Chemical Engineering Science 317 (2025) 122042

Contents lists available at ScienceDirect

Chemical Engineering Science

journal homepage: www.elsevier.com/locate/ces

Mapping process-product relations by ensemble regression—A granulation case study

Line Koleilat [a], Langdon Feltner [b], Donovan Stumpf [b], Paul Mort [b,*]

[a] Purdue University, Department of Agricultural and Biological Engineering, West Lafayette, 47907, IN, USA
[b] Purdue University, School of Materials Engineering, West Lafayette, 47907, IN, USA

ARTICLE INFO

Keywords:
Multi-effect regression
Image analysis
Size and shape characterization
Fluid bed granulation
Product/process mapping
Model-based optimization

ABSTRACT

Measurement and control of product-process relations are important for particulate products and processes, especially with respect to distributed product specifications, for example size and shape distributions. This paper shares an ensemble regression technique for linking process parameters with distributed granule size and shape characteristics obtained by dynamic image analysis. A case study of fluidized bed granulation is presented, linking process effects (inlet air flow, air temperature, binder spray rate, and binder content) with product attributes. The process was bounded using stable operating criteria defined by balancing wet binder addition with drying enthalpy of the fluidization airflow. Detailed mapping of size and shape trends were obtained using the regression of fully distributed image analysis data obtained from only a few pilot-scale runs, the key enabler being the simultaneous use of distributed size and shape data with process parameters in an ensemble regression model. Results suggest opportunities for optimization linking process and product objectives. More broadly, the methodology suggests opportunities for improved product/process modeling with distributed size and shape data.

1. Introduction

The ability to monitor and control particulate processes has advanced with the development of measurement technologies able to capture details of size and shape features, for example, dynamic image analysis (DIA) methods able to characterize particles in a flowing stream. Yet the reporting of these rich datasets is often reduced to only a few descriptive statistics, for example a mean size or selected percentiles in a distribution, e.g., $D10$, $D50$, $D90$. This degree of data reduction limits the utility of the data for trend-wise analyses, for example mapping product-process relations. This paper uses a case study linking granulation process parameters with product characteristics within thermodynamic process constraints. Rather than predicting granule characteristics with a dynamic model, for example, according to population balances (Gantt and Gatzke, 2005; Braumann et al., 2007), the current approach uses an ensemble regression approach to correlate process parameters with experimental measurements of size and shape distributions.

Historically, response surface methods have been used to evaluate the effects of key process parameters, for example the effects of binder flow rate and inlet air temperature on granule and tablet properties in top-spray fluidized bed granulation (Lipps and Sakr, 1994). Findings highlight the utility of quadratic regression models in capturing complex interactions between process variables and critical quality attributes like granule size, density, and drug release, with implications for optimizing pharmaceutical granulation and tableting. While reasonably good predictions for first-order distribution metrics such as geometric mean granule size were reported, poorer fits were noted for second-order metrics such as geometric standard deviation. Specifically, regression models for distribution spread often failed to reach statistical significance or showed low explanatory power, highlighting a limitation of the quadratic response surface methodology in capturing particle size distributions. This suggests that while the modeling approach could describe average trends effectively, it was less capable of accounting for distributions, possibly due to the limited number of input variables or the inherent complexity of granule growth mechanisms.

Modern machine learning approaches are rife with paths to better utilize high-resolution, distribution-level data, and have been applied extensively to predicting manufacturing process outcomes. Schmitt et al. (2022) created an ensemble machine learning model to predict dried particle size as a function of operational parameters across pilot and production scale spray dryers, achieving prediction errors between −7.7 % and 18.6 %. Oishi et al. (2020) applied Elastic Net regression with interaction terms to a high-dimensional dataset of 44 tablet formulations spanning six granulation methods, using 12 material attributes and one process parameter per sample. Their approach, like ours, leveraged data-rich inputs to model product–process relationships, but they noted

* Corresponding author.
E-mail address: pmort@purdue.edu (P. Mort).

https://doi.org/10.1016/j.ces.2025.122042
Received 4 April 2025; Received in revised form 20 May 2025; Accepted 12 June 2025
Available online 19 June 2025

reduced interpretability with many interaction terms and limited generalizability due to a fixed formulation and single active pharmaceutical ingredient.

These limitations highlight the need for models that integrate detailed, high-resolution, product data with robust and interpretable statistics, i.e., the ability to make stable predictions under data-scarce conditions or modest extrapolation while maintaining explicit linkages between process parameters and product features. In this work, we define "resolution" as the use of scalar size and shape metrics assessed on the individual granule level, rather than as a spatial or closure-based modeling resolution. Building on these developments, our approach combines the distribution-level richness of modern particle characterization with ensemble regression techniques that maintain these desirable properties while accommodating limited experimental trials.

We present an example of ensemble linear regression analysis linking distributed granule size and shape data with process parameters in fluidized bed granulation (FBG). The case study has a limited number of pilot-scale process runs that roughly explored the process design space without a strict design-of-experiment methodology. Products from a subset of runs were selected based on process stability (i.e., well-controlled fluidization and process yield) and characterized using dynamic image analysis (DIA). Compound weighted regression was used to fit the data to lognormal models, and extended to an ensemble regression, combining the detailed size and shape data with their process parameters. This ensemble regression enabled statistical fitting of multiple process parameters and their interactions with only a limited number of process trials. The result is a detailed relational model that leverages the wealth and resolution of the granule size and shape data.

2. Granulation case study

Powder granulation is commonly done to achieve and maintain mixing uniformity, improve flowability, facilitate compaction, and manage derivative product performance for various applications, for example, dispersion, dissolution and controlled release (Ennis et al., 1991; Iveson et al., 2001; Litster and Ennis, 2004; Rowe and Roberts, 1995; Kristensen and Schaefer, 1987; Rough et al., 2005; Sivarao et al., 2014; Liu et al., 2021). Whether for pharmaceutical, consumer goods, agricultural bio-processing, ceramic, or other material applications, it is frequently desired to map the relationships between granule attributes and process parameters in context of stable operation, i.e., a process design space. Regime mapping has an analogous objective, describing product-process relations across a broader range of processes and formulations. Although somewhat qualitative, regime maps aim to provide useful guidance for granulation based on physical mechanisms and material properties (Hapgood et al., 2003; Iveson et al., 2001; Kumar et al., 2016; Pohl and Kleinebudde, 2020; Kastner et al., 2013). In this case study, we narrow the focus to a specific formulation/process space and consider the effective use of experimental data for quantitative mapping of product-process relations.

Empirical mapping of product characteristics to process parameters requires sufficient data for rigorous statistical analysis (Kayrak-Talay and Litster, 2011; Kayrak-Talay et al., 2013; Veliz Moraga et al., 2015; Dan et al., 2024). Executing granulation experiments and/or collecting representative samples from a running granulation process can be challenging. Sample analysis typically involves data reduction, converting distributed data to descriptors such as a geometric mean size (d_g) and standard deviation (σ_g), or more simply a median with quantiles, e.g., $D10$, $D50$, $D90$. Drawing conclusions from such reduced data requires larger numbers of samples and even replicate experiments.

This communication considers direct analysis of fully distributed data as an effective approach to mapping distributed characteristics with limited process data. Quantitative trends in size and shape are mapped in terms of lognormal distribution descriptors obtained from weighted multi-effect linear regression of distributed data. The resulting maps are quantitative and show trends that relate the parameters of the FBG process to the geometric mean and standard deviations of granule size and shape features. Direct analysis of the distributed data avoids the data-loss associated with intermediate data reduction, provides a more efficient use of that data, and albeit empirical, enables more insightful use of the results by way of a quantitative product-process model. While the current study relies on empirical mapping of process-product relations, a reasonable next step can extend the approach to a dynamic flowsheet or digital twin model, for example using population balance equations describing growth (Braumann et al., 2007; Dan et al., 2024), and in more advanced instances, morphology of granules.

