, Japan) was used as an active pharmaceutical ingredient, SuperTab 11SD (DFE Pharma, Germany) was used as an excipient, and magnesium stearate (MgSt) (Taihei Chemical Industrial, Japan) was used as a lubricant. APAP and SuperTab were fed into mixer 1 (MG100, Powrex, Japan) by different feeders (LIW-300-P, Ishida, Japan) and mixed in mixer 1. Intermediate product 1 coming out from mixer 1 and MgSt were fed into mixer 2 and mixed. Intermediate product 2 coming out from mixer 2 was fed into the tablet press (FETTE 102i, Fette Compacting, Germany) and compressed and made into tablets, the final product.

|  |
| --- |
| 図形  中程度の精度で自動的に生成された説明 |
| Figure 2: Process flow diagram. |

* + 1. Experiments

A definitive screening design (DSD) with 19 runs (Jones et al., 2011) was adopted to determine the combinations of levels of nine variables listed in Table 1 because DSDs require a small number of experiments and accommodate second-order effects. The nine variables were grouped into four groups ($X\_{1-4}$). The measured MAs of the intermediate products 1 and 2 are listed in Table 2 and they were denoted as groups $Z\_{1}$ and $Z\_{2}$. The measured CQA was the disintegration time of tablets.

Table 1: Variables and their levels in the experiments.

|  |  |  |
| --- | --- | --- |
| Variable group  | Variable | Level |
| $$X\_{1}$$ | Mass fraction of APAP | 5, 10, 15 % |
|  | Production rate | 15, 20, 25 kg/h |
| $$X\_{2}$$ | Scraper blade rotation speed of mixer 1 | 30, 50, 70 rpm |
|  | Center blade rotation speed of mixer 1 | 500, 1500, 2500 rpm |
| $$X\_{3}$$ | Scraper blade rotation speed of mixer 2 | 30, 50, 70 rpm |
|  | Center blade rotation speed of mixer 2 | 100, 550, 1000 rpm |
| $$X\_{4}$$ | Force feeder blade rotation speed  | 10, 30, 50 rpm  |
|  | Ratio of pre-compression force to main compression force | 20, 40, 60 % |
|  | Main compression force | 5, 15, 20 kN |

Table 2: Intermediate product material attributes.

|  |  |
| --- | --- |
| Variable | Unit |
| Measured mass fraction of APAP | - |
| Standard deviation (SD) of measured mass fraction of APAP | % |
| Relative standard deviation (RSD) of measured mass fraction of APAP | % |
| Aerated bulk density | g/cm3 |
| Tapped bulk density | g/cm3 |
| Compressibility | - |
| Angle of repose | Degree |
| Angle of rupture | Degree |
| Angle of difference | Degree |
| Angle of spatula | Degree |
| Degree of agglomeration | % |
| Degree of dispersion | % |
| Particle size distribution, D10, D50, D90 | $$μm$$ |
| Basic flowability energy (BFE) | mJ |
| Flow rate index (FRI) | - |
| Specific energy (SE) | mJ/g |
| Cohesion | kPa |
| Unconfined yield strength (UYS) | kPa |
| Major principal stress (MPS) | kPa |
| Angle of internal friction (AIF) | Degree |
| Flow function coefficient (FF) | - |

* + 1. Modeling

The evaluation criteria of the model accuracy are root mean squared error (RMSE) and Q2.

|  |  |
| --- | --- |
| $$RMSE=\sqrt{\frac{1}{N}\sum\_{n=1}^{N}\left(y\_{n}-\hat{y}\_{n}\right)^{2}}$$ | (1) |
| $$Q^{2}=1-\frac{\sum\_{n=1}^{N}\left(y\_{n}-\hat{y}\_{n}\right)^{2}}{\sum\_{n=1}^{N}\left(y\_{n}-\overbar{y}\right)^{2}}$$ | (2) |

, where $N$ is the number of the samples, $y\_{n}$ is the $n$th actual value, $\hat{y}\_{n}$ is the $n$th predicted value, and $\overbar{y}$ is the mean of the output variable.

The modeling procedure is as follows:

1. Set the output variable of the model to be the disintegration time.
2. Select the model from PLS (Geladi et al., 1986) or random forest (RF) (Breiman, 2001).
3. Set $p=0$.
4. Set $m=0$.
5. Put the variables in the input variable set $p$ and in the nominal variable set $\acute{S}\_{p}$ in Table 3 into the input variable set $S\_{m}$.
6. Let one sample be test data, and the other be training data.
7. If PLS is selected in step 2, construct a PLS model using cross-validation. In the cross-validation, tune the number of latent variables. If RF is selected in step 2, construct a RF model using cross-validation. In the cross-validation, tune three parameters: the number of trees, the maximum depth of trees, and the number of features to consider when looking for the best split.
8. Calculate RMSE with test data.
9. Calculate each variable's permutation importance (PI) (Altmann et al., 2010).
10. Perform steps 6 through 9 using each sample as test data once.
11. Calculate the mean of RMSEs, referred to as RMSECV.
12. Calculate the mean of PIs for each variable, referred to as PICV.
13. Exclude the input variable with the smallest PICV among those not included in the nominal variable set $\acute{S}\_{p}$ from $S\_{m}$.
14. Update $m=m+1$
15. Perform steps 6 through 14 until $S\_{m}$ has the same number of input variables as that in $\acute{S}\_{p}$.
16. Select $S\_{m}$ with the smallest RMSECV and let the selected variable set be denoted $\tilde{S}\_{p}$.
17. Update $p=p+1$, and perform steps 4 through 17.
18. Perform steps 2 through 17 for PLS and RF.

