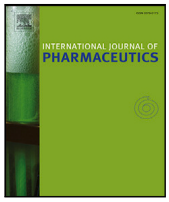
International Journal of Pharmaceutics 651 (2024) 123796

Contents lists available at ScienceDirect

International Journal of Pharmaceutics

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



Using AI/ML to predict blending performance and process sensitivity for Continuous Direct Compression (CDC)

O. Jones-Salkey [a,b,*], C.R.K. Windows-Yule [a], A. Ingram [a], L. Stahler [b], A.L. Nicusan [a], S. Clifford [b], L. Martin de Juan [c], G.K. Reynolds [b]

[a] School of Chemical Engineering, University of Birmingham, Edgbaston, Birmingham, UK
[b] Oral Product Development, Pharmaceutical Technology & Development, Operations, AstraZeneca, Macclesfield, UK
[c] Oral Product Development, Pharmaceutical Technology & Development, Operations, AstraZeneca, Gothenburg, SWE

ARTICLE INFO

Keywords:
Machine learning
Artificial intelligence
Mixing
Blending
Continuous Direct Compression
Prediction
Formulation

ABSTRACT

Utilising three artificial intelligence (AI)/machine learning (ML) tools, this study explores the prediction of fill level in inclined linear blenders at steady state by mapping a wide range of bulk powder characteristics to processing parameters. Predicting fill levels enables the calculation of blade passes (strain), known from existing literature to enhance content uniformity. We present and train three AI/ML models, each demonstrating unique predictive capabilities for fill level. These models collectively identify the following rank order of feature importance: RPM, Mixing Blade Region (MB) size, Wall Friction Angle (WFA), and Feed Rate (FR). Random Forest Regression, a machine learning algorithm that constructs a multitude of decision trees at training time and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees, develops a series of individually useful decision trees. but also allows the extraction of logic and breakpoints within the data. A novel tool which utilises smart optimisation and symbolic regression to model complex systems into simple, closed-form equations, is used to build an accurate reduced-order model. Finally, an Artificial Neural Network (ANN), though less interrogable emerges as the most accurate fill level predictor, with an r^2 value of 0.97. Following training on single-component mixtures, the models are tested with a four-component powdered paracetamol formulation, mimicking an existing commercial drug product. The ANN predicts the fill level of this formulation at three RPMs (250, 350 and 450) with a mean absolute error of 1.4%. Ultimately, the modelling tools showcase a framework to better understand the interaction between process and formulation. The result of this allows for a first-time-right approach for formulation development whilst gaining process understanding from fewer experiments. Resulting in the ability to approach risk during product development whilst gaining a greater holistic understanding of the processing environment of the desired formulation.

1. Introduction

A significant volume of research on continuous powder blending has focused on understanding how processing parameters affect performance using metrics such as residence time distributions and dispersion. Vanarase and Muzzio (2011), Vanarase et al. (2013), Holman et al. (2021), Van Snick et al. (2017), Hurley et al. (2022), while other – typically simulation-based – studied have demonstrated the differences that particle properties have on processing efficacy (Gao et al., 2011, 2012b,a; Sarkar and Wassgren, 2009, 2010; Toson et al., 2018). However, there is little work mapping physical experimental bulk powder characteristics to processing conditions in the blender. The only research done to the researchers' knowledge lay with that of Bekaert et al. (2022a,b), who investigated blend properties in a horizontal blender and the development of an empirical predictive model in an incline blender.

With the advent of readily-available, open-source Artificial Intelligence and Machine Learning tools, this study looks to leverage these methodologies to propose a novel framework to approach new pharmaceutical development projects. These methods allow the de-convolution of material properties and processing parameters by mapping processing outcomes to powder characteristics and processing parameters. This, in turn, allows for the prediction of key process outputs, such as the fill level within the blender and (thus) the strain experienced by the formulation, which has been shown to be a good predictor for blending performance (Vanarase and Muzzio, 2011; Jones-salkey et al., 2023a; Bekaert et al., 2022a).

* Corresponding author.
E-mail addresses: o.jones-salkey@hotmail.com, OXJ807@student.bham.ac.uk (O. Jones-Salkey).

https://doi.org/10.1016/j.ijpharm.2024.123796
Received 27 November 2023; Received in revised form 3 January 2024; Accepted 5 January 2024
Available online 6 January 2024
0378-5173/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

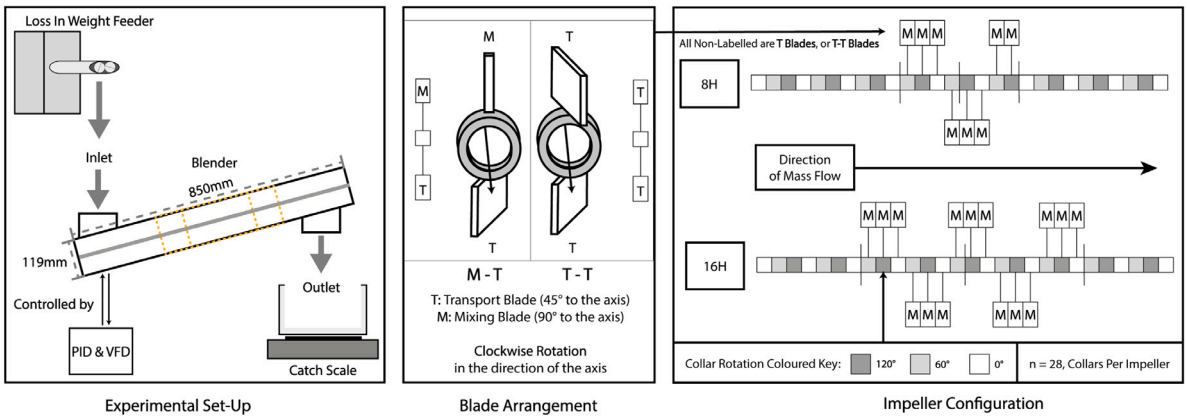


Fig. 1. The above figure graphically describes the Experimental Set-Up, Blade Arrangement, and Impeller configuration used for the experiments.

The utilisation of the three different models in the study ideally highlights the individual strengths of using each of the specific modelling methods. Random Forest Regression (RFR) can 'pull the logic' from the data by providing insights into feature importance and parameter relationships. Individual 'trees' extracted from the model may also provide useful guidance to system operators. PyTorch's Artificial Neural Networks (ANN), on the other hand, fit highly complex non-linear functions to the data to enable accurate and reliable predictions of the data surface, though are somewhat more of a 'black box', offering only limited physical insight regarding the outputs they produce. Finally, we use M^2E^3D, a tool that utilises evolutionary AI and symbolic regression to transform complex systems into simple equations (Nicusan and Windows-Yule, 2022a,b); the methodology of the tool is discussed further in Section 4.5 and accessible on GitHub: https://github.com/uob-positron-imaging-centre/MED.

Ultimately, by using these tools, this study unveils a correlation between fill level and optimum strain, which holds true for varied processing and material properties. The discussion focuses on how this relation develops by relating it to current literature and exploring space–time optimisation and optimum Peclet Numbers in specific regime-like behaviour. The study then discusses the use of such tools in an industrial setting by touching on how best to use this mapping ability to inform decisions to mitigate risk whilst navigating the complexities of drug project development.

2. Equipment

2.1. Prototype GEA, CDB-1, Linear Blender

A prototype of the commercially available GEA CDB-1 Linear Blender (GEA, Dusseldorf, Germany) was used for these experiments. The blender, made of stainless steel (316L), has an internal radius of 59 mm and a volume of 8.2L at a 15° incline to the horizontal. The equipment is driven by a motor paired with a proximity sensor (to measure the rotation of the impeller's shaft) and feed-back controlled variable frequency drive — allowing RPM to be set and maintained.

The shaft of the mixer has 28 removable collars each with two paddle blades the orientations of which can be adjusted from 0° (parallel to shaft axis) to 90° (orthogonal to shaft). Resulting in the ability, with the adjustable paddle blades, to simply change the angle and position of the collars to create a blade configuration. In this study, the 'number of mixing blades' refers to the helical organisation of 90° (to the axis) radial mixing blades from the centre of the mixer out towards the ends of the blender. In effect, the mixing blades are aligned in a helix along the shaft with 60° rotation between adjacent collars; thus increasing the number of mixing blades increases the size of the mixing zone. Ultimately, allowing the categorical variable (of a set configuration) to be considered continuous (as the number of blades varies the size of the mixing zone). The remaining blades in the configuration are aligned in transport orientation, at 45° to the axis. Fig. 1 displays these orientations and their positions in the configurations investigated; 8 mixing blades (8H) and 16 mixing blades (16H). These configurations are the same as the two configurations used in the PEPT study reported by Jones-salkey et al. (2023b).

Each blade configuration is examined across five blender speeds: 150 rpm, 225 rpm, 300 rpm, 375 rpm and 450 rpm. This choice allows the investigation across the full range of relevant blending speeds, exhibited by the CDB-1. Palmer et al. (2020) investigated in increments of 150 rpm, demonstrating a significant difference in mixing performance. In this work, we have reduced the increments to 75 rpm to provide better resolution and more reliable model fitting. Each configuration runs every material at 10 kg/h and 20 kg/h feed rate. Additionally, 3 materials (which are very commonly used filler excipients) run at 26.25 kg/h and 50 kg/h.

2.2. GEA, Compact Powder Feeder

The powder is fed to the blender using a Loss-In-Weight (LIW) GEA Compact Feeder (GEA, Dusseldorf, Germany). The LIW feeder holds roughly 2L of material and can operate in gravimetric (constant mass/time) and volumetric (constant screw speed, effectively constant volume/time) modes. Furthermore, the feeder is accompanied by a combination of a top-up ball valve, with varying volume inserts, and a 10L chute facilitating automatic system refills.

Gravimetric feeding operates by controlling screw speed to maintain a constant mass/time. Thus, a target feed rate can be maintained in the presence of powder flow disturbances, which may occur due to transient disturbances within the hopper or during powder transfer into the conveying screw volume. Therefore, gravimetric operation should be prioritised and the system's mechanical components should be configured to comfortably achieve the desired mass/time setpoint.

The calibration for the feeder is discussed in the following method Section 4.1. The calibration allows for each experiment to have the highest gravimetric feeding up-time, meaning the feeding of each powder is as accurate and reliable as possible.

2.3. Mettler Toledo, Catch Scale

The XS64000LX Mettler Toledo Balance (Mettler Toledo, Greifensee, Switzerland) was used to monitor the mass flow rate at the blender's outlet, recording mass (g) over time (s). Real-time response from the catch scale is used to monitor the mass flow rate from the outlet to ensure the system is operating at steady-state and to determine the residence mass at the end of the run. The scale is linked to Mettler Toledo's LabX software which records the mass at a frequency of 2 Hz, saving it to an output file. The position of the balance, relative to the

Table 1
Characterisation results for the ten materials and the surrogate formulation.

#) Material	Grade	Abbreviation	Basic Flowability Energy (BFE) [mJ]	Bulk Density [kg/L]	WFA @ 2 kPa, 0.4Ra	1/FFC @ 2 kPa	d10 [μm]	d50 [μm]	d90 [μm]
(1) MCC	Avicel PH-101	PH101	202	0.35	13.4	0.26	26.3	69.4	142
(2) MCC	Avicel PH-102	PH102	191	0.35	15.2	0.15	33.1	113	240
(3) MCC	Avicel PH-105	PH105	81.9	0.37	49.4	0.47	7	21	49
(4) MCC	Vivapur 350	VP350	124	0.86	3.87	0.01	377	432	509
(5) Lactose	Pharmatose 110M	P110M	119	0.73	10.2	0.14	78.9	153	219
(6) Lactose	Pharmatose 200M	P200M	128	0.60	44.5	0.31	21.8	57.8	122
(7) Lactose	Pharmatose 450M	P450M	100	0.47	11.4	0.63	3	21	56
(8) Lactose	Lactose FF316	316FF	274	0.61	36.5	0.08	39.8	92.9	173
(9) Mannitol	Pearlitol 200SD	PT200SD	297	0.54	14.5	0.03	40.3	143	238
(10) DCPA	DCPA A-Tab	ATAB	253	0.74	16.3	0.02	48.2	168	305
(F) Formulation	Surrogate	–	242	0.58	11.6	0.12	23	123	257

system, can be seen in Fig. 1. The fill level is then calculated from the residence mass and the bulk density of that powder, the calculation is discussed in the following Section 4.3.1.

3. Materials & characterisation

3.1. Material list

Nine commonly used commercial excipients were used in this study, comprising: Avicel PH-101, Microcrystalline Cellulose, abbreviated to 'PH-101' (DuPont, Cork, Ireland), Avicel PH-102, Microcrystalline Cellulose, abbreviated to 'PH-102' (DuPont, Cork, Ireland), Avicel PH-105, Microcrystalline Cellulose, abbreviated to 'PH-105' (DuPont, Cork, Ireland), Vivapur 350, Microcrystalline Cellulose, abbreviated to 'VP350' (JRS Pharma, Rosenberg, Germany), Fast Flo 316, Lactose Monohydrate, abbreviated to 'FF316' (Kerry, Tralee, Ireland), Pharmatose 110M, Lactose Monohydrate, abbreviated to 'PH110M' (DFE Pharma, Veghel, Holland), Pharmatose 200M, Lactose Monohydrate, abbreviated to 'PH200M' (DFE Pharma, Veghel, Holland), Pharmatose 450M, Lactose Monohydrate, abbreviated to 'PH450M' (DFE Pharma, Veghel, Holland), A-Tab, Dicalcium Phosphate Anhydrous, abbreviated to 'A-Tab' (Innophos, New Jersey, USA), and Pearlitol 200SD, Mannitol, abbreviated to 'PT200SD' (Roquette, Lestrem, France). In addition, the formulation utilises two additional materials: Powdered Paracetamol, abbreviated to 'pAPAP' (Mallinckrodt, USA) and Gycolys, Sodium starch glycolate (SSG), abbreviated to 'Gycolys' (Roquette, Lestrem, France). Furthermore, the drug surrogate formulation (F) in Table 1, is made up of 4 materials at the following weight percentages: Powdered Paracetamol (27% *w/w*), Avicel PH-102 (41% *w/w*), DCPA A-TAB (27% *w/w*), and Glycolys (5% *w/w*).