3. Experimental approach

Granules were prepared using a proxy pharmaceutical formulation of lactose monohydrate (200 mesh, Foremost Farms NF grade, Baraboo, WI, USA), microcrystalline cellulose (MCC, Avicel PH101, FMC Corp., Philadelphia, PA, USA), and a 25 % aqueous binder solution prepared using polyvinylpyrrolidone (Plasdone K-30, Ashland Chemical, Wilmington, DE, USA) powder dissolved in distilled water. Granules were produced using a pilot-scale bottom spray fluidized bed granulation system, the SolidLab 2 (SL2) Fluidized Bed Granulation System (Hüttlin GmbH, Syntegon Company, Schopfheim, Germany).

Six runs were performed within a stable operating space balancing the binder spray input with drying enthalpy. All runs transferred a premix of 2.55 kg lactose monohydrate and 0.30 kg microcrystalline cellulose into the SL2 where it was fluidized and granulated using a binder spray. The binder was sprayed using air-atomized nozzles with standard air cap configuration (2.3 mm) and 0.8 bar atomizing air pressure. Filter bag blowback pressure was constant at 3.2–3.5 bar, with the exception of run D where it was marginally increased to compensate for the build-up of elutriated powder at higher airflow conditions. Process and binder formulation parameters are summarized in Table 1. Note the air temperature was adjusted (T_{slot} instead of T_{inlet}) to compensate for thermal losses in the system, occurring primarily between the inlet air control thermocouple and the slots in the distributor plate (Koleilat et al., 2024).

The control of the FBG process is summarized in Fig. 1, where moisture content in the granulation process (Fig. 1a) was predicted using a mass and energy balance (Koleilat et al., 2025) and compared to the peak values measured experimentally via mass balance only (Fig. 1b). The SOR indicates a stable operating range for granulation based on having good agreement between the predictive model and experimental data, i.e., between about 7.5 % and 11 % peak moisture. While spraying the binder, the rate of moisture addition exceeds the drying rate associated with the heated airflow; the resultant increase in product moisture is associated with granulation. The product moisture reaches a peak at the end of the spray segment and then declines during drying. The peak moisture is an indicator of granule growth and must be balanced to maintain process stability in relation to the material properties—if the peak is too low, insufficient granulation may result in a dusty product; if the peak is too high, fluidization may be compromised, destabilizing the

Table 1
Summary of formulations and spray-segment process parameters used in each run. All runs used a 25 % aqueous PVP binder solution. The same parameters were used in the drying segment in all cases except run F where airflow was reduced to 90 m^3/h during drying.

Run	Binder soln, kg	PVP% dry basis (PVP)	Spray rate g/min (SR)	Airflow m^3/h (AF)	Superficial velocity m/s	Inlet T, °C	Slot T, °C (T)
A	0.67	5.5 %	50	90	0.35	90	70.1
B	0.80	6.6 %	50	100	0.39	80	63.7
C	0.80	6.6 %	50	80	0.31	85	65.6
D	0.80	6.6 %	50	130	0.53	60	50.2
E	0.60	5.0 %	50	90	0.35	65	52.0
F	0.60	5.0 %	72	120	0.47	65	53.5

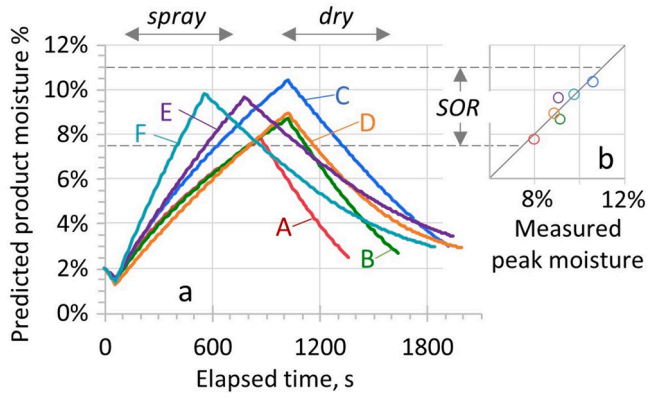

Fig. 1. Accumulated moisture in the fluid bed granulation process: a) predictive calculation using an enthalpy balance; b) parity plot with the measured peak moisture.

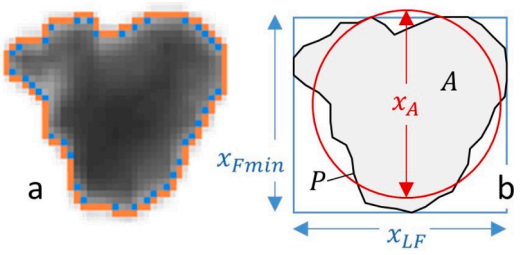

Fig. 2. Image analysis of a single granule: a) grayscale image overlaid (in color: orange = edges, blue = corners) to show pixels at the perimeter threshold; b) feature analysis including area, perimeter, and length measures.

process. Many studies have used accumulated moisture as a control for granulation (Saxena et al., 1979; Ochsenbein et al., 2019; Askarishahi et al., 2019).

3.1. Granule analysis

Granule products were characterized using dynamic image analysis (DIA) and analyzed for size and shape features. Samples from the fluid bed granulator were obtained by sequential riffling the full product from the FBG process, reducing the 3 kg batch to small (~ 5 g) samples suitable for DIA. The DIA system (SolidSizer, JM Canty, Lockport, NY, USA) has a controlled vibratory feed of dry granules, which then fall by gravity between a telecentric lens camera and an LED backlight. The camera magnification was set to 10 µm/pixel. Digital images of the back-lit particle projections were collected and analyzed using LabView NI-Vision (National Instruments, Austin, TX, USA). In Fig. 2, a grayscale image (8-bit grayscale, 0 = black; 255 = white) is overlaid with colored pixels indicating the perimeter threshold (214): orange indicates pixels with at least one edge on the perimeter; blue indicates pixels with at least one corner on the perimeter. The minimum Feret, x_{Fmin}, and its orthogonal, x_{LF}, are shown by the sides of the bounding box (blue) defined by x_{Fmin}. The area, A, and perimeter, P, are shown along with the area-equivalent circle (red).

The projected area-equivalent size, $x_A = \sqrt{(4A/\pi)}$, and Form Factor, $FF = 4\pi A/P^2$, are standard measures used to describe particle size and shape, respectively (International Standards Organization 9276-6, 2008). FF is a lumped shape metric that is sensitive to both elongation and perimeter irregularity, making it a logical choice for a comprehensive shape descriptor (Feltner et al., 2023). FF is bounded within $(0, 1]$, limiting the choice of distributions that can accurately capture the metric, for example truncated normal. We define an Inverse Form Factor, $iFF = (1 - FF)/FF$, as a transformation that maps iFF onto $(0, \infty)$, where $iFF = 0$ corresponds to a perfect circle. Since $ln(iFF)$ is unbounded, it is compatible with continuous distribution functions, such as the lognormal applied in this work.

