Journal of Pharmaceutical Sciences 112 (2023) 1427–1439

Contents lists available at ScienceDirect

Journal of Pharmaceutical Sciences

journal homepage: www.jpharmsci.org

Pharmaceutics, Drug Delivery and Pharmaceutical Technology

Developing a Virtual Flowability Sensor for Monitoring a Pharmaceutical Dry Granulation Line

Rexonni B. Lagare[a,*], Yan-Shu Huang[a], Craig Oh-Joong Bush[a], Katherine Leigh Young[a], Ariana Camille Acevedo Rosario[a], Marcial Gonzalez[b], Paul Mort[c], Zoltan K. Nagy[a], Gintaras V. Reklaitis[a]

[a] Davidson School of Chemical Engineering, Purdue University, West Lafayette, IN 47907, USA
[b] School of Mechanical Engineering, Purdue University, West Lafayette, IN 47907, USA
[c] School of Materials Engineering, Purdue University, West Lafayette, IN 47907, USA

A R T I C L E I N F O

Article history:
Received 18 September 2022
Revised 10 January 2023
Accepted 10 January 2023
Available online 14 January 2023

Keywords:
Continuous pharmaceutical manufacturing
Condition monitoring
Granule flowability
Size and shape distribution
Projection to latent structures
Dry granulation
Quality by design
Process analytical technology
Powder technology
Virtual sensor

A B S T R A C T

Current technologies to measure granule flowability involve at-line methods that can take hours to perform. This is problematic for a continuous dry granulation tableting line, where the quality assurance and control of the final tablet products depend on real-time monitoring and control of powder flowability. Hence, a real-time alternative is needed for measuring the flowability of the granular products coming out of the roller compactor, which is the unit operation immediately preceding the tablet press. Since particle analyzers have the potential to take inline measurements of the size and shape of granules, they can potentially serve as real-time flowability sensors, given that the size and shape measurements can be used to reliably predict flowability measurements.

This paper reports on the use of Partial Least Squares (PLS) regression to utilize distributions of size and shape measurements in predicting the output of three different types of flowability measurements: rotary drum flow, orifice flow, and tapped density analysis. The prediction performance of PLS had a coefficient of determination ranging from 0.80 to 0.97, which is the best reported performance in the literature. This is attributed to the ability of PLS to handle high collinearity in the datasets and the inclusion of multiple shape characteristics—eccentricity, form factor, and elliptical form factor—into the model. The latter calls for a change in industry perspective, which normally dismisses the importance of shape in favor of size; and the former suggests the use of PLS as a better way to reduce the dimensionality of distribution datasets, instead of the widely used practice of pre-selecting distribution percentiles.

Introduction

The basic operations in the continuous production of pharmaceutical tablets normally involve powder feeding, blending, and then powder compaction into tablets. Loss-in-weight feeders are often used to dose the different powder components of a tablet, in order to meet the correct composition after mixing the dosed components together in a continuous powder blender. Then, a tablet press compacts the powders into a pharmaceutical tablet with the correct physical characteristics of hardness, strength, and dimensions.[1] Because of the processes that a blend undergoes while inside the tablet press, the performance of the tablet press is dependent on the flowability of the pharmaceutical powder blend. For certain compositions, especially those with relatively high active pharmaceutical ingredients (API), the flowability can become insufficient enough to be detrimental to the quality of the final tablet products.[2,3]

To solve this problem, the tablet press is usually preceded by a granulation step, one of the primary goals of which is to improve the flowability by increasing the size of the particles.[4–6] During this process, the shape of the particles could be modified as well, which could either improve or diminish the flowability and other properties of the granules.[7] However, the changes in the size and shape distribution of the powder as it goes through a granulator, such as a roller compactor (RC), could vary. Even if the feed to the RC is a powder blend with consistently uniform properties, the quality of product granules could become inconsistent if the RC is unable to reliably maintain its targeted process parameters. This is likely to happen as the condition and performance of the RC and its associated controllers degrade

* Corresponding author.
E-mail address: rlagare@purdue.edu (R.B. Lagare).

https://doi.org/10.1016/j.xphs.2023.01.009

over time, leading to production of granules with high variability in its critical-to-quality attributes such as flowability.[8−11] Conversely, a perfectly running RC can still result in low quality granules if the powder blend feed to the RC is not consistent, possibly a result of faulty upstream operations. While certain levels of fluctuations in the quality of the product granules can be tolerated, corrective measures need to be implemented in a timely manner in case these fluctuations do go beyond the accepted limits. Hence, effective real-time monitoring is required.

One of the major challenges with monitoring granule flowability is the absence of a singular measure. It is a common practice to choose among the multiple methods available to characterize flowability depending on the process conditions experienced by the granule. In a tablet press, the granules are subjected to various flow regimes that include dynamic flow in the feed frame and orifice flow into the die during the dosing stage. Hence, measuring flowability to monitor the granules exiting the roller compactor require multiple flowability measurement methods.

Perhaps the biggest challenge with monitoring flowability is the absence of a real-time option. Flowability measurements are performed at-line, which would require a sample size reduction step to minimize sampling error. Over-all, performing these at-line flowability measurements could take hours, which makes it impractical for real-time monitoring. To circumvent this limitation, real-time measurements of the size and shape distribution have been a popular approach. [12−16] In fact, size and shape measurements from particle analyzers have been demonstrated to be successful in monitoring granulation processes in real-time, albeit for wet granulation.[14,16] However, this approach is limited by the unclear quantitative relationship that the size and shape distributions have on the behavior of the granules, albeit the qualitative relationship has long been established.

An interesting solution would be to utilize the size and shape measurements from the particle analyzers to predict granule behavior such as flowability.[14,16] The challenge is in the development of a predictive model that can embody the exact relationship of the size and shape distribution of the granules versus its flowability. More specifically, if the size and shape of a powder can be determined in real-time, would it be possible to infer its flowability based on that information alone?

The size and shape of constituent particles have long been established to be primary indicators of powder flow behavior,[6,17−19] but they have not been demonstrated to sufficiently regress against and reliably predict powder flowability. Many attempts have been made towards this goal, and the latest was in 2020,[20] which showed promising results. In that work, a linear regression model was used to regress the angle of repose and the Carr ratio, which are both measures of flowability, versus selected percentiles of the measured size distribution. For example, the angle of repose was regressed against the 10th percentile of the size distribution (i.e., the D10 value), and this resulted in the best performing model for angle of repose with a 0.65 coefficient of determination. The D10 of the size also resulted in the best model for predicting the Carr ratio, resulting in a 0.92 coefficient of determination. In that work, the circularity distribution of the powders was measured to provide insight on the shape but were not mentioned during the discussions on regression.

While demonstrably useful, this univariate approach to regression is limiting the success of the effort as it ignores the other percentiles that could offer useful information for predictions. Moreover, it ignores the probability that these percentiles could be highly-correlated with each other. This is especially true for granulation, where the formation of a larger particle from two or more smaller particles would result in the increase of the higher size percentiles, and the corresponding decrease of the lower size percentiles. Finally, the univariate approach fails to consider the effect of shape and its correlation with size. Such a relationship has been demonstrated in previous works and should be taken into account.

Aside from the regressors, the flowability measurements are also expected to be highly correlated. As an example, the rotary drum method for characterizing flow measures the angle of repose of the powders in the rotary drum as it spins at a certain RPM. Measures of the angle of repose are taken at various levels of RPM and this results in multiple angles of repose that are, as may be expected, correlated with each other—e.g., powders with good flowability would tend to have a low angle of repose regardless of the RPM level. Hence, both the regressors and the predicted variables are highly-correlated, and the regression method of choice should perform well with such datasets.

A good candidate for a regression model is projection to latent structures (PLS), which is a form of latent variable multivariate regression, and has been demonstrated to work well with highly-correlated datasets. It should be a clear upgrade over univariate linear regression and should be expected to perform better. This was evident when Yu *et.al.*(2011)[21] successfully used PLS to regress measurements of the Ring Shear Tester with size and shape distribution, reporting high coefficient of determination values of up to 0.82. However, using the Ring Shear Tester alone is inadequate for characterizing granule flowability in a tableting line because it has poor sensitivity for powders with decent flowability—i.e., flow function coefficients higher than 10.[22]

Since the RC granules are intended for the tablet press, flowability measurements should be able to characterize dynamic flow conditions at the feed frame and during die filling; these can be respectively achieved using the rotary drum flow and orifice flow test. [22,23] Furthermore, tapped density analysis can be included to characterize powder compression mechanics in the die and partly some of the quasi-static flow conditions that occurs in the tablet press.[17,19,24,25] It is thus significant to extend the demonstrated success of using PLS to predict these three flowability measurements that are relevant to tablet pressing, namely: rotary drum flow, orifice flow, and tapped density.

This paper reports the performance of PLS as a regression model for predicting multiple flowability measurements based on size and shape distribution, and it offers four unique contributions. First, it demonstrates the importance of shape-related measurements in developing a flowability sensor. Second, it shows the limitations of arbitrarily selecting percentiles (or D-values) to represent distributions of particle measurements. Third, it examines the predictability of flowability measurements based on size and shape distribution measurements. Lastly, it reveals that although the flowability measurement variable are correlated with each other, the three flowability tests under consideration also contributed unique information regarding flow behavior, justifying their joint consideration in the modeling effort.

