Journal Pre-proofs

Scale-independent solid fraction prediction in dry granulation process using a gray-box model integrating machine learning model and Johanson model

Kanta Sato, Shuichi Tanabe, Keita Yaginuma, Susumu Hasegawa, Manabu Kano

PII: S0378-5173(25)00193-0
DOI: https://doi.org/10.1016/j.ijpharm.2025.125357
Reference: IJP 125357

To appear in: *International Journal of Pharmaceutics*

Received Date: 22 December 2024
Revised Date: 28 January 2025
Accepted Date: 12 February 2025

Please cite this article as: K. Sato, S. Tanabe, K. Yaginuma, S. Hasegawa, M. Kano, Scale-independent solid fraction prediction in dry granulation process using a gray-box model integrating machine learning model and Johanson model, *International Journal of Pharmaceutics* (2025), doi: https://doi.org/10.1016/j.ijpharm.2025.125357

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2025 The Author(s). Published by Elsevier B.V.

Scale-independent solid fraction prediction in dry granulation process using a gray-box model integrating machine learning model and Johanson model

Kanta Sato[a,b]*, Shuichi Tanabe[a], Keita Yaginuma[a], Susumu Hasegawa[a], Manabu Kano[b]

[a] Formulation Technology Research Laboratories, Daiichi Sankyo Co., Ltd., 1-12-1, Shinomiya, Hiratsuka 2540014, Kanagawa, Japan

[b] Department of Informatics, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto 6068501, Kyoto, Japan

*Corresponding author at Formulation Technology Research Laboratories, Daiichi Sankyo Co., Ltd., 1-12-1, Shinomiya, Hiratsuka 2540014, Kanagawa, Japan

E-mail address: kanta.sato@daiichisankyo.com (K. Sato).

Journal Pre-proofs

Abstract

We propose a novel approach for predicting the solid fraction after roller compaction processes. It is crucial to predict and control the solid fraction, as it has a significant impact on the product quality. The solid fraction can be theoretically predicted by a first-principles model developed by Johanson. The Johanson model, however, cannot be directly used for solid fraction prediction after roller compaction because it requires the values of preconsolidation properties, i.e., pressure and solid fraction before compaction, which cannot be measured with standard equipment. In this work, we developed a statistical model that predicts the newly defined preconsolidation parameter, which reflects preconsolidation properties, from the powder's material properties and the process parameters. Then we integrated the statistical model with Johanson's first-principles model, resulting in a novel gray-box (hybrid) model for the solid fraction prediction. The preconsolidation parameter was universally available regardless of the roller compactor size. With past data on material properties, process parameters, and corresponding solid fraction, the statistical model predicted the preconsolidation parameter without roller compaction experiments for the target formulation. The gray-box model predicted the solid fraction across various roll speeds, including the high throughput conditions causing powder velocity gradients. This robustness results from satisfying the Johanson model's premise that the one-dimensional mass remains constant before and after compaction. These results demonstrate the advantage of the proposed gray-box model, which can be used across scales and formulations without introducing complex additional concepts to the roller compaction mechanism.

Journal Pre-proofs

Keywords

Solid fraction; Dry granulation; Johanson model; Process modeling; Hybrid model

Journal Pre-proofs

Abbreviations

D roller diameter [m]

F force factor [–]

K compressibility factor [–]

N_R roller speed [s^{-1}]

P_0 preconsolidation pressure [Pa]

P_{max} maximum surface pressure [Pa]

RC roller compaction

R_f roll force [N]

S roll gap at the point of minimum separation [m]

UC uniaxial die-punch compaction

V specific volume of the layer [m^3]

W roller width [m]

x distance in the spanwise direction from the point of minimum roll distance [m]

α nip angle [°]

γ solid fraction [–]

γ_0 preconsolidation solid fraction [–]

γ_R ribbon solid fraction [–]

δ_E effective angle of internal friction [°]

ΔL arc-length segment [m]

ζ preconsolidation parameter [$Pa^{-1/K}$]

θ angular roll position [°]

μ friction coefficient [–]

ν actual angle between the direction of major principal stress and tangent to roll surface [°]

ρ envelope density [kg/m^3]

Journal Pre-proofs

ρ_{bulk} bulk density [kg/m^3]

ρ_{true} true density [kg/m^3]

σ normal stress [Pa]

$\emptyset_W$ angle of wall friction [°]

Journal Pre-proofs

1. Introduction

Roller compaction (RC) is a granulation process used in the oral solid dosage forms manufacturing (Kleinebudde, 2004). This process transforms powder into granules with uniform size and density to improve flowability and prevent segregation. The powder transported by the screw is compressed by two counter-rotating rolls to form ribbons, which are subsequently milled into granules. Excessive compression can lead to a loss of granule compressibility, which poses a risk of low tablet tensile strength (Herting and Kleinebudde, 2007). The solid fraction, defined as the ratio of the envelope density to the true density, is recognized as a critical quality attribute (CQA) in the RC (Sun and Kleinebudde, 2016). It is influenced by the material properties and process parameters (Pérez Gago and Kleinebudde, 2017). Thus, understanding the process and the material properties is important to achieve the desired solid fraction in the RC process development.

Johanson (1965) developed a theoretical model that simplifies the RC process into a one-dimensional system and determines the relationship between roll force and maximum roll surface pressure based on material properties and equipment geometry. This model can determine the solid fraction after compaction based on the relationship between the solid fraction and the pressure before and after compaction. Unfortunately, the state before compaction, i.e., preconsolidation pressure (P_0) and preconsolidation solid fraction (γ_0), cannot be measured because the space between the screw and the roll is enclosed in standard equipment design. In this study, we propose a model that uses the Johanson model to predict the solid fraction without directly measuring P_0 and γ_0. In the remaining part, we explain how these parameters have been addressed in previous studies and the challenges associated with their measurement, thereby clarifying the necessity of our proposed approach.

Various efforts have been made to practically apply the Johanson model to predict the solid fraction. Reynolds et al. (2010) determined both γ_0 and the compressibility factor (K) from the data of bench-top uniaxial die-punch compaction (UC) tests, assuming P_0 of 1 MPa. They found that directly applying the maximum pressure of the UC tests to the RC process leads to an overprediction of the solid fraction. Bi et al. (2014) and Liu and Wassgren (2016) revealed that discrepancies in mass balance before and after compaction in one dimension result in an overprediction of the maximum roll surface pressure. To correct the maximum roll surface pressure, Liu and Wassgren (2016) introduced the concept of a mass correction factor. So et al. (2021) proposed a simplified Johanson model that does not require friction angles by employing rolls with non-smooth surfaces, in addition to the concept of mass correction. Since So et al. (2021) limited their study to three formulations under low roll speed conditions, the applicability of the mass correction factor to all formulations or high roll speeds remains uncertain. Beccaro et al. (2024) found that velocity gradient of the powder before compaction in the RC changes with pressure and proposed a pressure-dependent mass correction factor. Amini and Akseli (2020) predicted the solid fraction of the RC without the mass correction factor by setting specific P_0 and γ_0 for each formulation, which were determined based solely on the results of the UC experiments. Beccaro et al. (2024) and Amini and Akseli (2020) provided deep insights into linking formulation and process parameters with the Johanson

model, and successfully used the UC to reduce powder usage in early pharmaceutical development. However, these two studies were performed at a constant low roll speed, thus it is unclear whether these approaches can be scaled to high-throughput manufacturing.