Table 3: Variable groups containing input variable candidates.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Number of the input variable set  | $$X\_{1}$$ | $$X\_{2}$$ | $$Z\_{1}$$ | $$X\_{3}$$ | $$Z\_{2}$$ | $$X\_{4}$$ | Nominal variable set $\acute{S}\_{p}$ |
| 1 | ✓ |  |  |  |  | ✓ | Empty set |
| 2 |  |  |  |  | ✓ |  | $$\tilde{S}\_{1}$$ |
| 3 |  |  |  | ✓ |  |  | $$\tilde{S}\_{2}$$ |
| 4 |  |  | ✓ |  |  |  | $$\tilde{S}\_{2}$$ |
| 5 |  | ✓ |  |  |  |  | $$\tilde{S}\_{2}$$ |

* 1. Results and Discussions
		1. Experiments

Figure 3 shows the relationship between disintegration time and both the APAP mass fraction setpoint (left) and the main compression force (right). The disintegration time varied even when the APAP mass fraction setpoint was the same, thus, it is affected not only by the formulation but also by PPs such as the main compression force.

* + 1. Modeling

Table 4 shows the prediction results of disintegration time. The PLS models achieved better prediction accuracy than the RF models, therefore, the results of the PLS models are focused on in the following paragraph.

The comparison of RMSECV and Q2 between the input variable sets 1 and 2 in Table 4 shows that using MAs of intermediate product 2 significantly improves prediction accuracy. The comparison of RMSECV and Q2 among the input variable sets 2, 3, 4, and 5 in Table 4 shows that variables in $X\_{2}$,$X\_{3},$ and $Z\_{1}$do not significantly improve the prediction accuracy. The variables selected from the input variable set 2 are enough to accurately predict the disintegration time. This means that the equipment-independent model can be constructed for disintegration time prediction. The variables selected from the input variable set 2 are shown in Table 5 and the actual and predicted values are shown in Figure 4.

|  |  |
| --- | --- |
|  |  |
| Figure 3: Relationship between disintegration time and APAP mass faction setpoint (left) and main compression force (right). |

Table 4: Prediction results of disintegration time

|  |  |  |
| --- | --- | --- |
| Number of input variable set | RMSECV | Q2 |
| PLS | RF | PLS | RF |
| 1 | 122.2 | 99.8 | 0.18 | 0.50 |
| 2 | 48.2 | 99.7 | 0.86 | 0.36 |
| 3 | 48.2 | 99.7 | 0.86 | 0.36 |
| 4 | 48.2 | 99.7 | 0.86 | 0.36 |
| 5 | 45.5 | 99.7 | 0.85 | 0.36 |

Table 5: The variables selected from the input variable set 2 in Table 3.

|  |  |
| --- | --- |
| $$Z\_{2}$$ | $$X\_{4}$$ |
| RSD of measured mass fraction of APAP, MPS, SE, SI, UYS, angle of spatula, degree of agglomeration, compressibility, angle of difference, aerated bulk density | Main compression force |

|  |
| --- |
|  |
| Figure 4: Prediction results of disintegration time by the PLS model with variables in Table 5. |

* 1. Conclusions

Models for predicting the disintegration time of tablets were developed by using different sets of input variables with the data obtained in the experiments under different operating conditions in the CDC process. As a result, models independent of the mixer accurately predicted the disintegration time of tablets. Ten MAs and the main compression force were identified as important variables for accurate prediction. In the future, we will construct equipment-dependent models that accurately predict MAs. These models will be used together with the equipment-independent models constructed in this study to predict CQAs from PPs.

Acknowledgments

This work was supported by JST, the establishment of university fellowships towards the creation of science technology innovation, Grant Number JPMJFS2123, and by JSPS KAKENHI Grant Number JP21H01704.

References

A. Altmann, L. Toloşi, O. Sander, T. Lengauer, 2010, Permutation importance: a corrected feature importance measure. *Bioinformatics*. **26**, 1340–1347.

B. Bekaert, B. Van Snick, K. Pandelaere, J. Dhondt, G. Di Pretoro, T. De Beer, C. Vervaet, V. Vanhoorne, 2022, Continuous direct compression: Development of an empirical predictive model and challenges regarding PAT implementation. *International Journal of Pharmaceutics: X*. **4**, 100110.

L. Breiman, 2001, Random Forests. *Mach. Learn.* **45**, 5–32.

P. Geladi, B. R. Kowalski, 1986, Partial least-squares regression: a tutorial. *Anal. Chim. Acta* **185**, 1–17.

B. Jones, C. J. Nachtsheim, 2011, A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects. *J. Commod. Sci. Technol. Qual.* **43**, 1–15.

M. J. Kreiser, C. Wabel, K. G. Wagner, 2022, Impact of Vertical Blender Unit Parameters on Subsequent Process Parameters and Tablet Properties in a Continuous Direct Compression Line. *Pharmaceutics*. **14**, 278.