The materials list was selected according to use in commercial formulations and their relation to one another, with the intention of filling the characterisation space occupied by typical formulations. The first step was to select the more common formulation materials (i.e. PH102, 316FF and P200M), which had innately different material properties. Secondly, alternative grades of the same material were selected to evaluate the differences between the grades, allowing us to evaluate the effect of particle size distribution on both bulk characteristics and ultimate process outcomes. Lastly, after evaluating the characterisation space, additional materials were selected that either filled the characterisation spaces (e.g. PT200SD), provided an extreme value (e.g. VP350), or would be considered a good filler excipient for CDC formulations (e.g. ATAB), resulting in the final material list for the experiments.

3.2. Material characterisation & methods

The powders in Section 3.1 were subjected to a series of powder characterisation tests to determine intrinsic properties and flow behaviour. These are described below and the results are given in Table 1. All characterisation measurements were repeated in triplicate, from which the mean value is used.

Basic Flowability Energy (BFE) was measured using the Freeman FT4 Powder Rheometer (FT4, Freeman Technology, Tewkesbury, UK). BFE is, effectively, a measure describing the energy required for a stainless steel blade to initiate and sustain the bulk flow of a powder. Powders in the FT4 rheometer are exposed to a comparatively similar form of excitation to those in the CDB-1, suggesting that this measurement method may provide useful predictors of the powders' behaviours in the blender. For more information on the measurement technique see Freeman (2007) and for evaluation of the dynamics within the environment see Hare et al. (2015). In the present work, each powder went under four conditioning cycles and four test cycles.

Bulk density was measured using a 50 ml glass measuring cylinder and scale, where the powder was loosely deposited into the cylinder amounting to the maximum (50 ml) line. The powder mass is recorded and the bulk density is calculated (weight/volume). The samples were prepped by shaking the 250 ml holding vessel, ensuring the powder was not consolidated prior to pouring the powder into the measuring cylinder. Pouring the powder from the holding vessel, the cylinder is then filled to the measuring cylinder's 50 ml volume mark, the mass is recorded and used to calculate the bulk density. After each measurement, the powder was returned to the holding vessel before shaking the powder and taking another measurement. Furthermore, there was no manual intervention, agitation or extraction of powder while filling the cylinder — if the volume was not correctly poured the first time, the measurement was removed and repeated. Lastly, the bulk densities were all measured at the same time, negating the influence of humidity when comparing the results to one another.

Flow Function Coefficient (FFC, (using a consolidation pressure of 2 kPa) and Wall Friction Angle (WFA (0.4 Ra @ 2 kPa)) are two measurements describing powder cohesion and powder adhesion, respectively, characterised using the Schulze Ring Shear Tester (Dr. Dietmar Schulze, Wittenburg, Germany). For additional information on FFC and WFA see both Schwedes and Schulze (1990) and Schulze and Schulze (2021). Additionally, the FFC values were transformed into their reciprocal to linearise the measurement sensitivity. This is because it is difficult to discern differences between high values of FFC compared to low values of FFC.

Particle Size Distribution (PSD) was obtained from dynamic image analysis using the QICPIC Rodos (Sympatec GmbH, Clausthal-Zellerfeld, Germany), resulting in d10, d50 and d90 measurements. The Rodos powder disperser was set to 0.6 bar pressure and the M5 lens was used for all measurements.

It shows that the materials demonstrate a wide range of material properties (seen in Table 1), spanning those that would be appropriate for a typical pharmaceutical formulation (Megarry et al., 2019), offering a variation of suitably significant magnitude to successfully monitor their impact. The BFE can be varied from 81.9 to 297.2 mJ; Bulk Density from 0.346 to 0.86 kg/L; WFA from 3.87 to 49.4°; 1/FFC from 0.01 to 0.47; d10 from 3 to 367 μm; d50 from 21 to 432 μm; d90 from 49 to 509 μm.

VP350 has the highest bulk density, FFC and particle size distribution, and lowest WFA, whereas, Avicel PH-105 has the lowest BFE, lowest bulk density, and smallest particle size distribution, with the highest WFA. This is a rather striking result since both of these materials are different grades of the same material (MCC).

The characterisation techniques used in this study were selected based on their fulfilment of one or more of the following criteria:

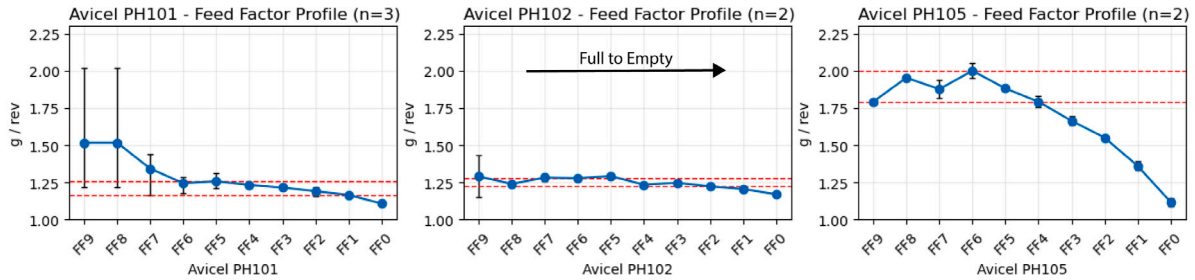


Fig. 2. Three example feed factor profiles, showcasing the profiles for the three different grades of the same material, Avicel (MCC). Error bars indicate the range of values from the duplicate/triplicate calibration runs (indicated by $n =$ in the graph's title). The red dashed lines indicate the feed factor at which the top-up occurred. The following windows were used for the three powders: FF1 to FF4 (0.8L top-up) for Avicel 101, FF2 to FF6 (0.8L top-up) for Avicel 102, and FF4 to FF6 (0.4L top-up) for Avicel 105.

First, they closely resemble the processing environment — for example, the basic flowability energy, BFE, which directly involves the motion of a rotating impeller through a non-compacted particulate medium. Second, they are extensively used in literature to describe a particular behaviour — for example the FFC, which is widely considered to be an accurate measurement technique for quantifying powder cohesion. Third, they are simple and/or widely accessible measurements that succinctly detail bulk powder behaviour, such as bulk density. Furthermore, the fewer measurement techniques used for an accurate prediction, the lower the number of material characterisation experiments required to make a prediction; as such, we have deliberately chosen parameters which encode a considerable amount of information about the system into a single parameter. The result is a set of measurements which can be performed with relative speed and ease, and require minimal use of proprietary equipment.

It is finally worth briefly noting that while for the comparatively free-flowing powders and formulations explored here, a simple poured bulk density measurement is found to be sufficient when exploring more cohesive powders, it may be wise to instead use the FT4's conditioned bulk density measurement, so as to ensure a more reproducible and representative measurement. For our materials, the two were found to be equivalent for all materials tested.

4. Methods

4.1. Feeder calibration

Before the feeder is used to take a measurement, it must be calibrated to ensure the system configuration can accurately dose the desired gravimetric (mass/time) feed rate. Therefore, duplicate/triplicate calibration runs are used to determine the feed factor profile and the hopper top-up strategy.

Feed factor is defined as the mass delivered (from the feeder spout) per screw revolution (g/rev). Due to the differences in the physical properties of the materials used, the loading of the conveying screw and behaviour within the hopper will vary with each material, so each one must be calibrated prior to each investigation.

The feed factor profile is gained through GEA's calibration method, which effectively measures the feed factors whilst emptying the hopper from full with volumetric feeding. The hopper is separated into 10% fill level intervals (F_n), the mass fill of the hopper is then divided into 10 slices in the vertical direction, and the average feed factor within each of those intervals is stored against that hopper fill level. The resultant profile (see Fig. 2) provides insight into the stability of powder dosing, at constant screw speed, with the expectation that free-flowing powders will have minimal change in feed factor across the full range of fill levels. Had a material being discharged suffered from disturbances to feeding, changes to either the gear ratio or screw type/pitch would have been tested to ensure a more consistent feed. Fortunately, the calibration was assessed for all materials and found that the same gearbox ratio of 63:1 and a pair of conveying screws with a concave pitch of 10 mm width were appropriate for all tests.

Lastly, the feed factors are also used to define the stable operating region for the hopper fill level for triggering a hopper top-up event. This was done by plotting mean feed factors from F_9 (full hopper) to F_0 (empty hopper) (100% into 10% intervals, respectively) and identifying the most stable region during discharge. From this, the time and volume of the top-up event were defined, for each material; this event was to be initiated during the stable period of the volumetric emptying and the volume of top-up was selected to be the total volume dispensed over the identified stable region. This was then validated and optimised to ensure the automatic refills occurred with the largest proportion of gravimetric feeding to volumetric feeding. For three different grades of the same powder, the refill window is indicated by red dashed lines in Fig. 2.

This methodology ensures that the system operates with the utmost accuracy and reliability. The stable operating region, describes the operating window that is most resistant to feeding variance under volumetric operation, signifying that the feeding is the least sensitive to the fill level within the hopper. This is because the powder above compresses the powder below, as explained in Engisch and Muzzio (2015). This means that during a top-up event, when the feeder switches from gravimetric to volumetric, the feeder will show a consistent and accurate feed in grams per revolution, which is indicated by the aforementioned feed factor profile.

4.2. Residence mass (RM) measurements

Residence mass (RM) is defined as the mass of powder within the blender's mixing volume during steady-state operation. Several studies discuss the importance of residence mass for improving the macro mixing potential of pharmaceutical blenders (Portillo et al., 2007, 2010; Van Snick et al., 2017). Notably, it is a measurement that can be directly, and physically gathered, without the use of sensors, detectors, or imaging devices. Furthermore, RM's utility is really described by its sensitivity to other processing parameters and material properties. For instance, increasing RPM decreases the RM (discussed in Section 5.1) and allows for the calculation of Mean Residence Time (using Eq. (1)). The result is a great variable for evaluating mixing due to the coupled attachment of material, RPM and MRT — two of the three variables are required to evaluate strain discussed in Section 4.3.2, using Eq. (3).

$$\text{Mean Residence Time (s)} = \frac{\text{Residence Mass (g)}}{\text{Feed Rate (g/s)}} \tag{1}$$

The experiment begins with setting the blender to the desired RPM for the experiment. Once the blender is at speed, the LabX software for the catch scale is then set to record the mass (over time) from the outlet of the blender. Simultaneously, the GEA Compact Feeder communicates with TwinCAT 2 software (Beckhoff, Henley-on-Thames, United Kingdom) through tags appropriate for operation e.g. weight in the hopper, gravimetric mode, and screw speed. After initiating the recording of data, for both dispensing and receiving through the catch scale, the feeder is then set to dispense gravimetrically at the desired setpoint for the experiment.

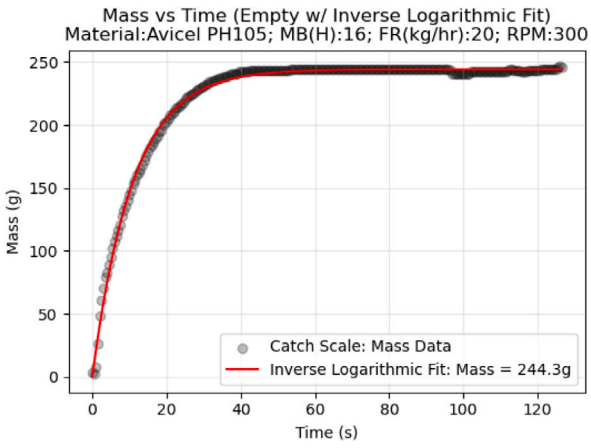


Fig. 3. Example of a '450 RPM Empty', where the offset (remaining mass in the blender) would then be added to the asymptote collected from the curve fitting to gain the final residence mass at steady state.

The powder is dispensed into the blender via the feeder. Steady state is then monitored and confirmed when the dispensing mass flow rate matches the mass flow rate received on the catch scale. While building mass to steady state, RPM is maintained by the feedback loop from the proximity sensor and variable frequency drive. Once steady state has been achieved all processing equipment is stopped simultaneously- and the residence mass within the blender is measured.

To measure the mass, in the most time-effective manner, the system is powered up to the fastest speed (450 RPM) with no powder flow into the system, and the resident mass is captured in the catch scale — demonstrated using Fig. 3. The result is mass removal which is a function of the RPM used.

It is known that 450 rpm, with no feeding, would not fully empty the blender so it was necessary to determine the residual mass. So it was assumed, and then checked, that this residual mass was independent of the speed history. Thus the residual mass was collected at the end of a run of five graduated rpm speeds (150 to 450 RPM) and treated as an 'offset', where it was manually added to the '450 RPM empty' residence masses. This was done for each of the feed rates and the blade configurations, despite the minimal difference being seen between changes in feed rate.