Sousa et al. (2020) observed that the roll speed affects the prediction results during equipment transfer. They demonstrated that the solid fraction can be kept constant during changes in equipment by maintaining a dimensionless parameter called the Midoux number that relates process parameters, geometric parameters, and material properties. Their study showed that high roll speeds overpredict the solid fraction. Luck et al. (2022) verified that the effect of roll speed on the solid fraction prediction is small for brittle materials and that roll speed affects the solid fraction prediction for plastic materials. They introduced the concept of compression dwell time to predict the solid fraction depending on the material deformation characteristics and roll speed. Li et al. (2024) proposed a model that combines the mass correction factor and the compression dwell time. It is based on the concept that the sensitivity of compression dwell time depends on the properties of the RC, just as the solid fraction of tablets increases with compression dwell time. They found that the mass correction factor depends on roll speed and material properties. Their study was conducted under constant pressure. Muthancheri et al. (2024) introduced the kinetics of compaction and proposed a correction to the Midoux model for the roll speed. While this represents an important advancement, building a model that considers compression dwell time still requires several RC experiments using the powder intended for predicting the ribbon's solid fraction.

There is no practical and consistent model for scaling up from the UC to the RC, scaling up to high-throughput conditions with a faster roll rotation speed, or scaling up by changing equipment. One of the main reasons for this is the lack of measurements of the preconsolidation properties P_0 and γ_0. Awasthi et al. (2023) utilized discrete element method (DEM) simulations to predict them and suggested that both P_0 and γ_0 vary with the process parameters. Muliadi et al. (2012) found a velocity gradient in the spanwise direction during compaction through finite element method (FEM) simulations. While both DEM and FEM simulations provide valuable insights into the process, their complexity results in long computational time, and the quantitative validation for given material properties remains challenging.

In this study, we developed a method to predict the solid fraction using the Johanson model, which cannot be used as is. The Johanson model needs P_0 and γ_0 to predict the solid fraction but these preconsolidation properties cannot be measured. Thus, we developed a statistical model to predict a newly defined preconsolidation parameter, which represents both P_0 and γ_0, from material properties and process parameters, and then we combined the statistical model and the Johanson model, resulting in the gray-box model for solid fraction prediction. In the following sections, we propose a concrete approach in the gray-box model to predict the solid fraction after the RC process without using the complex compaction mechanisms specific to each material and process.

2. Modeling approach for roller compaction process

In this section, we present the modeling framework for the roller compaction process. First, we outline the Johanson model, which serves as the basis for the white-box component. We then discuss the compressibility factor K and introduce the preconsolidation parameter ζ. Finally, we describe how these elements are integrated into the gray-box (hybrid) model to predict the solid fraction.

2.1. Johanson model

Fig. 1 shows a schematic diagram of the RC process. Johanson (1965) developed a model of the RC process, in which the powder is treated as a one-dimensional continuum and the maximum roll surface pressure (P_{max}) can be determined on the basis of the roll geometry and the roll force (R_f). This model divides the powder passing through the rolls into three regions. The first region, known as the slip region, is where particles rearrange due to weak pressure from the feeding mechanism, such as a screw. In the slip region, the powder moves more slowly than the rolls. The next region, called the nip region, is where the powder moves with the surface of rolls and is gradually compacted as a function of the roll angle (θ). The final region is the release region, where the powder is released from roll force. The angle at the boundary between the slip and nip regions is defined as the nip angle (α).

The normal and shear stresses in the slip region are determined by the wall friction angle of the powder relative to the roll surface. The stress gradient in the spanwise direction is given in Eq. (1).

$$\frac{\mathrm{d}\sigma_{slip}}{\mathrm{d}x} = \frac{4\sigma_{slip}\left(\frac{\pi}{2} - \theta - \nu\right)\tan\delta_E}{\frac{D}{2}\left[1 + \frac{S}{D} - \cos\theta\right]\left[\cot(A-\mu) - \cot(A+\mu)\right]} \quad (1)$$

where

$$\nu = \frac{1}{2}\left[\pi - \emptyset_w - \sin^{-1}\frac{\sin\emptyset_w}{\sin\delta_E}\right] \quad (2)$$

$$\mu = \frac{\pi}{4} - \frac{\delta_E}{2} \quad (3)$$

$$A = \frac{\left(\frac{\pi}{2} + \theta + \nu\right)}{2} \quad (4)$$

Here, x is the distance in the spanwise direction from the point of minimum roll separation, S is the roll gap at the point of minimum separation, and D is the roll diameter.

In the nip region, it is assumed that the mass of a differential volume at the nip angle α is the same as at any roll angle θ. The solid fraction (γ) and volume (V) per unit width of a differential layer at α and any θ in the nip region satisfy Eq. (5).

$$\frac{\gamma_{\alpha}}{\gamma_{\theta}} = \frac{V_{\theta}}{V_{\alpha}} \quad (5)$$

The volume V_{θ} at any roll angle θ is the product of the differential arc-length segment ΔL and the spanwise distance x as follows.

$$V_{\theta} = [S + D(1 - \cos\theta)]\cos\theta\,\Delta L \quad (6)$$

Johanson empirically observed a linear relationship between the logarithms of the solid fraction and the volume, leading to Eq. (7).

$$\frac{\sigma_{\alpha}}{\sigma_{\theta}} = \left(\frac{\gamma_{\alpha}}{\gamma_{\theta}}\right)^{K} = \left(\frac{V_{\theta}}{V_{\alpha}}\right)^{K} \quad (7)$$

Eq. (8) is derived from Eqs. (6) and (7).

$$\sigma_{\theta} = \sigma_{\alpha}\left[\frac{\left(1 + \frac{S}{D} - \cos\alpha\right)\cos\alpha}{\left(1 + \frac{S}{D} - \cos\theta\right)\cos\theta}\right]^{K} \quad (8)$$

The pressure gradient in the nip region can be obtained by differentiating Eq. (8) with respect to x, where $x = D/2\sin\theta$.

$$\frac{\mathrm{d}\sigma_{nip}}{\mathrm{d}x} = \frac{K\sigma_{nip}\left(2\cos\theta - 1 - \frac{S}{D}\right)\tan\theta}{\frac{D}{2}\left(1 + \frac{S}{D} - \cos\theta\right)\cos\theta} \quad (9)$$

Since α represents the angle at the boundary between the slip and nip regions, it can be determined by equating Eqs. (1) and (9) if K is known. Johanson also expressed the spanwise normal pressure at roll angle θ as $P_{\theta} = \sigma_{\theta}(1 + \sin\delta_E)$, deriving the relationship between the roll force and the maximum roll surface pressure.

$$R_f = \int_{\theta=0}^{\theta=\alpha} P_{\theta} W \frac{D}{2}\cos\theta\,\mathrm{d}\theta \quad (10)$$

Here, the pressure exerted by the rolls in the slip region is assumed negligible compared to that in the nip region. When α is sufficiently small, i.e., at the point of minimum roll distance, $\sigma_{\theta}(1 + \sin\delta_E)$ can be approximated as P_{max}, whereby roll force is expressed by Eq. (11).

$$R_f = \frac{P_{max} WDF}{2} (11)$$

F is the force factor, representing the integral of the roll force applied to the powder in the nip region.

$$F = \int_{\theta=0}^{\theta=\alpha} \left[\frac{\frac{S}{D}}{(1 + \frac{S}{D} - \cos\theta)\cos\theta} \right]^K \cos\theta \, \mathrm{d}\theta (12)$$

Assuming that the effective angle of internal friction δ_E is constant during compression, the ribbon solid fraction γ_R at $\theta = 0$ is calculated from Eq. (13).

$$\gamma_R = \gamma_0 \left(\frac{P_{max}}{P_0} \right)^{\frac{1}{K}} (13)$$

To derive γ_R, the Johanson model requires the preconsolidation solid fraction γ_0 and the preconsolidation pressure P_0. Although the ribbon compacted by the rolls usually undergo elastic recovery, Keizer and Kleinebudde (2020) found that the effect of the elastic recovery on the solid fraction prediction can be ignored since a certain amount of elastic recovery is expected under a certain pressure for the same formulation. Therefore, the elastic recovery was not considered in this study.

