



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Reduced-order hybrid modelling for powder compaction: predicting density and classifying diametrical hardness

Citation for published version:

Trower, M, Emerson, J, Yu, M, Vivacqua, V, Johnson, T, Stitt, H & Dos Reis, G 2023, 'Reduced-order hybrid modelling for powder compaction: predicting density and classifying diametrical hardness', *Powder Technology*. <https://doi.org/10.1016/j.powtec.2023.118745>

Digital Object Identifier (DOI):

[10.1016/j.powtec.2023.118745](https://doi.org/10.1016/j.powtec.2023.118745)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Powder Technology

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Reduced-order hybrid modelling for powder compaction: predicting density and classifying diametrical hardness

 Maia Trower^a, Joe Emerson^b, Mingzhe Yu^b, Vincenzino Vivacqua^b, Timothy Johnson^c, Hugh Stitt^b,  Gonçalo dos Reis^{a,d,*}

^a*School of Mathematics, University of Edinburgh, The Kings buildings, Edinburgh, EH9 3JF, UK*

^b*Johnson Matthey, PO Box 1 Belasis Avenue, Billingham, TS23 1LB, UK*

^c*Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh, EH14 4AS, UK*

^d*Centro de Matematica e Aplicacoes (CMA), Faculdade de Ciencias e Tecnologia, Campus da Caparica, Caparica, 2829-516, Portugal*

Abstract

Drawing on machine learning (ML) techniques and physics-based modelling, two feature-based reduced-order models are presented: one for the quantitative prediction of density and another for the classification of the diametrical hardness of pellets from a powder compaction process (pelleting). For interpretability, the models use as input only the parameters from a modified Drucker-Prager Cap (DPC) model calculated from process data monitoring and the applied maximal compression force. For quantitative density prediction, 8 features linked to first-principles models of powder compaction are generated, and the final model uses only 2. A critical part of the modelling, and also one of the main contributions, is a data augmentation step for the primary data set of this study by leveraging much smaller supplementary data sets that have measurements not present in the primary data set.

The final results imply a significant reduction in the quantity of data needed for model input and cut down the cost of data acquisition, storage, and computational time. Additionally provided is a detailed analysis of the impact and relevance of the generated features on the model performance.

The density prediction model, using only 2 features, reaches a mean absolute scaled error (MASE) of 12.9% and a mean absolute error (MAE) of 0.10 (where $r^2 = 0.975$). The scaled (diametrical) hardness classifier has an F1 score of 0.915 using 4 features.

*Corresponding author

Keywords: Powder compaction, Machine learning, Density modelling, Hardness modelling, Drucker-Prager Cap model

Nomenclature and Symbols

The following list describes acronyms and symbols that are used within the body of this document.

β	Internal friction angle (see DPC model)
\widehat{R}	Cap eccentricity (see DPC model)
d	Cohesion (see DPC model)
E	Young's modulus (see DPC model)
F_{max}	Maximum compression force
p_a	Evolution pressure (see DPC model)
p_b	Hydrostatic pressure yield stress (see DPC model)
r^2	Coefficient of determination
ν	Poisson ratio (see DPC model)
CQAs	Critical quality attributes
D-PIF	Density Physics-Informed Features
DEM	Discrete element method
DPC	Drucker-Prager Cap (model)
FEM	Finite element method
GPR	Gaussian process regression
H-PIF	Hardness Physics-Informed Features
LR	Linear regression
MAE	Mean absolute error
MASE	Mean absolute scaled error
ML	Machine learning
MLP	Multilayer Perceptron
mRMR	minimum Redundancy and Maximum Relevance
PLS	Partial least squares (regression)
SVM	Support vector machine (regression)

1. Introduction

Powder compaction is a ubiquitous manufacturing process found in a number of industries, from pelleting metallic materials to anode and cathode coatings for batteries [1, Table 1], to producing pharmaceutical products, detergents, ceramics and many others. For example, the production of chemical catalysts relies on this process, and the catalyst industry was valued at \$33.9 billion in 2019 [2]. This is expected to grow by an estimated 4.4% per year, indicating a significant economic incentive to improve the pelleting process. In addition to economic implications, controlling this process more effectively will reduce unsustainable waste and generate actionable insights for manufacturers in terms of quality control [3]. In the

context of this work, two key properties of the pelleting process are essential to monitor: pellet density and hardness.

The pelleting process typically involves 3 main stages [4]: die filling, powder compaction, and ejection. The die fill refers to the process of inserting the powder into the pellet mould, and the filling method aims to spread the powder evenly inside the mould. Often this is achieved using a paddle or a vibration feeder. Compaction of the powder further rearranges the particles and then deforms the material under increased pressure to form a pellet. Finally, the ejection stage applies a force to the underside of the pellet to remove it from the mould. Errors in any of these stages propagate through the manufacturing process and can produce pellets that are not fit for purpose.

Previous studies have primarily focused on physics-based constitutive models, for example plasticity models such as the Drucker-Prager Cap (DPC) model to describe plastic deformation [5]. Broadly speaking, the goal of this manuscript is to shed light on the relationships between elements of the manufacturing process and the resulting density and/or hardness and to contextualise these relationships by drawing on established first principles models such as the DPC model [6, 7].

There are two broad model paradigms in the literature: particulate models and continuum models. The discrete element method (DEM) views the powder as elastic-plastic spheres [8] and basic contact laws between particles allow the behaviour of the macroscopic pellet to be simulated [9]. However, this approach has its limitations when dealing with high deformation and multiple contacts [4]. An alternative is the Finite Element Model (FEM) methodology to simulate the powder compaction behaviour with a pressure-dependent yielding plastic model (modified Drucker-Prager Cap model). This method requires time-consuming mechanical testing for calibration, and its lab-scale model has extensive computing requirements. Another major limitation is that each simulation must be validated, which is a significant challenge [6, 7].

These observations motivate the need for a faster model for quality control, with real-time response. The solution to be explored comes from machine learning (ML) modelling and associated reduced-order models. ML modelling has already made significant contributions to compaction models in the pharmaceutical industry, addressing die filling performance [10], and relationships between input variables (such as process parameters and powder properties) and critical quality attributes (CQAs) [11, 12, 13]. Recent work has also shown that hybrid data-driven and theoretical approaches are of interest due to their interpretability. For example, the authors of [14] de-

rived a hybrid modelling approach which took into account the throughput of the tableting process, as this was found to impact density and hardness in addition to the compaction pressure.

In contrast to the previous pharmaceutical studies, the work contained within our paper concerns inorganic catalyst material, which is generally harder than pharmaceutical powders. However, the processing equipment and the underpinning physics is sufficiently similar to draw useful inferences from the pharmaceutical literature. This can be seen, for example, when comparing [15], which shows that compaction force was the most influential factor in the study of a pharmaceutical product, and [3], which mirrors this finding for inorganic catalyst material.

The goal in [3] was to understand the impact of the pelleting process parameters on the density and strength of the resulting pellets. This was achieved by developing a partial least squares (PLS) regression model and employing a sequential feature selection process to reduce the model inputs to the predictive sections of the time-series profiles of the various process parameters (including punch force, speed and displacement). The model utilised latent features from the PLS, which makes tractable causality analysis difficult to carry out. From this study, an accurate quantitative prediction of pellet density was achieved whereas the hardness prediction was less reliable.

In this manuscript, the aim is to build from [3] (see also [16]: it clarifies factual information regarding the machines used in [3]) by asking the question: can one improve the predictive capability of ML models by using physics-informed feature engineering¹? This approach offers the benefits of reducing the time-series data to a set of interpretable features with clear physical meaning while avoiding the random element of data-driven feature selection methods that add complexity to model interpretation. For the hardness modelling, it is argued that a classification approach is more suitable than the quantitative prediction of [3]. This is a credible alternative given the poor reliability of the hardness measurements in the region of average hardness for the available data.

This work is structured as follows: Section 2 gives information about the data used in this study and Section 3 provides details of physics-informed feature generation and the data augmentation procedure. The ML target

¹Physics-informed feature engineering refers here to the process of generating features from the available data based on first-principles physical models. In this work, this involves using the compaction data to estimate parameters of the DPC model.

problems and associated modelling process is described in Section 4. Section 5 presents the results and their discussion.

2. Data Description and data pre-processing steps

There are 3 data sets in this study, X_2 , Z and Y , where X_2 was studied in [3, 16] and was denoted there as X_2 . These 3 data sets are collected from different experiments on the same catalytic material: the powders have the same size, range, and characteristics. Nonetheless, since each experiment is different, each data set contains different measurements (with overlaps). Table 1 summarises the data sets and measurements contained in each.

This section provides a detailed description of the 3 data sets (see also [3, Section 2.1 and 2.2], [16]) and also outline the data pre-processing, including outlier removal and scaling. Specific descriptions are given afterwards.

Variables Recorded		Data set		
		X_2	Z	Y
Lower punch displacement	Upper punch displacement	✓	✓	✗
Lower punch force	Upper punch force			
Distance between punches	Punches force difference			
Lower punch linear speed	Upper punch linear speed			
Density (volumetric)		✓	✗	✓
Hardness: axial		✗	✗	✓
Hardness: diametrical		✓	✗	✓
Radial pressure		✗	✓	✗
Number of pellets (after data pre-processing)		771	140	140

Table 1: Variables recorded by data set and number of pellets (after data pre-processing)

2.1. Primary Data set X_2

2.1.1. Materials, Experimental Design and data collected

The primary data set, denoted X_2 , was generated by Johnson Matthey. The feed powder was a co-precipitated base metal plus inorganic support material catalyst precursor. This has not been calcined, so it will comprise essentially classic low-temperature, high surface area crystalline morphologies of the oxides or hydrated oxides. The feed has been pre-compacted using a roll compactor, and screened for under- and over-sized fractions. Representative data for the particle size are shown in Table 2. It is noted that the size distribution is rather broad, but this is influenced by production

	D_{10} (μm)	D_{50} (μm)	D_{90} (μm)
Average size	184	916	1850

Table 2: Solids size distribution

economics². The pelleting data were obtained using a STYL’One Evolution (Medelpharm SAS, France) compaction simulator [3, 16].

The STYL’One Evolution is designed to simulate production scale pelleting with an instrumented single-punch. The machine is fitted with an array of sensors to record data during the pelleting process (see Table 1).

In addition, an *instrumented die* (in the STYL’ONE Evolution) was fitted with sensors³ to record the diametrical stress and axial stress (further used for data sets Y and Z , discussed below) against other variables during the compaction/pelleting process, such as time, punch displacement and pellet volumetric density (the method is well described below and in [6]). The recorded data from instrumented die compactor were used to calibrate solids’ mechanical properties such as cohesion, internal friction angle, Young’s Modulus and Poisson ratio for developing the Drucker-Prager Cap model (DPC; detailed in the methods Section 3). The pelleting process involves four main steps, given in order: (a) filling of the die; (b) pre-compaction of the solids (rearranging the particles); (c) main compaction of the solids (forming the pellet); and (d) ejection of the pellet from the die. The measurements (described in Table 1) are recorded every 0.01 milliseconds (ms) over pre-compaction, main compaction, and ejection.