After a "450 RPM Empty" was done, the next RPM, with the same blade configuration and feed rate could then be done. Given the mass amounts are time-dependent, in particular the 450 RPM Empty, the removal of mass over time to the catch scale was recorded and the run stopped when the mass amount remained stable — see Fig. 3. To ensure each of these emptying processes was consistent an inverse log curve was fitted to the '450 RPM empty' residence masses, ensuring we could gather the mass at $t = \infty$. Ultimately, this meant that the residence mass values saw minor changes (in the order of a few grams), but the application of the same 'time-independent' methodology meant that the masses collected were consistently gathered.

The response, RM, serves as the backbone of this experiment, however, the response is best described as fill level (%), which can calculated from RM using Eq. (2).

Using fill level, instead of residence mass, results in simpler contextualisation when evaluating the powder relative to the blender's total volume. Providing a better sense of clarity when comparing different powders against one another, or considering powder behaviour in different blender systems/geometries.

In the discussion (Section 6.2) fill level is used, in place of RM, when evaluating the optimum process parameters to find the ideal regime behaviour across different materials.

4.3. Calculations

4.3.1. Fill level

Fill level [%], is defined as the percentage volume the powder occupies within the mixing volume.

$$\text{Fill Level (\%)} = \frac{(\text{Residence Mass}/\text{Bulk Density})}{\text{Mixer Volume}} \tag{2}$$

4.3.2. Strain

Strain (ω) can be considered as a measure of 'work done' to the powder. In the case of a continuous blender, it can be defined as the mean number of blade passes the powder flux would experience (Vanarase and Muzzio, 2011).

$$\text{Strain } (\omega) = \frac{FL \times \rho \times v_{blender}}{\dot{m}} \times \text{RPM} \tag{3}$$

The influence of the mixing blades is implicitly expressed in this equation due to how the mixing blades alter the fill volume and thus the mean residence time of the powder at a steady state. where RPM is the rotation rate of the agitator in rotations per minute; $\dot{m}$ is the feed rate in kg/h; FL is the fill level (as a fraction) of the powder at steady state; $v_{blender}$ is the volume of the blender in m^3; and ρ is the bulk density of the powder in kg/m^3.

4.4. Predictive modelling

4.4.1. Random forest regression

Despite the numerous insightful findings drawn from various studies using methods such as Partial Least Squares (PLS), Bekaert et al. (2022c), Dhondt et al. (2022), Vanarase et al. (2013), Bekaert et al. (2022a) the method presents a few key constraints. For instance, PLS is inherently linear and tends to obscure the intricate interplay of variables behind its principal components. In contrast, Random Forest Regression (RFR) offers the ability to capture non-linear relationships, providing a descriptive, 'logic-like' portrayal of the variables' interaction. When the primary focus is comprehending the system from both a material and processing perspective, it is crucial to understand these interactions and their resulting sensitivities.

RFR is a machine learning algorithm which combines the predictions of multiple decision trees to produce a more accurate and robust model. In this context, x represents an individual input data point (or a set of feature values) for which we want to predict an output, while y is the actual output or response value corresponding to that input. Hence, for each instance of x, we are trying to predict its corresponding y value. This can simply be depicted as multiple trees each making a prediction for a single x, and then aggregating those predictions to get $\hat{y}(x)$.

In this formulation, $\hat{y}(x)$ denotes the predicted value for the dependent variable based on the input x, aggregating the insights from all n decision trees. Mathematically, given n decision trees, the prediction of RFR for an instance x is given by:

$$\hat{y}(x) = \frac{1}{n}\sum_{i=1}^{n} T_i(x) \tag{4}$$

In the context of the equation, $T_i(x)$ represents the prediction of the i^{th} tree. Each tree is sculpted using bootstrapped data subsets, a technique where random samples are drawn with replacements from the original dataset, allowing some observations to be repeated in each sample. At every decision point, whether it is a split or a branching, a random selection of features (variables) is integrated to ensure that the trees are decor related and there is reduced inter-tree correlation. This inherent randomness can be observed in both the data samples and feature subsets. These techniques enhance the model's generalisation potential, making it more robust and accurate.

To train the model, the entire dataset was randomly split into a train:test split of 80:20 (203:51, datapoints respectively) and training

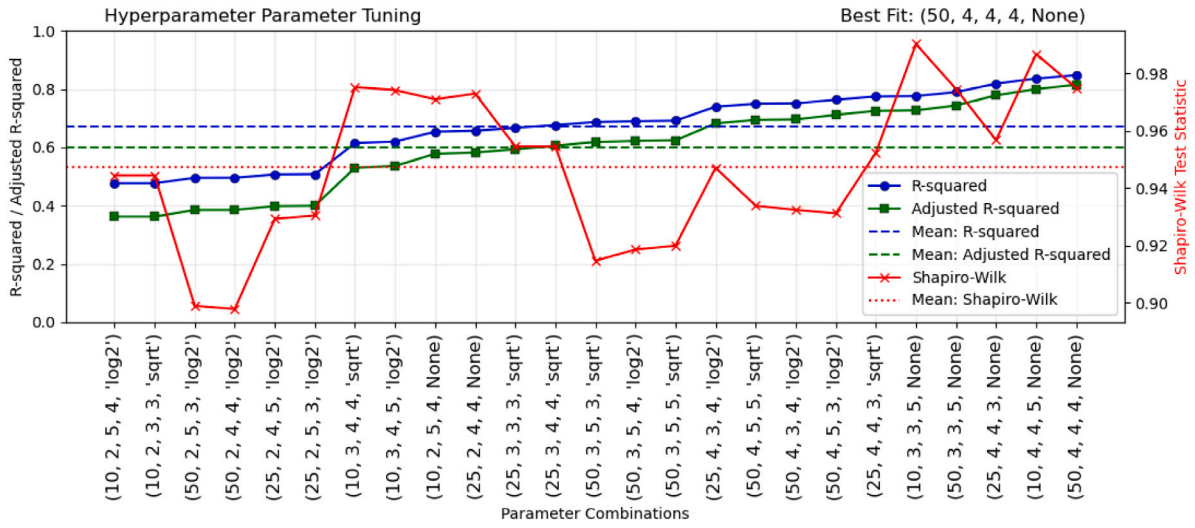


Fig. 4. Random Forest Regression (RFR) model parameter tuning. Results were ordered from worst (left) to best (right) r^2 value, with the secondary axis plotting the assessment of the residual's normal distribution using the Shapiro–Wilk Test (SWT)– higher SWT scores indicate a normal distribution. The x-axis shows the combination of the following 5 hyperparameters (in respective order): number of estimators, maximum depth, minimum number of samples to split, minimum number of samples for a leaf and maximum features.

data passed into the tuning stage of modelling. The RFR model was tuned by training approximately 240 models with different combinations of 'sensible' hyperparameter values: the number of estimators, max depth range, minimum samples to split, minimum samples for leaf, and maximum features range (for additional information on RFR see Sci-Kit's documentation: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html).

To clarify the function of the aforementioned (RFR) hyperparameters, the following explanation(s) is provided:

- **Number of Estimators**: Specifies the total number of decision trees within the forest. Increasing the number of trees can improve the model's performance, but there is a threshold beyond which accuracy gains are marginal, trading-off for an increased computational cost.
- **Max Depth Range**: Designates the maximum depth or levels each tree can achieve. Deeper trees can discern more intricate data relationships but risk overfitting. An optimal depth strikes a balance between bias and variance.
- **Minimum Samples to Split**: Specifies the minimum number of samples required for a decision tree split. Increasing this value can prevent the model from incorporating small patterns that may be noise, which serves as a regularisation technique.
- **Minimum Samples for Leaf**: Specifies the minimum number of samples required to be at a leaf node (end of tree branch). Increasing this value, like the prior hyperparameter, can prevent overfitting and enhance generalisation across training examples.
- **Maximum Features Range**: During optimal split determination, the number of features to consider is specified. Random feature selection for each split promotes tree decorrelation and model generalisation.

Sensibility – which is applicable to the rest of the study – is defined as the thought experiments that align the complexity of the model to the size and complexity of the dataset one is looking to model. This means that if one has a large and complex dataset, one could approach it with a complex multi-layered model, with values that will de-convolute the complexity found within the data. Alternatively, with a simpler dataset, utilising the same modelling framework or hyperparameters, one would be at risk of overfitting the data — rendering the model impractical. Therefore, thought is put into the hyperparameter selection to ensure the risk of overfitting is minimal, with the tuning operating at a 'sensible' range. The results from the initial 'sensible' hyperparameters provide feedback to this initial tuning phase and through iteration, the model can be refined to generate a suitable and robust model.

The resultant fitted models were then rank-ordered for r^2 (and adjusted) and plotted alongside the residual's frequency distribution, which is then quantified by the Shapiro–Wilk test (SWT); The Shapiro–Wilk test (SWT) is a statistical analytical test which checks whether a sample comes from a normally distributed population. The null hypothesis (1) assumes that the data is normally distributed, while the alternative hypothesis (0) assumes otherwise. A normal distribution would indicate a lesser propensity for the model to under- or over-predict values, while a narrow distribution indicates lesser error between the measured and predicted values; both a normal and narrow distribution are thus desirable.

The final selected model (Fig. 5), which held the following hyperparameter values: number of estimators (50), max depth range (4), minimum samples to split (4), minimum samples for leaf (4), and maximum features range (None), possessed a high r-squared value of 0.85 (showing a good fit) whilst demonstrating a good normal distribution of the residuals (quantified by the SWT, which was 0.97) — see Fig. 4. Any additional hyperparameters were held at SciKit Learn's default values.

The RFR model's performance is shown in Fig. 5, by the Parity Plot and Feature Importance graph. The Parity plot is the use of the model to predict the 'test' data, which is not used to train the model and shows a good r^2 fit of 0.848. However, there is considerable spread in places, leading to higher RMSE values — namely the cluster of points that are over-predicting at 10% fill. Furthermore, there is greater variability at the higher fill levels but this is potentially due to increased sparsity of data points at higher percentages. This sparsity at these higher fills, is a feature of the experiment, due to RPM highly dominating the system, irrespective of material or alternative process parameters (discussed in Section 5.1); the result is a higher number of points around lower fill levels due to increased RPMs.

The generated data from the full experimental dataset indicates that most operations should be at or below the 35% fill mark. However, further experimentation within these higher fill levels, through both material and process changes (e.g. higher feed rates), would provide the model with greater reasoning, and therefore, the improved capability to predict these fills.

The model was intentionally limited, in terms of accurate predictive capacity, to be able to derive the 'logic' from the system. In this simplified model, only the most important features will be shown, with the most contrasting changes between turning at each feature. This is

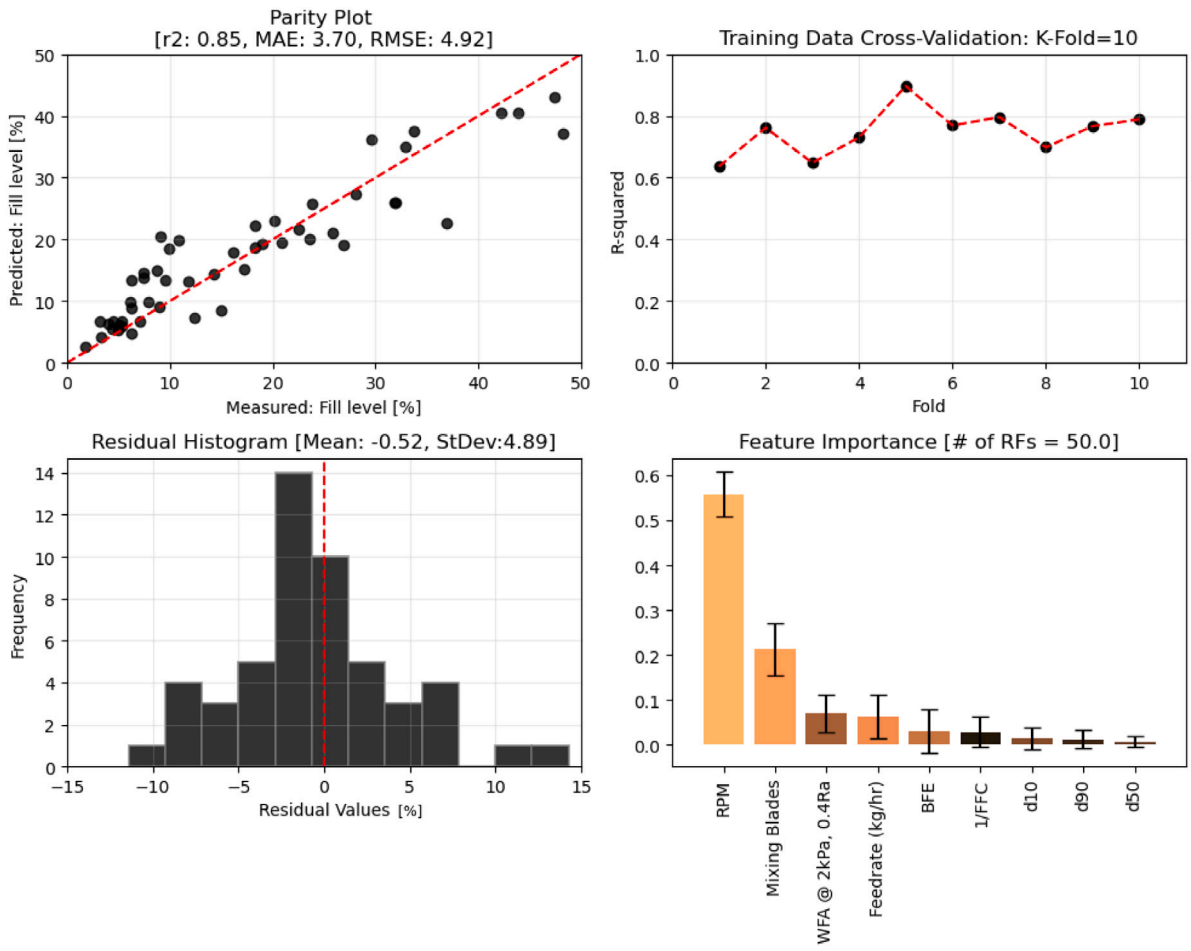


Fig. 5. Random Forest Regression (RFR) model performance. The top left graph shows a parity plot of the measured vs predicted fill level. Top right, training data (only) k-fold cross-validation. Bottom left, RFR's residual histogram. Finally, the bottom right shows the Feature Importance (FI) of each of the parameters on the RFR's predictive capability, higher indicates more importance. The error bars show the standard deviation of the FI across the $n = 20$ generated estimators.

done in contrast to having fewer samples to split and minimum samples for leaf – which would result in more and longer branches to reach the answer – giving rise to less granular predictions. Accordingly, these trees with higher complexity result in a cumbersome tree with many turns, inevitably resulting in the loss of 'what factors *really* matter'.