2.2. Compressibility factor *K*

To calculate the ribbon solid fraction γ_R, the compressibility factor K is required. In this study, the UC was used to determine K under the assumption that K is a material-specific constant independent of the equipment. While the preconsolidation properties such as P_0 and γ_0 may vary under different manufacturing conditions in the RC, P_0 and γ_0 are expected to remain constant under consistent test conditions in the UC since the preconsolidation properties before compaction are consistent for the same powder, and only the maximum pressure changes. The preconsolidation solid fraction γ_{0_UC} in the UC was defined as the bulk density ρ_{bulk} divided by the true density ρ_{true}. Additionally, ribbons produced by the RC often exhibit density gradients due to velocity profiles (Muliadi et al., 2012; Bi et al., 2014), whereas the UC applies uniform pressure, which reduces measurement errors. K can be determined for each powder blend based on the relationship between the applied compaction pressure in the UC $P_{\max\_UC}$ and the resulting solid fraction γ_{R_UC}. Eq. (14) is derived from taking the natural logarithm of both sides of Eq. (13).

$$\ln P_{\max\_UC} = K(\ln \gamma_{R\_UC} - \ln \gamma_{0\_UC}) + \ln P_0 (14)$$

Assuming the pressure applied to the powder before compression remains constant for the same powder in the UC, $\ln P_0$ can be neglected in the calculation of K from

experimental results. Therefore, K is obtained as the slope of the plot of $\ln P_{\max\_UC}$ against $\ln \gamma_{R\_UC} - \ln \gamma_{0\_UC}$.

2.3. Preconsolidation parameter ζ

As shown in Fig. 2, the region between the screw and rolls in the roller compactor is enclosed by side seals, which makes it difficult to directly measure P_0 and γ_0 during normal operation. In addition, P_0 and γ_0 depend on the material properties and process parameters, which means that these properties cannot be freely set before the experiment. According to the Johanson's theory that assumes equal mass before and after compaction, these two properties are interdependent under steady-state conditions; once one is determined, the other follows. Thus, we introduce a new parameter ζ, which represents the preconsolidation properties.

$$\zeta = \gamma_0 \left(\frac{1}{P_0}\right)^{\frac{1}{K}} (15)$$

Using ζ, the relationship between γ_R and P_{max} in Eq. (13) can be rewritten as follows.

$$\gamma_R = \zeta P_{max}^{\frac{1}{K}} (16)$$

The preconsolidation parameter ζ consolidates the effects of P_0 and γ_0 applied to the powder into a single variable under roller compaction operating conditions.

2.4. Gray-box model

We predicted the solid fraction through a gray-box modeling approach. The gray-box model, also called a hybrid model, combines a first principles model, known as a white-box model, with a statistical model, referred to as a black-box model. The gray-box model allows phenomena that cannot be fully captured by the white-box model to be accounted for (Ahmad et al., 2020). A white-box model expresses phenomenology in mathematical terms and is easy to interpret, but the system's internal structure is not always clear. In contrast, a black-box model is easier to build and does not require a deep understanding of the process, but it poses challenges in predicting results beyond the data points provided. The gray-box model combines the advantages of both models, offering a balance between accuracy and interpretability (Stosch et al., 2014). Recently, an integrated hybrid model for controlling roller compaction and ribbon milling operations was developed by Huang et al. (2024). In contrast, this study utilized a gray-box model (hybrid model) specifically to enhance the interpretability and accuracy of the roller compaction process.

The prediction of the ribbon solid fraction γ_R from the Johanson model or Eq. (16) alone is usually not possible because it requires ζ prior to the RC experiments. On the other hand, the prediction of γ_R from a statistical model alone lacks interpretability. In addition, the model might not work when the equipment scale is changed. In this study, a gray-box model is developed where ζ is determined by a statistical model and then γ_R is predicted by the Johanson model.

Fig. 3 shows the overall structure of the gray-box model. The prediction of γ_R involves three white-box models that calculate γ_R, P_{max}, and K, along with a black-box model that predicts ζ. γ_R is calculated using ζ, P_{max}, and K in white-box model 1 as described in Eq. (16). P_{max} is determined by white-box model 2, which corresponds to Eq. (11). It is influenced by both material parameters ($\emptyset_w$, δ_E) and process parameters (N_R, R_f, S), as well as geometric parameters (D, W). K is derived from white-box model 3, corresponding to Eq. (14), and is calculated from the results of the UC experiments. K depends on material properties (ρ_{true}, ρ_{bulk}) and process parameters such as the applied compaction pressure in the UC $P_{\max\_UC}$ and the resulting solid fraction γ_{0_UC}. ζ is predicted using the statistical model based on material properties (ρ_{true}, ρ_{bulk}, $\emptyset_w$, δ_E), process parameters (N_R, R_f, S), and geometric parameters (D, W), selected to maximize prediction performance.

Fig. 4 outlines the steps for building the black-box model that predicts ζ. First, UC experiments are conducted to determine K for each powder blend, as K is specific to each formulation. Then, P_{max} is calculated using the obtained K values, material properties, and process parameters, and RC experiments are performed under those conditions to obtain the experimental value of ζ. The collected data is used to train the statistical model, which predicts ζ based on the selected input parameters.

Once the statistical model is developed, it is integrated into the gray-box model to enable solid fraction prediction without requiring additional RC experiments. When predicting the outcome of a new formulation, UC experiments are required to determine K, but the results of RC experiments are not needed for prediction purposes; they are only used to validate the model's performance.

3. Materials and methods

3.1. Powder blends preparation

Nine formulations, labeled A to I, shown in Table 1, were used to confirm the robustness of the proposed method. Each formulation consisted of common pharmaceutical excipients with varying compaction properties. Among the nine formulations, only formulation I contained mefenamic acid as the model active pharmaceutical ingredient (API). Lactose, mannitol, dibasic calcium phosphate anhydrous (DCPA), and mefenamic acid are generally classified as brittle materials (Rowe and Roberts, 1995, Leuenberger, 1982). Corn starch and partially pregelatinized maize starch (PPMS) are classified as elastic materials, while microcrystalline cellulose

(MCC) is classified as a plastically deforming material (Bolhuis and Chowhan, 1996, Rowe and Roberts, 1995). Hydroxypropyl cellulose (HPC) and povidone were used as binders, crospovidone as a disintegrant, and magnesium stearate as a lubricant.

The supplier and material grade of the components used in this study were mefenamic acid (Cheng Fong Chemical, China), lactose (Dilactose® F, Freund Corporation, Japan, Pharmatose 200 M, DFE Pharma, Germany), mannitol (Pearlitol® 50 C, Roquette, France), DCPA (Fujicalin®, Fuji Chemical Industries, Japan), corn starch (Nihon Shokuhin Kako, Japan), PPMS (Starch 1500®, Colorcon, USA), MCC (CEOLUS® PH-101 and KG-1000, Asahi Kasei, Japan), HPC (NISSO HPC SL Fine powder, NIPPON SODA, Japan), povidone (Kollidon 30, BASF, Germany), crospovidone (Kollidon CL-F, BASF, Germany), and magnesium stearate (HyQual® 5712, SpecGx, USA).

The powder blend was prepared using a vertical granulator (VG-10, VG-25, or VG-50, POWREX, Japan) equipped with a horizontally rotating main blade at the bottom of the vessel and a vertically rotating high-speed cross-screw positioned above the main blade. The powder blend was mixed for 5 minutes using one of these machines. For formulation A, the mixing time was varied between 2 and 20 minutes to modify the powder blend properties. The rotation speed of the blades was adjusted to maintain a consistent peripheral speed, while the vessel size determined the specific speed setting to achieve uniform mixing intensity.