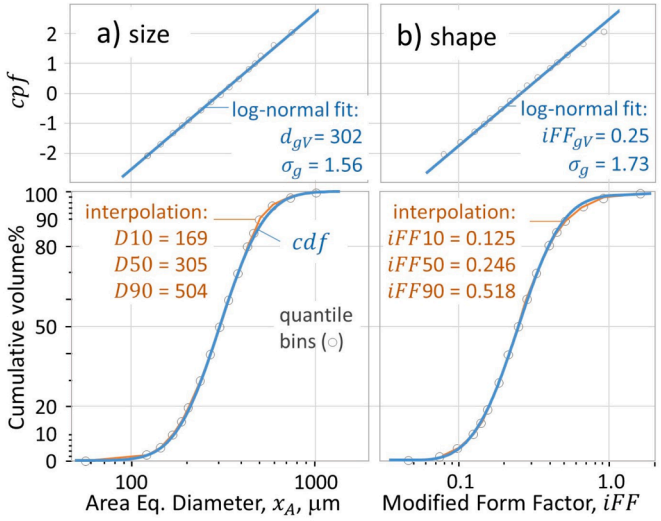

Fig. 3. Binning based on cumulative volume, $V = (\pi/6)x_A^3$, using quantile increments of run C data fitted using lognormal distributions: a) area equivalent size, x_A; b) modified Form Factor, $iFF = (1 - FF)/FF$. In lower plots with color, orange data points (binned, c_F) are connected by linear interpolation segments and overlaid by blue lognormal fits, cdf.

3.2. Distribution analysis

Typical DIA characterization includes data sets having thousands of particles imaged in random orientations in a dilute flow between a high-speed camera and [strobe-pulsed] backlight. There are various approaches to statistical analysis of these data, for example direct sampling approaches using statistical bootstrapping (Matsuyama, 2018). In the current work, data were sorted by the feature of interest, then classified into bins according to a scaled progression. Bins may be classified according to increments along the feature axis or the cumulative distribution axis. For consistency, volume-based bin increments were used throughout this paper. Fig. 3 illustrates binning based on quantile increments of the cumulative distribution data from run C; for graphical clarity, a total of 17 bins were equally spaced on 10 % increments across the center of the distribution (20 to 80 %), with progressively smaller increments representing the fine and coarse tails. In later ensemble analyses, higher binning resolution was used with 5 % increments across the center of the distribution, resulting in about 30 bins/sample.

While it is common in the literature to find size distribution data expressed in reduced form as interpolated quantiles ($D10$, $D50$, $D90$) and $Span = (D90 - D10)/D50$, these metrics are only marginally descriptive and are not sufficient for detailed modeling. On the other hand, fitted distribution functions (e.g., lognormal and stretched-exponential being the two most common) provide more useful reduced-order representations of distributed characteristics. The same statistical approach can be applied to shape distributions.

In the current work, measured size and shape data are described using lognormal distribution functions (Mort, 2023). For a given size or shape feature, binned data, $y = ln(feature)$, are regressed against their cumulative probabilities, $x = c_P = \sqrt{2} \cdot erfinv(2 \cdot c_F - 1)$, where c_P is the cumulative probability transform of the cumulative fractional data, c_F, evaluated within the domain (0,1). The result of the regression of y on x data is a cumulative probability function, $cpf = ln(d/d_g)/ln(\sigma_g)$, with the geometric mean, d_g, defined by the regression intercept, i.e., at $cpf = 0$; and the geometric standard deviation, σ_g, defined by the slope of the regression. The values of cpf indicate the number of geometric standard deviations away from the mean. Note that compound

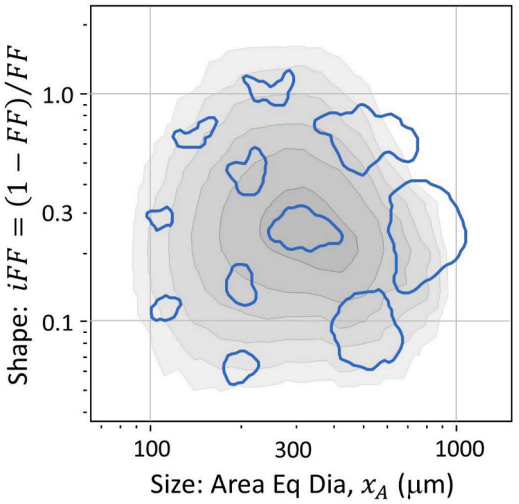

Fig. 4. Size and shape mapping of fluid bed granules: grayscale contours show volume-based density; blue traces are perimeter outlines of representative granules.

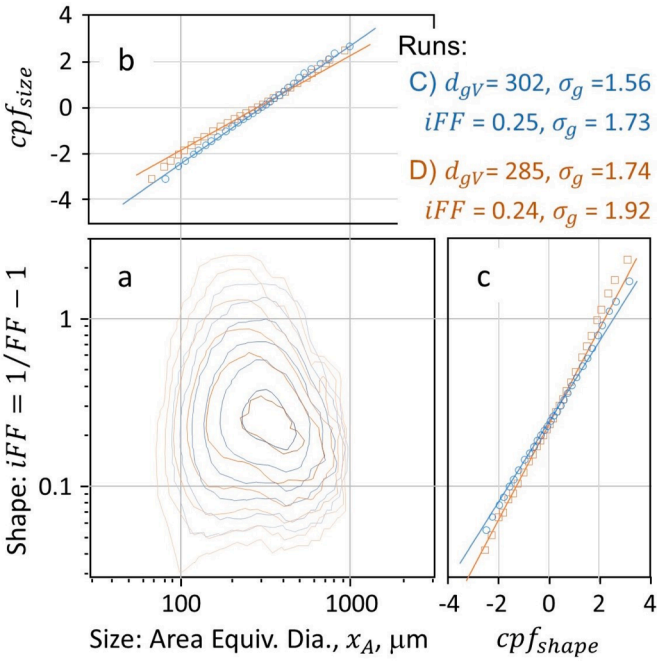

Fig. 5. Size and shape map with data reduced to density contours and distribution fitting using lognormal weighted regression. Two samples are shown in color: blue (C) and orange (D), having similar means (zero-cpf intercepts, d_{gV}) and different distribution breadths (slopes, $1/\sigma_g$).

weighting is required to minimize residuals with respect to the cumulative distribution function, $cdf = (1 + erf(cpf/\sqrt{2}))/2$, i.e., minimization of $\Sigma(cdf - c_F)^2$ (Mort, 2023; Kottler, 1950).

To exemplify lognormal fitting, the continuous cdf function can be easily compared with interpolated values; for the example of run C in Fig. 3 having a geometric mean, $d_{gV} = 302$ µm, the size (x_A) at $+1\sigma$ is $D84.1 \approx \exp(ln(d_{gV}) + ln(\sigma_g)) = 471$ µm, and the $+2\sigma$ size is $D97.7 \approx \exp(ln(d_{gV}) + 2ln(\sigma_g)) = 735$ µm. The 90th quantile, $D90$, is at $+1.28\sigma_g$, i.e., $D90 \approx \exp(ln(d_{gV}) + 1.28ln(\sigma_g)) = 534$ µm, in this case marginally higher than the interpolated value, visible by close inspection of Fig. 3a.

The same statistical approach was used to fit the distributed shape data. Note that the modified form factor, iFF, provides a well-matched fit to the lognormal distribution in Fig. 3b. Supplemental materials include image analysis data for all 6 runs along with binned ensemble datasets for the size and shape analyses shared in this report.

3.3. Multi-feature mapping

This section describes concurrent mapping of size and shape features on a per-run (i.e., per-sample) basis. It is useful to visualize how granule growth affects morphology. Fig. 4 is a map showing size on the abscissa and the shape on the ordinate for run C; the grayscale contours represent the volume-basis density of particles characterized by DIA in this sample (number of particles, $N = 21,808$). Perimeter traces (in blue) are overlaid to graphically illustrate the approximate size and shape of particles at various locations in the map. While not relating directly to the effect of process parameters, this style of mapping is a useful visualization of distributed granule characteristics.