Materials and Methods

Granules

The powders used in this work were Acetaminophen Grade 0048 (APAP) purchased from Mallinckrodt Pharmaceuticals and Avicel microcrystalline cellulose Grade PH-102 (MCC-102) from FMC BioPolymer. A powder blend was prepared by blending 10g APAP for every 90g of MCC-102 for 30 minutes in a 5L Tote blender. Using this powder blend as a feed to an Alexanderwerk WP-120 Roller Compactor (RC), six batches of granules were produced. The first three batches varied with the roll pressure used to form the granules, while the next three batches also varied with the roll pressure albeit with a larger screen size than the first three batches.

Table 1
Roller compactor process conditions for producing six batches of pharmaceutical granules.

Batch	Roll Pressure, bar	Granulator Screen Size, mm	Pre-Granulator Screen Size, mm	Roll Speed, RPM	Roll Gap, mm
1	30	1.0	2.5	4.0	2.0
2	60	1.0	2.5	4.0	2.0
3	90	1.0	2.5	4.0	2.0
4	30	1.6	2.5	4.0	2.0
5	60	1.6	2.5	4.0	2.0
6	90	1.6	2.5	4.0	2.0

All batches were produced with the gap control (feedback) function of the RC turned on. This means that the Roll Gap was kept at a constant value by constantly adjusting the speed of the feed screw that pushes the powders into the two counter-rotating rolls. On the other hand, the roll pressure was maintained via its own separate PID control. Table 1 lists the process conditions used to produce the batches of granules that were used to develop the flowability sensor.

Sample Collection and Sampling Size Reduction

The collection of granular samples started approximately after 1kg of throughput has been produced by roller compactor, ensuring that the collected granules represent the steady-state products of the unit operation. Since the volume requirements of each flowability tests are much smaller, an appropriate size reduction scheme is employed to minimize sampling error. This is achieved by using the PT100 sample divider by Retsch,[26] to subdivide the powders into eight equal subsamples. To exactly match the volume requirements of the flowability measurements, the operation of the sample divider was performed in multiple passes, recombining an appropriate number of subsamples so that the subsequent pass results into eight equal subdivisions that each has the desired subsample volume. Table 2 shows the sample volume requirements for each flowability measurement.

Flowability Measurements

There are three flowability measurements considered in this study: powder flow in a rotary drum, tapped density analysis, and powder flow through an orifice. As aforementioned, these measurements were considered jointly in order to match the flow conditions that a powder undergoes in a tablet press. Powder flow in a rotary drum was performed using the GranuDrum,[27] tapped density analysis was performed using the GranuPack,[28] and powder flow through an orifice was measured using Flodex.[29] Both GranuDrum and GranuPack were manufactured by GranuTools, while Flodex was manufactured by Teledyne Hanson.

GranuDrum

When taking flowability measurements using the GranuDrum, powder is loaded into a rotary drum with sides made of optical glass that allows the measurement of the angle of repose, which is the angle that the powder surface makes (from the horizontal) as the drum is rotated at a certain speed. This speed is then increased at 2RPM increments from 2RPM to 20RPM and the angle of repose is estimated each time. Multiple measurements are taken at each rpm setting, and the calculated mean and variance of those measurements are respectively included in the data output as the angle of repose and cohesion for that particular RPM, resulting in a data matrix of at least 20 columns.

Furthermore, since materials with good flowability would generally have low angle of repose while the opposite is true for powders with poor flowability, each of these columns in the data matrix are expected to be highly correlated with each other.

Table 2
Sample volume requirements for each flowability test.

Flowability Measurement	Volume
1. GranuDrum	50 mL
2. GranuPack	25 mL
3. Flodex	34 mL

GranuPack

For flowability measurements using the granuPack, the powder sample of known mass is introduced into a tube and a diabolo is added on top of the powder to ensure a flat surface at the top of the powder. As the tube is tapped at a precise intensity, the instrument monitors the changes in the position of the diabolo, which is then translated to a corresponding change in the powder volume and bulk density. This results in a compaction curve of number of taps versus bulk density (or volume), and several parameters are automatically calculated to represent flowability characteristics: bulk density, tapped density, Hausner ratio, Carr ratio, and two kinetic parameters. While this data output is much less than that of the GranuDrum, these characteristics are still expected to be correlated with each other, as powders with good flowability are expected to have low Hausner and Carr ratios while the opposite is true for powders with poor flowabiity.

Flodex

The Flodex involves loading a known mass of powder into a cylindrical flat bottom hopper with a spring-loaded latch that can be activated to release the powders. A concentric annular disk with a specific orifice diameter is placed just above the bottom latch, so powder flows through the orifice upon its release. Because of the annular disk, not all of the powder will be drained out of the container. If the powder is not too cohesive, it will form an inverted cone shape, the angle of which could be used to characterize flowability. This angle is called the drained angle of repose, and it could be estimated based on the mass of residual powder and the size of the orifice using Equation (1), which is based on simple geometry and symmetry of a cylinder.[22]

$$\phi_d = \arctan\left[\frac{24M_{ret}}{\pi\rho_{bulk}\left(2D_h^3 - 3D_oD_h^2 + D_o^3\right)}\right] \tag{1}$$

The annular disks could be switched out to change the orifice size, which ranges from 2.0mm to 20mm. Finding the orifice size through which a powder can feely fall is another flowability characteristic that can be measured with Flodex. This is referred to as the flowability index,[23] which will henceforth be referred to in this paper as the orifice flow index to avoid confusion with other flowability indices. Another similar flowability measure is the jamming onset, which is the largest orifice size at which the powder jams.[22]

Particle Size and Shape Measurements

The size and shape of the granule produced by the roller compactor were measured using the SolidSizer,[30] manufactured by Canty

Process Technology. The SolidSizer was chosen primarily because of its ability to measure all the particles that are loaded into its hopper. Since its capacity is higher than the volume requirements of all considered flowability tests, it can provide the size and shape measurements of all the particles that are responsible for the results of each flowability tests. This minimizes sampling error, which is important when trying to establish the predictability of flowability based on size and shape measurements.

The software for the SolidSizer has been modified by Canty Process Technology to export the full dataset of measurements for each particle, making it possible to extract the distributions of size and shape-related parameters, which would prove to be critical in predicting flowability measurements. It is for this reason that the SolidSizer, despite being an at-line measurement method, was chosen for this study. Other commercially available particle analyzers could have been used, like the $Eyecon_2$[31] from Innopharma, that can be readily deployed in-line to provide real-time measurements of a continuous process. However, the predictive performance of such particle analyzers would be limited, since most are focused on the size distribution and less importance is given to the shape distribution. In some devices, only the central tendencies of the shape distribution—e.g., the mean and standard deviation—are reported, which would not have been ideal per the results discussed in Section 3.3.

Size and Shape Characterization

The powders that are loaded into the hopper of the SolidSizer exit to a slightly tilted vibratory chute. This chute transports the particles to the lower end of the chute where it falls off as a curtain of particles. With a light source on one end of the fall-off, a high-speed camera on the other end captures back-lit 2D projections of the particles. An illustration of the SolidSizer and a schematic of its working mechanism are respectively shown in Figs. 1 and 2.

The camera images are then analyzed by the JM Canty Image Analysis software, which detects the edges of particle projections and extracts features relevant to its size and shape. As shown in Fig. 3, an ellipse can be fitted around the particle edges to deduce relevant primary features such as the major diameter, minor diameter, area, and perimeter.

The size distribution of each granule sample is based on the diameter of a sphere with an equivalent volume as the measured particle. To get a volume from a 2-dimensional image, the geometric average between the major and minor diameter of each particle is computed to get a tertiary diameter, and the resulting dimensions are used to compute the volume of the ellipsoid, which approximates the volume of the particle.

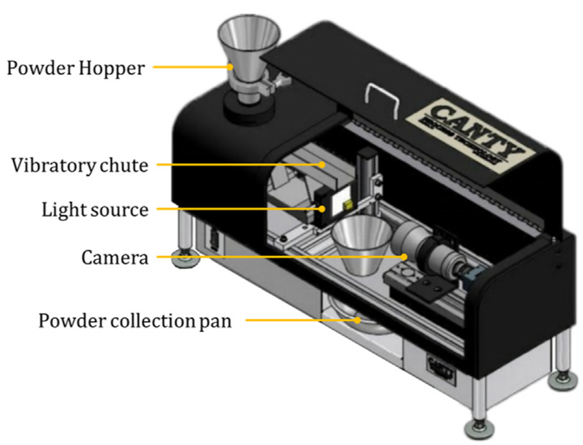

Figure 1. Labelled parts of the SolidSizer.

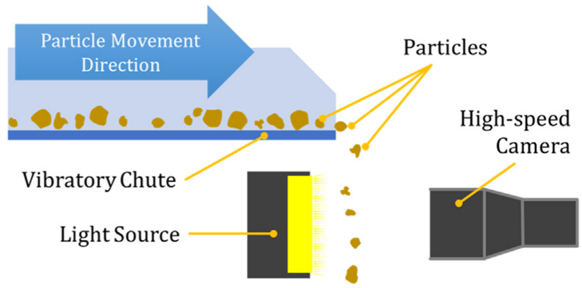

Figure 2. Schematic of SolidSizer measurement.

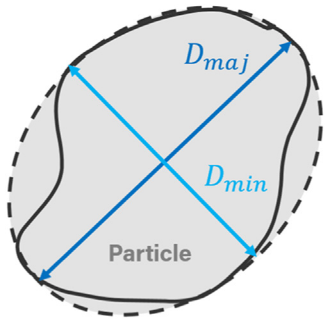

Figure 3. Fitting an ellipse around a detected particle.