The true density of each formulation was measured using a helium gas pycnometer AccuPyc II 1340 (Micromeritics Instrument, USA) equipped with a 3.5 cm^3 chamber.

The bulk density was measured by gently introducing the powder into a dry 100 mL graduated cylinder without compacting, then reading the apparent volume to the nearest graduated unit. The bulk density measurement was repeated twice, and the average value was used.

The effective angle of internal friction δ_E and the wall friction angle φ_W were measured using an FT-4 powder rheometer (Freeman Technology, UK). The standard test method was used with a pre-shear stress of 9 kPa. The measurement was repeated three times, and the average value was used.

Table 2 shows the material properties of each powder blend prepared.

3.2. Uniaxial die-punch compaction (UC) tests

The compression profiles of the powder blends were determined using the UC. A 300 mg sample of each powder blend was loaded into a 10 mm diameter round die and compressed at 10 mm/min using a flat punch in the Autograph AG-I (Shimadzu, Japan). Compression tests were conducted under maximum test pressures ranging from 2.5 MPa to 100 MPa. Since each powder blend required a different minimum pressure for tablet formation, compressibility factors were calculated based on data from the tablet formation range. The mass and thickness of the tablets were measured at least 24 hours after

compaction using the tablet testing instrument WHT-3MJ (PHARMA TEST, Germany) accounting for elastic recovery. After the measurement, the solid fraction of each tablet was calculated.

3.3. Roller compaction (RC) tests

Two roller compactors were employed: FP90x30s, equipped with a 30 mm wide and 90 mm diameter knurled roller, and FP125x40, equipped with a 40 mm wide and 125 mm diameter knurled roller (FREUND TURBO, Japan). Both pieces of equipment operate in two modes: the gap control and the screw speed control. In the gap control mode, roll force, roll speed, and roll gap are set, while the screw speed is automatically adjusted. In the screw speed control mode, roll force, roll speed, and screw speed are set, allowing the roll gap to vary. In this study, all experiments were conducted in the screw speed control mode. The experiments were performed using FP90x30s for formulations A to G and FP125x40 for formulations B to D, H, and I, producing ribbons under various process parameters. Table 3 provides the range, data count of the process parameters, and solid fractions in the RC experiments. The ribbon solid fractions were calculated from the envelope density (ρ), and the true density (ρ_{true}) according to Eq. (17).

$$\gamma_R = \frac{\rho}{\rho_{true}} \quad (17)$$

The envelope densities were measured using Geopyc 1365 (Micromeritics Instrument, USA). Each measurement was performed three times, and the average value was used.

3.4. Statistical model

The statistical models were constructed to predict ζ, a parameter that cannot be directly observed in the RC process. In this study, partial least squares regression (PLSR) models (Wold et al., 2001) and gaussian process regression (GPR) models (Rasmussen and Williams, 2006) were compared.

PLSR is a linear regression method using a latent variable $\mathbf{T}$ that is a linear combination of the matrix of input variables $\boldsymbol{X}$.

$$\mathbf{T} = \mathbf{XW} \quad (18)$$

where, $\mathbf{W}$ is the weight matrix, which defines the linear combination of the original input variables that yields each latent variable. $\mathbf{T}$ is selected to maximize the covariance between $\mathbf{T}$ and the vector of an output variable $\mathbf{y}$. This allows $\mathbf{y}$ to be explained by fewer latent variables than the number of input variables, even when many input variables exist. The PLSR model can be written as:

$$\mathbf{X} = \mathbf{TP}^{\mathrm{T}} + \mathbf{E} \quad (19)$$

$$\mathbf{y} = \mathbf{Td} + \mathbf{f} (20)$$

where, $\mathbf{P}$ is the loading matrix, $\mathbf{d}$ is the regression coefficient vector, and $\mathbf{E}$ and $\mathbf{f}$ represent the residuals of $\mathbf{X}$ and $\mathbf{y}$, respectively. The predicted value of the output variable $\hat{y}$ can be expressed using the regression coefficient vector $\mathbf{b}$ as follows:

$$\hat{y} = \mathbf{Xb} = \mathbf{XW}(\boldsymbol{P}^T\mathbf{W})^{-1}\boldsymbol{d} (21)$$

GPR is a nonlinear regression method that models the nonlinear relationship between inputs and outputs using kernel functions. The choice of the kernel functions determines the patterns and complexity that the model can capture. In this study, the Matérn kernel and white kernel were used. Specifically, GPR assumes that y follows a gaussian distribution with a covariance matrix $\mathbf{K}$. The elements of $\mathbf{K}$ are

$$K_{ij} = k(x_i, x_j) (22)$$

where $k(x_i, x_j)$ is the kernel function. Let $\mathbf{x}^*$ be a new input and $\mathcal{D} = (\mathbf{X},\mathbf{y})$ be the training data; the predictive distribution of y^* is given by:

$$p(\mathrm{y}^* \mid \mathbf{x}^*,\mathcal{D}) = \mathcal{N}\left(\mathrm{k}_*^T\mathbf{K}^{-1}\mathrm{y}, k_{**} - \mathrm{k}_*^T\mathbf{K}^{-1}\mathrm{k}_*\right) (23)$$

where k_* is the covariance vector between $\mathbf{x}^*$ and k_{**} is the variance at $\mathbf{x}^*$. y^* is normally distributed and its predictive distribution mean $\mathrm{k}_*^T\mathbf{K}^{-1}\mathrm{y}$ and variance $k_{**} - \mathrm{k}_*^T\mathbf{K}^{-1}\mathrm{k}_*$.

Using the above approaches, we define ζ in a data-driven manner as:

$$\zeta = f_{bb}\left(\rho_{true}, \rho_{bulk}, \emptyset_W, \delta_E, N_R, R_f, S, D, W, K\right) (24)$$

where f_{bb} is built by either PLSR or GPR.

The number of latent variables for the PLSR model and the hyperparameters of the kernel functions for the GPR model were optimized using 10-fold cross-validation to minimize the root mean squared error (see Section 3.5) between ζ calculated from the experimental data and ζ predicted by the statistical models.

The selection of input variables for the statistical models was optimized by maximizing the prediction accuracy in the cross-validation in Section 3.5. First, all parameters used in the first principles model were considered candidate input variables. For the PLSR model, the correlation coefficients between each input variable and ζ were calculated, and the parameters with low correlation coefficients were sequentially removed. For the GPR model, the permutation importance (PI) (Fisher et al., 2019) was assessed, and parameters with low PI were sequentially removed. PI measures the importance of a variable by randomly shuffling its values and observing the resulting decline in model performance. This process was repeated 10 times to calculate the variable importance.

3.5. Performance evaluation

The following methods were used to compare and evaluate the performance of the models. To evaluate the applicability of the models to the experimental data, the coefficient of determination (R^2) and root mean squared error (RMSE) were calculated as follows.

$$R^2 = 1 - \frac{\sum_{i=1}^{I} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{I} (y_i - \bar{y})^2} (25)$$

$$RMSE = \sqrt{\frac{1}{I} \sum_{i=1}^{I} (y_i - \hat{y}_i)^2} (26)$$

where I is the number of samples, $\hat{y}_i$ is the *i*-th predicted value.

The RMSE of cross-validation (RMSECV) represents the RMSE of ζ predicted by the statistical model and γ_R predicted by the gray-box model during cross-validation. Prediction accuracy for data not used in model building was assessed using the RMSE of prediction (RMSEP). A white-box model for predicting γ_R using ζ fitted by least squares to the experimental results from Eq. (16), was used as a comparison to the gray-box model. The RMSE of γ_R predicted by the white-box model alone is referred to as the RMSE of calibration (RMSEC).