4.4.2. Artificial Neural Network (PyTorch)

Artificial Neural Networks (ANN or NN) have demonstrated a great ability to make predictions from data on a wide range of applications. The following ANN model was developed from the guidance provided by Goodfellow et al. (2016) comprehensive book on machine learning (ML); the book covers foundational elements of ML in detail, with Chapters 5 and 7 playing a pivotal role in this model's creation.

ANNs function by propagating input data through layers of artificial neurons. These neurons apply a linear transformation followed by a non-linear activation function. Mathematically, each layer's output, h, is computed as:

$$h = \sigma(W x + b) \tag{5}$$

In Eq. (5), σ represents the activation function empowering the model to capture intricate relationships between inputs and outputs. The term W stands for the weights, which dictate the significance of each input. The bias b adjusts the neuron's output, while x is the input to the neuron. As data is fed into the network, it undergoes transformations by each layer. The process, known as forward propagation, continues until the final layer produces the network's output or prediction. The connections between neurons, characterised by the weights and biases, determine the strength and nature of the relationships the network captures. However, after a single forward propagation, the network's predictions may not be accurate. This is where *backpropagation* comes into play. Backpropagation is an optimisation algorithm used for minimising the error in the predictions. It works by computing the gradient of the loss function with respect to each weight by the chain rule and iteratively adjusting the weights and biases in the direction that reduces the error. An *epoch* refers to one complete forward and backward pass of all the training examples. By running multiple epochs, the network continues to refine its weights and biases, aiming to enhance its prediction accuracy over time.

ANNs are complex systems that require iteration to configure and fine-tune model performance. The following hyperparameters (used for the model) are detailed and described below:

- **Architecture**:
  This study employs PyTorch to construct an Artificial Neural Network. Given the dataset's size, this architecture was chosen to be relatively simple to prevent overfitting, which occurs when the model becomes too closely tailored to the training data and performs poorly on unseen data.
  - **Input Layer**: Comprises 9 nodes, corresponding to the number of features in the dataset.
  - **Hidden Layers**: Two layers with 41 and 14 nodes respectively. These layers process and transform the input data.
  - **Output Layer**: A single-node layer, providing the final prediction.

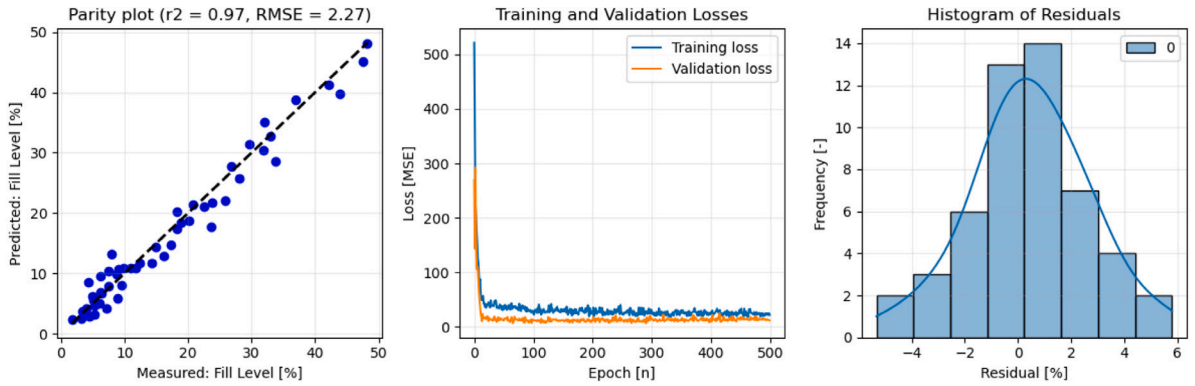


Fig. 6. ANN training and performance: Left, parity plot of the measured vs. predicted fill level percentage. Centre, training and validation losses (in MSE (Mean Squared Error)) of the model vs. the number of epochs. Right, histogram of the residuals (measured-predicted). The relative root mean squared error was 12%.

- **Hidden Layer Nodes (Linear)**: Each node applies a linear transformation to the data, based on the weights and biases. Using the following equation: $h = \sigma(Wx + b)$. These nodes within the network form the basis of the prediction.
- **Activation Function (ReLU)**: Defined as $\sigma(z) = \max(0, z)$. It introduces non-linearity into the network, allowing the model to understand and represent complex relationships in the data.
- **Dropout Rate (0.0155)**: A regularisation method where a fraction of neurons are randomly deactivated during each training iteration. This helps in preventing the model from over-relying on specific neurons and mitigates overfitting.
- **Learning Rate (0.0185)**: Determines the magnitude of adjustments made to the network's weights during training. A smaller learning rate means slower convergence but possibly more precise outcomes, while a larger rate could speed up the training but risk overshooting the optimal solution.
- **Epochs (500)**: Refers to the number of times the entire dataset is presented to the network during training. More epochs can allow for better learning, up to a point, after which overfitting might occur.
- **Batch Size (32)**: The number of data samples processed before the model's internal parameters are updated. Using batches is more memory-efficient than using the entire dataset.
- **Optimiser (ADAM)**: An optimisation algorithm that adjusts the weights of the network. ADAM is known for its adaptive properties, where individual learning rates are computed for each parameter.
- **Loss Criterion (Mean Squared Error (MSE))**: Measures the squared differences between the predicted and actual values. The goal during training is to minimise this value.
- **Data Normalisation (MinMaxScaler)**: Transforms features to lie in a range between 0 and 1. This ensures that no particular feature dominates the learning process due to its scale, leading to more balanced and efficient training.
- **Other Hyperparameters**: Any parameters not explicitly mentioned were kept at PyTorch's default settings to maintain stability and general best practices in training.

The Python package OpTuna was utilised to methodically search for the best combination of two vital parameters: learning rate and dropout rate. The learning rate determines how quickly the model updates its weights in response to the calculated error, while the dropout rate is a technique where certain neurons are randomly "dropped" or deactivated during training to enhance model robustness. OpTuna's search spanned across 100 combinations, bounded within $1e^{-5}$ to $1e^{-1}$ for the learning rate and 0.0 to 0.5 for the dropout rate. The performance of the model was gauged through a combination of metrics, visualised in Fig. 6.

First, a key indicator of a good model includes a steady decrease in the training and validation loss (centre), indicating that the model is learning, and a close match between these two loss curves, suggesting minimal overfitting. The training loss provides a measure of the model's error on the data it is actively learning from, effectively the loss seen at its current epoch. The validation loss measures the error on a separate subset of data not used during training (unseen data at the current epoch), offering insights into the model's ability to generalise to new, unseen data.

Second, once the model has been trained on the 'training data' (80% of the dataset) the model is asked to predict the remaining 'test' (20%) data, the result is the parity plot (left) — which further validates/indicates the model's ability to predict unseen data. The fit of the model can be quantified by an RMSE of 2.27 and a high r^2, which in this case is exceptional with 0.97.

Third, a residual histogram plots the distribution of the actual data minus the predicted data. A narrow and normal distribution here would conclude that the model has minimal bias under or over-predicting across the range of predicted test values. In addition, the frequency distribution of the residuals is greatly improved compared to the RFR's.

Lastly, the developed ANN was investigated using SHAP (SHapley Additive exPlanations) analysis, a method rooted in game theory (Lundberg and Lee, 2017); further details and documentation can be found at https://github.com/slundberg/shap. Effectively SHAP treats each discrete data point as a player and allows the 'game' to unfold to determine which features are the most 'useful' for predicting the response variable. The resultant is a graph (Fig. 7) similar to that found in the RFR's Feature Importance (Fig. 5).

The top-performing features match in the same rank order for both models; highlighting the reliability of these factors in the prediction of fill level. The remaining features differ in rank order suggesting the models utilise the remainder of the variables differently. Alternatively, it suggests that the variables (and/or the differences between them) are statistically insignificant compared to the important ones.

4.5. Symbolic regression — M^2E^3D

Multiphase Materials Exploration via Evolutionary Equation Discovery — or M^2E^3D, further abbreviated simply to MED, is a data-driven tool for discovering analytical equations modelling the underpinning physics or correlations of industrial systems that are too complex to be modelled from first principles (Nicusan and Windows-Yule, 2022a,b). The core algorithm of M^2E^3D is its symbolic regression engine, the *SymbolicRegression.jl* library written in the Julia programming language (Cranmer et al., 2020; Bezanson et al., 2017), which discovers analytical expressions f mapping given input values X to a measured output $y = f(X)$; the general algorithmic steps are:

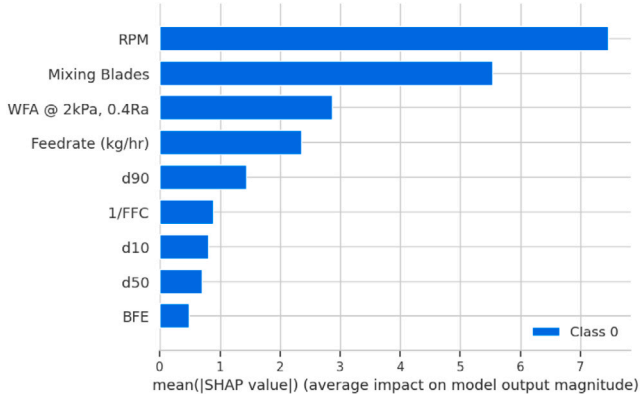


Fig. 7. ANN SHAP analysis, showing the mean of the modulus SHAP Values. Effectively, the average impact of each of the variables on the output magnitude. This is then ranked-ordered from the most influential feature to the least. The scale quantifies the relative importance of the variables.

1. The user selects the operators available to the symbolic regression engine, including common binary operators (addition, multiplication, etc.) and unary operators (log, sin, etc.) or custom functions — allowing users to incorporate some domain knowledge into the equation search, e.g. if the process outputs are modelled against system geometry, trigonometric functions might be included.
2. The operators, along with inputs and fitting coefficients are combined in a tree to form equations (see Fig. 8).
3. The free coefficients are fitted against the measured data y and the residuals are computed $f(X) - y$.
4. An evolutionary algorithm is used to minimise both residuals **and equation complexity** (the latter measured simply as the number of branches in the equation tree) - this multi-objective optimisation problem is necessary to discover sensible, physical, interpretable equations rather than overfitted chains of functions.
5. The process of equation generation - fitting - evaluating - evolving is continued until the best-performing equations at each level of complexity are found to stabilise (i.e. not change between generations) - and hence a Pareto front of equation complexity-accuracy is found.

The main advantages of the technique include:

- **Interpretability**: the equations found give insight into the underlying physics governing the system, while coefficients bear physical units which, again, allow physical interpretation of the correlations found.
- **Robustness against noise**: as equation complexity is also minimised, the most persistent underlying physical terms are identified, naturally removing measurement noise.
- **Requiring few measurements**: if the system is governed by underlying physical laws whose inputs are included in the training data, identifying them requires $O(10)$ data points, which is orders of magnitude lower than what surrogate models and neural networks typically require.
- **Ability to solve the inverse problem**: oftentimes it is not just a predictive model that is desired, but the ability to dynamically change a process input to achieve a fixed output (e.g. fill level); this is possible by rearranging the equations found and can be integrated in continuous on-line process control schemes.
- **Allowing some degree of extrapolation**: unlike virtually all surrogate modelling approaches, if an underlying physical law is identified (e.g. Stokes' law in fluid flow), it can be extrapolated beyond the input training data range to a point (e.g. in the previous example, until turbulence becomes significant).