The compaction simulator was used to produce 200 pellets (200mg per pellet) for each of the six experimental runs, and each run used different solids feeder setups. Two different solids feeder systems were used: (a) a force feeder, which pushes solids over the die to allow it to fill; and (b) a vibration feeder, which vibrates so that solids fall into the die. The force feeder was used in both a left and right orientation, which changes the angle of the blade that pushes the solids over the die. The feeders were operated at different speeds and the vibration feeder at different vibration intensities. The feeder set-ups for each experiment are detailed below:

- Force feeder
 1. Paddles rotating left, speed 40%

²It is impractical in this case to require a homogeneous particle size distribution, as this would be too expensive for this industrial application. In particular, the low pass yield of the upstream roll compaction would result in prohibitive costs.

³For clarification, “instrumented die” is a piece of unit in the machine.

- 2. Paddles rotating left, speed 70%
- 3. Paddles rotating right, speed 70%
- Vibration feeder
 - 1. Speed 20%
 - 2. Speed 38%
 - 3. Speed 70%

The pre-compression thickness was kept constant, the main compaction depth was set at 2.9mm, and the machine speed was set at 25 cycles per minute. Furthermore, the *SOTAX ST50* (SOTAX AG, Switzerland), a semi-automatic pellet hardness tester (diametric crush strength test), was used to measure four properties of the pellets: weight, diameter, thickness and hardness.

Although the experiment generated data for 1,200 pellets, data set X_2 contains only compaction measurements for around 800 pellets as they are compressed, and the measurements recorded are shown in Table 1.

2.1.2. Data pre-processing

Significant amounts of data processing is required for data set X_2 . Initially, various summary statistics and non-numerical entries were present in X_2 , which were identified and removed. Exploration of the data showed that pellets did not undergo compression at the same time. The stages in the process occurred at a range of times (see Fig. 1a and 1b). Plotting the time series profiles for all variables revealed that the lower punch force shows the pre-compression, main compression, and ejection stages most clearly. Therefore the pellets are aligned based on the time at which this variable reaches its maximum, which corresponds to aligning pellets so that the main compression occurs at a pre-determined time point. This shift is applied for all variables so that the main stages of the process happen at easily identifiable time points.

It is clear that there are large time intervals at which no compression or ejection is occurring. For example Fig. 1b shows that any measurements recorded before 2300ms or after 3000ms do not contain useful information. Removing these data points reduces the size of X_2 very significantly (and improves the performance of the models discussed in later sections). Additionally, it is possible to remove all measurements in the time intervals between compression and ejection events without negatively impacting the performance of the models. **That is, when no compression or ejection events are occurring, the behaviour of the punches have no impact on the resulting density or hardness of the pellet as the punches themselves are no longer in**

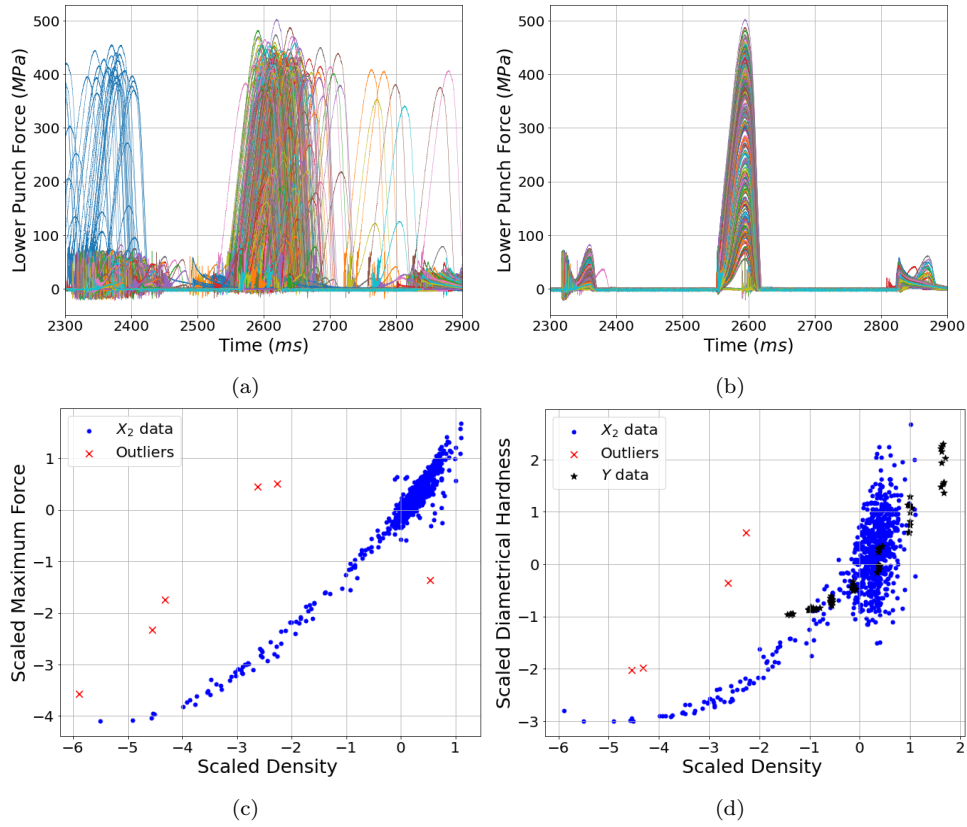


Figure 1: Lower punch force profile for X_2 : (a) unaligned and (b) aligned by main compression time. (c) Data mismatches (red cross) identified using scaled density and maximum compression data from all experiments in X_2 . (d) Data mismatches in X_2 (red cross) identified using scaled density and scaled diametrical hardness, with corresponding data points from Y overlaid (black star).

contact with the material. Therefore, removing the measurements recorded at these time intervals does not remove useful information from the model, hence not increasing the prediction error. Using this method reduces the number of time points by more than 70%, from approximately 70,000 to fewer than 19,000.

A further step in the pre-processing operations for the data is to identify and remove outliers. Inspection of feature profiles after alignment allows clear outliers to be easily identified, but an initial linear regression on the variables indicated that further significant outliers were present. Therefore, assuming that the data points are normally distributed, all points outwith a 95% confidence interval are identified as outliers and removed.

Due to the method of data collection, it is highly likely that there are mismatches in the compaction data and the response variables. The volumetric density (in g/cm^3) and hardness (in Newtons, N) of the pellets were measured by collecting compacted pellets from the simulator and manually transferring them to a hardness tester. During the transfer, any changes to the order of the compacted pellets will result in a pellet being associated with a density and hardness measurement that is not correct. To identify these, one observes that the maximum compression force, denoted F_{max} , is highly correlated with the density of the pellet, and thus an inspection of Figs. 1c and 1d shows clear identifiable mismatches in the data that are then removed. After removing the outliers, the data set X_2 contains 771 observations (see Table 1). Throughout, the density and hardness measurements are scaled to have zero mean and unit variance.

Lastly, Figure 1d evidences a scatter effect in the hardness measurements and this will influence the modelling choices. The issue is discussed in Section 4.3 and includes a literature perspective.

2.2. Secondary data sets: Z and Y

Two further smaller data sets, Z and Y , were also provided by Johnson Matthey. These data sets correspond to independent experiments, each with data from 140 pellets (after pre-processing), that were prepared using the same feed material as for data set X_2 – it is the same catalytic material: the powders have the same size, range, and characteristics.

The secondary compaction data set, Z , contains largely the same compaction information as X_2 (see Table 1) and underwent the same data pre-processing steps as X_2 . There are *critical* differences, namely that Z contains: (a) an additional measurement of the radial pressure in the die during compaction (measured from a different machine); (b) significantly fewer observations than X_2 (140 pellets compared to 771 of X_2); (c) there are 20

pellets for each of 7 pre-selected compression forces (hence averages can be computed); and, (d) the pellets in Z *do not* have any hardness or density measurement associated to them.

The other supplementary data set Y contains results of axial and diametrical hardness tests, and sufficient additional information to compute the [volumetric](#) density for each of its 140 pellets. Note that Y contains the same amount of pellets as Z , again 20 per pre-set compression force, and of these 20: 10 pellets have an axial hardness measurement associated and the other 10 have a diametrical hardness measurement. It is not possible to record both hardness measurements for each pellet due to the destructive nature of the test. Again, Y contains significantly fewer observations than X_2 , and Z and Y are from different experiments. For emphasis, data set Y *only contains* hardness and (computed) density measurements, but in contrast to X_2 , Y has both axial and diametrical hardness in Y (see Table 1). Finally, like Z , there are multiple pellets with the same (relative) density and thus averages may be used (see Fig. 4 below).

The experimental description of the Y data set confirms the pellets underwent the same type of compression force profiles as X_2 and Z (but such profiles are not provided). Importantly, the measurements in Y cannot be matched directly to the compaction measurements from Z , but instead Y contains data about pellets of the same material as those in X_2 and Z spanning a larger range of densities. Nonetheless, it is possible to approximately match the data in Y with the measurements in Z using the procedure detailed in Section 3.2, and it is these approximate pairings that are later used to enhance X_2 . This procedure is a primary contribution of this work.

2.2.1. Further remarks on data set Y and Z

The data in Z are collected using a gravity feeder in the Styl'One compaction simulator (see Section 2.1), with an instrumented die capable of measuring the radial pressure during compaction. The diameter of the pellets was 11.28 mm, and for each given tablet thickness (or compaction pressure) twenty pellets were produced. Seven pressure levels were explored, and no hardness or density information was recorded for these tablets.

Data set Y contains only hardness measurements, again for seven distinct pressure levels. For each pressure level, twenty pellets are tested; 10 pellets are used to measure diametrical hardness, and the remaining 10 for axial hardness. This is because each of the tests are destructive so cannot be performed on the same pellet. Diametrical hardness was measured using the Brazilian test, which is standard practice in the pharmaceutical industry [17]. The test applied increasing compressive force across the diameter of

the pellet to induce a fracture. The pressure at which the tablet fractures indicates the diametrical or tensile strength of the pellet. Similarly, the axial hardness is measured by exerting a compressive force at a 90 degree angle to the diameter of the pellet. Again, pressure at which the fracture occurs can be used to compute the axial strength of the pellet. Both of these measurements are recorded in Y .