Comparative mapping with fitted distributions is illustrated in Fig. 5, having contour overlays for two samples, runs C and D. The contour map (Fig. 5a) shows the interpolated density in geometric increments. The linear regression plots (Fig. 5b and c) show lognormal fitting against the cumulative probability, cpf. In this case, regressions were done independently for each sample. While both samples have similar geometric mean values, their distribution breadths are different, with run C being more narrowly distributed in both size and shape. Graphically, the breadth is indicated by the reciprocal slope of the regression.

3.4. Multi-effect analysis using ensemble regression

Regression fitting of individual data sets was extended to multi-effect ensemble regression including process parameters in the FBG case study. The graphical convention of Fig. 3, representing feature values on the abscissa and cumulative distributions (cdf, cpf) on the ordinate, is inverted, predicting the scalar feature of interest, $y = ln(feature)$, as a function of a matrix of process parameters, $\boldsymbol{X}$, and the cumulative probability, c_P, associated with the feature distribution. In the case of size, the feature of interest is made dimensionless, dividing by a unit measure, i.e., microns. Shape features are dimensionless by definition. The model can be expressed in matrix form, where $\boldsymbol{A}$ is the coefficient vector governing the log of the geometric mean, $ln(d_{gV}/\mu m) = \boldsymbol{A}^\top \boldsymbol{X}$, and $\boldsymbol{B}$ is the coefficient vector governing the log of the geometric standard deviation $ln(\sigma_g) = \boldsymbol{B}^\top \boldsymbol{\Delta X}$. The overall predicted value at any cumulative probability c_P is given by:

$$y = \boldsymbol{A}^\top \boldsymbol{X} + c_P \times (\boldsymbol{B}^\top \boldsymbol{\Delta X})$$

$$\boldsymbol{X} = \begin{bmatrix} 1 \\ AF \\ T \\ PVP \\ SR \\ (AF - \bar{AF})(T - \bar{T}) \end{bmatrix}, \qquad \boldsymbol{\Delta X} = \begin{bmatrix} 1 \\ AF - \bar{AF} \\ T - \bar{T} \\ PVP - \bar{PVP} \\ SR - \bar{SR} \end{bmatrix}$$

Here, AF, T, PVP, and SR are the FBG process parameters. The overbars (e.g., $\bar{AF}$) indicate the dataset-wide mean for that parameter, ensuring that interaction terms like $(AF - \bar{AF})(T - \bar{T})$ are mean-centered.

From a matrix algebra perspective, the ensemble regression model is dimensionally consistent. The vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ have dimension 6×1 and 5×1 respectively. Their transposed products with the corresponding input vectors yield scalar outputs for both $ln(d_{gV}/\mu m)$ and $ln(\sigma_g)$, ensuring that the final prediction, y, is also scalar-valued. Each remaining entry in $\boldsymbol{A}$ must have reciprocal units of its corresponding input in $\boldsymbol{X}$, such as AF^{-1}, T^{-1}, etc., such that their product yields a dimensionless argument to the logarithm. Likewise, the coefficients in $\boldsymbol{A}$ must carry units reciprocal to those in $\boldsymbol{\Delta X}$, ensuring that $\boldsymbol{B}^\top \boldsymbol{\Delta X}$ remains dimensionless, and the full expression $c_P \times (\boldsymbol{B}^\top \boldsymbol{\Delta X})$ is a valid logarithmic offset.

By including the c_P term in the regression model, the ensemble method is able to integrate process parameters with rich, distribution-level data. The c_P term accesses the ~ 30 binned data points per sample,

Table 2
Comparison of distribution fitting, independent (per run) and multi-effect (ensemble, combined runs) for size and shape features.

Run	Independent, per run				Multi-effect, ensemble			
	Size, x_A		Shape, iFF		Size, x_A		Shape, iFF	
	d_{gV}, µm	σ_g	iFF_{gV}	σ_g	d_{gV}, µm	σ_g	iFF_{gV}	σ_g
A	221	1.67	0.32	1.99	218	1.69	0.32	2.00
B	255	1.67	0.31	1.87	250	1.67	0.31	1.85
C	302	1.56	0.25	1.73	300	1.57	0.25	1.75
D	285	1.74	0.24	1.92	280	1.76	0.24	1.94
E	249	1.56	0.22	1.92	248	1.55	0.22	1.92
F	249	1.71	0.20	1.89	249	1.71	0.20	1.90

Table 3
Size, x_A, and shape, iFF, model coefficients obtained from ensemble multi-effect regression analyses (Fig. 6). AF is airflow, m^3/h; T is slot temperature, °C; PVP is the fractional amount of binder in the dry-basis formulation; SR is the spray rate of the 25 % PVP binder solution, g/min; and c_P is the cumulative probably transform of the binned data.

Fit parameters	Fit coefficients (|t-ratio|)			
A	Size: $ln(d_{gV}/\mu m)$		Shape: $ln(iFF_{gV})$	
Intercept	$6.04E+0$	(233)	$-3.13E+0$	(106)
AF	$-7.48E-3$	(61)	$9.18E-3$	(66)
T	$-1.77E-2$	(73)	$2.89E-2$	(105)
PVP	$1.51E+1$	(70)	$-2.16E+0$	(9)
SR	$7.14E-3$	(39)	$-1.33E-2$	(63)
$(AF-\bar{AF})(T-\bar{T})$	$-4.94E-4$	(35)	$6.64E-4$	(41)
B	Size: $ln(\sigma_g)$		Shape: $ln(\sigma_g)$	
c_P	$5.07E-1$	(280)	$6.34E-1$	(307)
$c_P \times (AF-\bar{AF})$	$3.43E-3$	(21)	$3.07E-3$	(17)
$c_P \times (T-\bar{T})$	$4.86E-3$	(13)	$4.42E-3$	(11)
$c_P \times (PVP-\bar{PVP})$	$-1.43E+0$	(4.0)	$-7.70E+0$	(19)
$c_P \times (SR-\bar{SR})$	$-7.54E-4$	(2.4)	$-5.01E-3$	(14)

said binned data derived from the DIA reduction of ~ $20k$ granule images per sample. The use of compound-weighted linear regression ensures that each bin contributes appropriately to the overall model, enabling the simultaneous estimation of both distribution mean and spread as functions of process parameters.

Ensemble multi-effect regression of the full sample set (runs A-F) included the distributed size and shape features from each run as well as the airflow, temperature, binder content, and spray rate parameters given in Table 1. The results are shown in Fig. 6 as parity plots with ANOVA summary statistics (JMP, Version 16.0.0. SAS Institute Inc., Cary, NC). The parity plots show excellent predictability over the centers of the probability distributions, i.e., within +/- 1 geometric standard deviation (the 16th and 84th quantiles), achieved with the compound weighed regression; note that divergent points may occur at the tails of the distribution (e.g., < 10 %, > 90 %) which have reduced weighting in the regression.

The ensemble fit quality for both size and shape are well aligned with individually fitted samples (Table 2), demonstrating that the fitting approach is consistent within the explored formulation and operating space. The model coefficients for the multi-effect regression (Table 3) show the effects of airflow, inlet temperature, binder content, and their interactions, including interactions with size and shape distributions. Each model has 11 statistically significant coefficients (10 degrees of freedom plus the intercept). The t-ratio is a measure of how many standard errors the regressed coefficients are away from 0; high absolute values of the t-ratio indicate high confidence in the statistical significance of the associated parameter. All estimates are highly significant, i.e., having a statistical probability > $|t|$ < 0.01 % in all cases except for x_A: $c_P \times SR$, prob > $|t|$ < 0.2 %.