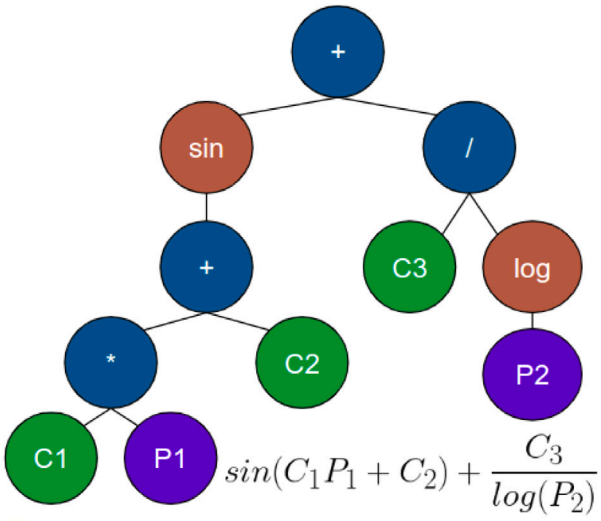


Fig. 8. M^2E^3D representation of an example analytical model (bottom right), built as a tree with its nodes (circles) representing input parameters (purple), constants which need fitting (green), binary operators (blue) and unary operators (orange).

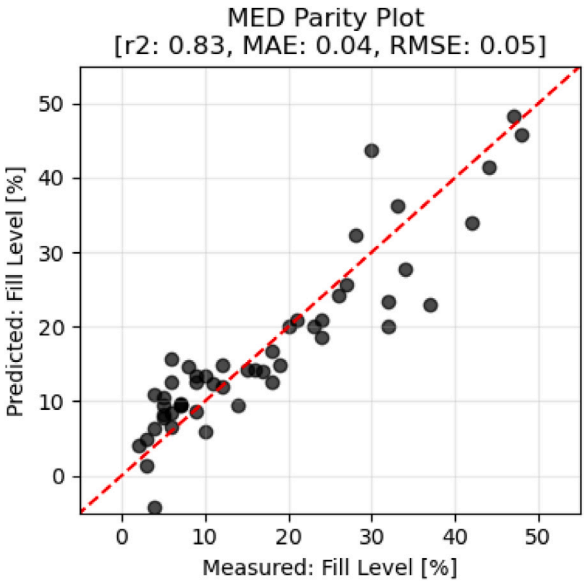


Fig. 9. M^2E^3D Parity Plot, measured vs predicted.

Both the algorithm and the codebase for M^2E^3D can be found on GitHub: https://github.com/uob-positron-imaging-centre/MED)

M^2E^3D has been trained on the same 80:20 train-test split data as the other AI/ML methods and produced an equation (Eq. (6)) whose specific form and interpretation are discussed below. The equation was then used to predict the test data and produced the parity plot shown in Fig. 9. It is recognised that there is a single prediction with a negative fill level, this would be considered erroneous and scrapped (since a negative fill level is impossible), but displays the potential limitation of selecting an equation which is not complex enough. M^2E^3D could also be told to use different or more complex operators and re-trained, this may improve the tool's ability to generate more suitable equations — for the training of this model, it was limited to binary operators.

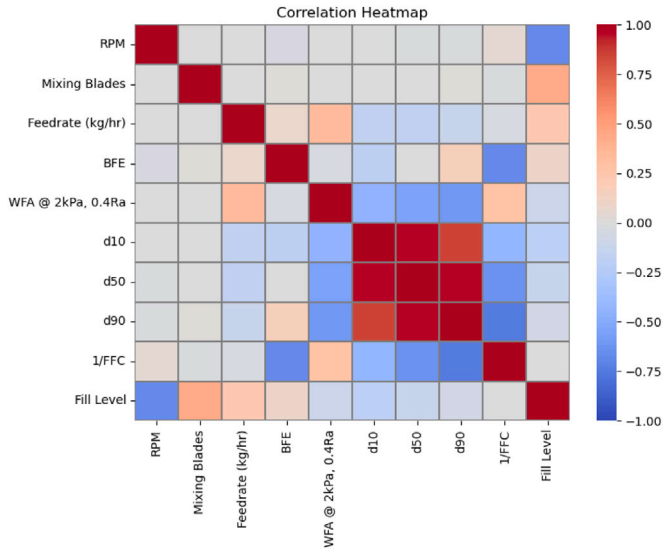


Fig. 10. Correlation Heatmap (−1 to +1) between process parameters, material properties and response variable (fill level). The colour bar indicates the positive correlation (+1, Red) and negative correlation (−1, Blue).

5. Results

5.1. General data trends

The correlation between the response and variables (normalised between +1 and −1) was calculated and plotted as a heat map 10. The correlations between the response variable, Fill Level, and the other variables highlight the effect that variables have on the response. For instance, it shows that RPM is highly negatively correlated with Fill Level, whereas Mixing Blades are moderately positively correlated with Fill Level. It is of particular interest how the material properties are related to the Fill Level — but there are no correlations of a moderate magnitude that point to a specific characteristic being a significant influence.

Plotting the fill level against RPM (shown in Fig. 11) produces a clear, negative linear trend, and in some cases resembles a reverse Sigmoidal curve. Fig. 11, shows the fill level response against the three processing parameters: RPM, Mixing Blades, and Feed Rate. As mentioned, RPM dominates the system, reducing the fill level, irrespective of both material properties and mixing blade configuration. A higher feed rate sees a translation upwards, effectively increasing the fill level – but not necessarily changing the shape of the curve – this is seen to have an increased effect for a higher mixing blade configuration (16H). Perhaps this is due to a decrease in axial transport in the centre of the mixing volume, and so there is less direct transport of the powder from the inlet to the outlet. Moreover, increasing the higher mixing blade configuration, resulted in higher fill levels across the rpm range but had a lesser effect at higher RPMs — again suggesting RPM is the most influential parameter on fill level. These observations agree well with prior experimental and numerical studies (Portillo et al., 2008, 2010; Vanarase and Muzzio, 2011; Van Snick et al., 2017; Palmer et al., 2020; Zheng et al., 2022).

5.2. Model: Random forest regression

Evaluating the general data trends with a single material property provides no significant distinction or correlation. This may be due to the difficulty of discerning a true pattern with a single bulk characteristic — as it would be considered too granular. Therefore, we can better understand their behaviour by modelling how these bulk characteristics interact with other material characteristics or process parameters.

The decision tree (35 of 50), seen in Fig. 12, showcases an example of the logic output. The tree can be interpreted by moving down the tree from the initial top node. Each node then passes another if statement, questioning *if* the assessed sample possesses values greater or less than what the node is asking, which will decide the following node. The final box (called a *'leaf'*) contains the predicted (value) Fill Level %, for the specific path of logic gates used to reach the leaf.

The results align with Section 5.1, insofar as saying that both RPM and Mixing Blades show great significance on the final fill level as these parameters dictate the breadth of the fill levels, sitting at Tier 1 and 2 of the tree. Interestingly, BFE, WFA, d10 and RPM show on the 3rd tier but these are shown to be dependent on the fill level at this point. Furthermore, this is reflected in the initial correlation plot (Fig. 10), meaning that at higher fill levels BFE and WFA are shown to be factors to make a further distinction on the fill level.

Furthermore, when evaluating the positional importance of a node it is also important to consider the colour of the node — which highlights the mean fill level of all samples at the current node. It is shown at higher fill levels (where motion is likely more frictionally dominated), particle physical properties influence the prediction, demonstrated by the position and colour association of BFE, WFA and FFC. Whereas at lower fill levels (where motion is more likely collisionally dominated) the rotation rate seems to dominate; intuitively described by the increased influence of processing parameters due to the reduction of powder volume and the resultant increase in both the frequency and intensity of particle–wall collision and reduction of particle–particle contacts.

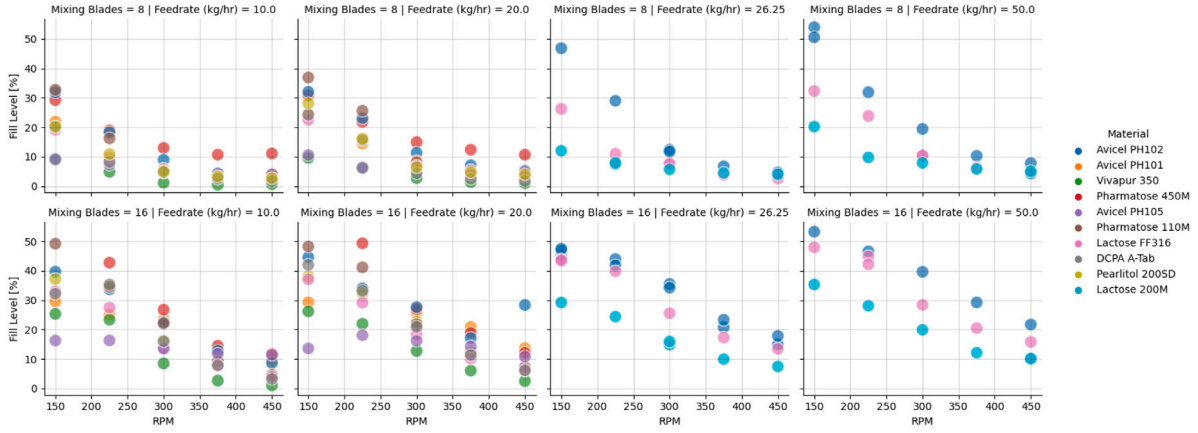


Fig. 11. Scatter graphs showing the fill level response of the materials used for the experiment with respect to the three processing parameters: RPM, Mixing Blades, and Feedrate (kg/h).

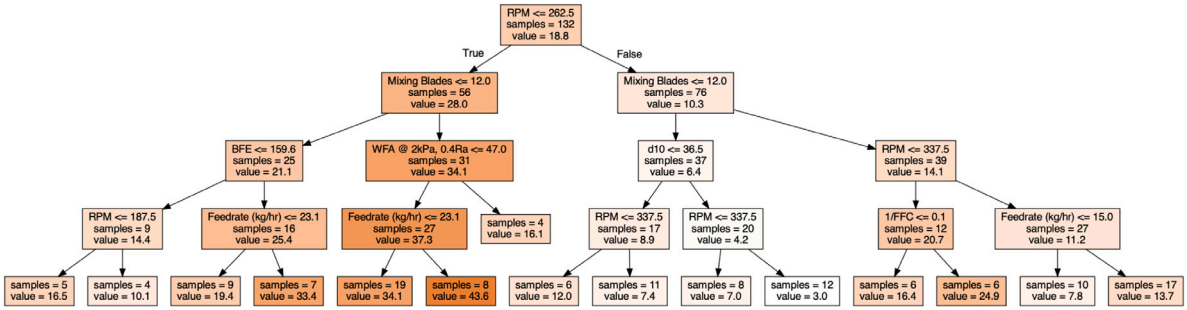


Fig. 12. Decision Tree (35 of 50) from the Random Forest Regression Model, with darker colours indicating high Fill Level. Samples refer to the number of data points, and value is the mean Fill Level for those samples – in percentage (%) – for that node.

5.3. Model: PyTorch ANN feature analysis & contour mapping

The wall friction angle (WFA @ 2 kPa, 0.4 Ra) is shown to be the material characteristic with the greatest feature importance in the model, which holds true for both ANN and RFR models. The practical application of understanding this is formulation adjustment. Different excipients can be paired with the existing blend or powder, in order to modify the final powder characteristic. The result of this allows a quality-by-design approach, as the prediction allows for the understanding of how sensitive a particular property is to processing parameters. This is especially important when building in robustness and early mitigation of risks associated with potential batch-to-batch variation. This operational sensitivity is best shown in Fig. 13, where the two highest-ranking operational parameters have been plotted against the range of WFA values used to train the model. Thus, the sensitivity of the material property can be effectively gauged.

The model shows that the effect of WFA on fill level is, generally, lower at rpms greater than 300 — agreeing with the RFR decision tree (Fig. 12). At these RPMs, the fill level is lower, typically ranging from 6%–20%, which complements the previous discussion point (from the previous section) on collision-based bulk motion/mixing. The mixing blade configuration comes into play by holding up more mass, promoting a higher fill level. This interaction, which promotes more 'frictional' mixing, results in a more complex interaction between the WFA and the discussed process parameters. This is best demonstrated by the change of shape and colour across the three contour plots.

The application of this knowledge in a practical setting suggests that: if there were (for example) a perturbation from the feeder and thus a slight transient change in the formulation's WFA (due to compositional change), this combination of parameters would provide greater confidence that the fill level (and thus the desired process outcome) will be preserved. This methodology and philosophy build in key quality by design aspects to harmonise formulation development and operational robustness.

5.4. Model: M^2E^3D's symbolic regression

While ANNs and random forest regressors can be interrogated to some extent – as has been demonstrated in the preceding sections – they remain, to some extent 'black boxes'. M^2E^3D's methodology (discussed in Section 4.5), meanwhile, produces complete, closed-form equations to describe the behaviours of the systems to which it is applied. For the present data set, the M^2E^3D equation which provided the highest 'score' was the following:

$$\text{Fill Level} = 100 \times \left(0.0032 \times \text{Feed Rate} - 0.003 \times \text{WFA} - 0.0004 \times \text{d10} + \frac{4.284 \times \text{Mixing Blades}}{\text{RPM}} \right) \quad (6)$$

The linear equation can also be viewed as separate terms (seen in (8)), where the following terms can be used as descriptors for their relation to the fill level. Furthermore, these equation terms can be evaluated using an ANOVA. To do this, the M^2E^3D Equation's terms were gathered and fit into an OLS (ordinary least squares) model before assessing the statistical significance of each of the terms.