3. Modelling and Methods: data augmentation

The goal of Section 3.1 and Section 3.2 is to use the secondary data sets Z and Y to augment X_2 , via the Drucker-Prager Cap (DPC) model. These dependencies are summarised in Table 3.

In this section, it is shown how to compute estimates of DPC parameters as functions of the relative density of the pellets. [Here, the *relative density* is defined as the ratio of the actual density of the pellet to the maximum density of the pellet \(measured when the maximum compression force is applied\)](#). Note that in Sections 4-6, the text refers to the *scaled* density of X_2 , which is *distinct from* the relative density of X_2 , Z and Y . The scaled density refers to the actual density scaled to have a zero mean and unit variance, and this scaling is done using Python’s *StandardScaler*. Scaling is also applied to the diametrical hardness of X_2 .

3.1. Feature generation: DPC model parameters in Z and Y

The Drucker-Prager Cap model is a plasticity model that is capable of representing the densification and hardening of a powder during compaction [18]. In this model, the powder is viewed as a continuum, and properties such as cohesion, internal friction, and material stresses are averaged over all particles – the validity of this representation is a generally accepted approach [19]. The idea is to express the properties of the material as the compaction progresses as a function of a specified state variable – typically the chosen state variable is the relative density, $\tilde{\rho}$.

Fig. 2 shows the components of a DPC model with the relevant parameters and an in-depth discussion can be found in [18]. The two distinct segments, the cap surface F_c and the shear failure surface F_s , represent the two behaviours expected during powder compression. A third surface, F_t , is introduced to smooth the transition between F_s and F_c , and hence is known as the transition surface. These surfaces are described in (A.1)-(A.3).

When the hydrostatic pressure p is low, the model shows that the powder will follow a Mohr-Coulomb shear failure line. Shear failure will occur when the stresses (p, q) lie on the Mohr-Coulomb line (where q is the von Mises

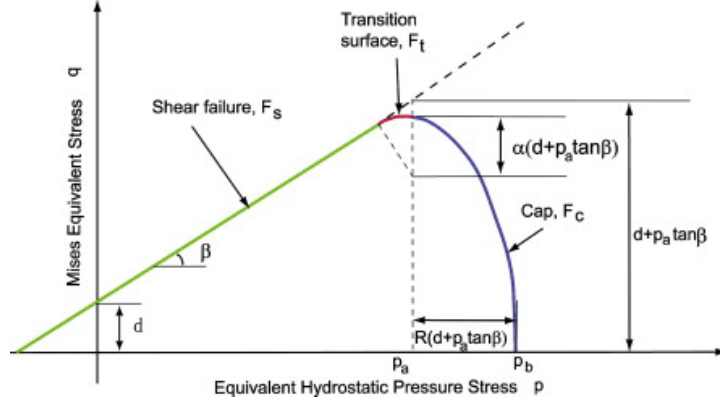


Figure 2: Drucker-Prager Cap Model: yield surface in the $p - q$ plane [18, Fig. 5] (reproduced with permission from Elsevier)

equivalent stress). Note that throughout this section, stress variables are measured in megapascals, MPa . When $p > p_a$, the behaviour of the material is described by the cap surface F_c . This represents the hardening of the material as a result of compaction; if the state (p, q) lies on the surface F_c , the powder undergoes densification and hardening. This is caused by plastic deformation in the material. The transition surface, F_t , exists solely to smooth the transition between F_c and F_s . Note that these DPC parameters, summarised in Table 3, are typically density-dependent and are determined as a function of the volumetric plastic strain; [20] corroborates that DPC model-based material properties for compaction properties of dry granulated powders are density-dependent.

DPC Parameter	Symb.	Data set
Cohesion	d	Y
Internal friction angle	β	Y
Cap eccentricity	\hat{R}	Z (needs: d, β)
Evolution pressure	p_a	Z (needs: d, β)
Hydrostatic pressure yield stress	p_b	Z (needs: d, β)
Young's modulus	E	Z
Poisson ratio	ν	Z

Table 3: Drucker-Prager Cap parameters [21] and the supplementary data set from which they are computed to then enhance data set X_2 . Note that the radius of the transition surface parameter α is chosen arbitrarily within a specified range and hence is not included as a density-dependent feature.

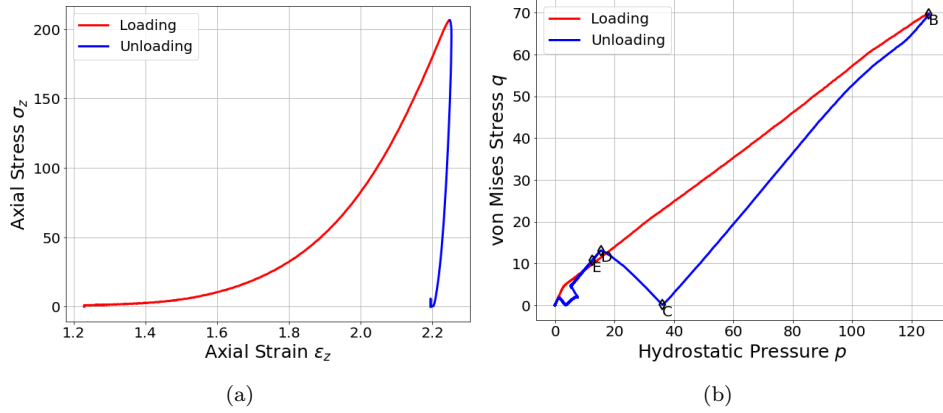


Figure 3: (a) Relationship between axial stress and strain on the powder during compaction of pellets in Z . (b) Relationship in Z data between the hydrostatic pressure, p and the von Mises equivalent stress, q . For both figures, loading is shown in red and unloading in blue.

To determine the parameters of the DPC model, one begins by computing the stress and strain of the powder as it undergoes compaction. These parameters can be inferred directly using the compaction data in Z , and are then transformed into stress invariants as follows [22, Chapter 10]:

$$\text{axial stress} = \sigma_z = \frac{F}{A}, \quad \text{axial strain} = \epsilon_z = \log \tilde{\rho}, \quad (1)$$

$$\text{radial stress} = \sigma_r, \quad p = \frac{1}{3}(\sigma_z + 2\sigma_r) \quad \text{and} \quad q = \frac{1}{\sqrt{3}}\|\sigma_z - \sigma_r\|. \quad (2)$$

In (1)-(2), F refers to the compression force exerted on the powder, A is the surface area of the pellet, and $\tilde{\rho}$ is the relative density. These are all computed from the data in Z .

Shear failure surface parameters d, β

Consider the DPC model shown in Fig. 2. The shear failure surface is determined by the cohesion, d , and the internal angle of friction, β , which can be computed using Y as follows ⁴:

$$d = \frac{(\sqrt{13} - 2)\sigma_t\sigma_c}{\sigma_c - 2\sigma_t} \quad \text{and} \quad \beta = \arctan \left[3 \left(\frac{\sigma_c - d}{\sigma_c} \right) \right]. \quad (3)$$

⁴Note that the below formula for d and β in [18, Eqs (23) and (24)] have a small typo in the signs. Equation (3) is the correct version (as are [22, Equations 10.34a and 10.34b]).

In (3), σ_c and σ_t denote the compression and tensile stress respectively, defined as

$$\sigma_c = \frac{4h_u}{\pi D^2} \quad \text{and} \quad \sigma_t = \frac{2h_d}{\pi D t}, \quad (4)$$

where D is the diameter of the pellet and t is the measured thickness. The force h_u refers to the axial hardness, and h_d to the diametrical hardness. Additionally, the internal friction angle β is converted from radians to degrees.

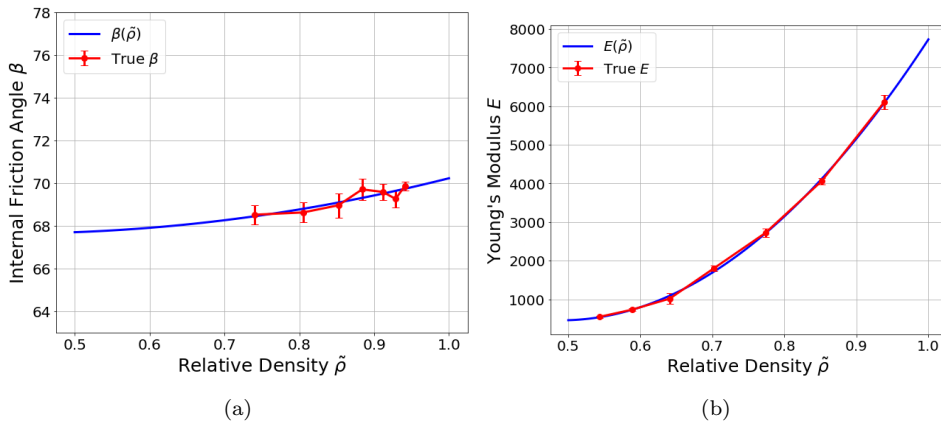


Figure 4: (a) Internal friction angle β as function of the relative density $\tilde{\rho}$ computed from Y with 95% confidence intervals in red, and 2nd order polynomial interpolant $\beta(\tilde{\rho})$ in blue. (b) Young's modulus as function of $\tilde{\rho}$ and computed using Z with 95% confidence intervals in red and 2nd order interpolant $E(\tilde{\rho})$ in blue.

Fig. 4a illustrates the relationship between the relative density of a given pellet and the internal friction angle β , and a similar computation is undertaken for the cohesion d . These computations aim to determine the parameters as functions of the relative density, $d(\tilde{\rho})$ and $\beta(\tilde{\rho})$. The approximate function $\beta(\tilde{\rho})$ is also shown in Fig. 4a. One can then use these approximations to find estimates of d and β for the pellets in Z by: computing the relative density $\tilde{\rho}$ for each Z pellet, and then evaluating $d(\tilde{\rho})$ and $\beta(\tilde{\rho})$ at each of these values. These estimates are essential for the determination of the cap parameters in the proceeding section.

Cap parameters \hat{R} , p_a , and p_b

The cap surface is described by the parameters \hat{R} , p_a , α , and p_b (see Table 3). Again, following [18, Section 4] and as described above, one estimates the values of \hat{R} , p_a and p_b for Z data via the functions $\hat{R}(\tilde{\rho})$, $p_a(\tilde{\rho})$ and $p_b(\tilde{\rho})$

determined by Y . Combining these with the shear surface parameters d and β , the cap parameters can be used to fully determine the DPC model via data set Z . All cap parameters are computed by following the formulae in [22, Chapter 10] (see (A.7)) and, just like for $d(\tilde{\rho})$ and $\beta(\tilde{\rho})$, one produces estimated functions for these.