Ensemble regression solves for both product distributions (i.e., the cdf) and process parameter effects (airflow, temperature, binder amount, and spray rate) simultaneously. The regression intercept along

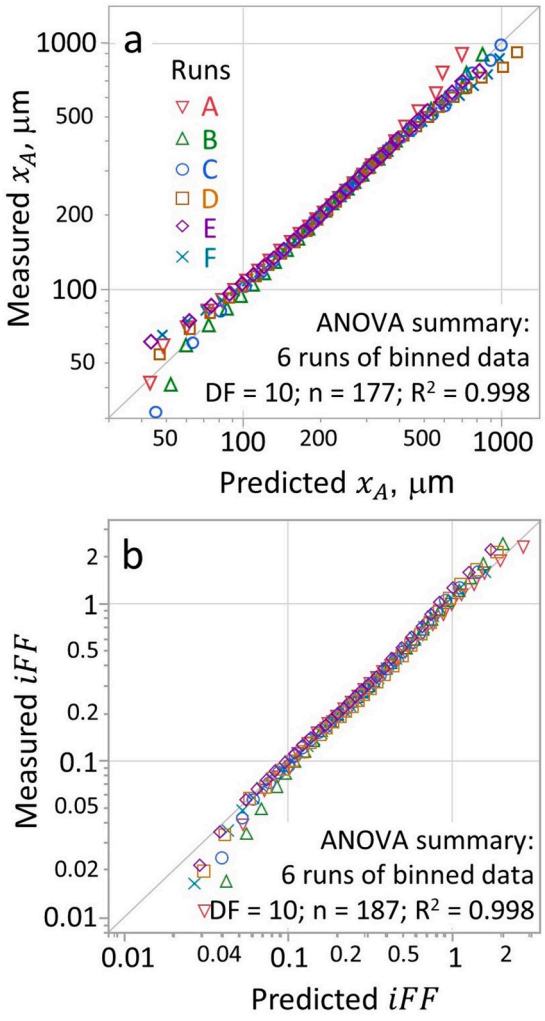

Fig. 6. Multi-effect regression and summary of binned data combining runs A-F: a) size; b) shape distributions.

with non-c_P terms apply to the geometric mean, and the terms having c_P dependence apply to the geometric standard deviation. In this example, six terms affect geometric means (both size and shape), and five affect geometric standard deviations.

3.5. Model validation

Cross-validation methods are often used when data are insufficient to run a train-test split; in this case, its use is motivated by the sparsity of process data. A series of holdout splits was applied to the ensemble particle size data (ensemble number of bins, $N = 178$), where the data were randomly assigned as either training or validation subsets over a range of split ratios (training, validation): (90,10), (70,30), (50,50), (40,60), (30,70), (20,80), (15,85), and (10,90). In each case, the root-average-squared-error (RASE) was calculated for both subsets as shown in Fig. 7a. A low training RASE relative its validation RASE is an indication of overfitting. The convergence of training and validation RASE values indicates higher confidence in the model, shown as “robust” in the figure.

The solid lines in Fig. 7a represent best-fit stretched-exponential trends of the holdout split RASE data, $y_{train} = y_\infty(1-\exp(-(x/x^*)^m))$, where y is a proxy for RASE, $y_\infty = 0.0132$, $x^* = 10.8$ %, and $m = 1.18$. While there are no hard definitions of overfitting or robustness, the cross-validation results suggest reasonable limits can be proposed. In this case, overfitting is defined by $y_{train} < 0.95 \cdot y_\infty$, and robust fitting by

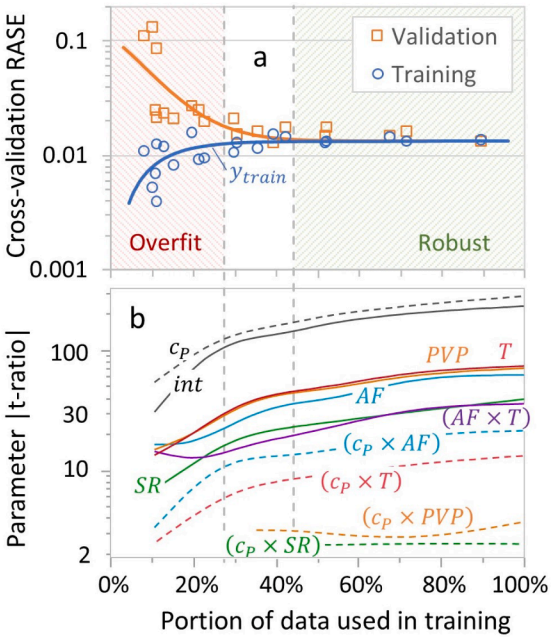

Fig. 7. Holdout validation: a) Root-Average-Squared-Error (RASE) and b) P-ratio trends of parameter fits as functions of randomized splitting of training and validation data. Solid trendlines in (b) apply to geometric mean size; dashed lines apply to the geometric standard deviation.

$y_{train} > 0.995 \cdot y_{\infty}$. A robust threshold of about 44 % of 178 data points for an 11-parameter model (including the intercept) suggests a data/parameter ratio of at least about 7:1 for robust fitting of the current system.

In Fig. 7b, ensemble regression with distributed data increases the statistical significance of coefficients, especially those describing the breadth of said distributions. Conversely, as the fraction of distributed data used on model training is reduced, there is a trend toward reduced t-ratios of the fitting parameters. In the current work, t-ratios show relatively stable behavior in the robust range, and show a significant drop-off below the overfitting threshold.

4. Results and discussion

Predictive models obtained from multi-effect regression were used to create contour maps for graphical display of size and shape distributions as a function of process and formulation inputs. The models are further employed with an iterative solver to suggest optimal conditions for producing uniform products within various operating constraints.

Contour maps were used to illustrate size and shape predictions in terms of inlet air flow, air temperature, and binder level (Figs. 8 and 10, respectively). The masked areas of the contour maps indicate ineffective operating conditions that were either too dry (peak product moisture < 7.5 %) or too wet (peak product moisture > 11 %) for the formulations used in this study. In over-wet conditions, excessive moisture led to bed collapse and channeling, resulting in poor fluidization and uneven granule growth. In over-dry conditions, there was insufficient moisture for granule formation, resulting in excessive fines and dust. The unmasked areas indicate a stable balance between the spray rate and drying enthalpy for granulation, illustrated by the stable operating range (SOR) in Fig. 1.

4.1. Granule size mapping

An array of contour maps (Fig. 8) shows the effect of inlet air flow, air temperature, binder amount, and binder spray rate on the predicted geometric mean size, d_{gV}, and standard deviation, σ_g. The contour curvature of the predicted d_{gV} comes from the interaction of temperature and airflow on drying enthalpy. The individual interactions of the size distribution with temperature and airflow are also apparent but as straight [dashed] lines due to the lack of a statistically-significant three-way empirical correlation between size c_P, temperature, and airflow.

Note the Fig. 8 maps generated for 50 g/s spray rate (i.e., bottom row) are based on 5 of the 6 experimental runs, while the higher spray-rate conditions rely on only one run (F) at 5 % dry-basis PVP; hence, the maps on upper right are extrapolated from the model. Although a more thorough and balanced DOE should include a more representative distribution of experimental conditions, the ensemble regression model provides a reasonable prediction of a limited operating space at the high-high (upper right) condition. Indeed, balancing wetting and drying with an extended high spray rate can be challenging, hence the model-guidance of operating with higher drying enthalpy limits growth and narrowing of the size distribution.

Table 4
Minimization of σ_g by model-based optimization of process parameters (airflow, AF, m^3/h; slot temperature, T, °C) as a function of binder content (dry-basis PVP) and spray rate, SR, of 25 % binder solution, g/min.