The shape distribution is based on more than one quantity since there multiple ways to characterize shape. As listed in Table 3, shape-related variables like the aspect ratio, eccentricity, form factor, and elliptical form factor can all be deduced from the primary features measured by the SolidSizer.

The aspect ratio and the eccentricity both reflect changes in the elongation of the particles. The form factor may also be considered as circularity, and is the square of the ratio of the area equivalent spherical perimeter and the actual perimeter. It incorporates information pertaining to the nature of the surface of the particle such as its roundness, roughness, and interlocking tendency. Similarly, the elliptical form factor also considers the perimeter of the particle image but compares it with an area equivalent elliptical perimeter instead of circular. This gives the ability to properly distinguish a smooth elongated shape from an irregular but equiaxed shape, which is prone to interlocking and would have poor flowability.[32]

Size and Shape Distribution Characterization

In order to have a consistent number of variables in the regressor dataset without having to make assumptions about the nature of the distribution, the percentiles were computed from the size and shape cumulative distribution of each granule sample. At least 20 variables are included in the regressor dataset by including every 5th percentile from 5 to 100.

Furthermore, the span of each size and shape distribution is calculated by the following formula:

Table 3
Shape-related variables considered for shape distributions.

Shape-related variable	Formula	
Aspect ratio	D_{min}/D_{maj}	**Equation 2**
Eccentricity	$\sqrt{1-\left(\frac{D_{min}}{D_{maj}}\right)^2}$	**Equation 3**
Form factor	$4\pi Area/Perimeter^2$	**Equation 4**
Elliptical form factor	$\beta\pi Area/Perimeter^2,\ \beta=\left(\frac{1.5(Aspect\ \ Ratio+1)}{\sqrt{Aspect\ \ Ratio}}-1\right)^2$	**Equation 5**

$$Span = \frac{D_{90} - D_{10}}{D_{50}} \tag{6}$$

This allows the characterization of the breadth of each distribution, which is an important feature that could affect flowability.[33]

Projection to Latent Structures and the NIPALS Algorithm

Projection to latent structures or partial least squares (PLS) is the regression method of choice because of the high collinearity that are expected among the variables in the X and Y datasets (i.e., datasets containing the regressor and response variables respectively). PLS works best with such a highly-correlated dataset as it can project them into lower dimensional space with much fewer dimensions, and in a manner that maximizes the correlation between the X and the Y datasets. Furthermore, PLS modeling can reveal insights on the relationships among the X space, among the Y space, and between X and Y; this is done via plots such as the loadings plot and variable importance to the projection plot.[34−37]

Principal components analysis (PCA) and partial least squares modeling, including visualization of results, were performed using the PyPhi python module developed by Dr. Salvador Garcia. This module is open source and can be accessed via Github at https://github.com/salvadorgarciamunoz/pyphi.

Results and Discussion

The flowability measurements from the Flodex, GranuDrum, and GranuPack have a total of 26 variables, which can be combined into a single Y dataset for PLS regression. Merging the Y datasets is warranted because they are all related to granule flowability, which means that the variables in the datasets are expected to be highly collinear (correlated). This is confirmed by a principal components analysis (PCA) on the Y dataset (see Fig. 4), which shows that only five principal components can already account for 90% of the total variation in the entire dataset.

This dimensionality reduction is even more significant for the X dataset, which contains a total of 84 variables that includes the percentiles and the spans of the distributions of size and shape—i.e., equivalent volume diameter, eccentricity, form factor, and elliptical form factor. Principal components analysis on the X dataset as shown in Fig. 5 suggests high collinearity among the variables as more than

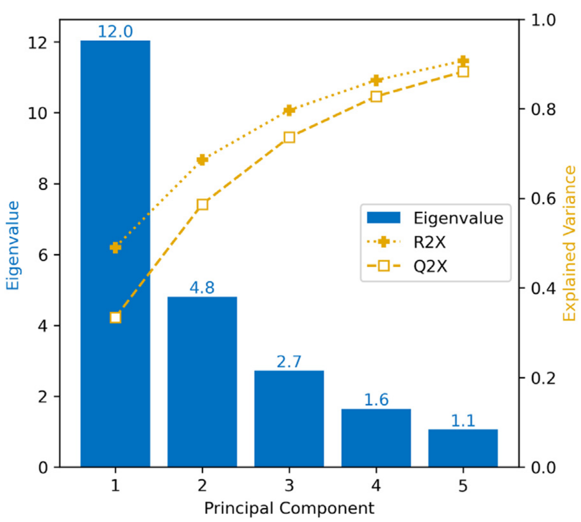

Figure 4. Principal components analysis on the combined Y dataset.

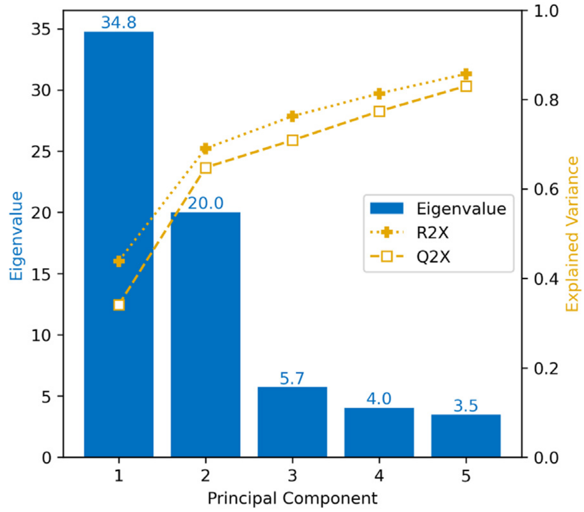

Figure 5. Principal components analysis on the X dataset.

80% of the total variation in the dataset can be explained by five principal components.

Although the number of principal components is limited to five in both Fig. 4 and Fig. 5, it can definitely be increased to improve the explained variance. However, as can be seen from the figures, this improvement tapers off. At some point, any additional component, which increases the complexity of the model, can no longer be justified, and could even lead to data overfitting. Conversely, less principal components could be employed to utilize a simpler model, but this comes at a cost to model accuracy. This relationship between the number of principal components and the explained variance is applicable not only to PCA, but to PLS as well, as can be seen from Fig. 6.

Predictability of Flowability Measurements

The demonstrated collinearity among variables is not only limited to within the datasets, but also across each other. This is clear when training a PLS model that maximizes the correlation between the X and the Y datasets. The resulting model, which is shown in Fig. 6, reveals that the 100 combined variables of the X and Y datasets can be reduced to 7 principal components. Furthermore, the results of

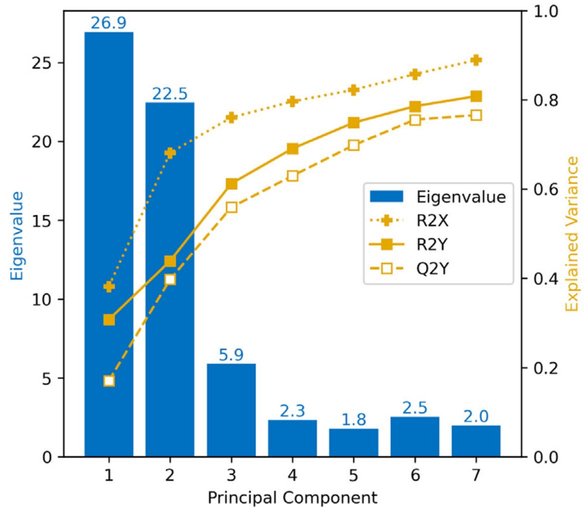

Figure 6. Effect of principal components on the PLS model to predict flowability measurements from size and shape distribution.

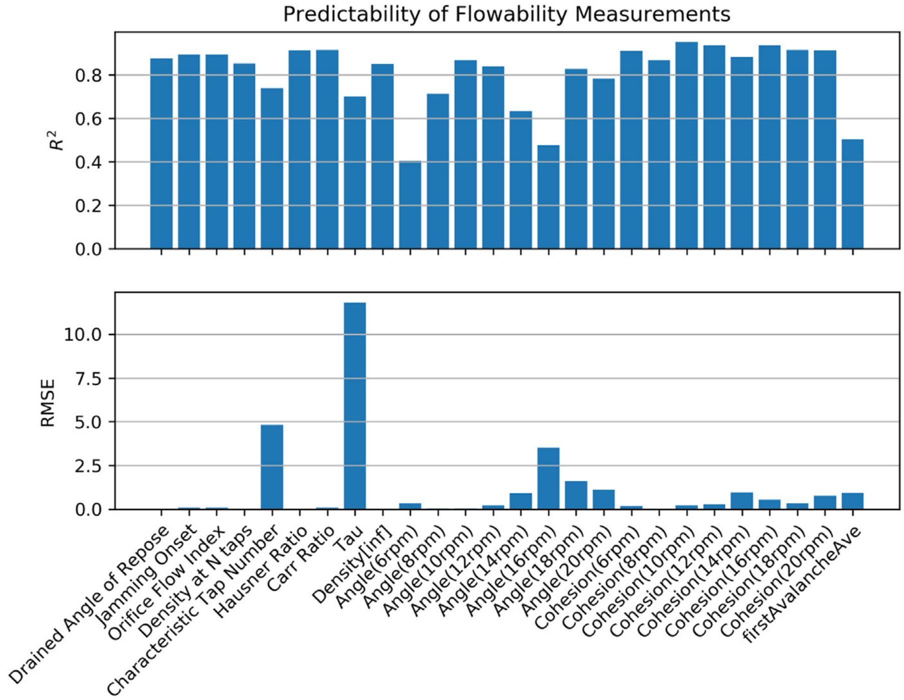

Figure 7. Predictability of flow measurement variables from the PLS model.

modeling with 7 principal components show that 89% of the variations in X is responsible for 81% of the variations in Y. As a check against overfitting, cross-validation was implemented, where the model still accounted for 77% of the variations in the Y space of the cross-validation datasets.