Note that, since the parameters are density dependent, increasing the compression force will change the shape and magnitude of the DPC surfaces. Fig. 4b shows the relationships between the Young’s modulus and the relative density of the pellets in Z , and it is clear that there is a strong positive correlation. This provides good evidence that computing estimates for the DPC parameters in the data set X_2 as inputs to a model is likely to generate a well-performing predictor.

Young’s modulus E and Poisson ratio v

Using (1)-(2), one is able to compute the Young’s modulus and the Poisson ratio of the powder for pellets in Z . This is done by following [18, Section 4] – see (A.4)-(A.5). Note that the results are dependent on the pellet’s relative density $\tilde{\rho}$. Therefore, compressing the powder at varying maximum compression forces generates a range of values for both the Young’s modulus and Poisson ratio for the same material, so that these properties can be expressed as functions of the relative density. These estimated functions are denoted as $E(\tilde{\rho})$, for the Young’s modulus, and $v(\tilde{\rho})$, for the Poisson ratio. Fig. 4b shows the true values of E computed using Z and the polynomial estimate $E(\tilde{\rho})$.

3.2. Enhancing Data set X_2

Using the procedure outlined in the previous section, it is straightforward to compute the cohesion and friction angle from Y . One can then use the estimated functions $\beta(\tilde{\rho})$ and $d(\tilde{\rho})$ to find an approximation for β and d for each pellet in Z . Then, these parameters are used to find the cap parameters for Z pellets, for which one then produce polynomial estimates as discussed above. Along with the estimated polynomials $E(\tilde{\rho})$ and $v(\tilde{\rho})$, one now has the following functions of the relative density: $d(\tilde{\rho})$, $\beta(\tilde{\rho})$, $\hat{R}(\tilde{\rho})$, $p_a(\tilde{\rho})$, and $p_b(\tilde{\rho})$. *Contingent on a justification provided below, one finds the DPC parameters for X_2 simply by evaluating these functions on the relative density of the pellets in X_2 .*

Inspection of X_2 in Fig. 5a reveals that the relative densities of the pellets in this primary data set X_2 range from 0.84 to 0.99. From this information one can retrospectively analyse the polynomial estimates and confirm that a good fit for the true data in this range is indeed observed. For example,

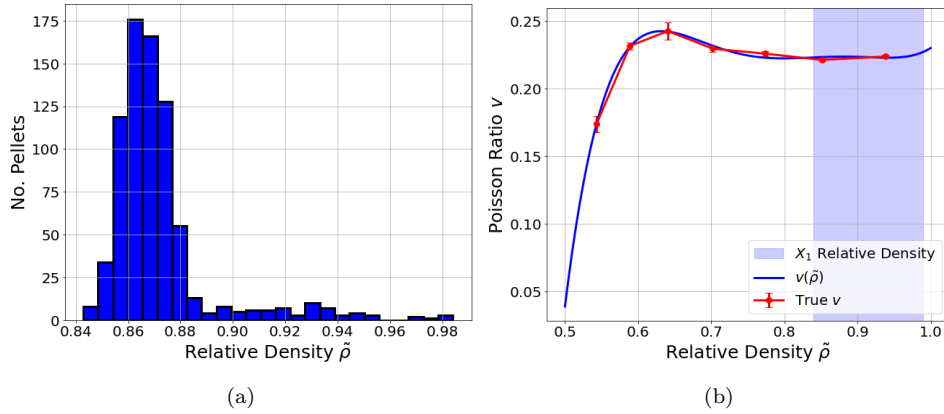


Figure 5: Illustration of data enhancement for X_2 using estimated parameters from Z and Y . (a) Spread of densities of pellets in primary data set X_2 showing a density range contained in the interval $[0.84, 0.99]$. (b) Poisson ratio v as function of relative density $\tilde{\rho}$ computed using Z with polynomial estimate $v(\tilde{\rho})$; mean-value observations from Z (red dots) with 95% confidence intervals and connected via linear interpolation; in blue is a 5th order interpolant that defines $v(\tilde{\rho})$; banded region refers to the pellet density range in X_2 shown in (a).

in Fig. 5b, the polynomial fits the true Poisson ratio well in this range, but has a steep negative gradient as $\tilde{\rho}$ is decreased below 0.6. Although this is not likely to be a good fit for real data due to the physical interpretation of v , for this work one is not concerned with such fit outside of the range of interest determined by X_2 .

At this point, one has generated, via extrapolation, estimates for all features displayed in Table 3 for X_2 (with the exception of α , which is arbitrarily chosen). These estimated features encode information from a significant proportion of the original features in X_2 , which themselves have no established physical interpretation from a first-principles standpoint.

This physics-informed augmentation procedure is motivated by an interest in contextualising the input features to the machine learning models developed next, so that the predictions can be clearly linked to first-principles models of powder compaction. The real gain of this methodology is that one is able to extract an increased value out of data that already exists without the need for additional testing or data generation.

In conclusion, this section shows how to augment X_2 by including the DPC parameters of Table 3. To highlight the gain of the augmentation procedure note that in the next modelling step only the density, hardness, maximum compression F_{max} of X_2 , and the 7 DPC estimated parameters

(justified by the strong relation seen in Fig. 1c) are considered. All other force measurements of X_2 (see Table 1) are discarded: this reduces the dimension of the feature space from more than 70,000 features to just 8.

4. Modelling and Methods: Machine learning modelling

The next sections describe: (i) a feature selection mechanism (Section 4.1) used to ascertain which features highlighted in Table 3 are most relevant to the model and lead to parsimonious models; and, (ii) a description of the machine learning modelling carried out for the task of *quantitative prediction of density* (Section 4.2) and task of *classification of hardness* (Section 4.3).

Our modelling gain: an alternative approach to that used in this paper for the prediction of density would be to inject the estimated DPC parameters into the DPC model itself and solve it for the density. Using ML and statistics gives further interpretability to feature importance loadings and the small number of parameters needed to predict/estimate density (model parsimony). The sole use of DPC abdicates the ability to fine-tune the model scores on train/test and fit/overfit balance and – as is shown below – density can be estimated with a small number of parameters, much smaller than the 7 DPC parameters estimated in the previous section.

4.1. Feature selection process: mRMR

Effective feature selection has a number of benefits, including improved model interpretability [23]. In this work, the Minimum Redundancy and Maximum Relevance (mRMR) selection (introduced in [24]) is deployed. It chooses a subset of the original features based on the maximising the correlation with the target variable (relevance) and minimising shared information within the selected subset (redundancy). Specifically, the algorithm computes the mutual information between each of the original features and the target, and subtracts from this the mean of the pairwise mutual information between the features [25] (see (A.8)). The implementation used is that of [26]. Crucially, mRMR performs better than feature selection mechanisms based solely on Pearson’s correlation.

4.2. Quantitative prediction model for Density

A number of machine learning models for density prediction were developed and analysed, including classical linear regression (LR), partial least squares regression (PLS) [27, 3], support vector machine regression (SVM) [28], and Gaussian process regression (GPR) [29, 30]. All models performed similarly but GPR was able to reduce overall model variance and over-fitting,

capture non-linear relationships in the data, and had the best performance (both in sample and out of sample) in terms of the metrics considered. For brevity, only GPR is discussed as it is not in the scope of this work to compare model performances.

Model description. GPR takes a Bayesian approach to predictive modelling and generates a probabilistic output; a prior is specified, which one then uses with the training data to define a posterior function. The GPR model then finds a probability distribution over all posteriors that fit the data, and predictions can be made using each of these. The mean of the predictions is taken as the point prediction for the GPR model, and one can also use the standard deviations to generate the prediction intervals of each prediction [31]. This provides a reliable uncertainty quantification element for every density prediction, which is a major advantage of this model.

Input features and prediction target. It is known that GPR performs poorly when too many input features (more than 20-30) are used [29]. Using only the DPC estimates and the maximum force F_{\max} as features, a total of 8 features are used – this set is referred to as the *Density Physics-Informed Features* (D-PIF) set. Clearly, this set of features generates a model that is in the efficiency region of GPR with respect to number of input variables. The feature selection mechanism of Section 4.1 reduces these further. The prediction target is the scaled density of the pellets in X_2 .

Model performance metrics. Two metrics are considered: the mean absolute scaled error (MASE) given in (A.9), and the classical statistical coefficient of determination, r^2 . The models’ evaluations is based on the scores on an independent testing set, and metrics are also computed during model training to identify and correct overfitting.

Details on model training and tuning. The data are split into training and testing sets in a 70/30 ratio respectively. The model error is estimated using 5-fold cross-validation [32, Section 7.10]. A grid search on GPR kernels is performed, and the kernel minimising the MASE of the model is implemented. This kernel is the sum of a Matérn kernel with a length scale of 1000 and a smoothing parameter ν of 1.5, and a white noise kernel with noise level 0.5.

4.3. Classification model for Diametrical Hardness

The results from [3, 16], the evident scatter in hardness measurements seen in Figure 1d and initial experimentation showing poor MASE and r^2 metrics, indicate that the *quantitative prediction* of hardness is a significantly more challenging problem than a quantitative prediction of density.

A perspective on the high scatter in hardness data. The literature indicates that catalyst crush strength data are subject to significant scatter [33, 34, 35]. Variability in excess of 25% is commonly seen in published data sets, even where the pellets are taken from an ostensibly homogeneous sample. This is irrespective of the test method used, whether horizontal crush strength (as used for the data set analysed herein), cutting or the three point bend test [36]. The key origin of the variance arises from the random alignment of microcracks, that are propagated in pellet fracture, with the tensile stress lines imposed during the test [37] and the data are commonly fitted using the fracture theory derived Weibull distribution [34, 35, 38, 39]. “Repeat” tests from comparable pellets may however equally fit a Gaussian distribution in accordance with the noted random microcrack orientation [35, 40, 41]. Moreover, it has been shown that a large number of test samples (30-50) is required to confidently fit the parameters of the distribution [38], implying that production environment strength data will always suffer high scatter or low signal to noise ratio which will be difficult to fit by regression without risk of over-fitting

Modelling choices. In order for the hardness model to function as a quality control tool, it is thus not necessary to generate a point prediction of the hardness. A *classification model* is a natural option if one is interested only in knowing whether or not a pellet falls within an acceptable hardness range.

Input features and prediction target. The input data for the classification model are the 8 features used for the quantitative prediction of density discussed in Section 4.2, and additionally the predicted density generated by the GPR model that is denoted by the symbol Γ . A total of 9 features are used initially – this collection is referred to as the *Hardness Physics-Informed Features* (H-PIF) set. The mRMR feature selection algorithm is also applied to these features.

Next, a region of acceptability must be established. Fig. 1d (also Fig. 9c) shows that the majority of pellets have a scaled hardness measurement of approximately -1 to 1 , and in-field informed discussion indicated that this cluster would be a reasonable acceptability region for hardness in X_2 .