The scikit-learn Python library (Pedregosa et al., 2011) was employed for this task. All statistical calculations in this study were performed using Python 3.8.18. The calculations were performed on a computer with 128 GB of RAM and an Intel® Xeon® CPU E5-2667 v4 @ 3.20 GHz.

3.6. Data usage for model training and testing

In this study, different subsets of the experimental data were used for training and testing in each analysis step (Sections 4.2–4.4). In Section 4.2, cross-validation was performed for the statistical model using all nine formulations, where the dataset was divided into ten folds, and the model was trained on nine folds while evaluating its predictive accuracy on the remaining fold. This was repeated for all the folds that were split. In Section 4.3, the training dataset consisted of FP90x30s data, while the model's predictive performance was assessed using FP125x40 data, which was not included in the training. In Section 4.4, each formulation was excluded in turn from the training set, and the model's predictive performance was evaluated using the left-out formulation, which was not used in training. This was repeated to build statistical models for each formulation.

4. Results & Discussions

4.1. Calculation of K from the UC and ζ from the RC

To calculate ζ, the preconsolidation parameter, from the RC test dataset, the compressibility factor K was first calculated for each powder using Eq. (14) based on the UC test results. Table 4 shows the calculated K, ranging from 3.5 to 8.7. The γ_{R_UC} range obtained from the UC covered most of the γ_R range observed in the RC. Up to 10 UC samples were used to calculate K, with a total of approximately 3 grams of powder sufficient for the analysis. As shown in Fig. 5, the slopes of the log-log plot of pressure versus density from the UC and the RC represent the compressibility factor. Formulations exhibiting high compressibility factors, such as A, F, G, H, and I, contained a high proportion of brittle materials like mefenamic acid, lactose, and mannitol. In contrast, formulations with low compressibility factors, such as C, D, and E, contained a high proportion of plastic materials like MCC. These results suggest that the ratio of brittle to plastic materials affects the compressibility factor. Despite containing a large amount of brittle DCPA, formulation B showed a low compressibility factor. This might be due to the presence of MCC KG-1000, which is a fibrous and highly formable grade. Even small amounts of MCC KG-1000 can cause plastic deformation to dominate over brittle fracture during compaction. As shown in Fig. 5, the compressibility factor calculated from the UC was used to derive ζ from the RC experimental results for each powder and process parameter. This assumption is based on the premise that the compressibility behavior remains consistent across different equipment. Since the compressibility factor is calculated using the density after elastic recovery, it is important to note that ζ from each RC dataset also reflects the effects of elastic recovery.

The relationship between ζ and the material properties of the powders and the RC process parameters is shown in Fig. 6. ζ was strongly influenced by material properties derived from formulations, particularly correlated with compressibility factor, bulk density, and true density. However, none of the material properties were sufficient to explain the variability in ζ within each formulation. Process parameters had some impact on ζ. The roll gap appears to affect ζ, but it does not fully explain the differences within a formulation. Roll force and roll speed might affect ζ. A comparison of the plots of FP90x30s and FP125x40 showed no significant differences in the range of process parameters or the distribution of ζ for the same formulation across different equipment.

4.2. Validation of statistical model and gray-box model

Fig. 7 shows the impact of the number of input variables on the predictive performance of ζ in cross-validation when the PLSR and GPR models are employed as statistical models, based on the RC manufacturing data from nine formulations. It also shows how the number of input variables influences the prediction accuracy of γ_R in the gray-box model. For both ζ and γ_R, the GPR model outperformed the PLSR model for all numbers of input variables. For PLSR, the input variables were removed stepwise in ascending order of absolute values of their correlation coefficients with ζ, starting with the smallest: roll speed, effective angle of internal friction, wall friction angle, roll diameter, roll force, roll gap, true density, and bulk density. For GPR, the input variables

were removed stepwise in ascending order of PI: roll speed, wall friction angle, effective angle of internal friction, roll diameter, bulk density, roll force, roll gap, and true density. Consequently, the GPR model was employed in subsequent studies as a part of the gray-box model. When building the gray-box model across formulations and equipment, seven input variables were selected to maximize the prediction performance: true density, bulk density, effective angle of internal friction, roll diameter, roll force, roll gap, and compressibility factor. Fig. 8 illustrates the PI of the selected input variables for predicting ζ. Compressibility factor, bulk density, true density, and roll gap which were observed to correlate with ζ in Fig. 6, exhibited high PI. Fig. 9 shows that the GPR model with the selected seven input variables accurately predicts the preconsolidation parameter ζ with R^2 of 0.995. ζ could be predicted only by the variables used in the Johanson model calculation. It should be noted that the input variables might need to be re-selected when the new dataset becomes available in the future.

Fig. 10 demonstrates the prediction results of γ_R by the gray-box model and the white-box model that used ζ fitted by the least-squares for each powder. The white-box model provided poor predictions since it did not account for variations in process parameters, although it was able to capture the difference between formulations. The gray-box model significantly outperformed the white-box model in prediction performance for all data points regardless of formulations and process parameters. The gray-box model accurately predicted γ_R over a wide range of roll speeds from 6 to 20 min^{-1} using different formulations.

Interestingly, the effects of roll speed observed by Luck et al. (2022) and Li et al. (2024) on the solid fraction prediction did not need to be considered in the gray-box model. In this study, the one-dimensional mass of the powder is assumed to be the same before and after roll compaction based on the Johanson model. We used the UC that does not change the mass before and after compaction and derived the compressibility factor by considering the elastic recovery rate. Since the powder mass before and after compaction is also preserved in the RC, the elastic recovery rates are expected to change according to the pressure level regardless of the compression dwell time. Despite the simple theoretical basis that does not explicitly incorporate compression dwell time and roll speed, the predictive performance of our proposed gray-box model is comparable to the modified Midoux model by Muthancheri et al. (2024). The gray-box model was able to predict the solid fraction even at nearly the upper and lower limits of the roll speed of the equipment.

4.3. Impact of the RC equipment change

Fig. 11 shows the prediction results of the statistical model and the gray-box model in the case where the training data was obtained from the smaller RC equipment FP90x30s and the test data from the larger RC equipment FP125x40. Formulations B, C, and D, manufactured on both pieces of equipment, were used to evaluate the effect of the equipment size on the prediction accuracy of ζ. To investigate the cross-scale applicability of ζ, the roll diameter was excluded from the input variables. Table 5 shows the PI of the input variables, the prediction accuracy in cross-validation, and the accuracy for the scaled-up data obtained using FP125x40. The statistical model was built

using FP90x30s data only. Three input variables, roll force, roll gap, and true density, were selected to predict ζ^*. The R^2 of ζ^* and γ_R in cross-validation was greater than 0.8 regardless of the number of input variables, when the statistical model's input variables included bulk density, effective angle of internal friction, or compressibility factor. On the other hand, when these variables were included, the prediction accuracy for the FP125x40 data was low, with the R^2 of ζ^* and γ_R below 0.2. Formulations B, C and D each used a single powder at each scale. As a result, the trained statistical model did not account for differences in material properties between the powders within the same formulation. Among the material properties, only true density was selected, leading to improved prediction accuracy for the FP125x40 data. The R^2 exceeded 0.7 for ζ^* and 0.8 for γ_R. The relationship between the selected input variables and the prediction accuracy did not depend on the scale but on the variation in the powders used. True density exhibited a high PI among the material properties due to its formulation-specific values since the formulations differed only in the ratio of their components.