These percentages of explained variance reflects the good predictive performance of the PLS model on the combined flowability measurement variables, but it does not show the individual predictabilities of each variable, which are not the same. As shown in Fig. 7, majority of the flowability variables have decent predictability with coefficient of determination (R^2) values higher than 0.80. However, there are variables with poor predictability—i.e., they have low coefficient of determination and high values of root mean square error (RMSE). These are the tapped density analysis dynamics-related variables (i.e., Tau and Characteristic Tap Number) from the GranuPack flowability tests, and most of the Angle (of Repose) values from the GranuDrum flowability tests. For the GranuDrum tests, it seems that only cohesion values and dynamic angle of repose measured at select RPM settings can be included in the flowability sensor model, albeit there is currently no explanation for these noteworthy observations.

Collinearity in the Datasets

The successful reduction in dimensionality achieved by the PLS model is proof of high collinearity among variables[34–36], which can be corroborated by investigating the loadings in the X and Y datasets, starting with Fig. 8 for the first principal component. Within X, the size percentiles (e.g., Dv5, Dv10, etc.) are positively correlated with the percentiles of the form factor (e.g., FFDV5, FFDV10, etc.) and the elliptical form factor (e.g., EFFDV5, EFFDV10, etc.), while being negatively correlated with the percentiles of the eccentricity as well as the span of the form factor and elliptical form factor. This supports evidence from previous studies [38,39] that shows powder particles becoming smoother and more spherical at larger sizes, while becoming more elongated, rougher, and more prone to interlocking at smaller sizes.

Analogous observations can be made for the Y space, as shown in Fig. 9, where the drained angle of repose, orifice flow index, and cohesion can be seen to be inversely correlated with the Hausner ratio, Carr ratio, and the dynamic angle of repose. Finally, the correlations are also apparent among variables in X and Y. By simultaneously observing Fig. 8 and Fig. 9, the size and form factor can be seen to have a direct relationship with the Hausner ratio and the Carr ratio while having an indirect relationship with the drained angle of repose, cohesion, and orifice flow index.

Attributing Predictive Performance Improvement

The predictive performance of the PLS model has an overall coefficient of determination (R^2 value) of 0.81, outperforming the most recent attempt [20] to link measurements of size and shape with flowability. This improvement can be attributed to three major reasons: using a PLS model to account for collinearity, using a combination of size and shape to predict flowability, and using all measures of shape that are relevant to flowability instead of just using one.

Limitations of Pre-Selecting Percentiles Amid Collinearity

The ability of a PLS model to account for collinearity among variables means that there is no need to pre-select which variables should be used as input variables into the model. Often, when distributions are involved, certain percentiles (e.g., 10^{th}, 50^{th}, and 90^{th}) are used to represent the distribution, and these are the ones considered for analysis or for model building.[20,40] This is especially true for the 50^{th} percentile, which is commonly referred to as "D50". Unfortunately, a lot of information is lost when doing this.

The predictor variables may be ranked based on their influence to the projection, which is calculated from the weighted sum of squares of the PLS-weights with the weights calculated from the explained variance per variable in the Y-space for each PLS component.[34] This results in a plot called the VIP plot, which stands "Variable Influence to the Projection" plot. The VIP plot in Fig. 10 shows that the D50 does not necessarily rank the highest. Instead, the most important variables include a combination of size and shape percentiles.

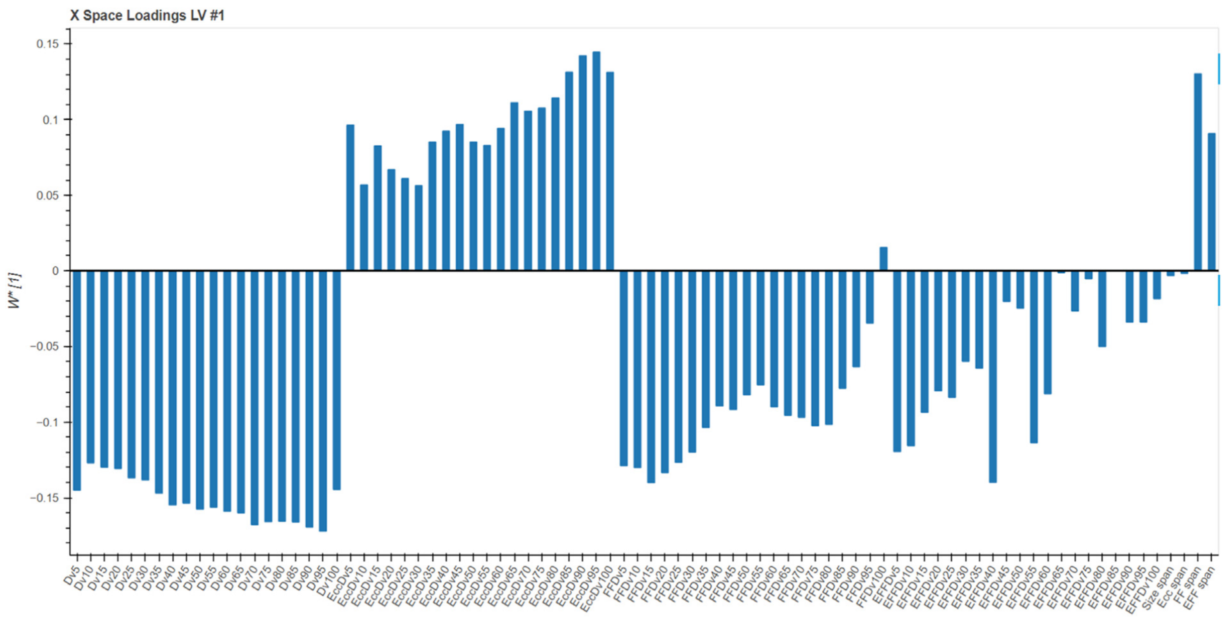

Figure 8. Loadings in the X-space for the first principal component of the PLS model.

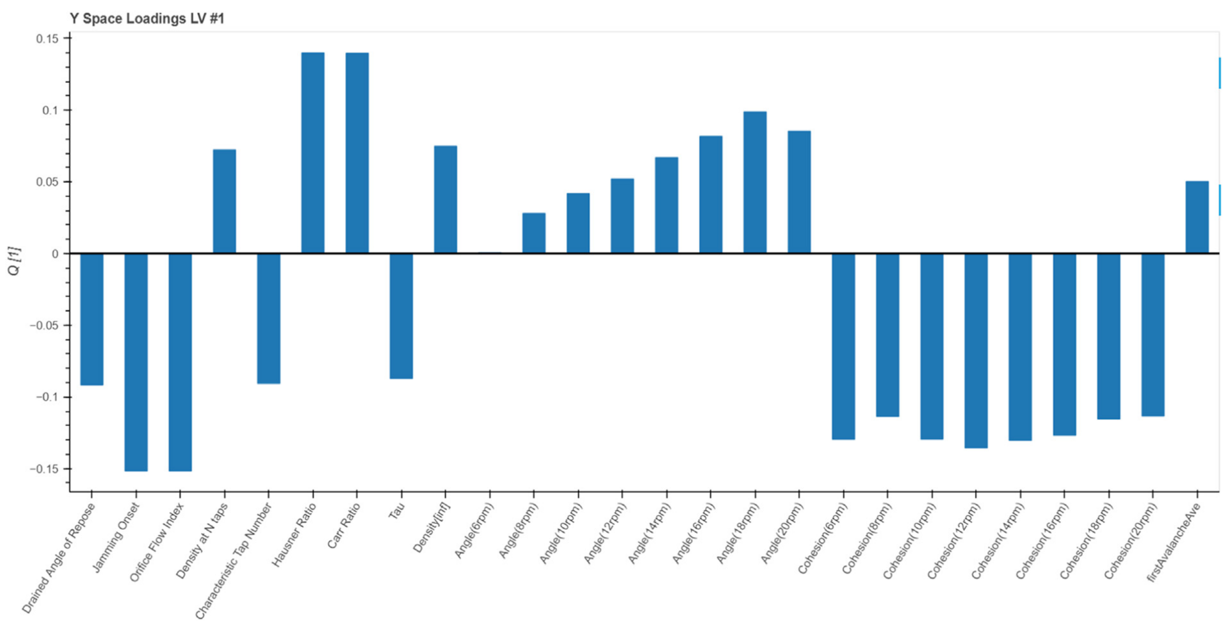

Figure 9. Loadings in the Y-space for the first principal component of the PLS model.