Model description. As in Section 4.2, several classification methods were implemented and analysed. Only the Multilayer Perceptron (MLP) model is discussed here as this achieved the best performance according to the relevant metrics (see below). The MLP model belong to the class of (deep) Neural Network models [42]: it contains an input layer, through which the features are passed to the hidden layers. These hidden layers are composed of perceptrons, which are simple artificial neurons that take an input and

calculate a weighted sum, before applying an activation function, and then producing an output [43]. The final layer of a MLP model is known as the output layer. This model is trained efficiently using backpropagation to compute the gradient of the error in the network [42], and iterates standard gradient descent to converge to a minimal loss.

Model performance metrics. Analyses of classification models require a choice of metric to evaluate against. By default, many classifiers use an accuracy score but in practise alternative metrics can be significantly more insightful. An alternative metric, the F1 score, is often more informative as it balances the precision and the recall of the model, which provides a more complete picture of the model performance [43]. This is the metric against which the MLP model is evaluated, and is calculated as the harmonic mean of the precision and the recall. Equation (5) below gives the definition:

$$\begin{aligned} \text{precision} &= \frac{\text{true positives}}{\text{true positives} + \text{false positives}}, \\ \text{recall} &= \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}, \\ \text{F1} &= 2 \left(\frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \right). \end{aligned} \tag{5}$$

This metric gives a high score to classifiers that perform well in terms of both precision and recall [43]. This is useful since one would like to minimise both the false negatives and the false positives. False negatives refer to pellets that are classed as not acceptable when in fact they are fit for purpose; these should be minimised because the model will be used to flag problems in the compaction process, and a high rate of false negatives reduces trust in the model and may obscure genuine errors. On the other hand, false positives are pellets that are not acceptable, but the model determines that they are. These misclassified pellets would be at risk of fracturing during packaging or transportation, or may be too hard to dissolve or react when used. Both errors in the model have significant impact and the F1 score in equation (5) provides a good solution balancing the precision-recall trade-off.

Details on model training and tuning. A number of parameters must be specified prior to training the MLP model. These choices are detailed in Table 4. Since a logistic activation function is used, the cross-entropy loss (also known as the logistic loss) function is minimised [44]. With this setup, the model is trained and outputs a binary classification of ‘1’ for a pellet that falls in the acceptability region, and ‘0’ otherwise.

Number of hidden layers	1
Size of hidden layers	50
Learning rate	0.01
Activation function	Logistic
Epochs	100

Table 4: MLP parameters for hardness classification.

5. Results

As discussed throughout the previous sections, understanding feature interpretability and relevance is a primary motivation for this work. In particular, the aim is to identify the smallest possible subset of physically meaningful features that produces a well-performing, parsimonious model according to the specified metrics. For each problem the feature selection and relevance is firstly discussed and the results of the model itself are then presented.

5.1. Model for density prediction

5.1.1. Feature selection

Fig. 6a shows the Pearson correlation between the 8 features and both targets (density and hardness); this measure indicates how strongly related the features are in terms of their linear relationship, and suggests which may be most relevant for predicting a given target [45]. This correlation analysis is limited to linear relationships in the data and more complex nonlinear relationships are not highlighted here. One sees that the strongest positive correlation in the variables is between the density and the maximum compression force, F_{max} , with the strongest negative correlation occurring between density and the Young’s modulus, E .

A further notable finding from Fig. 6a is that the Poisson ratio, ν , has low correlation with the density. This suggests that this feature is the least relevant of the D-PIF set when building a predictive model for density predictions.

Section 4.1 describes how the Pearson correlation can quantify the relevance of each feature and weight this against the measured redundancy in order to select the best feature set. Fig. 6 displays the effect of feature selection on the model performance on the left (Fig. 6b) and the mRMR feature relevance of all D-PIF on the right (Fig. 6c).

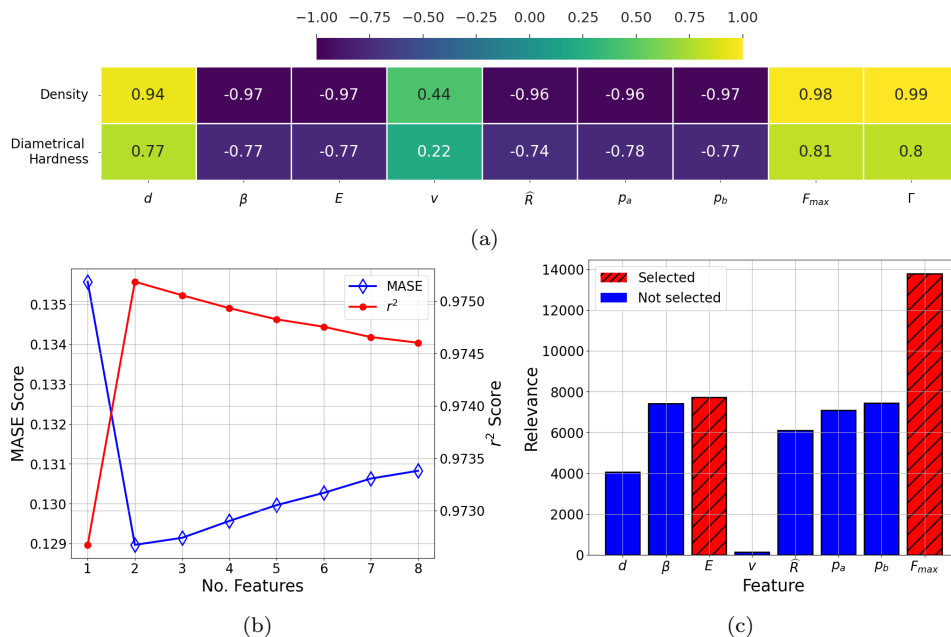


Figure 6: (a) Pearson correlation between D-PIF and H-PIF features and target labels. The last element, Γ , is the predicted density from the GPR model that belongs to the H-PIF set. (b) Optimal number of features for GPR model against MASE (red, on left y -axis) and r^2 (blue, on right y -axis). Both of these metrics are computed using the test set. (c) D-PIF feature relevance using mRMR feature selection for GPR model (right). Feature relevance plots show the features selected by mRMR for the model (red). See Table 3 for feature symbols and descriptions.

For density prediction, selecting only 2 features improves both the MASE and r^2 score for the GPR model. As suggested by Fig. 6a, the most relevant features are the Young’s modulus, E , and the maximum compression force F_{max} . These are exactly the two features selected by mRMR, which indicates that there is minimal shared information (redundancy) between these variables. The selection of F_{max} is in line with the findings of [15, 3], and the strong correlation highlighted by Fig. 1c. This latter feature has a clear physical interpretation and is easily controlled during manufacture.

The Young’s Modulus is a pellet property that is known to cross-correlate with pellet porosity [46, 47]. When the feed material is constant, as it is in this manuscript, Young’s Modulus makes for a (mathematically) good “soft sensor” or marker for density. Zhang et al. [47] in a univariate analysis of experimentally derived data note additionally that the hydrostatic

yield stress (p_b), internal friction b and Cap Eccentricity (R , related to the evolution pressure, p_a) all show a significant monotonic relationship with pellet density and will therefore be statistically confounded. This may explain why the feature importance ranking in this study indicates only minor discrimination between their influence (Fig.6c).

5.1.2. Model performance

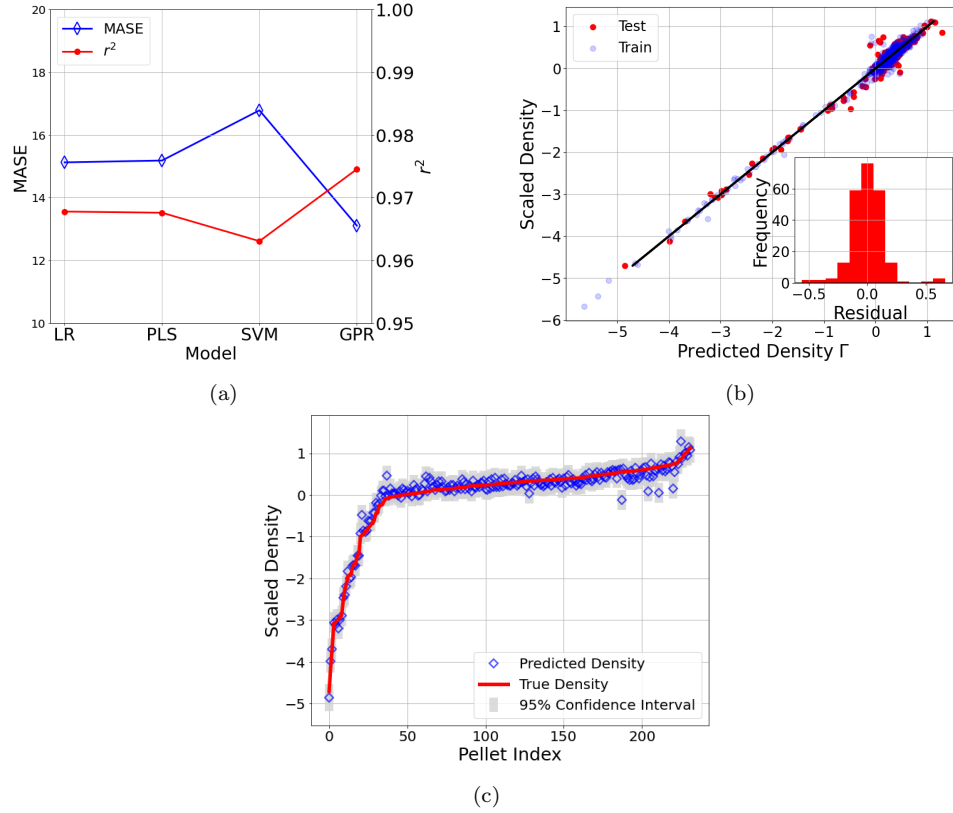


Figure 7: (a) Scores on test set by model (LR, PLS, SVM, GPR) for scaled density prediction against MASE (red, on left y -axis) and r^2 (blue, on right y -axis). (b) GPR parity plot for testing (red) and training (blue) data, with residual histogram inset. (c) GPR mean predictions and 95% prediction intervals (on test set).

Fig. 7a summarises the performance of several models for density prediction for both MASE and r^2 , trained on only the estimated DPC features (Table 3) and maximum compression force. The results from LR, PLS and SVM models are displayed without detail and for reference only. It is clear that the GPR model is the better performing model in terms of the MASE

score on the independent testing, achieving a 13.1% scaled error. Comparatively, the simple LR model has an MASE of 15.1%. The GPR parity plot is displayed in Fig. 7b and compares the density prediction to the actual measured density. The black line indicates a 1:1 parity and provides a reference point for the quality of the predictions. That is, the closer to the line a point falls, the closer the predicted density is to the true value. Residual errors are shown in the histogram inset in Fig. 7b, and it is clear that these errors are distributed approximately normally with zero mean – this is indicative of a good model fit. Fig. 7c shows the prediction intervals for GPR, which are smaller than those of the other models implemented. This indicates that the prediction generated by GPR has less uncertainty associated with it. The relatively low level of uncertainty also promotes explainability in the results.