The predicted ζ^* was not affected by the equipment scale. As illustrated in Fig. 6, the distribution of ζ^* remains consistent across scales, which indicates that ζ^* is a universal variable applicable across different scales. In the Johanson model utilized in the gray-box model, the influence of scale is captured by the roll diameter and roll width. This consistency allows information obtained prior to equipment transfer to be directly applied for predicting outcomes after the transfer. This study revealed that the proposed framework can significantly reduce the resources required for scale-up or scale-down of equipment.

4.4. Formulation independent prediction of ζ

Fig. 12 shows the prediction performance of ζ by PLSR and GPR and γ_R by the gray-box model when using the FP90x30s data of formulations A to G. For PLSR, the input variables were removed stepwise in ascending order of absolute value of their correlation coefficients with ζ, starting with the smallest: roll speed, effective angle of internal friction, wall friction angle, roll force, roll gap, true density, and bulk density. For GPR, the input variables were removed stepwise in ascending order of PI: wall friction angle, roll speed, effective angle of internal friction, compressibility factor, roll force, roll gap, and true density. In this section, GPR with four variables, true density, bulk density, roll force and roll gap, was finally adopted as the statistical model. Fig. 13 shows the prediction accuracy of the statistical model and gray-box model for formulations not included in the training data set. ζ was predicted using a statistical model that excluded the target formulation to be predicted from the training data, and then a gray-box model predicted γ_R using the predicted ζ. The model building and predictions were repeated for each formulation. The RMSEP of ζ was 0.011 and the R^2 was 0.990. The RMSECV of ζ was 0.011 and the R^2 was 0.992 in the cross-validation with all the formulations, indicating that the statistical model was able to predict ζ for any combination of the material properties and the RC process parameters. The predicted ζ for formulation B tended to be lower than the measured values. Since formulation B has the lowest ζ and the highest true density, this suggests that the prediction accuracy

has decreased beyond the range of the training data. When manufacturing data from the RC process are not available for the formulation for which the solid fraction is to be predicted, the following information is required to predict ζ for the RC process applied to that formulation: i) the material properties and process parameters of the formulation to be predicted, ii) a data set covering the range of i), and iii) γ_R corresponding to the data set in ii). The gray-box model can predict γ_R using predicted ζ from past experimental results for new formulations, which differs from many previously proposed approaches that predict γ_R using only the material properties and process parameters from a formulation to be predicted.

5. Conclusions

In this study, the gray-box model that combines the white-box model and the black-box model was developed to predict the solid fraction, which is a critical quality attribute (CQA) in the roller compaction (RC) process. The white-box model relies on the Johanson model. Since the preconsolidation pressure and preconsolidation solid fraction in the Johanson model are difficult to measure during normal operation, the Gaussian process regression (GPR)-based statistical model predicts the newly defined preconsolidation parameter ζ, which represents these preconsolidation properties. This gray-box model is applicable for wide-range throughput manufacturing conditions since the gray-box model satisfies the principle proposed by Johanson that one-dimensional powder mass is conserved before and after compaction. Since ζ is a parameter independent of the RC size, the gray-box model can predict the solid fraction for RC sizes different from the one used to build the statistical model. The statistical model was able to predict ζ given the past experimental results for different formulations, including the material properties, process parameters of the RC, and the corresponding solid fraction. Therefore, the solid fraction can be predicted even when there are no RC experimental results for the formulation being predicted. The gray-box model retains the original Johanson model's structure, offering clear interpretability and practical applicability. By utilizing existing experimental data, it can significantly reduce the time and raw materials needed for the manufacturing process development.

CRediT authorship contribution statement

Kanta Sato: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft. **Shuichi Tanabe:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Writing – review and editing. **Keita Yaginuma:** Conceptualization, Methodology, Software, Writing – review and editing. **Susumu Hasegawa:** Methodology, Project administration, Resources, Supervision, Writing – review and editing. **Manabu Kano:** Conceptualization, Methodology, Project administration, Supervision, Writing – review and editing.

Declaration of competing interest

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

Acknowledgements

The study was funded by a research grant from Daiichi Sankyo Co., Ltd.

Data availability

The data that have been used are confidential.

References

1. Ahmad, I., Ayub, A., Kano, M., Cheema, I.I., 2020. Gray-box soft sensors in process industry: current practice, and future prospects in era of big data. Processes 8, 243. https://doi.org/10.3390/pr8020243.

2. Amini, H., Akseli, I., 2020. A first principle model for simulating the ribbon solid fraction during pharmaceutical roller compaction process. Powder Technol. 368, 32–44.

   http://dx.doi.org/10.1016/j.powtec.2020.04.022.

3. Awasthi, S., Gopireddy, S.R., Kako, D., Tanabe, S., Nakagawa, H., Miyajima, M., Profitlich, T., Scherließ, R., Urbanetz, N.A., 2023. Combined DEM and Johanson model for ribbon density prediction in a roller compactor. Powder Technol. 419, 118296.

   https://doi.org/10.1016/j.powtec.2023.118296.

4. Beccaro, L., Facco, P., Dhenge, R.M., Khala, M.J., Cenci, F., Bezzo, F., Barolo, M., 2024. Accelerating pharmaceutical tablet development by transfer of powder compaction equipment across types and scales. Int. J. Pharm. In Press, 124904.

   https://doi.org/10.1016/j.ijpharm.2024.124904

5. Bi, M., Alvarez-Nunez, F., Alvarez, F., 2014. Evaluating and modifying Johanson's rolling model to improve its predictability. J. Pharm. Sci. 103, 2062–2071. https://doi.org/10.1002/jps.24012.

Journal Pre-proofs

6. Bolhuis, G.K., Chowhan, Z.T., 1996. Materials for direct compression, in: Alderborn, G., Nystrom, C. (Eds.), Pharmaceutical Powder Compaction Technology. Marcel Dekker Inc., New York, pp. 429–491.

7. Fisher, A., Rudin, C., Dominici, F., 2019. All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. J. Mach. Learn. Res. 20, 1–81.

8. Herting, M.G., Kleinebudde, P., 2007. Roll compaction/dry granulation: effect of raw material particle size on granule and tablet properties. Int. J. Pharm. 338, 110–118. https://doi.org/10.1016/j.ijpharm.2007.01.035.

9. Huang, Y.-S., Lagare, R.B., Bailey, P., Sixon, D., Gonzalez, M., Nagy, Z.K., Reklaitis, G.V., 2024. Hybrid model development and nonlinear model predictive control implementation for continuous dry granulation process. Comput. Chem. Eng. 183, 108586.

   https://doi.org/10.1016/j.compchemeng.2024.108586.

10. Johanson, J.R., 1965. A rolling theory for granular solids. J. Appl. Mech. 32, 842.

    https://doi.org/10.1115/1.3627325.

11. Keizer, H.L., Kleinebudde, P., 2020. Elastic recovery in roll compaction simulation. Int. J. Pharm. 573, 118810.

    https://doi.org/10.1016/j.ijpharm.2019.118810.

12. Kleinebudde, P., 2004. Roll compaction/dry granulation: pharmaceutical applications. Eur. J. Pharm. Biopharm. 58, 317–326.

    https://doi.org/10.1016/j.ejpb.2004.04.014.

13. Leuenberger, H., 1982. The compressibility and compactibility of powder systems. Int. J. Pharm. 12, 41–55.

    https://doi.org/10.1016/0378-5173(82)90132-6.

14. Li, J., Tseng, Y., Paul, S., 2024. A modified mechanistic approach for predicting ribbon solid fraction at different roller compaction speeds. Int. J. Pharm. 660, 124366.

    https://doi.org/10.1016/j.ijpharm.2024.124366.