Table 2
ANOVA Table for the generated M^2E^3D equation (see Eq. (6)) based on the split terms (See Eq. (7) and Eq. (8)), showing the sum squared variance (sum_sq), degrees of freedom (df), F-Statistic (F) and P-Statistic (PR(>F)).

	sum_sq	df	F	PR(>F)
Term1	0.38	1	7.1e+15	0.00e+00
Term2	0.35	1	6.8e+15	0.00e+00
Term3	0.29	1	6.2e+15	0.00e+00
Term4	2.96	1	2.0e+16	0.00e+00
Residual	0.00	249	nan	nan

The ANOVA results can be found in Table 2. Each term was found to have significance in predicting fill level, which was expected as M^2E^3D ensured that every term was significant. However, by assessing the sum squared variance of the prediction based on each term's sensitivity – quantified as the change in the predicted fill level in response to a standardised perturbation, such as a 1% or 10% increase from the mean value of each term – we can determine their relative importance in predicting fill level. The rank order is as follows: $t4 \gg t1 > t2 > t3$. This indicates that the fourth term is responsible for the largest range of variability in the prediction. These findings align with the other modelling (in Sections 5.2 and 5.3) methods as they show RPM and Mixing Blades to be the most important features for predicting.

$$\text{'}Predicted = 100 \times (t1 + t2 + t3 + t4)\text{'} \quad (7)$$

$$\begin{aligned} &Where: \\ &[Predicted] \rightarrow \text{Fill Level} \\ &[t1] \rightarrow (0.0032 \times \text{Feed Rate}) \\ &[t2] \rightarrow (-0.003 \times \text{WFA}) \\ &[t3] \rightarrow (-0.0004 \times \text{d10}) \\ &[t4] \rightarrow \left(\frac{4.284 \times \text{Mixing Blades}}{\text{RPM}} \right) \end{aligned} \quad (8)$$

6. Discussion

6.1. Blind formulation validation and model comparison

The described models (Sections 5.2–5.4) have been developed with the same single excipient data. In this section, we test whether – despite this considerable limitation in their training data – they can accurately predict the behaviours of a full pharmaceutical blend. The same experiments were run using a 4-component blend which served as a surrogate for a real drug formulation — the active compound of the campaigned drug is replaced with powdered paracetamol, ensuring the

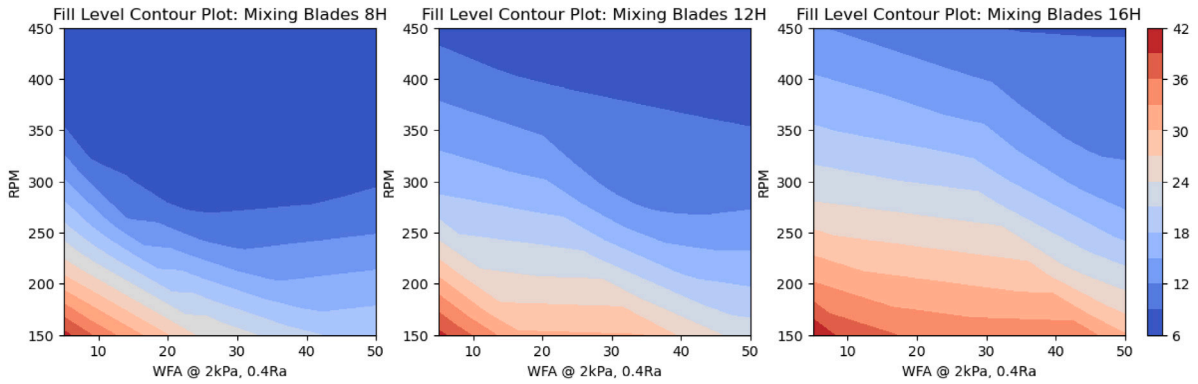


Fig. 13. Fill Level Contour Plot Prediction using the ANN as a function of RPM and WFA. The (shared) colour bar indicates the fill level in percentage for the 3 plots. The held values for the contour plot are as follows: RPM (150–450), MB (8, 12, 16), Feed Rate (22 kg/h), BFE (188 mJ), WFA (5–50), d10 (61 μm), d50 (122 μm), d90 (206 μm), 1/FFC (0.189).

final characteristics of the surrogate blend remain representative of the campaign. The surrogate was validated by ensuring the bulk properties of the surrogate formulation and the true formulation still possessed similar bulk physical characteristics (notably the bulk density and FFC), after the proprietary API was switched out. The bulk characteristic behaviours of this blend were gathered using the same methodology described in Section 3.1, and included in the final row of Table 1.

The experiments measured the fill level of the surrogate formulation at 3 speeds (250, 350, 450 rpm) at 15 kg/h feed rate and 16 mixing blades. The blend's material characteristics were then measured and used to predict the fill level. The results can be seen in Fig. 14. The models were all, to varying degrees of precision, able to predict the fill level of the blend and the variation thereof with RPM. This means that, despite the complexity of a multi-component formulation, there is a correlation between the formulation's characteristics and the resultant fill level. This is an important and potentially exciting finding, as it suggests that the models developed may be used to predict the behaviours of new formulations with no additional training.

The prediction accuracy ranked the model's Mean Absolute Error (MAE) in the following order ANN (1.4), M^2E^3D (3.7), and then RFR (4.6). The contour plots between the ANN and M^2E^3D are similar, displaying similar fills in the same areas. ANN and M^2E^3D differ in terms of linearity, M^2E^3D generally presents a very linear progression of contours — with tighter packing (of boundaries) in the lower corner and wider packing at the maximum. ANN showcases complex shapes and curvature, which has been shown to be advantageous for prediction. RFR, on the other hand, shows almost categorical boundaries. highlighting that despite the granularity of the model, there are some key features which allow for reliable fill level prediction. Ultimately, the following can be said: using each of the predictive methods provides a trade-off between accuracy and conceptual understanding. By using all three tools, however, one can achieve both highly accurate prediction, as well as strong mechanistic insight.

It is finally important to note that data-driven models have the highest predictive accuracy within the bounds of their training data — that is to say, while our models show impressive predictive capabilities within the parameter ranges upon which they were trained, this same level of accuracy cannot be assured outside this parameter space. Nonetheless, as we conduct more experiments, the model can be retrained, and the predictive confidence boundary correspondingly expanded. Furthermore, if the blender used were not fixed in inclination, it could be another processing variable for the training data. Thus making it possible to evaluate the mixing conditions across a wide set of inclinations, improving applicability for other commercial blenders.

6.2. Predicting mixing performance

Strain, determined as the number of blade passes through a given powder (calculated using Eq. (3)) has been demonstrated to be a strong predictor of good mixing in several prior publications (Vanarase and Muzzio, 2011; Portillo et al., 2008; Palmer et al., 2020; Vanarase et al., 2013; Zheng et al., 2022).

Using the validated ANN, we can input a given blend's material properties to calculate the strain from the predicted fill level. The resultant strain values across the range of processing parameters can then be plotted as a set of surface plots in Fig. 15. The detailed landscape describes the interaction between the formulation's properties and the range of processing parameters in the context of mixing performance. The surface plots seen in Fig. 15 are generated for the surrogate formulation (detailed in Table 1). The surface plots are then coloured by the ANN's prediction of the fill level at those same conditions. The fact that the model is able to predict the surrogate accurately (demonstrated in Section 6.1) provides confidence in the surface's shape and position of the colours.

It is immediately interesting to note that, irrespective of processing conditions, the optimum strain seems to consistently be found at or near a fill level of 1/3. The correlation suggests that the material in the blender has reached an idealised balance between the RPM and Mean Residence Time. To speculate on exactly why this occurs; it seems to suggest that the powder is, at this fill level, exhibiting regime-like behaviour which promotes ideal mixing conditions.

To better describe the hypothesis of optimum strain occurring around 1/3 fill level, it is useful to consider similar surrogate systems which show regime-like behaviour. There are two situations that can be conceptually related: rotating drums and fluidised beds.

- **Rotating Drums** utilise a single continuous motion within a batch vessel to mix the granular media within. Thus, when different processing conditions are used, the granular media (e.g. powder) within the drum demonstrate different bulk motion behaviours, which are described as regimes. Increasing the rotating speed of the drum, results in a progressive pulling of the granular media up the side of the drum in the direction of motion, resulting in regimes labelled as (from low to high RPM) rocking, rolling, cascading, cataracting, and centrifuging (to name a few) (Morrison et al., 2016). The transitions between these regimes are known to be affected not only by drum RPM, but also fill level — as is the mixing achieved thereby Arntz et al. (2008). If mixing is intended for the media, motions which cause high levels of granular dispersion (cascading/cataracting) result in improved mixing. In the context of the incline linear blender, the described 1/3 fill level suggests that this fill may be the balance for adequate powder volume for the powder to mix into (increasing MRT, and therefore macro mixing), whilst providing a sufficient mixing rate (RPM).

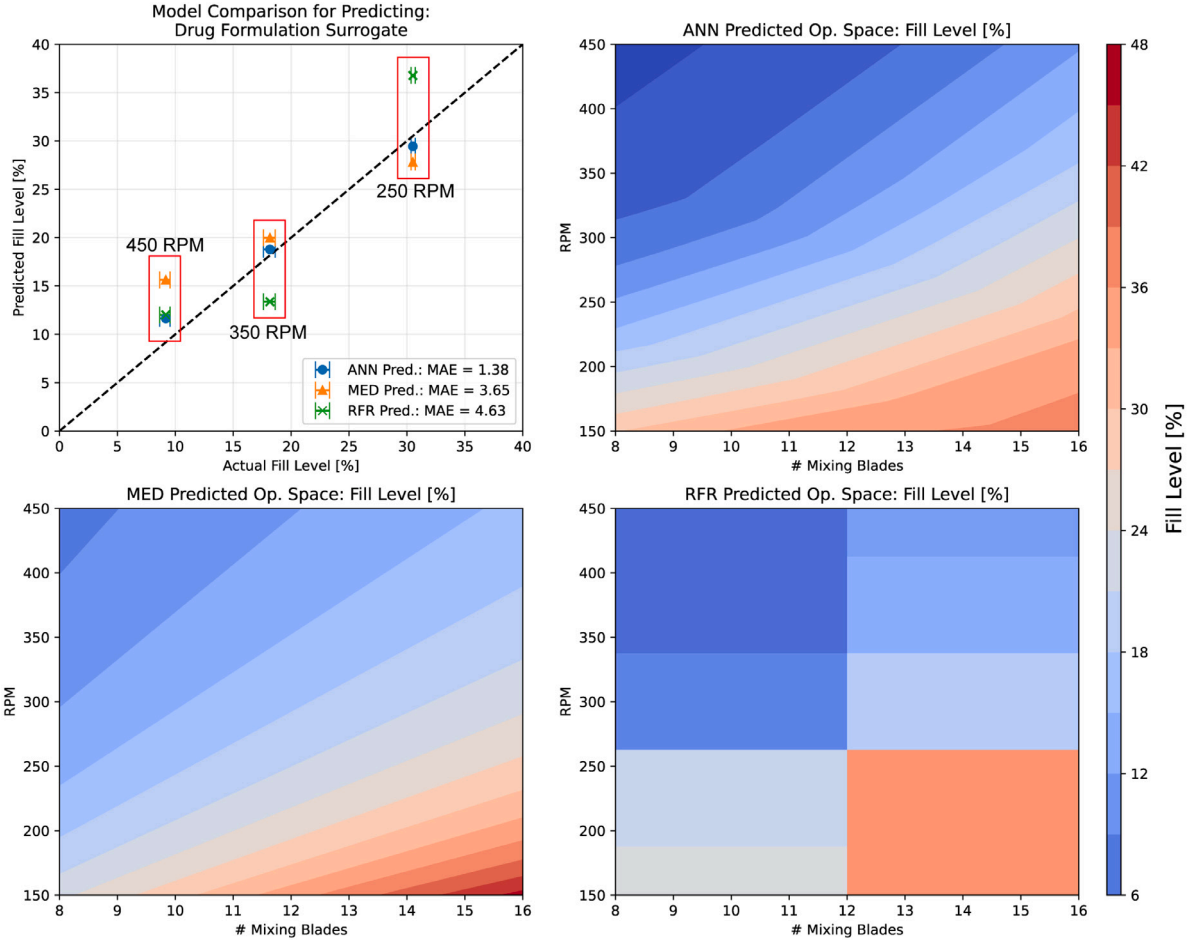


Fig. 14. Parity Plot between the experimental and predicted fill levels. The parity plot shows the three different models predicting the fill level. Each experiment was repeated 3 times, the mean is shown and the bars indicate the range of values (on the *x*-axis). Each model demonstrates the operating space between the RPM and mixing blades. The three models are ranked by the MAE (Mean Absolute Error) prediction of the experiment. Subsequently, an example of the design space showing the two most important processing parameters is plotted for each of the models. The model's contour plots are indicated by the graph's title. Furthermore, all contour graphs share the same colour bar and number of gradient steps (12) between the min and max.