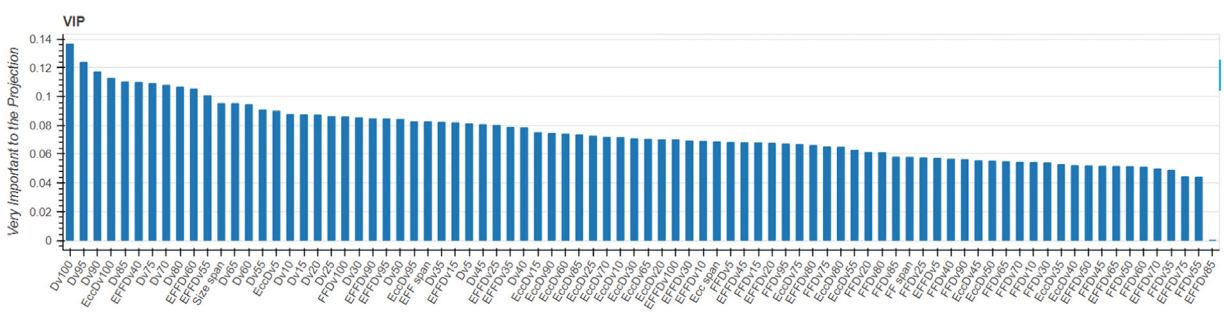

Figure 10. Ranking of X variables in order of influence to the prediction of Y variables.

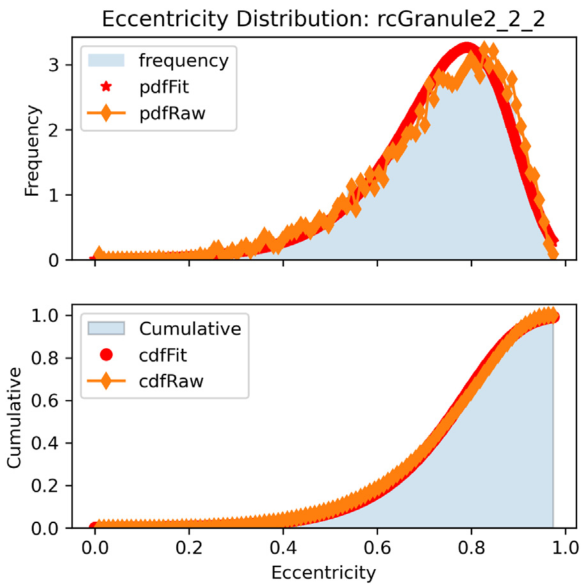

Figure 11. Eccentricity data fitted to a Weibull distribution.

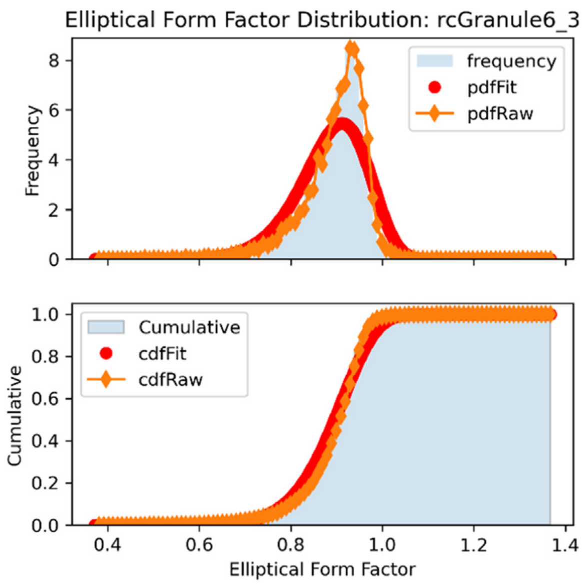

Figure 12. Elliptical form factor data fitted to a Weibull distribution.

Moreover, the relative magnitude of their importance suggests that most of these variables need to be taken altogether in order to achieve the current predictive capability. Hence, pre-selecting percentiles should be avoided since it prevents the optimum use of information from a dataset; instead, methods like PCA or PLS should be considered to transform the dataset into one with a much lower number of dimensions while retaining maximum information.

An alternative to reducing the dimensionality of a distribution data is to fit it to a continuous distribution such as a Gaussian curve, essentially reducing the number of variables to just the mean and the variance. However, real distributions rarely follow a bell curve and fitting one would lead to significant loss of information. Although other distributions are available such as the Weibull distribution that can better fit real distributions of particle size and shape for powders,[41] they often do not fit perfectly, leading to substantial loss of information. Figures Fig. 11 and Fig. 12 respectively show both a good and a bad fit for a Weibull distribution fitted to one of the granule shape distributions. Ultimately, poor fits can lead to poor performance, which is why the use of the highly collinear percentiles with PLS proved to be a robust alternative, since the latter can reduce the dimensionality while retaining most information.

Importance of Shape in Predicting Flowability

The VIP plot in Fig. 10 shows the relative importance of shape measurements in predicting flowability. Shape factors like eccentricity and elliptical form factor are among the most important variables alongside certain percentiles of size. This ranking is based on a PLS model that is trained using data from granules produced using two different screen sizes for the mill.

Interestingly, the importance of shape becomes even more pronounced if the PLS model is trained with datasets from granules produced from only a single screen size for the mill. Training a model this way is sensible if the intended purpose of the flowability sensor is for process control, since the mill screen size is not usually changed in the middle of an operation. As shown in Fig. 13, even more percentiles of the elliptical form factor rank the highest of importance, albeit they still need to be taken altogether with the other predictor variables in order to achieve the current predictive performance.

Limiting the datasets to granules produced at a constant screen size (i.e., granules 1, 2, and 3 in Table 1) means the variations in the granules can only be attributed to changes in the roll pressure during production. The increased emphasis on the shape for this limited dataset implies that changes in the roll pressure strongly affects the shape of the granules, particularly the elliptical form factor. In contrast, varying the screen size tends to affect the size more, albeit the shape is also affected. These observations about the effect of these two process parameters of the RC on the resulting granule size and shape are previously unreported in literature.

Limiting Datasets to Improve Predictive Performance

Another interesting observation when limiting the training datasets to granules produced at a constant screen size is the improved predictive performance of the PLS model. As shown in Fig. 14, such a model with 7 principal components can explain at least 97% of the variance in X and Y, which is a significant improvement over the

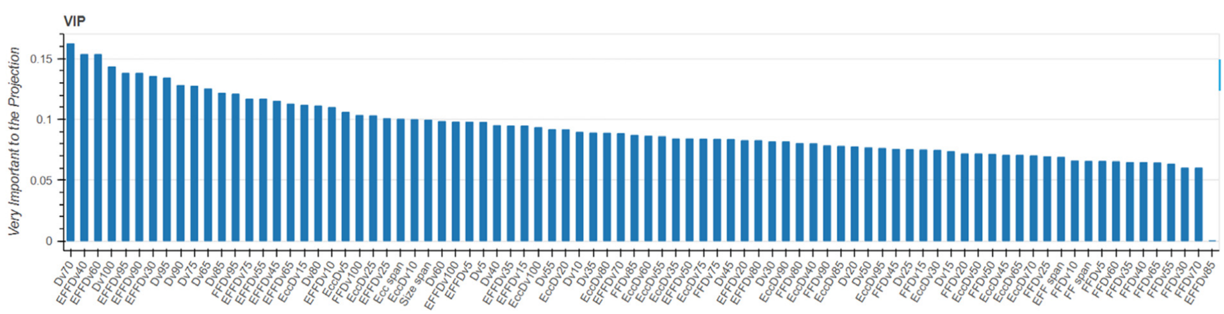

Figure 13. VIP plot of PLS model trained with limited dataset (from Granules made with 1.0mm mill screen size).

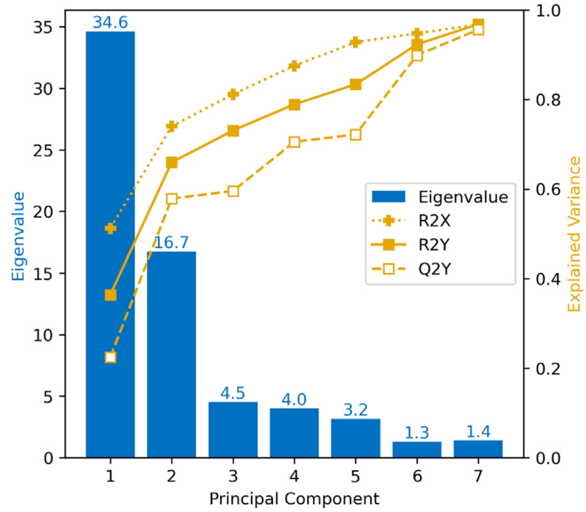

Figure 14. PLS model results of a limited dataset (granule batches 1-3).

model developed from a dataset that involves multiple screen sizes (Fig. 6). This means that varying the roll pressure during granule production induces changes in granule flowability that can be mostly explained by the currently considered measures of size/shape—i.e., diameter, eccentricity, form factor, and elliptical form factor. On the other hand, changing the screen size might have induced changes in the granules that are not totally reflected by these size and shape quantities. This warrants an investigation on additional shape (or size) quantities that can complete information on the physical characteristics of the granules. Introducing these quantities, into the predictor dataset might further improve the performance of a model trained with a more diverse dataset.

A potential implication of the aforementioned observations when developing a flowability sensor is to implement a hierarchical strategy, where several PLS models could be developed for specific screen sizes that might be used for the RC. During operation, the appropriate PLS model would then be selected according to the screen size that would be used. This does not mean, however, that one should completely abandon the notion of training a PLS model with datasets including multiple screen sizes. When the intended purpose is fault detection, having a PLS model that spans various screen sizes would be more suitable for detecting maintenance-related issues. An example would be the detection of the reinstallation of the incorrect screen during maintenance, or perhaps the detection of the gradual blockage of the mill screen with granules as the RC mill is continuously operated.