15. Liu, Y., Wassgren, C., 2016. Modifications to Johanson's roll compaction model for improved relative density predictions Powder Technol. 297, 294–302.

    https://doi.org/10.1016/j.powtec.2016.04.017.

16. Luck, M., De Saeger, M., Kleinebudde, P., 2022. Influence of roll speed during roll compaction and its effect on the prediction of ribbon solid fraction. Pharmaceutics 14, 2399.

https://doi.org/10.3390/pharmaceutics14112399.

17. Muliadi, A.R., Litster, J.D., Wassgren, C.R., 2012. Modeling the powder roll compaction process: comparison of 2-D finite element method and the rolling theory for granular solids (Johanson's model). Powder Technol. 221, 90–100.

https://doi.org/10.1016/j.powtec.2011.12.001.

18. Muthancheri, I., Rousselin, M., Espinose, A., Sanchez, N., Authelin, J.-R., 2024. Modified roller compaction model to account for roll speed effect on powder compaction in dry granulation process. J. Pharm. Sci. 113, 2484–2491.

https://doi.org/10.1016/j.xphs.2024.05.006.

19. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

20. Pérez Gago, A., Kleinebudde, P., 2017. MCC–mannitol mixtures after roll compaction/dry granulation: percolation thresholds for ribbon microhardness and granule size distribution, Pharm. Dev. Technol. 22, 764-774.

https://doi.org/10.3109/10837450.2016.1163388.

21. Rasmussen, C.E., Williams, C.K.I., 2006. Gaussian process for machine learning. MIT Press.

22. Reynolds, G., Ingale, R., Roberts, R., Kothari, S., Gururajan, B., 2010. Practical application of roller compaction process modeling. Comput. Chem. Eng. 34, 1049–1057.

https://doi.org/10.1016/j.compchemeng.2010.03.004.

23. Rowe, R.C., Roberts, R.J., 1995. The mechanical properties of powders. Adv. Pharm. Sci. 7, 1–62.

https://doi.org/10.1016/S0065-3136(06)80003-8.

24. So, C., Leung, L.Y., Muliadi, A.R., Narang, A.S., Mao, C., 2021. Simplifying Johanson's roller compaction model to build a “virtual roller compactor” as a predictive tool – theory and practical application. Int. J. Pharm. 601, 120579.

https://doi.org/10.1016/j.ijpharm.2021.120579.

25. Sousa, R., Valente, P.C., Nakach, M., Bardet, L., Wacquet, M., Midoux, N., Authelin, J.-R., 2020. Roller compaction scale-up made simple: an approximate analytical solution to Johanson's rolling theory. J. Pharm. Sci. 109, 2536–2543.

https://doi.org/10.1016/j.xphs.2020.05.004.

26. Sun, C.C., Kleinebudde, P., 2016. Mini review: mechanisms to the loss of tabletability by dry granulation. Eur. J. Pharm. Biopharm. 106, 9–14.

https://doi.org/10.1016/j.ejpb.2016.04.003.

27. von Stosch, M., Oliveira, R., Peres, J., Feyo de Azevedo, S., 2014. Hybrid semi-parametric modeling in process systems engineering: past, present and future. Comput. Chem. Eng. 60, 86–101.

https://doi.org/10.1016/j.compchemeng.2013.08.008.

28. Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst. 58, 109–130.

https://doi.org/10.1016/S0169-7439(01)00155-1.

Journal Pre-proofs

Table 1 Composition of nine formulations in mass percentage.

Ingredient	Formulation								
	A	B	C	D	E	F	G	H	I
Mefenamic acid	–	–	–	–	–	–	–	–	25
Lactose Dilactose	–	–	–	–	–	72	78	78	–
Lactose Pharmatose	–	–	–	–	–	–	–	–	42
Mannitol	72	–	–	–	–	–	–	–	–
DCPA	–	78	49	20	14	–	–	–	–
Corn starch	–	–	–	–	–	–	–	–	24
PPMS	–	–	–	–	30	–	–	–	–
MCC PH-101	18	–	–	–	–	8	20	–	–
MCC KG-1000	–	20	49	78	54	–	–	20	–
HPC	8	–	–	–	–	18	–	–	–
Povidone	–	–	–	–	–	–	–	–	4
Crospovidone	–	–	–	–	–	–	–	–	5
Magnesium stearate	2	2	2	2	2	2	2	2	–

Journal Pre-proofs

Table 2 Material properties of each powder blend.

Formulation	Blending time (min)	ρ_{true} (kg/m^3)	ρ_{bulk} (kg/m^3)	δ_E (°)	$\emptyset_W$ (°)
A	2	1477	649	30	10
	4	1477	676	26	9
	5	1477	645	30	9
	7	1477	662	31	9
	8	1477	662	30	10
	10	1477	658	30	10
	20	1477	658	30	10
B	5	2379	459	29	9
	5	2379	439	29	7
C	5	2042	389	34	9
	5	2042	490	33	7
D	5	1728	341	34	8
	5	1728	292	39	7
E	5	1648	395	21	7
F	5	1471	602	23	6

Journal Pre-proofs

G	5	1541	505	20	8
H	5	1538	595	30	7
I	5	1424	427	36	13

Journal Pre-proofs

Table 3 Range of roller compaction parameters.

Formulation	Roller compactor	R_f (kN/cm)	S (mm)	N_R (min^{-1})	γ_R (–)	Data count
A	FP90x30s	4.4–11.9	0.76–1.04	7–20	0.63–0.85	18
B	FP90x30s	3.5–11.0	0.75–0.95	6–20	0.31–0.56	13
	FP125x40	8.7–14.0	0.75–1.08	6–20	0.40–0.64	7
C	FP90x30s	3.5–11.0	0.78–0.94	6–20	0.38–0.57	7
	FP125x40	4.7–14.0	0.77–0.89	6–20	0.46–0.62	8
D	FP90x30s	3.5–11.0	0.77–0.89	6–20	0.43–0.65	12
	FP125x40	4.7–14.0	0.75–0.98	8–20	0.46–0.75	8
E	FP90x30s	3.5–11.0	0.77–0.91	6–20	0.52–0.66	12
F	FP90x30s	3.5–11.0	0.76–1.01	8–20	0.65–0.88	10
G	FP90x30s	3.5–11.0	0.79–1.02	6–20	0.66–0.86	7
H	FP125x40	5.8–14.0	0.75–0.98	8–20	0.68–0.87	6
I	FP125x40	4.7–14.0	0.79–0.94	6–20	0.78–0.88	8

Journal Pre-proofs

Table 4 The UC test results and the derived compressibility factor K.

Formulation	Blending time (min)	P_{max} (MPa)	γ_R (–)	K (–)
A	2	5.0–100	0.59–0.88	8.15
	4	5.0–100	0.62–0.89	7.89
	5	5.0–100	0.63–0.89	8.10
	7	5.0–100	0.62–0.90	7.68
	8	5.0–100	0.61–0.89	7.78
	10	5.0–100	0.62–0.86	8.74
	20	5.0–100	0.62–0.90	7.77
B	5	5.0–100	0.32–0.62	4.22
	5	5.0–100	0.32–0.63	3.91
C	5	5.0–100	0.33–0.69	3.89
	5	2.5–100	0.31–0.70	3.67
D	5	7.5–100	0.39–0.77	3.69
	5	5.0–100	0.33–0.76	3.53
E	5	5.0–100	0.41–0.80	4.31
F	5	2.5–100	0.51–0.88	6.61

Journal Pre-proofs

G	5	7.5–100	0.53–0.83	5.51
H	5	7.5–100	0.54–0.83	5.89
I	5	2.5–100	0.58–0.86	8.69

Journal Pre-proofs

Table 5 Prediction performance (R^2 and RMSE) of the preconsolidation parameter ζ and the ribbon solid fraction γ_R for FP125x40 data based on GPR model trained with FP90x30s data. Prediction performance and permutation importance (PI) are shown for different numbers of input variables, with variables incrementally removed starting from those with the lowest PI values.