Table 5 summarises the findings. It indicates that choosing the 2 features (F_{\max} and E) suffices for improved model performance in terms of MASE. This is likely due to correlations and noise present in the other features, which can cause overfitting, so removing the extra features improves the model’s ability to generalise to testing data. The number of features seems to have no significant impact on the model’s r^2 score. Finally, the last row of Table 5 is discussed in remark 5.1.3.

Features	MASE (%)	MAE	r^2
2 D-PIF features (Section 5.1.1)	12.9	0.10	0.975
All 8 D-PIF features	13.1	–	0.975
8 (mRMR selected) compaction features	14.3	–	0.970

Table 5: Comparison of GPR scores on test set for different feature sets. MAE reported only for the best model for comparison purposes. All models take less than 15 seconds to train. See Section 5.1.1 for further details; last row discussed in Section 5.1.3.

5.1.3. Remark: using statistical features Vs physics-informed features

The last row of Table 5 shows, for comparison purposes, the metrics when using only compaction data as input, where ‘compaction data’ refers to statistical features generated from only the original data in X_2 (see Table 1); this is in line with what was done in [3]. An in-depth discussion of said statistical features is out of the scope of this work, but the summary Table A.6 provides an overview. It is not feasible to train a GPR model on all compaction features due to the computational effort required, so the mRMR feature selection mechanism is used to shrink the feature space to 8 elements. [The 8 features selected from this set of input features are also detailed in A.6.](#)

Selecting 8 features from the compaction data gives a clear comparison for the model performance to when using the D-PIF set. Although the model with compaction features has relatively good performance with a MASE of 14.3% and r^2 of 0.970, these scores show that the D-PIF input generates better predictions. Crucially, Table 5 demonstrates that one is able to derive physically relevant features from existing data and train a maximally explainable model, whilst also improving upon model performance.

5.2. Model for Hardness classification

5.2.1. Feature selection

The analysis for feature selection for the hardness classifier is similar to that presented in Section 5.1.1 for the density prediction task. Fig. 6a shows that, for this data, the D-PIF features are more significantly related to the density than to the hardness. This may relate to the observation that a quantitative hardness prediction is a more difficult problem than for density [3], as the data does not contain a sufficient amount of information. The Poisson ratio, ν , is weakly correlated with the hardness compared to the other H-PIF features; the same was found when considering density correlations.

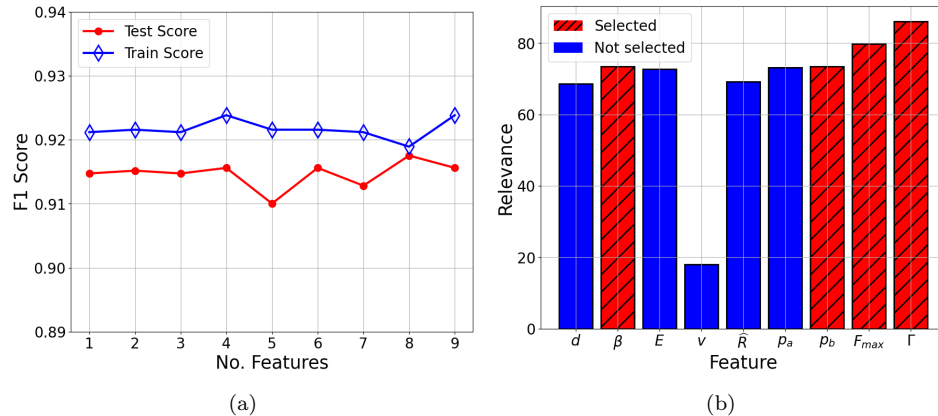


Figure 8: (a) F1 score (y -axis) as function of the number of features (x -axis) used for training; Test score in red dots; Train score in blue crosses. (b) H-PIF feature relevance including the predicted density from the GPR model, Γ , using mRMR feature selection for the MLP classifier. Feature relevance plots show the features selected by mRMR for the model (red). See Table 3 for feature descriptions.

This observation is also found in Fig. 8b, which illustrates that the Poisson ratio has minimal mRMR relevance, so it is clear that this feature is not informative for either prediction task. The F1 score for the classification

model is not impacted significantly by reducing the number of features (see Fig. 8a). Therefore one chooses 4 features, in order to reduce the feature space by 55%. Again, F_{max} is selected. The selection of F_{max} and Young’s modulus, E , for the model is particularly interesting as F_{max} is a quantity that is easily monitored and adjusted *in situ* while Young’s modulus is a pellet property.

From a theoretical point of view, Griffiths’ Fracture Theory [46], and the eponymous equation which shows the fracture tensile stress $\sigma_F \propto \sqrt{E}$, it is not surprising that Young’s modulus demonstrates a strong influence. Evidently though, the particle-particle powder friction, β , and onset of plastic deformation during compaction, p_b , are able to capture this important relationship statistically just as well (see Figure 8b).

5.2.2. Model performance

Training the MLP classifier yields an F1 score of 0.915 on the independent test set, which signifies that there are very few false negatives or false positives identified. This indicates that the model has very good predictive capabilities in terms of accurately classifying pellets according to their hardness.

It is known that machine learning models can be susceptible to overfitting [48], which is characterised by scores that are significantly better on the training data than on the validation set. This occurs when the model learns the training data too well, for example by learning the noise present in those data. Two ways to reduce the likelihood of overfitting are to reduce the complexity of the model, and to reduce the number of observations in the training set. A single hidden layer was chosen in Section 4.3 for simplicity, and Fig. 9a shows that overfitting does not appear to be a problem. However, one notes that using 242 training instances produces a model that generalises well to the validation set, with a sufficiently narrow confidence interval on the F1 score. It is clear that additional training data has little impact on the performance of the classifier so one selects 242 pellets randomly to train the MLP model. This promotes computational efficiency in the model without compromising on performance.

Fig. 9b shows the confusion matrix for the trained MLP model. This illustrates the errors made by the model, and it is noteworthy to observe that the most common error is a false positive. That is, when the model makes an error, it is more likely that this mistake classifies an out of specification pellet as acceptable. The false positives identified in the validation set are shown in Fig. 9c. This highlights the fact that the majority of false positives lie above the acceptability region, with a cluster of such errors falling just

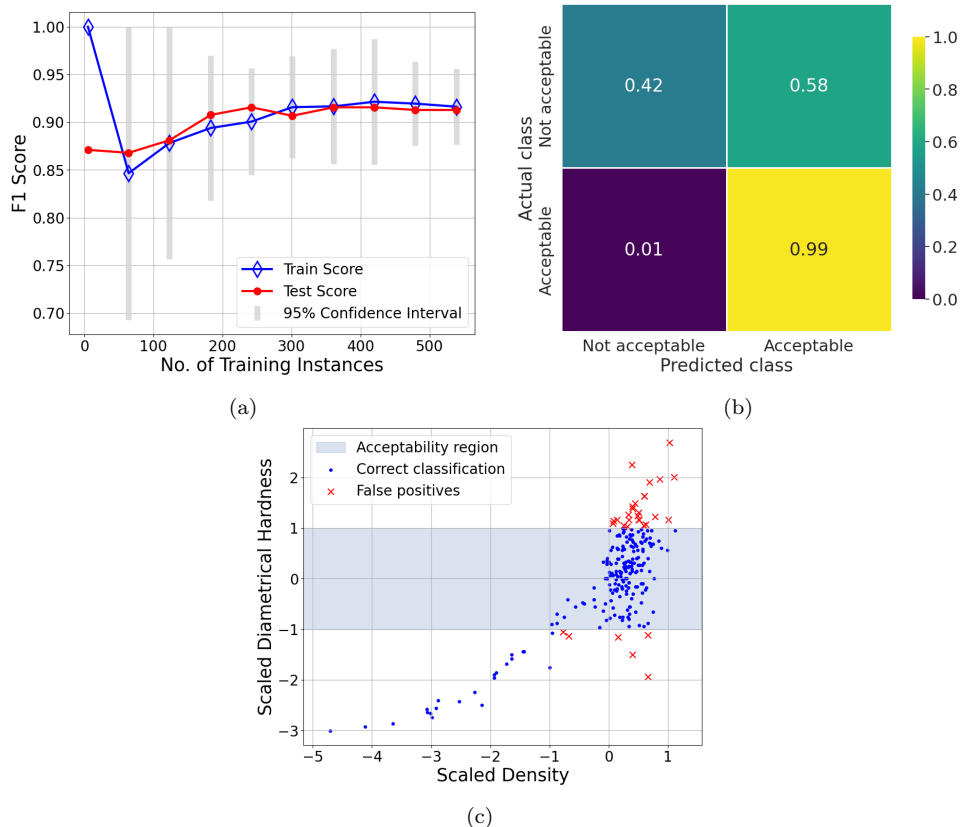


Figure 9: (a) MLP model sensitivity to training data set size by F1 score with 95% confidence intervals. (b) Confusion matrix for MLP hardness classifier. (c) Pellets falsely identified as acceptable (red crosses) by MLP model, pellets correctly identified (blue dots), and acceptability region shaded.

outwith the boundary.

5.3. Comparison with the literature

The main data set of this work, X_2 , is also the data set X_2 from [3] but in [3] neither Z or Y were used. In [3], the authors use Partial Least Squares (PLS) Regression, a covariance based regression model, to quantitatively predict scaled density and hardness. The input for the model are the complete force profiles of X_2 (see the first four lines in Table 1), and during model training the inputs are reduced via PLS's Variable Importance in Projection scores. Critically, the features generated by the PLS are latent variables, which diminishes the understanding of why certain parts of the

time-series are selected and how the selected features relate to rejected features. For quantitative prediction of scaled density, their PLS model with 4 latent variables reports MAE between 0.09 and 0.18, and $r^2 = 0.86$. Comparatively, the 2-feature GPR model in this work reports an MAE of 0.10, and $r^2 = 0.975$ (and MASE of 12.9). No comparison can be established between the hardness models as here one deals with a *classification* problem while there it is a *quantitative prediction* problem.

The main advantage that the model here offers in terms of the *quantitative prediction of density* is a construction based on physics-informed features ultimately motivated by an interest in contextualising the input features to the models. The models in [3] cannot provide for this.