Relationships Among Variables

The relationship between the flow characteristics of a powder and the size and shape of its constituent particles, is already well-known. However, the loading coefficients of a resulting PLS can reveal additional insights regarding these relationships, which is interesting because multiple shape measurements have not been previously considered altogether as predictors for flow properties. As already mentioned, the datasets from granules produced at varying roll pressures (constant screen size) resulted in a PLS model that utilized shape-related variables with greater importance. The loading coefficients of the first principal component, as shown in Fig. 15, indicates that a decrease in particle size strongly corresponds with an increase in eccentricity an increase in the form factor and elliptical form factor. This suggests that variations in the roll pressure that lead to lower particle sizes tend to make the granules more elongated (higher eccentricity values) and would make the shape rougher and/or more irregular (lower form factors). As expected,[42,43] these changes in the predictor variables corresponded to poor flowability trends in the Y space loadings (Fig. 16), which showed higher Hausner and Carr ratios and a higher drained angle of repose.

Remarkably, however, loadings in that same Y space also show trends that indicate better flowability; cohesion values went down and both the jamming onset and orifice flow index decreased. This contradicting information proves that although the different flowability measurements are highly correlated and overlapping, each one provides unique information that is necessary to fully characterize flowability of granules produced from roll compaction.

Similar observations can be made for the loading coefficients of the second principal component shown in Figs. 17 and 18. The size distribution became broader as the higher size percentiles increase while the lower percentiles decrease. This can be attributed to the

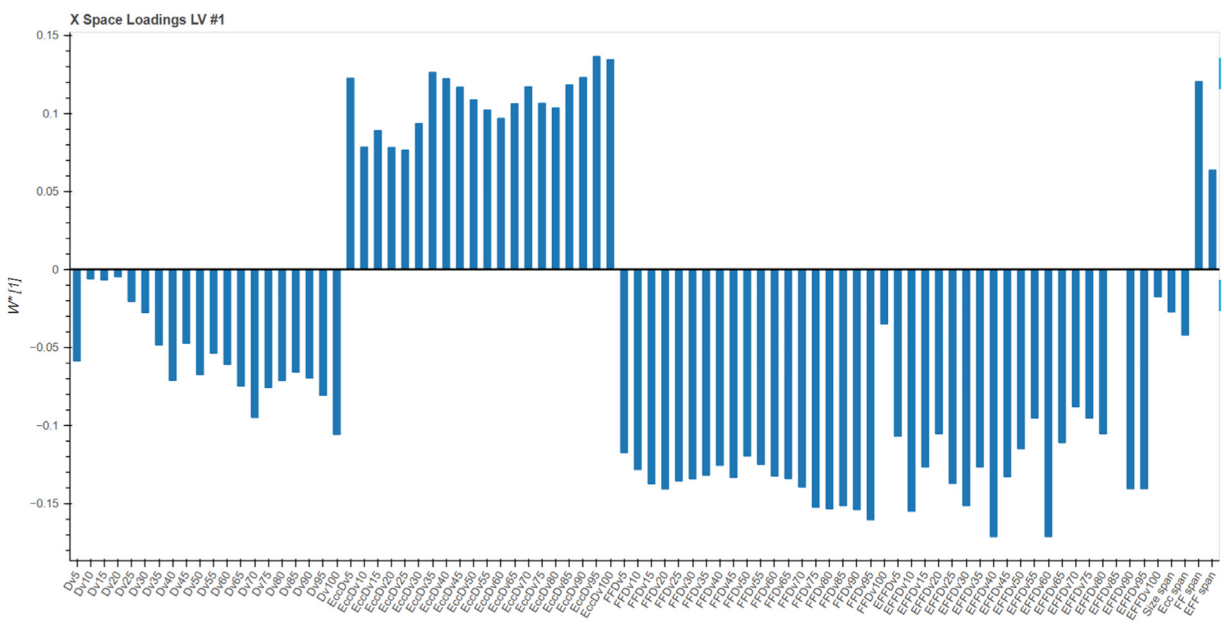

Figure 15. X-space loadings of first principal component for PLS model trained with limited dataset (granules batches 1-3).

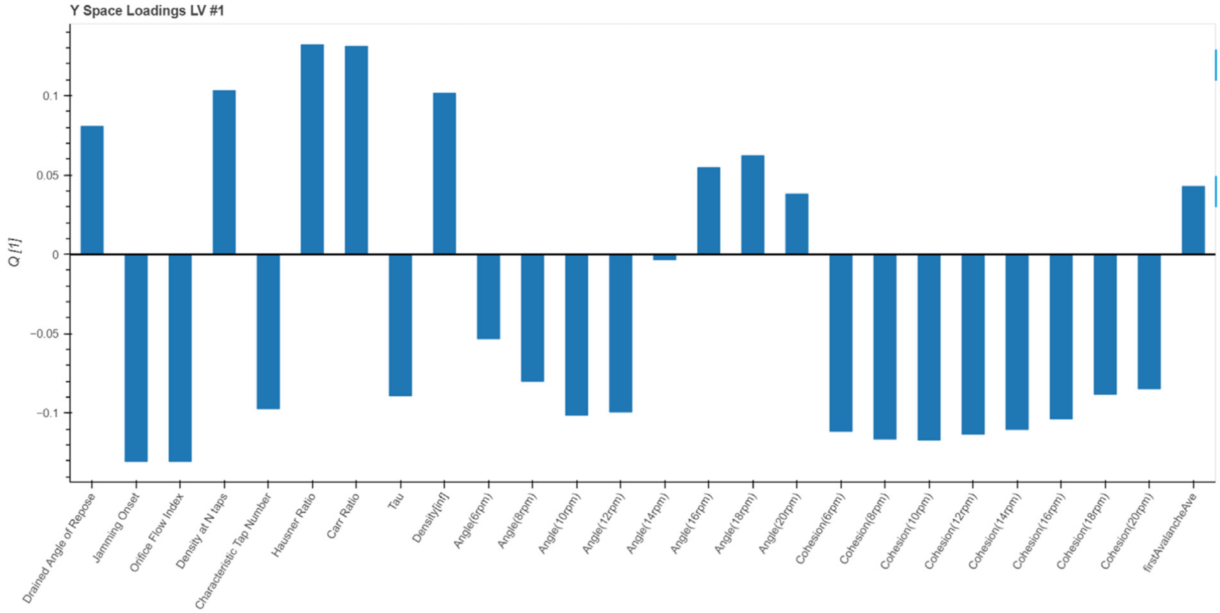

Figure 16. Y-space loadings of first principal component for PLS model trained with limited dataset (granule batches 1-3).

process in the roller compactor that leads to generation of finer particles, which occurs during ribbon formation and milling, as well as agglomeration of particles into larger granules, which occur during ribbon formation. Changes in the shape are less pronounced for the second principal component, although larger percentiles of the elliptical form factor have corresponding significant increases.

On the surface, a broader size distribution should lead to poor flowability [33], which can be seen from the higher cohesion values. However, variables in the Y dataset also showed trends indicating improved flowability, like a significantly lower drained angle of repose, lower Hausner and Carr ratios, and lower dynamic angle of repose; most likely, the improved flowability trends can be attributed to increased values of the elliptical form factor. Similar to the first principal component, these contradictions support the notion that the flowability of granules produced by a roller compactor require multiple flowability measurements, even though they are highly correlated.

The uniqueness of information provided from each flowability measurement is also apparent when comparing the performance of PLS models when all the flowability measurements are combined into a single Y dataset, versus when different PLS models are developed for each flowability measurement. When the columns of the Y dataset—i.e., the individual flowability measurements—are highly correlated, they tend to have an overlapped latent space, leading to improved prediction performance when taken altogether as a single Y dataset during PLS modeling.[36]

However, the results in Fig. 19 show slightly worse results for a singular PLS model versus separate PLS models for each flowability

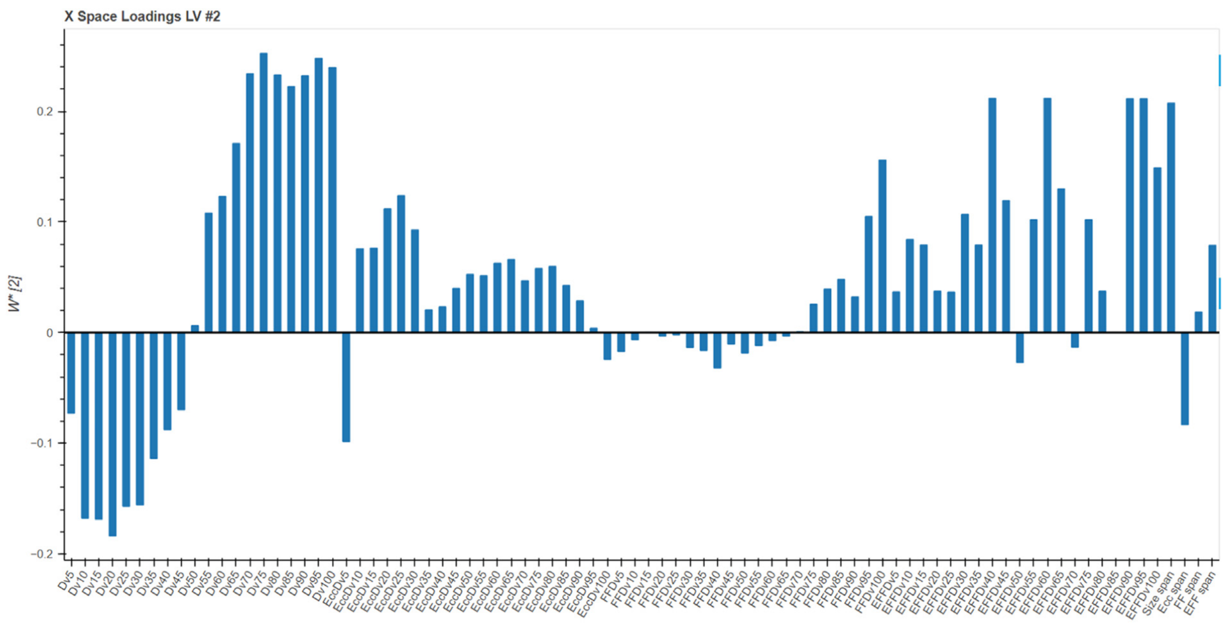

Figure 17. X-Space loadings of second principal component for PLS model trained with limited dataset (granules batches 1-3).