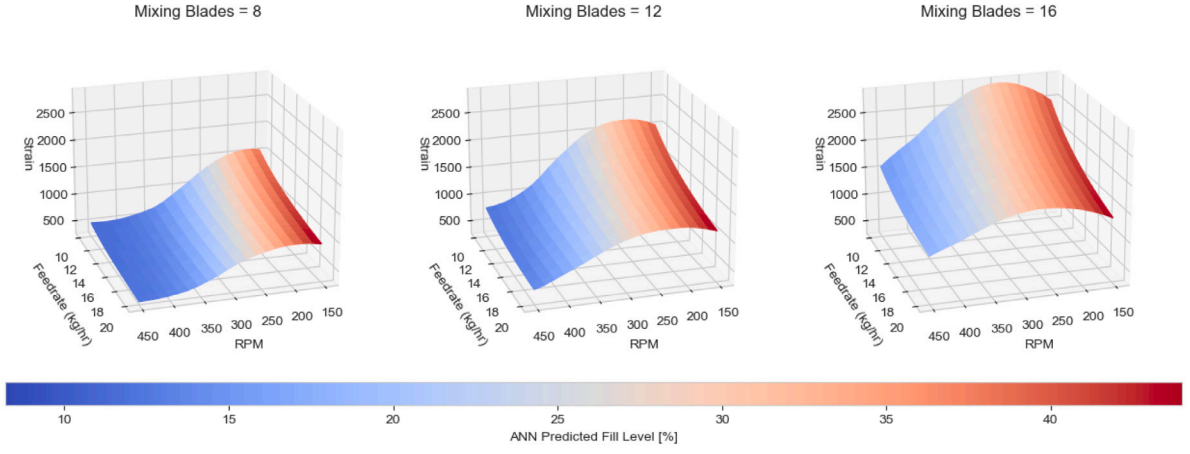


Fig. 15. Surface plots showing the strain by the three processing parameters for the drug surrogate formulation. The continuous parameters (those that can change during operation) share the same plot as the strain. The colour bar indicates the ANN predicted fill level, the three graphs share the same colour scale — allowing comparison.

- **Vibrofluidised Beds** utilise mechanical agitation in the form of vibration to cause an effective reduction in gravity, invoking fluid-like or even gaseous motion of dicrete solid particles (Rosato and Windows-Yule, 2020). To achieve this, a certain amount of agitation (i.e. a suitably high vibration strength) is necessary for the powder bed to reach fluidisation. Similarly, in the context of the inclined linear blender, the rpm agitator also has a mechanical agitation influence. The suggestion for the incline linear blender is that at a suitably high rpm (Froude Number), the motion of the impeller causes the powder to reach this fluidised state, in

which the effects of both gravity and inter-particular cohesion are sufficiently overcome to markedly reduce the resistance to inter-particular mixing.

Thus the hypothesis describes the resulting regimen as a combination of the two situations but in the context of continuous powder blending, where the powder flux receives enough work through agitation, leading to semi-fluidised behaviour. This overcomes the effects of gravity and inter-particulate cohesion of the powder, therefore promoting the local displacement and replacement of powder leading to improved micromixing. When paired with higher 'H' mixing blade configurations (otherwise described as larger mixing zones) there is an increase in axial slippage, which promotes the similarly described re-constitutive behaviour as the cascading regime seen with rotating drums. This, in turn, explains that for a given powder (or formulation), feed rate and blade configuration there is an optimum rpm, at which the powder reaches this regime; which coincides with the blender being 1/3 full. Either side of this apparent optimal fill level leads into contact-dominated mixing (for higher rpms) or friction-dominated mixing (lower rpms) — similar to that outlined by Zheng et al. (2022).

However, to attain this regime at a higher strain value- and therefore improved mixing, blade configuration is key, as it will manage the bulk mass — due to its role of promoting both axial and radial transport. In the context of this study, that would be the 16H configuration (when compared to 8H and 12H), which should, by these standards, supply adequate transport at the base (inlet), improved 'slippage' and radial dispersion in the mixing region, and lastly enough transport at the latter part of the barrel to remove sufficiently blended powder. This is supported further by the 'folding region' seen in the Positron Emission Particle Tracking (PEPT) study of Jones-salkey et al. (2023b) which describes localised spatial residence time retention of powder in the former part of the barrel, just prior to the mixing zone.

While it is possible that this is a specific feature of the helical mixing blade configuration (H) arrangement, the findings implicitly suggest that alternative blade configurations can form different regime-like behaviour. Nevertheless, in this instance, it would be sensible to offer the use of the H-configuration in order to exploit the described feature. It would be interesting to evaluate alternative configurations, like those discussed in Van Snick et al. (2017), to see if they present varied regime-like behaviour; and if not, would the configurations they describe present different processing sensitivities to those regimes?

To expand upon the insights from Fig. 15, let us examine how the proposed theory compares with the broader literature on this subject.

Building on the work of Kushner and Schlack (2014), who explored the scaling of blending powdered lubricant in both bin and tubular batch blenders, they highlighted a parameter K that could scale blending across vessel sizes spanning five orders of magnitude, given a constant Froude number of 0.4. Here, K denotes the 'amount of mixing', which is the product of the 'number of rotations (r)', the 'batch blender's characteristic length scale (L)', and the 'fraction of headspace ($F_{\text{headspace}}$)'.

K can be broken down into two components: 'mixing intensity' ($L \times F_{\text{headspace}}$) and 'mixing duration' (r). Specifically, the characteristic length roughly indicates the mixing region and the headspace gives the powder room to move, facilitating mixing. These combine to show the dispersive potential of each rotation. Meanwhile, the number of rotations conveys the number of times the powder undergoes dispersive rotations. Fortunately, for batch systems, these two components are decoupled, meaning 'mixing intensity' can be compensated by increasing 'mixing duration', and vice versa.

Transitioning this concept from batch to continuous processes leads to the formulation of the strain equation (strain = rpm × mean residence time), where rpm and mean residence time encapsulate the mixing intensity and duration, respectively. In continuous systems, however, since the MRT is a function of RPM (assuming a constant feed rate), one cannot independently adjust intensity and duration, emphasising that a balance is required.

Critical to these discussions is the underlying principle that regardless of scaling in batch or varied continuous setups primarily depends on fostering particle–particle interactions — the real catalyst for mixing. This becomes evident in bulk powder flow regimes, especially visible in granular mixing within rotating drums.

Relevant studies by Van Snick et al. (2017), Palmer et al. (2020), Bekaert et al. (2022a), and Zheng et al. (2022) offer deeper insights. The first three emphasise that maximising strain results in better content uniformity in both horizontal and inclined (15°) linear powder blenders. The latter provides context on regimes.

Interestingly, Bekaert et al. (2022a), using a horizontal mixer, found no interaction between the number of blade passes and blend properties, which contrasts the findings of this study. The interpretation is that the inclination of the blender seems to play a pivotal role in powder transport, which is best illustrated by materials showing different fill levels at the same rpm- see Fig. 11. This highlights the importance of understanding how the material properties interact with the geometrical differences.

Van Snick et al. (2017) indicated that in the same blender used in this study (15° linear blender), accurate and consistent content uniformity was achieved at intermediate residence mass and an intermediate rpm. This finding was reinforced when comparing two different blade configurations under identical conditions.

Moreover, Palmer et al. (2020) demonstrated that after achieving a certain strain, there is only a minimal further increase in blend uniformity, effectively stating that there is an effective 'ideal strain' for a formulation which achieves optimal mixing for a minimal time- and energy-expenditure. Thus, by utilising this tool we can predict the maxima and work back towards this 'effective' strain. Such insights enable researchers to pinpoint the optimal strain for increased processing efficiency whilst ensuring consistent content uniformity. Furthermore, an additional graph shows a V-shaped dip in the Peclet Number (Pe) plotted against fill level, with the lowest point (of this V-shaped dip) at around 1/3 fill level. Pe number is a ratio which indicates a balance between advective and diffusive transport rates; which could be effectively described as a balance of micro-to-macro mixing. A Pe of one, seems to occur at the fill level that provides the highest strain (demonstrated in Fig. 15).

Lastly, Zheng et al. (2022) utilised calibrated DEM simulations to identify two distinct regimes, which transition at an intermediate RPM of 250. This aligns with the first logic gate in the RFR decision tree shown in Fig. 12; whilst it is important to note, the model has similarly outlined this RPM across the ten different excipients and remains comparable to the simulation.

In conclusion, the body of literature, when analysed in tandem with the findings from Fig. 15, and the related discussions provide a comprehensive rationale of the behaviour being exhibited under these conditions. The significance of this should provide both researchers and industry professionals with an initial look into continuous blending regimens.

6.3. Application: Identifying the strain window

One potential practical application of these tools could be to adjust the process to compensate for a large disturbance. For example, APC could be used to reduce the RPM, and therefore increase the MRT, in response to a large disturbance. Following the disturbance, the RPM can be returned back to the steady-state set-point.

Referring to Palmer et al. (2020), the lowest rpm on the APC should match the most suitable point on the exponential decay curve, which could be effectively termed 'Strain Point (ω_P)'. This represents the minimum blade passes needed for desired content uniformity. Given that cohesive species are given an effective shear rate (tip-speed) to sufficiently break particle agglomerates/aggregates. This leads to questions for future research: *"what is the minimum tip speed required for effective blending of cohesive powders"* and *"are two **equal** strain values*

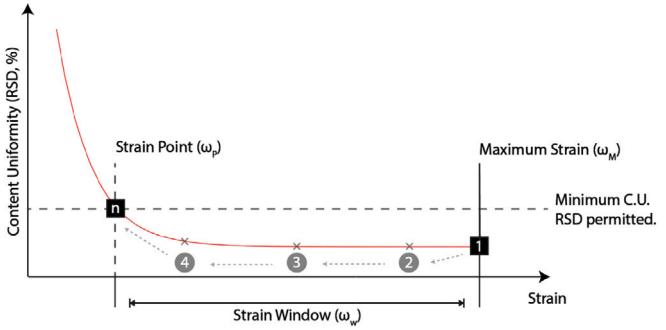


Fig. 16. Graphical Representation of the Strain Window (ω_W). Points 2, 3, and 4, indicate guided experiments to find the Strain Point (ω_P) — or the expansion of the current Strain Window (ω_W).

with different RPMs similar for mixing?" especially concerning values above or below the specific shear requirement for a formulation.

The predictive toolkit highlighted in Figs. 13, 14, and 15 enables the identification of 'Maximum Strain' (ω_M) for specific processing parameters, verifiable through subsequent experimentation. Upon confirming the attainment of minimum shear, our focus shifts to optimising the 'Strain Window' (ω_W), which is the difference between ω_M and ω_P, as depicted in Fig. 16. We can direct our experiments to maintain content uniformity around the ideal RPM, as illustrated by points 2 and 3 in the figure. These experimental points establish a strain window from the lowest acceptable strain value to ω_M. Further exploration, as seen at point 4, could widen this window. For a comprehensive understanding of the process limits, additional guided experimentation, towards point n (ω_P), may be performed to identify the threshold of failure.

It is noteworthy that theoretically, the toolkit could streamline experimentation to just three pivotal points. This top-down approach defines a steady-state set-point and establishes bounds for upper and lower RPM ranges. Similarly, it can assess the impact of other variables like feed rate changes on blend properties.

The deployment of these modelling tools is instrumental in developing robust control strategies, offering a nuanced understanding of the process and its operational parameters up to the point of failure. This outlines a safer operating domain, ensuring the product meets critical quality attributes. The use of these tools allows this understanding to be achieved with reduced experimentation, which is particularly important as extensive large-scale experimentation can slow development as well as impact on cost and sustainability. Furthermore, they can verify the consistency of processes or formulations, valuable for modifications like reformulation or changes to processing set-points. This process similarity can be corroborated through a series of validation trials, affirming the model's accurate prediction of the operational landscape.

7. Conclusion

In this paper, we have developed several machine learning/artificial intelligence models for predicting the performance of a continuous pharmaceutical blender, whilst also facilitating new insight into the fundamental behaviours thereof.

The methods used include RandomForestRegression (RFR), which extracts logic from the data; a novel symbolic regression and evolutionary AI tool (M^2E^3D), which produces complete, closed-form equations, effectively creating a reduced order model of the data; and Artificial Neural Networks (ANN), which provide more accurate predictions at the expense of a less-interpretable model. All three models independently identified the agitator speed (RPM), the number of mixing blades, the wall friction angle (WFA, 2 kPa with 0.4 Ra coupon) and the feed rate (kg/h) as the most important features for the prediction of fill level, and thus the strain experienced by the powder — an important predictor for mixing. By interrogating the models further, we were also able to gain valuable insight into the dynamics and mechanics of the blender explored, through prediction of the operating space.

In terms of pure predictive capability, the ANN demonstrated the greatest strength, both for the single components on which it was trained, as well as in further blind validation, predicting the behaviour of a 4-component formulation to a high degree of accuracy. Its ability to accurately predict the fill level of a previously unseen formulation, containing an API not included in its training data set, shows considerable promise for such an approach, in the future, to be used as a means of reducing the time and cost currently associated with bringing new drugs to market.

It is worth noting that we have discovered an interesting correlation between the maximum strain and 1/3 fill level (during steady state) in the prediction of the operating space. This correlation applies regardless of formulation or variable processing parameters, and it represents a crucial set point for blending processes that use a 15-degree incline blender. By achieving this, 1/3 fill level, set point during operation, a balance between advective and diffusive mixing can be attained. It is of further interest if other fill-level, and thus space–time, optimisations exist for other blending platforms and inclinations.

The magnitude of strain can be increased if the process is intensified — in effect simultaneously increasing both RPM and MRT. Yet, since RPM and MRT are coupled, increasing RPM will decrease MRT. Therefore, it is important to consider both the feed rate and the blade configuration. Decreasing the feed rate will increase MRT while decreasing the number of radial mixing blades will decrease MRT.