6. Conclusions

In this manuscript two predictive models were presented as well as a reproducible mechanism to enhance data sets by leveraging smaller ones from independent experiments. Regarding the models, the first is for the quantitative prediction of pellet density and the second is for the classification of pellet hardness as ‘in-specification’ or ‘out-of-specification’. The analysis provided demonstrates that it is possible to derive physically relevant features from existing data and train a maximally explainable model, whilst also improving upon model performance – the model selection of compaction forces and peak force is in-line with either DPC literature or early data-driven modelling [3, 16]. The models developed improve upon significantly existing findings and provide insights into the relationships between compaction process parameters and pellet CQAs (density and hardness). This is a step towards optimisation of the pelleting process parameters to achieve reduced variability in pellet density and hardness. Critically, the models were built consistently with relevant quality control metrics of high-recall (minimising false negatives) and improved precision (minimising false positives). Our models are the most competitive data-driven physics-informed reduce-order models found in current literature (to the best of the authors’ knowledge), seen by either low errors, better explainability or a reduced amount of training data.

From an industrial perspective, the models developed here are inline with the trend of sustainable advanced manufacturing aiming to capture complex materials’ properties and mechanical processing requirements. These models deliver on the possibility of controlling the process from design to manufacturing for less well-studied material, and thus provide savings of

the high-value materials, energy consumption, working risks and other eco-environmental factors of practical trial and error. The mantra being the same as in [3]: “Control of quality leads to improved economics and sustainability”.

From a practical perspective, the information provided by these models can be deployed to supervisor plant operators or embedded in an automated control system to guide and monitor the production process ensuring on-specification outputs. From a holistic point of view, such models help overcome the challenges of scale-up and scale-down by being digital-twin component proxies to the multi-variables across the production scales. Left open is further research into the impact of the processes upstream of pelletizing, such as roller compaction and granulation, and the influence of formulation and manufacturing process disturbances. These would help to bridge the gap between R&D and commercial scale production.

Appendix A. Methods

Appendix A.1. Drucker-Prager Cap Equations

The equations for the DPC surfaces described in Section 3 are given in this Appendix. See Table 3 for descriptions of the parameters used in the next equations (A.1)–A.3. The shear failure surface, F_s , is written as

$$F_s = q - p \tan \beta - d = 0. \quad (\text{A.1})$$

The cap surface, F_c , is described by the following equation

$$F_c = \sqrt{(p - p_a)^2 + \left[\frac{\widehat{R}q}{1 + \alpha - \alpha/\cos \beta} \right]^2} - \widehat{R}(d + p_a \tan \beta) = 0. \quad (\text{A.2})$$

Finally, the transition surface, F_t , is

$$F_t = \sqrt{|(p - p_a)|^2 + \left[q - \left(1 - \frac{\alpha}{\cos \beta} \right) (d + p_a \tan \beta) \right]^2} - \alpha(d + p_a \tan \beta) = 0. \quad (\text{A.3})$$

One finds next the formulae to calculate the Young’s modulus and the Poisson ratio. These quantities require the calculation of the bulk modulus K and shear modulus G , which is done by solving the equations below

$$K + \frac{4}{3}G = \frac{\sigma_{zB} - \sigma_{zC}}{\epsilon_{zB} - \epsilon_{zC}} \quad \text{and} \quad \frac{2G}{\sqrt{3}K} = \frac{q_B}{p_B - p_C}. \quad (\text{A.4})$$

Recall that σ_z and ϵ_z are the axial stress and strain respectively, and then σ_{zB} is σ_z at point B in Fig. 3a. The quantities σ_{zC} , ϵ_{zB} and ϵ_{zC} are defined similarly, namely, as the stress and strain at points B and C . Also, p_B and p_c are the hydrostatic pressure p at points B and C from Fig. 3b respectively.

The Young's modulus, E , and the Poisson ratio, ν , are computed through

$$E = \frac{9GK}{3K + G}, \quad \text{and} \quad \nu = \frac{3K - 2G}{2(3K + G)}. \quad (\text{A.5})$$

Define $\hat{\alpha}$ as

$$\hat{\alpha} = 1 + \alpha - \alpha / \cos \beta.$$

Then the cap equations are as shown:

$$p_a = \frac{-3q_B - 4d\hat{\alpha}^2 \tan \beta + \sqrt{9q_B^2 + 24dq_B\hat{\alpha}^2 \tan \beta + 8\hat{\alpha}^2 (3p_Bq_B + 2q_B^2) \tan^2 \beta}}{4\hat{\alpha}^2 \tan^2 \beta}, \quad (\text{A.6})$$

$$\hat{R} = \sqrt{\frac{2\hat{\alpha}^2 (p_B - p_a)}{3q_B}} \quad \text{and} \quad p_b = \hat{R}d + p_a (1 + \hat{R} \tan \beta). \quad (\text{A.7})$$

Appendix A.2. Machine Learning elements

Feature generation: statistical feature from compaction data

Observing the compaction data in X_2 , it is clear that the force profiles of different pellets have varying skews, and the gradient of the compression force at each time point is steeper when the maximum force recorded is larger. For these reasons, one generates statistical features from the curve profiles in X_2 (e.g., Fig. 1b) that correlate with the density variable. Table A.6 overviews said generated features.

Table 5 contains the results of applying the ML model to the statistical features. In particular, the mRMR algorithm is applied to a set of features containing exactly the measurements from dataset X_2 and the features in Table A.6. For comparison with the main result of this work, 8 features are selected, and these are given below, in order of relevance (as determined by mRMR):

1. Upper punch force skew,
2. Lower punch displacement 2597.4,
3. Lower punch displacement 2597.3,
4. Lower punch displacement 2597.8,

Feature	Formula
Skew	$\frac{\sum_i^N (X_i - \mu)^3}{(N-1)\sigma^3}$
Kurtosis	$\frac{\sum_i^N (X_i - \mu)^4}{(N-1)\times\sigma^4}$
Area under curve	$\int f(x)dx$
Squared feature	$f(x)^2$
Derivative (first and second order, force only)	$\frac{df}{dx}, \frac{d^2f}{dx^2}$

Table A.6: Generated features for X_2 . In the above, μ is the mean value of the feature, with standard deviation σ , and the number of observations is N . For each variable, $f(x)$ represents the profile of the variable over all relevant time points x .

5. Lower punch displacement 2597.9,
6. Lower punch displacement 2597.7,
7. Lower punch displacement 2597.5,
8. Lower punch displacement 2597.6.

Note that the number refers to the time at which the measurement is recorded (in ms), and that in this case maximum compression occurs at time 2595.1. That is, the punch force features selected appear to relate to post-compression relaxation.

Details on the mRMR Feature Selection algorithm

The mRMR algorithm utilises the F-statistic, $F(\cdot, \cdot)$ and the Pearson correlation, $\rho(\cdot, \cdot)$, to rank features according to their importance score, f^{mRMR} as shown [23]:

$$f^{\text{mRMR}}(X_i) = F(Y, X_i) - \frac{1}{|S|} \sum_{X_s \in S} \rho(X_s, X_i), \quad (\text{A.8})$$

where X_i for $i \in \{1, \dots, m\}$ are the m original features, Y is the target, and S is the set of selected features. This version of mRMR is known as FCD and is implemented for continuous features.

To select K features, one computes the score of each candidate feature using (A.8) and then adds the feature with the highest score to the set of selected features, S . This selected feature is removed from the candidate set. These steps are performed iteratively until $|S| = K$, at which point the algorithm terminates and the output S is presented as the selected feature subset.

Performance metrics

The metrics considered in this study are the coefficient of determination, r^2 , mean absolute error (MAE), mean absolute scaled error (MASE). They are defined as follows [49, Section 3.4]:

$$\left\{ \begin{array}{l} r^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\ \text{MAE}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \\ \text{MASE}(y, \hat{y}) = 100 \times \frac{\text{MAE}(y, \hat{y})}{1/n \sum_{i=1}^n |y_i - \bar{y}|} \end{array} \right. , \quad (\text{A.9})$$

where y_i, \hat{y}_i are the actual and predicted values for sample i respectively, n is the number of samples, and \bar{y} is the mean of the actual values.

References

- [1] J. Zhu, W. Li, T. Wierzbicki, Y. Xia, J. Harding, Deformation and failure of lithium-ion batteries treated as a discrete layered structure, *International Journal of Plasticity* 121 (2019) 293–311.
- [2] G. V. Research, Catalyst market size, share & trends analysis report by raw material (chemical compounds, zeolites, metals), by product (heterogeneous, homogeneous), by application, by region, and segment forecasts, 2020 - 2027, <https://www.grandviewresearch.com/industry-analysis/catalyst-market> (2020).
- [3] J. Emerson, V. Vivacqua, H. Stitt, Data-driven modelling of a pelleting process and prediction of pellet physical properties, *Johnson Matthey Technology Review* 66 (3) (2022) 1245. doi:10.1595/205651322X16499427403168.
- [4] C. Sinka, Modelling powder compaction, *KONA Powder and Particle Journal* 25 (2007) 4–22.
- [5] I. Aydin, B. Briscoe, N. Ozkan, Modeling of powder compaction: A review, *MRS Bulletin* 22 (12) (1997) 45–51. doi:10.1557/S0883769400034746.
- [6] J. Cunningham, I. Sinka, A. Zavaliangos, Analysis of tablet compaction. I. characterization of mechanical behavior of powder and powder/tooling friction, *Journal of pharmaceutical sciences* 93 (8) (2004) 2022–2039.

- [7] I. Sinka, J. Cunningham, A. Zavaliangos, Analysis of tablet compaction. II. finite element analysis of density distributions in convex tablets, *Journal of pharmaceutical sciences* 93 (8) (2004) 2040–2053.
- [8] J.-F. Jerier, B. Hathong, V. Richefeu, B. Chareyre, D. Imbault, F. Donze, P. Doremus, Study of cold powder compaction by using the discrete element method, *Powder Technology* 208 (2011) 537–541. doi:10.1016/j.powtec.2010.08.056.
- [9] P. A. Cundall, O. D. Strack, A discrete numerical model for granular assemblies, *geotechnique* 29 (1) (1979) 47–65.
- [10] V. K. Ojha, S. Schiano, C.-Y. Wu, V. Snášel, A. Abraham, Predictive modeling of die filling of the pharmaceutical granules using the flexible neural tree, *Neural Computing and Applications* 29 (7) (2018) 467–481.
- [11] G. Shi, L. Lin, Y. Liu, G. Chen, Y. Luo, Y. Wu, H. Li, Pharmaceutical application of multivariate modelling techniques: a review on the manufacturing of tablets, *RSC Adv.* 11 (2021) 8323–8345. doi:10.1039/D0RA08030F.
- [12] G. M. Khalid, A. Usman, Application of data-intelligence algorithms for modeling the compaction performance of new pharmaceutical excipients, *Future Journal of Pharmaceutical Sciences* 7 (12 2021). doi:10.1186/s43094-021-00183-w.
- [13] J. Djuris, S. Cirin-Varadjan, I. Aleksic, M. Djuris, S. Cvijic, S. Ibric, Application of machine-learning algorithms for better understanding of tableting properties of lactose co-processed with lipid excipients, *Pharmaceutics* 13 (5) (2021). doi:10.3390/pharmaceutics13050663.
- [14] Z. Huang, S. C. Galbraith, B. Cha, H. Liu, S. Park, M. H. Flamm, M. Metzger, A. Tantuccio, S. Yoon, Effects of process parameters on tablet critical quality attributes in continuous direct compression: a case study of integrating data-driven statistical models and mechanistic compaction models, *Pharmaceutical Development and Technology* 25 (10) (2020) 1204–1215. doi:10.1080/10837450.2020.1805760.
- [15] N. Millen, A. Kovačević, J. Djuriš, S. Ibrić, Machine learning modeling of wet granulation scale-up using particle size distribution characterization parameters, *Journal of Pharmaceutical Innovation* 15 (4) (2020) 535–546.