	5 % PVP				7 % PVP			
SR	AF	T	d_{gV}	σ_g	AF	T	d_{gV}	σ_g
40					64	64	345	1.49
50	65	58	280	1.46	79	67	320	1.57
60	72	59	280	1.51	88	72	294	1.65
70	84	62	272	1.58	100	75	254	1.73
80	93	64	264	1.62	112	78	215	1.81
90	101	66	248	1.68				

4.2. Optimization on size distribution

Notice that the narrowest distributions (i.e., minimizing σ_g) occur at the wet SOR limit. This observation suggests an approach to optimized mapping using a combination of binder addition parameters at the wet limit, for example assuming a fixed composition of binder and overlaying contours representing various spray rates (Fig. 9), where the locus of points having the minimum σ_g represents an optimal set of spray-rate dependent process conditions. As the spray rate increases, the anticipated production benefit of reduced batch cycle time can be evaluated in context of the predicted change in granule growth and broadening of the size distribution along the optimized path (Table 4).

Airflow has multiple effects on the process, contributing to fluidization, granulation dynamics, and drying. Combinations of airflow and temperature are necessary to provide drying enthalpy to the process, and are optimized by approaching the upper (wet) peak moisture limit. In addition to the minimum σ_g tangent at the wet peak moisture limit, where the slope of the tangent line depended on both airflow and temperature, the general trend of narrowing distributions favored enthalpies comprising lower airflow and higher temperature. Mechanistically, one can posit that the lower airflow provides more uniform mixing of the fluidized powder with the binder spray, avoiding excessive elutriation away from the spray zone.

4.3. Granule shape mapping

Fluidized bed processing had measurable effects on granule shape, in some cases having an even higher statistical significance than size (t-ratios, Table 3). Fig. 10 shows a shape map (iFF) analogy of the center panel condition of the size maps in Fig. 8. The model trend shows more rounding and narrowing of the shape distribution with increasing peak moisture content, again with an optimum depending on a combination of process inputs. The thin-line contours show the tangent to the wet limit, suggesting an optimum condition for a narrow shape distribution. For reference, granule shapes corresponding to iFF values are shown in Fig. 4.

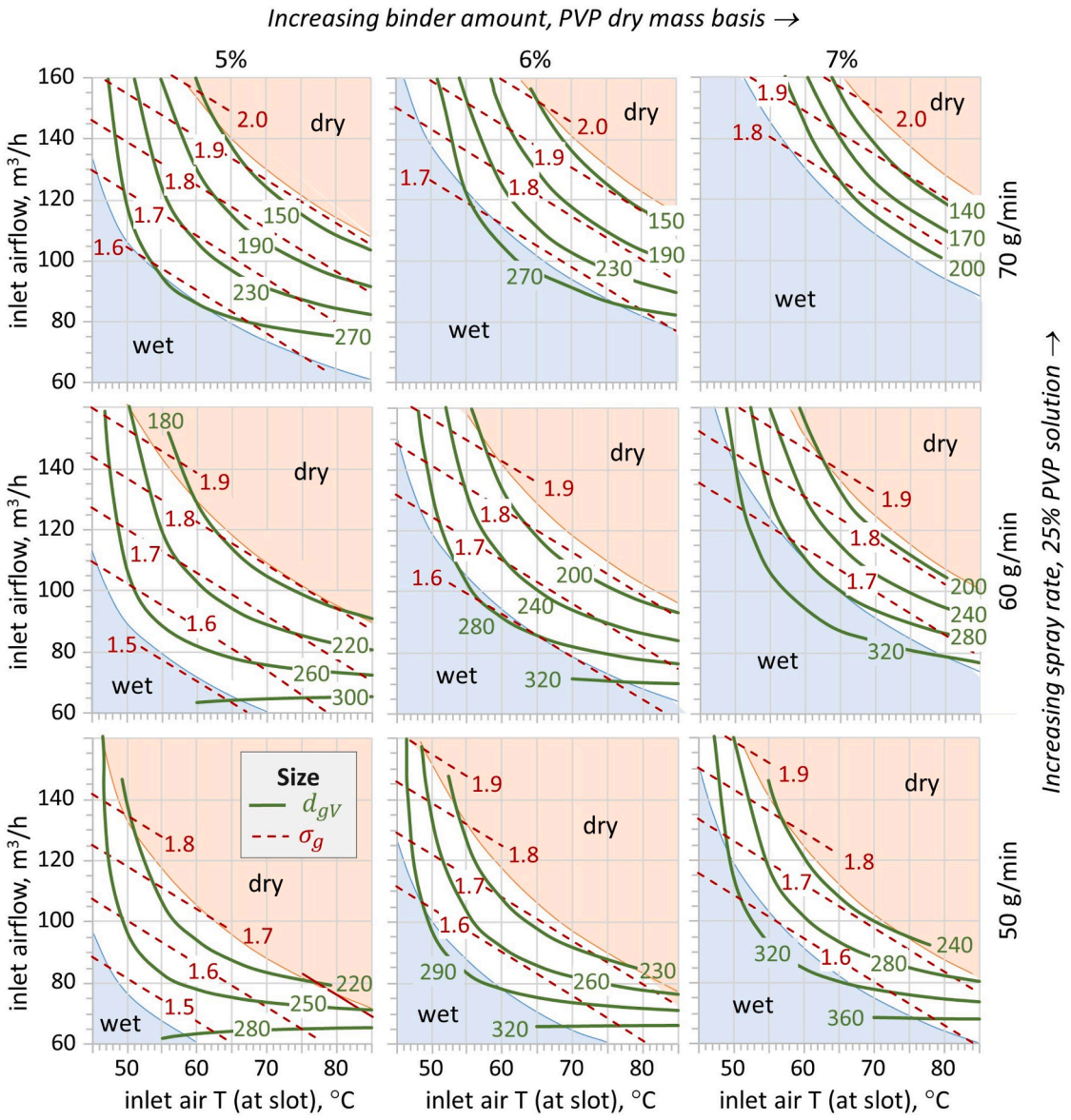

Fig. 8. Size maps showing effects of process and formulation parameters (inlet airflow, air temperature, binder amount, and binder spray rate) on the granule size distribution (d_{gV}, σ_g), where size is the area-equivalent diameter x_A, μm, on a volume basis. The stable operating zone between over-dry and over-wet zones is defined by the enthalpy balance between the aqueous binder solution spray rate and moisture evaporation.

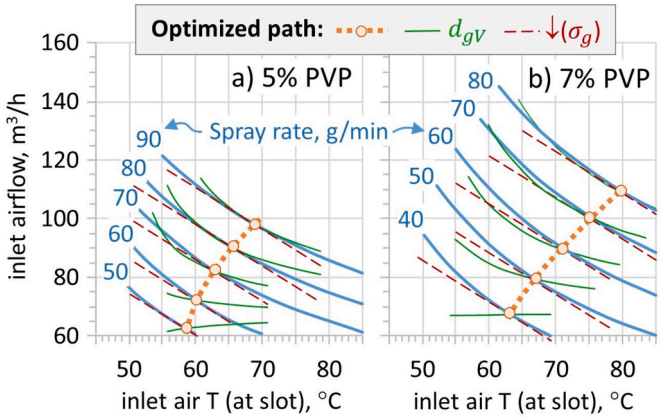

Fig. 9. Combined mapping of multiple spray rates, optimized for the minimum distribution breadth, σ_g, tangent to the wet enthalpy limit (11 % product moisture at the end of the spray segment).

Shape characterization is emerging as a practical capability and may become more relevant in context of its correlation with bulk properties such as granular flow, packing, compressibility, and dispersion/dissolution. The use of ensemble regression modeling, while empirical, may provide a useful tool for integrating shape into process control and optimization thinking.