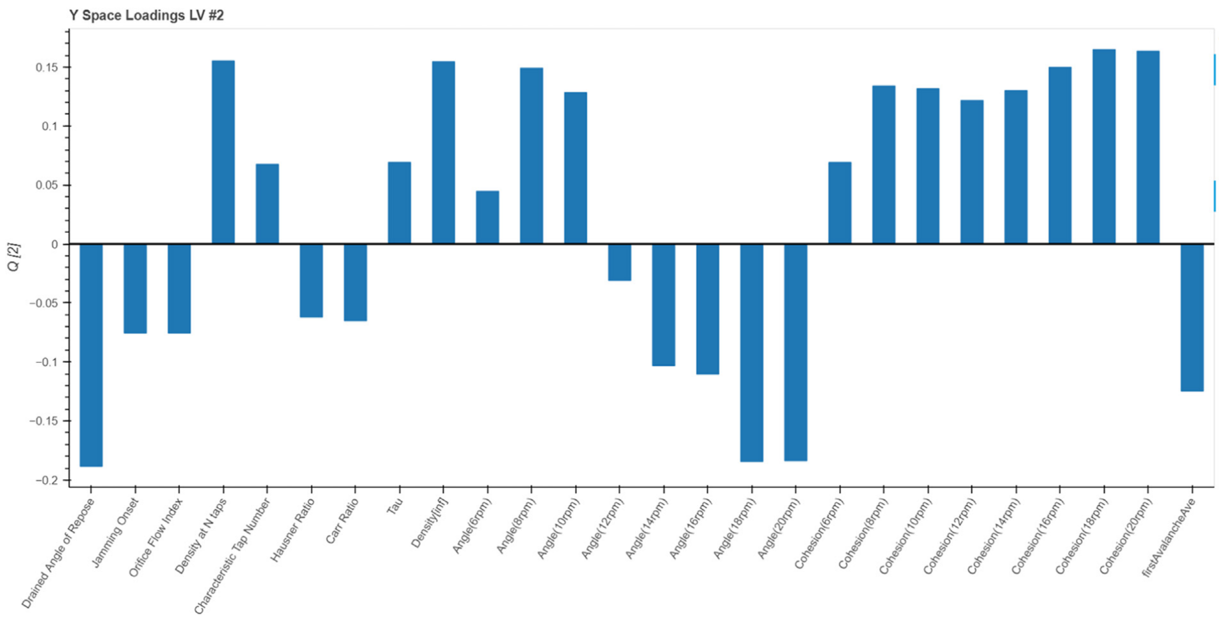

Figure 18. Y-space loadings of second principal component for PLS model trained with limited dataset (granules batches 1-3).

measurement, suggesting that despite these overlaps in the latent spaces, each flowability measurement offers unique information relating to granule flowability. Hence, they should be included in the model.

Conclusions

Particle analyzers that are deployed in-line could fill the technological gap for effectively monitoring a continuous pharmaceutical dry granulation line. By extracting the full distributions of relevant size and shape parameters from the particle measurements, real-time predictions of granule flowability could be achieved.

The relevant size and shape parameters might differ depending on the application, which would affect which flowability measurements would be relevant for monitoring. For monitoring a roller compactor in a dry granulation line, flowability was characterized according to the flow conditions that the product granules may experience in the tablet press, which were via three different methods: flow in a rotary drum, flow through an orifice, and tapped density. PLS analysis revealed that although the measurement variables from these methods are highly-correlated with and among each other, all of them are necessary for a complete characterization of the flow behavior of a powder.

PLS proved to be an effective regression method for predicting granule flowability based on the size and shape distributions of granules. Its

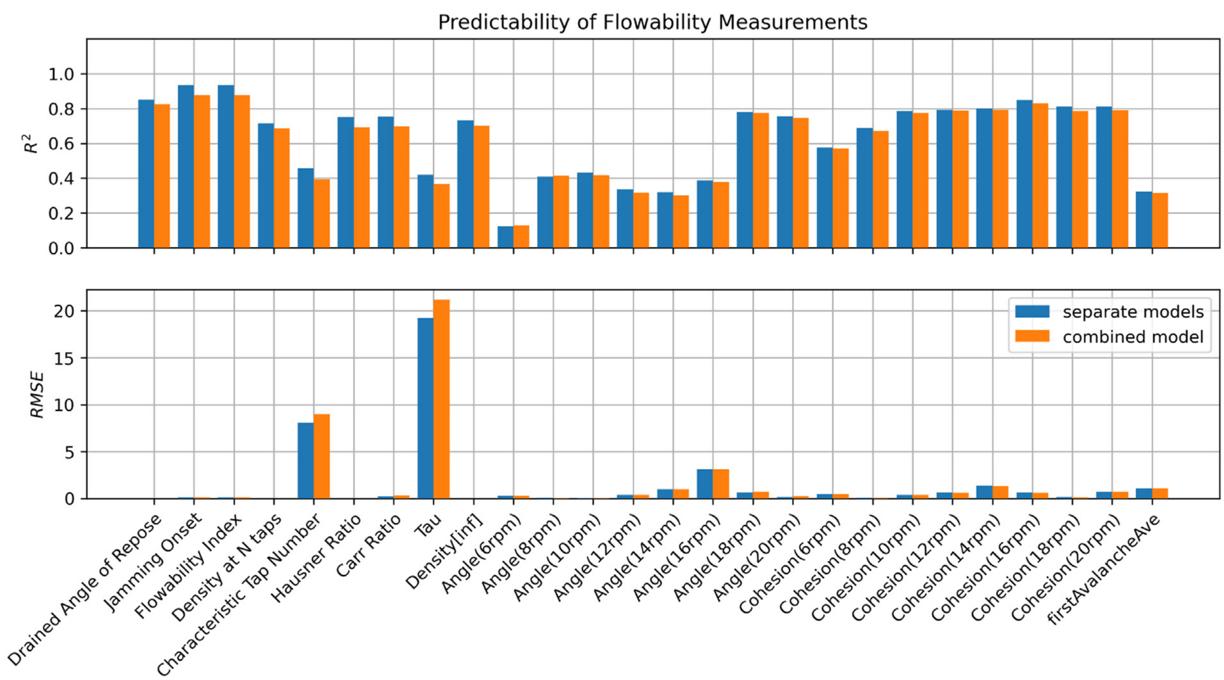

Figure 19. Comparison of predictive performance between a singular PLS model (Trained with a combined Y dataset) and separate PLS models (one model for each flowability test).

ability to greatly reduce the dimensionality of a highly-correlated dataset, while maintaining maximal information, makes it possible to input entire distribution percentiles into a regression model without resorting to pre-selecting percentile variables or assuming a distribution form that might be incorrect. PLS excelled because of its ability to handle high collinearity, which was present in both the predictor and the response variable datasets, ultimately resulting in coefficient of determination values ranging from 0.80 to 0.97, depending on the dataset used to train the PLS model. Another important feature that had a major contribution to this modeling performance is the inclusion of entire distributions of multiple relevant shape descriptors in the predictor variable dataset. The addition of the form factor and the elliptical form factor to supplement the information provided by eccentricity proved to have a significant positive impact to the performance as evidenced by the loadings plot and the VIP plot.

Often, particle analyzers focus on particle size and treat particle shape to be of diminished importance. For instance, the Eyecon_2[31] is a promising instrument for inline measurements of particle size and shape. While its data output includes the full distribution of size, it only gives the mean and standard deviation of the eccentricity. Since shape distributions are not necessarily Gaussian, this leads to loss of information that ultimately results in poor predictive performance of the flowability. This dismissal of particle shape measurements is mostly driven by the lack of appreciation of shape by industry. Monitoring particle size and shape by itself is hardly useful unless it is used to predict more process-relevant quantities. Hopefully, the findings reported in this study about the importance of shape would improve the appreciation and focus of the powder processing industry towards shape measurements.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We acknowledge Mariana Araujo da Conceicao, David Sixon, Lily Waterman, and Dr. Clairmont Clementson for their valuable contributions in granule characterization.

We also acknowledge Dr. Salvador Garcia for his guidance and advice on Partial Least Squares and the use of the PyPhi Python Package.