- [16] J. Emerson, V. Vivacqua, H. Stitt, Erratum: Data-driven modelling of a pelleting process and prediction of pellet physical properties: Control of quality leads to improved economics and sustainability, *Johnson Matthey Technology Review* 66 (3) (2022) 245–245.
- [17] I. C. Sinka, K. G. Pitt, A. C. F. Cocks, Chapter 22 the strength of pharmaceutical tablets, in: A. D. Salman, M. Ghadiri, M. J. Hounslow (Eds.), *Particle Breakage*, Vol. 12 of *Handbook of Powder Technology*, Elsevier Science B.V., 2007, pp. 941–970. doi:[https://doi.org/10.1016/S0167-3785\(07\)12025-X](https://doi.org/10.1016/S0167-3785(07)12025-X). URL <https://www.sciencedirect.com/science/article/pii/S016737850712025X>
- [18] L. Han, J. Elliott, A. Bentham, A. Mills, G. Amidon, B. Hancock, A modified Drucker-Prager Cap model for die compaction simulation of pharmaceutical powders, *International Journal of Solids and Structures* 45 (10) (2008) 3088–3106. doi:10.1016/j.ijsolstr.2008.01.024.
- [19] J. C. Cunningham, Experimental studies and modeling of the roller compaction of pharmaceutical powders, Ph.D. thesis, Drexel University (2005).
- [20] L. Perez-Gandarillas, A. Mazor, O. Lecoq, A. Michrafy, Compaction properties of dry granulated powders based on drucker–prager cap model, *Powder Technology* 337 (2018) 43–50.
- [21] V. Buljak, S. Baivier-Romero, A. Kallel, Calibration of Drucker-Prager Cap constitutive model for ceramic powder compaction through inverse analysis, *Materials* 14 (2021) 4044. doi:<https://doi.org/10.3390/ma14144044>.
- [22] J. P. K. Seville, C.-Y. Wu, *Particle Technology and engineering: An engineer’s guide to particles and powders: Fundamentals and computational approaches*, Elsevier/Butterworth-Heinemann, 2016.
- [23] Z. Zhao, R. Anand, M. Wang, Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform (2019). doi:10.48550/ARXIV.1908.05376.
- [24] C. Ding, H. Peng, Minimum redundancy feature selection from microarray gene expression data, *Journal of bioinformatics and computational biology* 3 (02) (2005) 185–205.

- [25] A. Tsanas, Relevance, redundancy, and complementarity trade-off (rrct): A principled, generic, robust feature-selection tool, *Patterns* 3 (5) (2022) 100471. doi:<https://doi.org/10.1016/j.patter.2022.100471>.
- [26] S. Mazzanti, mrmr, <https://github.com/smazzanti/mrmr> (2021).
- [27] V. E. Vinzi, W. W. Chin, J. Henseler, H. Wang (Eds.), *Handbook of partial least squares*, Springer Handbooks of Computational Statistics, Springer, Heidelberg, 2010, concepts, methods and applications. doi: 10.1007/978-3-540-32827-8.
- [28] A. J. Smola, B. Schölkopf, A tutorial on support vector regression, *Statistics and Computing* 14 (2004) 199–222.
- [29] C. E. Rasmussen, Gaussian processes in machine learning, in: O. Bousquet, U. von Luxburg, G. Rätsch (Eds.), *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2 - 14, 2003, Tübingen, Germany, August 4 - 16, 2003, Revised Lectures*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 63–71. doi:10.1007/978-3-540-28650-9_4.
- [30] C. E. Rasmussen, C. K. I. Williams, *Gaussian processes for machine learning*, Adaptive Computation and Machine Learning, MIT Press, Cambridge, MA, 2006.
- [31] J. Wang, An intuitive tutorial to Gaussian processes regression, arXiv preprint arXiv:2009.10862 (2020).
- [32] T. Hastie, R. Tibshirani, J. Friedman, *The elements of statistical learning*, 2nd Edition, Springer Series in Statistics, Springer, New York, 2009, data mining, inference, and prediction. doi:10.1007/978-0-387-84858-7.
- [33] S. B. Bhowmik, *Physical characteristics of LT shift catalysts.*, University of London, University College London (United Kingdom), 1991.
- [34] Y. Li, J. Zhao, L. Chang, Factors analysis for mechanical strength in pelleting process of fe-based high temperature shift catalyst, in: *Studies in Surface Science and Catalysis*, Vol. 63, Elsevier, 1991, pp. 145–153.
- [35] M. Zakeri, A. Samimi, M. S. Afarani, A. Salehirad, Interaction between weibull parameters and mechanical strength reliability of industrial-scale water gas shift catalysts, *Particuology* 32 (2017) 160–166.

- [36] Y. Li, D. Wu, J. Zhang, L. Chang, D. Wu, Z. Fang, Y. Shi, Measurement and statistics of single pellet mechanical strength of differently shaped catalysts, *Powder technology* 113 (1-2) (2000) 176–184.
- [37] Y. Li, X. Li, L. Chang, D. Wu, Z. Fang, Y. Shi, Understandings on the scattering property of the mechanical strength data of solid catalysts: A statistical analysis of iron-based high-temperature water-gas shift catalysts, *Catalysis today* 51 (1) (1999) 73–84.
- [38] D. Wu, Y. Li, J. Zhang, L. Chang, D. Wu, Z. Fang, Y. Shi, Effects of the number of testing specimens and the estimation methods on the weibull parameters of solid catalysts, *Chemical engineering science* 56 (24) (2001) 7035–7044.
- [39] Y. Li, D. Wu, J. Zhang, L. Chang, D. Wu, Z. Fang, Y. Shi, Measurement and statistics of single pellet mechanical strength of differently shaped catalysts, *Powder technology* 113 (1-2) (2000) 176–184.
- [40] A. Samimi, M. Zakeri, B. Maleki, D. Mohebbi-Kalhor, Experimental and statistical assessments of the mechanical strength reliability of gamma alumina catalyst supports, *Particuology* 21 (2015) 74–81.
- [41] J. M. Sonnergaard, Distribution of crushing strength of tablets, *European journal of pharmaceutics and biopharmaceutics* 53 (3) (2002) 353–359.
- [42] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, 2016, <http://www.deeplearningbook.org>.
- [43] G. Aurélien, *Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems*, O’Reilly Media, Sebastopol, CA, 2017.
- [44] H. Masnadi-shirazi, N. Vasconcelos, On the design of loss functions for classification: theory, robustness to outliers, and SavageBoost, in: D. Koller, D. Schuurmans, Y. Bengio, L. Bottou (Eds.), *Advances in Neural Information Processing Systems*, Vol. 21, Curran Associates, Inc., 2008, p. 8 pages.
URL <https://proceedings.neurips.cc/paper/2008/file/f5deaeae1538fb6c45901d524ee2f98-Paper.pdf>
- [45] J. J. Berman, Chapter 4 - understanding your data, in: J. J. Berman (Ed.), *Data Simplification: Taming Information with Open Source*

Tools, Morgan Kaufmann, Boston, 2016, pp. 135–187. doi:<https://doi.org/10.1016/B978-0-12-803781-2.00004-7>.

- [46] F. Knudsen, Dependence of mechanical strength of brittle polycrystalline specimens on porosity and grain size, *Journal of the American Ceramic Society* 42 (8) (1959) 376–387.
- [47] B. Zhang, M. Jain, C. Zhao, M. Bruhis, R. Lawcock, K. Ly, Experimental calibration of density-dependent modified drucker-prager/cap model using an instrumented cubic die for powder compact, *Powder Technology* 204 (1) (2010) 27–41. doi:<https://doi.org/10.1016/j.powtec.2010.07.003>.
URL <https://www.sciencedirect.com/science/article/pii/S0032591010003402>
- [48] S. Lawrence, C. Giles, Overfitting and neural networks: Conjugate gradient and backpropagation, in: *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, Vol. 1, 2000, pp. 114–119 vol.1. doi:10.1109/IJCNN.2000.857823.
- [49] R. J. Hyndman, G. Athanasopoulos, *Forecasting: principles and practice*, OTexts, 2018.

Funding

This research received partial funding by Johnson Matthey under an iCASE PhD studentship.

M. Trower is a PhD student in EPSRC’s MAC-MIGS Centre for Doctoral Training. MAC-MIGS is supported by the UK’s Engineering and Physical Science Research Council (grant number EP/S023291/1).

G. dos Reis acknowledges partial support from the *Fundação para a Ciência e a Tecnologia* (Portuguese Foundation for Science and Technology) through the project UIDB/00297/2020 and UIDP/00297/2020 (Center for Mathematics and Applications, CMA/FCT/UNL).

Acknowledgements

Both M. Trower and G. dos Reis would like to thank Calum Strange and Rasheed Oyewole Ibraheem (both from the University of Edinburgh) for the helpful discussions around machine learning modelling and its deployment.

All the authors thank Dr. Karen Huang (Principal Scientist at Johnson Matthey) for the data sets Y and Z .

Author contributions

All authors provided domain expertise, edited and reviewed the manuscript. **M.T.:** methodology, software, visualisation, data curation; writing, review and editing.

J.E., M.Y.: conceptualisation, methodology; writing, review and editing.

V.V.: conceptualisation, methodology; review.

T.J., H.S.: supervision

G.d.R.: conceptualisation, methodology, supervision, visualisation; writing, review and editing.

Competing interest declaration

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

The data used for this research is proprietary to Johnson Matthey, and remains confidential outwith. For commercial reasons, the specific type of catalyst cannot be disclosed including compositional, morphological, or intended reaction duty characteristics. However, we have added some details concerning the nature of the material.

Additional information

This work has no supplementary information file.