5. Perspective and outlook

In the current work, ensemble modeling combines regression of distributed product characteristics with process parameters. We chose fluidized bed granulation as a case study based on its complex process-product interactions, where practical bounds on process parameters assist the empirical modeling approach. The authors had the fortuitous opportunity to use pilot-scale FBG equipment having robust control and operational capabilities. Although pilot-scale trials require larger material quantities, they provide greater confidence in manufacturing-scale operations. Clearly, it is an advantage to have confidence in process modeling of design spaces on a manufacturing-scale in the context of

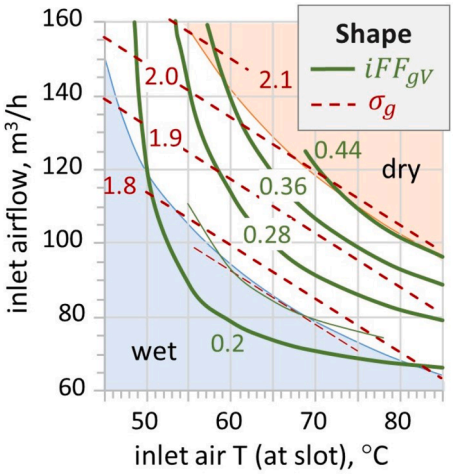

Fig. 10. Mapping of granule shape distribution at $PVP = 6\,\%$ (dry mass basis) and $SR = 60g/min$. The wet-limit tangent (narrow dashed line) indicates a narrow shape distribution, i.e., $\sigma_g = 1.76$, at $AF = 82m^3/h$ and $T = 68°C$.

limited pilot-scale data, motivating the case study shared in this work. This has practical utility, especially for industries where it is necessary to define a design space in a regulatory filing.

More broadly, recent work in ensemble regression modeling focuses on reducing uncertainty associated with individual models. Combining models as an ensemble is a strategy used across a wide range of applications and is consistent with the recent surge of interest in machine learning and artificial intelligence. Weighted linear regression is a foundational method for supervised machine learning and has a long history in particulate systems, for example, ranging from workplace safety and aerosols (Hatch and Choate, 1929; Heintzenberg, 1994) to photographic film emulsions (Kottler, 1950), and the more recent focus on microplastics in the environment (Schirrmeister et al., 2024). The ensemble modeling approach described herein is broadly applicable to distributed systems where large data sets can be modeled using cumulative distribution functions.

In industrial powder processing, size and shape distributions affect flow and packing behavior, especially in contexts where particle body forces exceed cohesion (Mort, 2015; Sandler and Wilson, 2010). Additionally, there is an extensive legacy relating size and shape characterization with geophysics (Ulusoy, 2023). Combining ensemble regression tools with granular mechanics is a logical step forward.

6. Conclusions

Geometric distributions are prevalent across particulate processing of chemicals, foods, pharmaceuticals and other products. Distributed particle size and shape features affect the stability and control of industrial processes including granulation, milling, crystallization, mixing, handling, classification, and separations. It is common practice to describe such distributions by log-linear regression of binned data, for example using lognormal or Weibull distribution models. In this paper, we presented a case study using dynamic image analysis to characterize products made by fluidized bed granulation, using a limited number of pilot-scale process trials. Each product had a rich dataset comprising large numbers of particle images that were analyzed and then condensed into quantile bins. Ensemble multi-effect regression combined the binned data with the process parameters of each trial to simultaneously solve for product distributions (e.g., lognormal geometric mean and standard deviation) as a function of process parameters.

In the fluid bed granulation case study, sparse process data were combined with rich product characterization to create statistically robust empirical models of granule size and shape distributions. Six process trials were included having airflow, air temperature, binder amount, and binder spray rate as independent parameters. Each product was characterized using dynamic image analysis, collecting and analyzing on the order of 20k particle images per sample, which were condensed into ~ 30 bins distributed according to cumulative distribution increments, both for size (x_A) and shape (iFF) features. The cumulative data were transformed to a cumulative probability function, cpf, suitable for weighted linear regression analysis, said regression revealing strong statistical correlation of eleven fit parameters, six relating to the distribution geometric mean (d_{gV}, iFF_{gV}) and five relating to the geometric standard deviation (σ_g). The resulting lognormal size model was combined with first-principal enthalpy constraints to map out a detailed design space, including some areas that were well represented by the experimental data and others that were extrapolated. Shape regression was enabled by using a modified form factor compatible with lognormal fitting. The shape model offers the potential for further insight detailing the design space.

Symbols

Symbol(s)	Meaning
Fluid bed granulation (FBG)	
A,B,C,D,E,F	run identifiers
AF	airflow, m³/h
T	slot temperature, °C
PVP	binder amount, dry basis, wt.%
SR	binder spray rate, g/min
SOR	stable operating range, cumulative moisture
Regression data, distribution analysis	
$\boldsymbol{A}, \boldsymbol{B}$	matrices of regression coefficients
c_F	cumulative fraction, data
c_P	cumulative probability, data transform
cdf	cumulative distribution function, fitted
cpf	cumulative probability function, fitted
x, y	scalar data values
$\boldsymbol{X}, \boldsymbol{\Delta X}$	matrix data, FBG process parameters
Size and shape, subscript V denotes volume basis	
A	projected area, image analysis
d_g	geometric mean size
FF	form factor, $FF = 4\pi A/P^2$
iFF	inverse form factor, $iFF = 1/FF - 1$
σ_g	geometric standard deviation
P	projected perimeter, image analysis
x_A	area equivalent size, $x_A = \sqrt{4A/\pi}$

CRediT authorship contribution statement

Line Koleilat: Writing – original draft, Methodology, Investigation, Data curation, Conceptualization; **Langdon Feltner:** Writing – review & editing, Methodology; **Donovan Stumpf:** Writing – review & editing, Methodology; **Paul Mort:** Writing – review & editing, Writing – original draft, Supervision, Software, Resources, Methodology, Formal analysis, Conceptualization.

Data availability

Shared as supplemental material

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.

Acknowledgments

Materials used in the study were kindly provided by Foremost Farms, FMC, and Ashland Chemicals; details cited in the paper. Advice and guidance on the operation of the SolidLab2 fluidized bed granulator was kindly provided by Christian Karl Paasche (Dipl. Ing.) of Syntegon GmbH.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.ces.2025.122042.

References

Askarishahi, M., Salehi, M.-S., Maus, M., Schröder, D., Slade, D., Jajcevic, D., 2019. Mechanistic modelling of fluid bed granulation, part II: eased process development via degree of wetness. Int. J. Pharm. 572, 118836. https://doi.org/10.1016/j.ijpharm.2019.118836

Braumann, A., Goodson, M.J., Kraft, M., Mort, P.R., 2007. Modelling and validation of granulation with heterogeneous binder dispersion and chemical reaction. Chem. Eng. Sci. 62 (17), 4717–4728. https://doi.org/10.1016/j.ces.2007.05.028

Dan, A., Paul, S., Vaswani, H., Grząbka-Zasadzińska, A., Thakkelapally, A., Li, J., Koyel, S., Deanne, R., Tseng, Y.-C., Ramachandran, R., 2024. Quantitative analysis of the effects of multi-component formulation parameters on granule and tablet properties via a combined population balance and statistical predictive model. Powder Technol. 435, 119391. https://doi.org/10.1016/j.powtec.2024.119391

Ennis, B.J., Tardos, G., Pfeffer, R., 1991. A microlevel-based characterization of granulation phenomena. Powder Technol. 65, 257–272. https://doi.org/10.1016/0032-5910(91)80189-P

Feltner, L., Korte, E., Bahr, D.F., Mort, P., 2023. Particle size and shape analyses for powder bed additive manufacturing. Particuology. https://doi.org/10.1016/j.partic.2023.09.001

Gantt, J.A., Gatzke, E.P., et al., 2005. High-shear granulation modeling using a discrete element simulation approach. Powder Technol. 156 (2), 195–212. https://doi.org/10.1016/j.powtec.2005.04.012

Hapgood, K.P., Litster, J.D., Smith, R., 2003. Nucleation regime map for liquid bound granules. AlChE J. 49 (2), 350–361. https://doi.org/10.1002/aic.690490207

Hatch, T., Choate, S.P., 1929. Statistical description of the size properties of non uniform particulate substances. J. Frankl. Inst. 207 (3), 369–387. https://doi.org/10.1016/S0016-0032(29)91451-4

Heintzenberg, J., 1994. Properties of the log-normal particle size distribution. Aerosol. Sci. Technol.. https://doi.org/10.1080/02786829408959695

International Standards Organization 9276-6, 2008. Iso 9276-6:2008(en), representation of results of particle size analysis—Part 6: descriptive and quantitative representation of particle shape and morphology.