This work was supported by the United States Food and Drug Administration under grant U01FD006487-01. The views expressed by the authors do not necessarily reflect official policies of the Department of Health and Human Services; nor does any mention of trade names, commercial practices, or organization imply endorsement by the United States Government.

References

1. Su Q, Moreno M, Ganesh S, Reklaitis G v, Nagy ZK. Resilience and risk analysis of fault-tolerant process control design in continuous pharmaceutical manufacturing. *J Loss Prev Process Ind*. 2018;55:411–422. https://doi.org/10.1016/j.jlp.2018.07.015.
2. Su Q, Bommireddy Y, Gonzalez M, Reklaitis G v, Nagy ZK. Variation and risk analysis in tablet press control for continuous manufacturing of solid dosage via direct compaction. *Computer Aided Chemical Engineering*. 44. Elsevier B.V.; 2018:679–684. https://doi.org/10.1016/B978-0-444-64241-7.50108-7.
3. Su Q, Ganesh S, Moreno M, et al. A perspective on Quality-by-Control (QbC) in pharmaceutical continuous manufacturing. *Comput Chem Eng*. 2019;125:216–231. https://doi.org/10.1016/j.compchemeng.2019.03.001.
4. Hsu SH, Reklaitis G v, Venkatasubramanian V. Modeling and control of roller compaction for pharmaceutical manufacturing. Part I: process dynamics and control framework. *J Pharm Innov*. 2010;5(1-2):14–23. https://doi.org/10.1007/s12247-010-9076-0.
5. Becker-Hardt M. *The compaction people continuous dry granulation by Roller compaction an introduction to the Alexanderwerk Roller compaction process*.; 2018.
6. Prescott JK, Barnum RA. On powder flowability. *Pharm Technol*. 2000:60–84. Published online.
7. Freeman T, vom Bey H, Hanish M, Brockbank K, Armstrong B. The influence of roller compaction processing variables on the rheological properties of granules. *Asian J Pharm Sci*. 2016;11(4):516–527. https://doi.org/10.1016/j.ajps.2016.03.002.
8. Venkatasubramanian V, Rengaswamy R, Yin K, Kavuri SN. A review of process fault detection and diagnosis: part I: quantitative model-based methods. *Comput Chem Eng*. 2003;27(3):293–311.
9. Wu H, Yu Z, Wang Y. Real-time FDM machine condition monitoring and diagnosis based on acoustic emission and hidden semi-Markov model. *Int J Adv Manuf Technol*. 2017;90(5-8):2027–2036. https://doi.org/10.1007/s00170-016-9548-6.
10. de Beer T, Burggraeve A, Fonteyne M, Saerens L, Remon JP, Vervaet C. Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes. *Int J Pharm*. 2011;417(1-2):32–47. https://doi.org/10.1016/j.ijpharm.2010.12.012.
11. Lee SL, O'Connor TF, Yang X, et al. Modernizing pharmaceutical manufacturing: from batch to continuous production. *J Pharm Innov*. 2015;10(3):191–199. https://doi.org/10.1007/s12247-015-9215-8.
12. Hogan E. *Particle size analysis of 2-5 mm particles of varying colour from coating and granulation processes using process analytical technology*. www.innopharmalabs.com/technology.
13. Patel P, Godek E, O'callaghan C, Jones I. *Real-time prediction of polymer-coated multiparticulate dissolution using process analytical technology*. www.innopharmalabs.com/technology.
14. Kiernan L, Hogan E, Todaro V, Healy AM, Greene A, Jones I. *Real-time in-line monitoring of the impact of process parameter changes on critical quality attributes in fluid bed granulation using process analytical technology*. Accessed September 12, 2022. https://www.innopharmatechnology.com/.
15. McAuliffe MAP, Omahony GE, Blackshields CA, et al. The use of PAT and off-line methods for monitoring of roller compacted ribbon and granule properties with a view to continuous processing. *Org Process Res Dev*. 2015;19(1):158–166. https://doi.org/10.1021/op5000013.
16. Richter DrIM, O'Callaghan C. *Unique advantages of process analytical technology in twin-screw granulation*. www.innopharmalabs.com/technology.
17. Santomaso A, Lazzaro P, Canu P. Powder flowability and density ratios: the impact of granules packing. *Chem Eng Sci*. 2003;58(13):2857–2874. https://doi.org/10.1016/S0009-2509(03)00137-4.
18. Bodhmage A. *Correlation between Physical Properties and Flowability Indicators for Fine Powders*. University of Saskatchewan; 2006.
19. Schulze D. *Understanding and measuring powder flowability*.; 2002.
20. Kudo Y, Yasuda M, Matsusaka S. Effect of particle size distribution on flowability of granulated lactose. *Adv Powder Technol*. 2020;31(1):121–127. https://doi.org/10.1016/j.apt.2019.10.004.
21. Yu W, Muteki K, Zhang L, Kim G. Prediction of bulk powder flow performance using comprehensive particle size and particle shape distributions. *J Pharm Sci*. 2011;100(1):284–293. https://doi.org/10.1002/jps.22254.
22. Mort P. Characterizing flowability of granular materials by onset of jamming in orifice flows. *Pap Phys*. 2015;7. https://doi.org/10.4279/PIP.070004.
23. Gioia A. Intrinsic flowability: a new technology for powder-flowability classification. *Pharm Technol*. 1980. Published online February.
24. Traina K, Cloots R, Bontempi S, Lumay G, Vandewalle N, Boschini F. Flow abilities of powders and granular materials evidenced from dynamical tap density measurement. *Powder Technol*. 2013;235:842–852. https://doi.org/10.1016/j.powtec.2012.11.039.
25. Lagare RB, da Conceicao MA, Rosario ACA, et al. Development of a virtual sensor for real-time prediction of granule flow properties. *Computer Aided Chemical Engineering*. 51. Elsevier; 2022:1081–1086.
26. Retsch. Sample divider PT 100. Accessed September 12, 2022. https://www.retsch.com/products/assisting/sample-divider/pt-100/function-features/?gclid=EAIaIQobChMIkveRrdOS-gIV7BXUAR19bwtzEAAYASAAEgLjB_D_BwE.
27. GRANUTOOLS. GranuDrum rotating drum: our powder tester for cohesion. Accessed September 12, 2022. https://www.granutools.com/en/granudrum.
28. GRANUTOOLS. GranuPack: high resolution tapped density analyzer. Accessed September 12, 2022. https://www.granutools.com/en/granupack.
29. TELEDYNE HANSON. Flodex. Accessed September 12, 2022. https://www.teledynehanson.com/physical-testing/flodex.
30. Canty Process Technology. Lab Solidsizer. Accessed September 12, 2022. https://www.jmcanty.com/product/lab-solidsizer/.
31. Innopharma Technology. Eyecon2. Accessed December 21, 2022. https://www.innopharmatechnology.com/products/eyecon2tm.
32. Mort P, Holder J. Measurement of agglomerate shape and shape distribution using on-line image analysis. *AIChE Annual Meeting Particle Technology Forum*. 2002.
33. Kalman H. Bulk densities and flowability of mono-sized, binary mixtures and particle size distributions of glass spheres. *Powder Technol*. 2022;397. https://doi.org/10.1016/j.powtec.2021.117086.
34. Wold S, Sjostrom M, Eriksson L, Sweden S. PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab Syst*. 2001;58:109–130. www.elsevier.comrlocaterchemometrics.
35. Geladi P, Kowalski BR. *Partial Least-Squares Regression: a Tutorial*. 186. Elsevier Science Publishers B.V; 1986.
36. Burnham AJ, Macgregor JF, Vivero R, Viveros R. Latent variable multivariate regression modeling. *Chemom Intell Lab Syst*. 1999;48(2):167–180.
37. Ng KS. *A simple explanation of partial least squares*.; 2013. Accessed September 12, 2022. https://users.cecs.anu.edu.au/~kee/pls.pdf.

38. Buckland HM, Saxby J, Roche M, et al. Measuring the size of non-spherical particles and the implications for grain size analysis in volcanology. *J Volcanol Geotherm Res*. 2021:415. https://doi.org/10.1016/j.jvolgeores.2021.107257.
39. Califice A, Michel F, Dislaire G, Pirard E. Influence of particle shape on size distribution measurements by 3D and 2D image analyses and laser diffraction. *Powder Technol*. 2013;237:67–75. https://doi.org/10.1016/j.powtec.2013.01.003.
40. Park SY, Galbraith SC, Liu H, et al. Prediction of critical quality attributes and optimization of continuous dry granulation process via flowsheet modeling and experimental validation. *Powder Technol*. 2018;330:461–470. https://doi.org/10.1016/j.powtec.2018.02.042.
41. Bailey AG, Balachandran W, Williams TJ. The Rosin-Rammler size distribution for liquid droplet ensembles. *J Aerosol Sci*. 1983;14(1):39–46.
42. Boschini F, Delaval V, Traina K, Vandewalle N, Lumay G. Linking flowability and granulometry of lactose powders. *Int J Pharm*. 2015;494(1):312–320. https://doi.org/10.1016/j.ijpharm.2015.08.030.
43. Kaerger JS, Edge S, Price R. Influence of particle size and shape on flowability and compactibility of binary mixtures of paracetamol and microcrystalline cellulose. *Eur J Pharm Sci*. 2004;22(2-3):173–179. https://doi.org/10.1016/j.ejps.2004.03.005.