Given that maximising throughput is often desired, the blade configuration should be adjusted to include more radial mixing blades to retain higher fill levels (at the same RPM) and the RPM should be increased — thus intensifying the process. Similarly, this resolves the issue for highly agglomerated species, where using a higher RPM provides a higher tip speed, which then acts to break the highly agglomerated species. There may be slight differences between the clinical commercial equipment and the prototype (used in this study), however, the fundamental understanding remains the same. To validate this observation, we performed additional runs using various blade configurations, feed rates, and surrogate formulations (as detailed in Section 2.1). These runs were not included in the model training. However, they have shown that the maximum strain is also consistently observed at a fill level of 1/3. In addition, these findings were checked against a backlog of development runs for other formulations using the GMP-compliant CDC-50 from GEA and saw the same relationship of around a 1/3 fill level delivering the highest strain values.

Ultimately, the predictive tools described in this publication can find this optimum range of processing parameters for a newly characterised formulation and then evaluate both the operational and formulation stability of that formulation within the blender. Said tools thus stand to both mitigate risk and reduce wastage of time, energy and materials during product development through prediction of the variability at these early phases, meaning a directed first-time-right approach to experimentation, validation and process establishment, carrying benefits both economic and environmental.

CRediT authorship contribution statement

O. Jones-Salkey: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **C.R.K. Windows-Yule:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Conceptualization. **A. Ingram:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Conceptualization. **L. Stahler:** Writing – original draft, Methodology, Investigation, Data curation. **A.L. Nicusan:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis. **S.**

Clifford: Validation, Supervision, Resources, Project administration, Methodology, Investigation, Conceptualization. **L. Martin de Juan:** Supervision, Project administration, Investigation, Conceptualization. **G.K. Reynolds:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Andrew Ingram reports financial support, administrative support, equipment, drugs, or supplies, and writing assistance were provided by AstraZeneca. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

Experimental work was carried out as part of an Engineering Doctorate programme funded by EPSRC, United Kingdom through the Centre for Doctoral Training in Formulation Engineering (grant no. EP/L015153/1), and from AstraZeneca plc, United Kingdom. The computations described in this paper were performed using the University of Birmingham's BEAR Cloud service, which provides flexible resource for intensive computational work to the University's research community. See http://www.birmingham.ac.uk/bear for more details.

References

Arntz, M., den Otter, W.K., Briels, W.J., Bussmann, P., Beeftink, H., Boom, R., 2008. Granular mixing and segregation in a horizontal rotating drum: a simulation study on the impact of rotational speed and fill level. AIChE J. 54 (12), 3133–3146.

Bekaert, B., Grymonpré, W., Novikova, A., Vervaet, C., Vanhoorne, V., 2022a. Impact of blend properties and process variables on the blending performance. Int. J. Pharm. 613 (September 2021), http://dx.doi.org/10.1016/j.ijpharm.2021.121421.

Bekaert, B., Van Snick, B., Pandelaere, K., Dhondt, J., Di Pretoro, G., De Beer, T., Vervaet, C., Vanhoorne, V., 2022b. Continuous direct compression: Development of an empirical predictive model and challenges regarding PAT implementation. Int. J. Pharm. X 4 (December 2021), 100110. http://dx.doi.org/10.1016/j.ijpx.2021.100110.

Bekaert, B., Van Snick, B., Pandelaere, K., Dhondt, J., Di Pretoro, G., De Beer, T., Vervaet, C., Vanhoorne, V., 2022c. In-depth analysis of the long-term processability of materials during continuous feeding. Int. J. Pharm. 614 (December 2021), 121454. http://dx.doi.org/10.1016/j.ijpharm.2022.121454.

Bezanson, J., Edelman, A., Karpinski, S., Shah, V.B., 2017. Julia: A fresh approach to numerical computing. SIAM Rev. 59 (1), 65–98.

Cranmer, M., Sanchez-Gonzalez, A., Battaglia, P., Xu, R., Cranmer, K., Spergel, D., Ho, S., 2020. Discovering symbolic models from deep learning with inductive biases. In: NeurIPS 2020. arXiv:2006.11287.

Dhondt, J., Bertels, J., Kumar, A., Van Hauwermeiren, D., Ryckaert, A., Van Snick, B., Klingeleers, D., Vervaet, C., De Beer, T., 2022. A multivariate formulation and process development platform for direct compression. Int. J. Pharm. 623 (June), 121962. http://dx.doi.org/10.1016/j.ijpharm.2022.121962.

Engisch, W.E., Muzzio, F.J., 2015. Feedrate deviations caused by hopper refill of loss-in-weight feeders. Powder Technol. 283, 389–400. http://dx.doi.org/10.1016/j.powtec.2015.06.001.

Freeman, R., 2007. Measuring the flow properties of consolidated, conditioned and aerated powders—a comparative study using a powder rheometer and a rotational shear cell. Powder Technol. 174 (1–2), 25–33.

Gao, Y., Ierapetritou, M., Muzzio, F., 2012a. Periodic section modeling of convective continuous powder mixing processes. AIChE J. 58 (1), 69–78. http://dx.doi.org/10.1002/aic.12563, arXiv:0201037v1, http://doi.wiley.com/10.1002/aic.12563, https://onlinelibrary.wiley.com/doi/10.1002/aic.12563.

Gao, Y., Muzzio, F.J., Ierapetritou, M.G., 2012b. Optimizing continuous powder mixing processes using periodic section modeling. Chem. Eng. Sci. 80, 70–80. http://dx.doi.org/10.1016/j.ces.2012.05.037, URL https://www.sciencedirect.com/science/article/pii/S0009250912003181?casa_token=ekRMR8ElnJ4AAAAA:8ju5z1WMyR3MIO9ry04pJwNXZfq3L-kUGr-Fa7ns3R-YNT9dWj6RiZYo6YCaRAoIIssTEKIr_PM.

Gao, Y., Vanarase, A., Muzzio, F., Ierapetritou, M., 2011. Characterizing continuous powder mixing using residence time distribution. Chem. Eng. Sci. 66 (3), 417–425. http://dx.doi.org/10.1016/j.ces.2010.10.045.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, http://www.deeplearningbook.org.

Hare, C., Zafar, U., Ghadiri, M., Freeman, T., Clayton, J., Murtagh, M., 2015. Analysis of the dynamics of the FT4 powder rheometer. Powder Technol. 285, 123–127.

Holman, J., Tantuccio, A., Palmer, J., van Doninck, T., Meyer, R., 2021. A very boring 120 h: 15 million tablets under a continuous state of control. Powder Technol. 382, 208–231. http://dx.doi.org/10.1016/j.powtec.2020.12.073, URL https://linkinghub.elsevier.com/retrieve/pii/S003259102031250X.

Hurley, S., Tantuccio, A., Escotet-espinoza, M.S., Flamm, M., Metzger, M., 2022. Development and use of a residence time distribution (RTD) model control strategy for a continuous manufacturing drug product pharmaceutical process. Pharmaceutics 14 (2), http://dx.doi.org/10.3390/pharmaceutics14020355.

Jones-salkey, O., Chu, Z., Ingram, A., Windows-yule, C.R.K., 2023a. Reviewing the Impact of Powder Cohesion on Continuous Direct Compression (CDC) Performance, No. Cdc. pp. 1–48.

Jones-salkey, O., Nicusan, A.L., Windows-yule, C.R.K., Ingram, A., Werner, D., 2023b. Application of Positron Emission Particle Tracking (PEPT) for the evaluation of powder behaviour in an incline linear blender for Continuous Direct Compression (CDC). Int. J. Pharm. 645 (May), 123361. http://dx.doi.org/10.1016/j.ijpharm.2023.123361.

Kushner, J., Schlack, H., 2014. Commercial scale validation of a process scale-up model for lubricant blending of pharmaceutical powders. Int. J. Pharm. 475 (1–2), 147–155. http://dx.doi.org/10.1016/j.ijpharm.2014.08.036.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems. pp. 4765–4774.

Megarry, A.J., Swainson, S.M., Roberts, R.J., Reynolds, G.K., 2019. A big data approach to pharmaceutical flow properties. Int. J. Pharm. 555 (October 2018), 337–345. http://dx.doi.org/10.1016/j.ijpharm.2018.11.059.

Morrison, A.J., Govender, I., Mainza, A.N., Parker, D.J., 2016. The shape and behaviour of a granular bed in a rotating drum using Eulerian flow fields obtained from PEPT. Chem. Eng. Sci. 152, 186–198. http://dx.doi.org/10.1016/j.ces.2016.06.022.

Nicusan, A.-L., Windows-Yule, K., 2022a. PyMED: Multiphase Materials Exploration via Evolutionary Equation Discovery. Zenodo, http://dx.doi.org/10.5281/zenodo.7215239.

Nicusan, A.-L., Windows-Yule, K., 2022b. M2E3D: Multiphase Materials Exploration via Evolutionary Equation Discovery.

Palmer, J., Reynolds, G.K., Tahir, F., Yadav, I.K., Meehan, E., Holman, J., Bajwa, G., 2020. Mapping key process parameters to the performance of a continuous dry powder blender in a continuous direct compression system. Powder Technol. 362, 659–670. http://dx.doi.org/10.1016/j.powtec.2019.12.028.

Portillo, P.M., Ierapetritou, M.G., Muzzio, F.J., 2008. Characterization of continuous convective powder mixing processes. Powder Technol. 182 (3), 368–378. http://dx.doi.org/10.1016/j.powtec.2007.06.024, URL https://www.sciencedirect.com/science/article/pii/S0032591007003142?pes=vor&entityID=https%3A%2F%2Fidp.bham.ac.uk%2Fshibboleth#fig1.

Portillo, P.M., Muzzio, F.J., Ierapetritou, M.G., 2007. Hybrid DEM-compartment modeling approach for granular mixing. AIChE J. 53 (1), 119–128. http://dx.doi.org/10.1002/aic.11054, arXiv:0201037v1, URL http://doi.wiley.com/10.1002/aic.11054.

Portillo, P.M., Vanarase, A.U., Ingram, A., Seville, J.K., Ierapetritou, M.G., Muzzio, F.J., 2010. Investigation of the effect of impeller rotation rate, powder flow rate, and cohesion on powder flow behavior in a continuous blender using PEPT. Chem. Eng. Sci. 65 (21), 5658–5668. http://dx.doi.org/10.1016/j.ces.2010.06.036.

Rosato, A., Windows-Yule, K., 2020. Influence of key parameters. Segregation in Vibrated Granular Systems. Academic Press, pp. 197–218. http://dx.doi.org/10.1016/b978-0-12-814199-1.00018-4, URL https://www.sciencedirect.com/book/9780128141991/segregation-in-vibrated-granular-systems#book-description.

Sarkar, A., Wassgren, C.R., 2009. Simulation of a continuous granular mixer: Effect of operating conditions on flow and mixing. Chem. Eng. Sci. 64 (11), 2672–2682. http://dx.doi.org/10.1016/j.ces.2009.02.011, URL https://www.sciencedirect.com/science/article/pii/S0009250909001080.

Sarkar, A., Wassgren, C., 2010. Continuous blending of cohesive granular material. Chem. Eng. Sci. 65 (21), 5687–5698. http://dx.doi.org/10.1016/j.ces.2010.04.011.

Schulze, D., Schulze, D., 2021. Flow properties of bulk solids. In: Powders and Bulk Solids: Behavior, Characterization, Storage and Flow. Springer, pp. 57–100.

Schwedes, J., Schulze, D., 1990. Measurement of flow properties of bulk solids. Powder Technol. 61 (1), 59–68.

Toson, P., Siegmann, E., Trogrlic, M., Kureck, H., Khinast, J., Jajcevic, D., Doshi, P., Blackwood, D., Bonnassieux, A., Daugherity, P.D., am Ende, M.T., 2018. Detailed modeling and process design of an advanced continuous powder mixer. Int. J. Pharm. 552 (1–2), 288–300. http://dx.doi.org/10.1016/j.ijpharm.2018.09.032, URL https://www.sciencedirect.com/science/article/pii/S0378517318306835?via%3Dihub.

Van Snick, B., Holman, J., Vanhoorne, V., Kumar, A., De Beer, T., Remon, J.P., Vervaet, C., 2017. Development of a continuous direct compression platform for low-dose drug products. Int. J. Pharm. 529 (1–2), 329–346. http://dx.doi.org/10.1016/j.ijpharm.2017.07.003, URL http://www.sciencedirect.com/science/article/pii/S0378517317306026, https://linkinghub.elsevier.com/retrieve/pii/S0378517317306026.

Vanarase, A.U., Muzzio, F.J., 2011. Effect of operating conditions and design parameters in a continuous powder mixer. Powder Technol. 208 (1), 26–36. http://dx.doi.org/10.1016/j.powtec.2010.11.038.

Vanarase, A.U., Osorio, J.G., Muzzio, F.J., 2013. Effects of powder flow properties and shear environment on the performance of continuous mixing of pharmaceutical powders. Powder Technol. 246, 63–72. http://dx.doi.org/10.1016/j.powtec.2013.05.002.

Zheng, C., Li, L., Nitert, B.J., Govender, N., Chamberlain, T., Zhang, L., Wu, C.Y., 2022. Investigation of granular dynamics in a continuous blender using the GPU-enhanced discrete element method. Powder Technol. 412 (July), 117968. http://dx.doi.org/10.1016/j.powtec.2022.117968.