Iveson, S.M., Wauters, P. A.L., Forrest, S., Litster, J.D., Meesters, G. M.H., Scarlett, B., 2001. Growth regime map for liquid-bound granules: further development and experimental validation. Powder Technol. 117 (1), 83–97. Granulation and Coating of Fine Powders. https://doi.org/10.1016/S0032-5910(01)00317-5

Kastner, C.A., Brownbridge, G. P.E., Mosbach, S., Kraft, M., et al., 2013. Impact of powder characteristics on a particle granulation model. Chem. Eng. Sci. 97, 282–295. https://doi.org/10.1016/j.ces.2013.04.032

Kayrak-Talay, D., Dale, S., Wassgren, C., Litster, J., 2013. Quality by design for wet granulation in pharmaceutical processing: assessing models for a priori design and scaling. Powder Technol. 240, 7–18. https://doi.org/10.1016/j.powtec.2012.07.013

Kayrak-Talay, D., Litster, J.D., 2011. A priori performance prediction in pharmaceutical wet granulation: testing the applicability of the nucleation regime map to a formulation with a broad size distribution and dry binder addition. Int. J. Pharm. 418 (2), 254–264. https://doi.org/10.1016/j.ijpharm.2011.04.019

Koleilat, L., Paasche, C.K., Wade, J., Hanson, J., Wassgren, C., Mort, P., 2024. Flow field analysis of a toroidal flow fluidized bed granulator. Chem. Eng. Res. Des. 208, 359–368. https://doi.org/10.1016/j.cherd.2024.07.003

Koleilat, L., Paasche, C.K., Wade, J., Hanson, J., Wassgren, C., Mort, P., 2025. Fluid bed granulation—process optimization. Powder Technol. 449, 120358. https://doi.org/10.1016/j.powtec.2024.120358

Kottler, F., 1950. The distribution of particle sizes. J. Frankl. Inst. 250 (4), 339–356. https://doi.org/10.1016/0016-0032(50)90926-4

Kristensen, H.G., Schaefer, T.S., 1987. Granulation: a review on pharmaceutical wet-granulation. Drug Dev. Ind. Pharm. 13 (4–5), 803–872. https://doi.org/10.3109/03639048709105217

Kumar, A., Dhondt, J., Vercruysse, J., De Leersnyder, F., Vanhoorne, V., Vervaet, C., Remon, J.P., Gernaey, K.V., De Beer, T., Nopens, I., 2016. Development of a process map: a step towards a regime map for steady-state high shear wet twin screw granulation. Powder Technol. 300, 73–82. https://doi.org/10.1016/j.powtec.2015.11.067

Lipps, D.M., Sakr, A.M., 1994. Characterization of wet granulation process parameters using response surface methodology. 1 top-spray fluidized bed. J. Pharm. Sci. 83 (7), 937–947.

Litster, J., Ennis, B., 2004. The Science and Engineering of Granulation Processes. Springer Dordrecht. https://doi.org/10.1007/978-94-017-0546-2

Liu, B., Wang, J., Zeng, J., Zhao, L., Wang, Y., Feng, Y., Du, R., 2021. A review of high shear wet granulation for better process understanding, control and product development. Powder Technol. 381, 204–223. https://doi.org/10.1016/j.powtec.2020.11.051

Matsuyama, T., 2018. An application of bootstrap method for analysis of particle size distribution. Adv. Powder Technol. 29 (6), 1404–1408. https://doi.org/10.1016/j.apt.2018.03.002

Mort, P., 2015. Characterizing flowability of granular materials by onset of jamming in orifice flows. Papers Phys. 7. https://doi.org/10.4279/pip.070004

Mort, P., 2023. Analysis and graphical representation of particle size distributions. Powder Technol. 420, 118100. https://doi.org/10.1016/j.powtec.2022.118100

Ochsenbein, D.R., Billups, M., Hong, B., Schäfer, E., Marchut, A.J., Lyngberg, O.K., 2019. Industrial application of heat- and mass balance model for fluid-bed granulation for technology transfer and design space exploration. Int. J. Pharm. 1, 100028. https://doi.org/10.1016/j.ijpx.2019.100028

Oishi, T., Hayashi, Y., Noguchi, M., Yano, F., Kumada, S., Takayama, K., Okada, K., Onuki, Y., 2020. Creation of novel large dataset comprising several granulation methods and the prediction of tablet properties from critical material attributes and critical process parameters using regularized linear regression models including interaction terms. Int. J. Pharm. 577, 119083.

Pohl, S., Kleinebudde, P., 2020. A review of regime maps for granulation. Int. J. Pharm. 587, 119660. j.ijpharm.2020.119660

Rough, S.L., Wilson, D.I., York, D.W., 2005. A regime map for stages in high shear mixer agglomeration using ultra-high viscosity binders. Adv. Powder Technol. 16 (4), 373–386. https://doi.org/10.1163/1568552054194186

Rowe, R.C., Roberts, R.J., 1995. 1 The mechanical properties of powders. In: Ganderton, D., Jones, T., McGinity, J. (Eds.), Advances in Pharmaceutical Sciences. Academic Press. Vol. 7, pp. 1–IV. https://doi.org/10.1016/S0065-3136(06)80003-8

Sandler, N., Wilson, D., 2010. Prediction of granule packing and flow behavior based on particle size and shape analysis. J. Pharm. Sci. 99 (2), 958–968. https://doi.org/10.1002/jps.21884

Saxena, S.C., Chatterjee, A., Patel, R.C., 1979. Effect of distributors on gas-solid fluidization. Powder Technol. 22 (2), 191–198. https://doi.org/10.1016/0032-5910(79)80026-1

Schirrmeister, S., Kurzweg, L., Gjashta, X., Socher, M., Fery, A., Harre, K., 2024. Regression analysis for the determination of microplastics in sediments using differential scanning calorimetry. Environ. Sci. Pollut. Res. Int.. https://doi.org/10.1007/s11356-024-33100-8

Schmitt, J.M., Baumann, J.M., Morgen, M.M., 2022. Predicting spray dried dispersion particle size via machine learning regression methods. Pharm. Res. 39 (12), 3223–3239.

Sivarao, S., Said, R.b., Rostam, O., Hafiz, M.Z., Hassan, L.H., Shaaban, A., Dimin, M.F., 2014. Top spray fluidized bed granulated paddy urea fertilizer. Appl. Mech. Mater.. https://doi.org/10.4028/www.scientific.net/AMM.606.137

Ulusoy, U., 2023. A review of particle shape effects on material properties for various engineering applications: from macro to nanoscale. Minerals 13 (1). https://doi.org/10.3390/min13010091

Veliz Moraga, S., Villa, M.P., Bertín, D.E., Cotabarren, I.M., Piña, J., Pedernera, M., Bucalá, V., 2015. Fluidized-bed melt granulation: the effect of operating variables on process performance and granule properties. Powder Technol. 286, 654–667. https://doi.org/10.1016/j.powtec.2015.09.006