Number of input variables	PI						Prediction accuracy of ζ				Prediction accuracy of γ_R			
							FP90x30s (cross-validation)		FP125x40		FP90x30s (cross-validation)		FP125x40	
	K	ρ_{bulk}	δ_E	ρ_{true}	R_f	S	RMSECV	R^2	RMSEP	R^2	RMSECV	R^2	RMSEP	R^2
6	0.04	0.11	0.98	0.45	0.59	1.47	0.009	0.870	0.020	0.099	0.036	0.869	0.192	0.127
5	−	0.15	0.83	0.53	0.59	1.46	0.009	0.871	0.021	-0.002	0.036	0.869	0.191	0.136
4	−	−	0.49	0.81	0.57	1.43	0.009	0.876	0.022	-0.091	0.035	0.876	0.199	0.064
3	−	−	−	0.07	0.60	1.50	0.010	0.849	0.011	0.743	0.039	0.839	0.088	0.817
2	−	−	−	−	0.54	1.51	0.009	0.871	0.014	0.584	0.038	0.853	0.118	0.669

Journal Pre-proofs

Journal Pre-proofs

Fig. 1 Schematic diagram of the roll compaction process, in which powder exists in three regions: the slip region, the nip region, and the release region.

Fig. 2 Photographs near the rolls of the roller compactor: (a) before the side seals are assembled and (b) after full assembly.

Fig. 3 The procedure for predicting the ribbon solid fraction γ_R using the gray-box model.

Fig. 4 Preparation and development of the black-box model for predicting the preconsolidation parameter ζ.

Fig. 5 Log-log plots of maximum surface pressure versus solid fraction divided by the preconsolidation parameter from UC and RC. The slope represents the compressibility factor K.

Fig. 6 Pairplots showing the relationship among material attributes, RC process parameters, and the preconsolidation parameter ζ.

Fig. 7 Comparison of the PLSR and GPR models in R^2 and RMSECV for (a) the preconsolidation parameter ζ and (b) the ribbon solid fraction γ_R using different numbers of input variables.

Fig. 8 Permutation importance (PI) of selected input variables based on the GPR model for predicting the preconsolidation parameter

Journal Pre-proofs

ζ.

Fig. 9 Prediction results of the preconsolidation parameter ζ by the GPR model.

Fig. 10 Prediction results of the ribbon solid fraction γ_R by (a) the gray-box model and (b) the white-box model.

Fig. 11 Validation of scale dependency: prediction results of (a) the preconsolidation parameter ζ and (b) the ribbon solid fraction γ_R using data from larger RC equipment FP125x40 for modeling and data from smaller FP90x30s for evaluation.

Fig. 12 Comparison of the PLSR and GPR models in R^2 and RMSECV for (a) the preconsolidation parameter ζ and (b) the ribbon solid fraction γ_R using different numbers of input variables when using only the FP90x30s data.

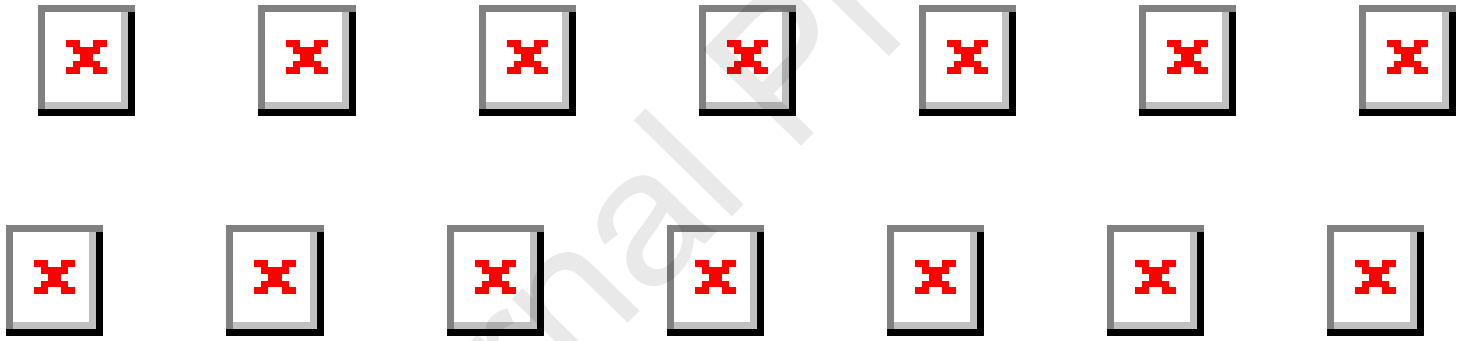
Fig. 13 Validation of the formulation independence: prediction results of (a) the preconsolidation parameter ζ by the GPR model and (b) the ribbon solid fraction γ_R by the gray-box model, both built without the formulation to be predicted.

Highlights

- A gray-box model was developed for solid fraction prediction of roller compaction.
- The black-box model predicts the preconsolidation parameter in the Johanson model.

Journal Pre-proofs

- The gray-box model predicts solid fraction without experiments on larger equipment.
- The gray-box model applies across manufacturing conditions including high-throughput.
- The gray-box model leverages past roller compaction knowledge in new formulations.



Journal Pre-proofs

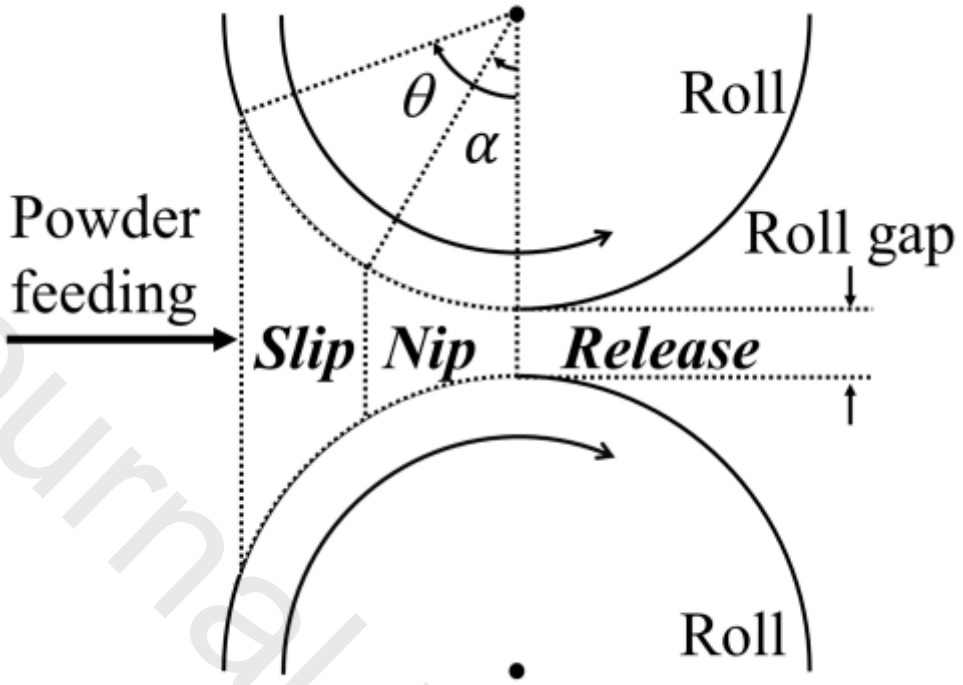
Declaration of interests

☐ 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.

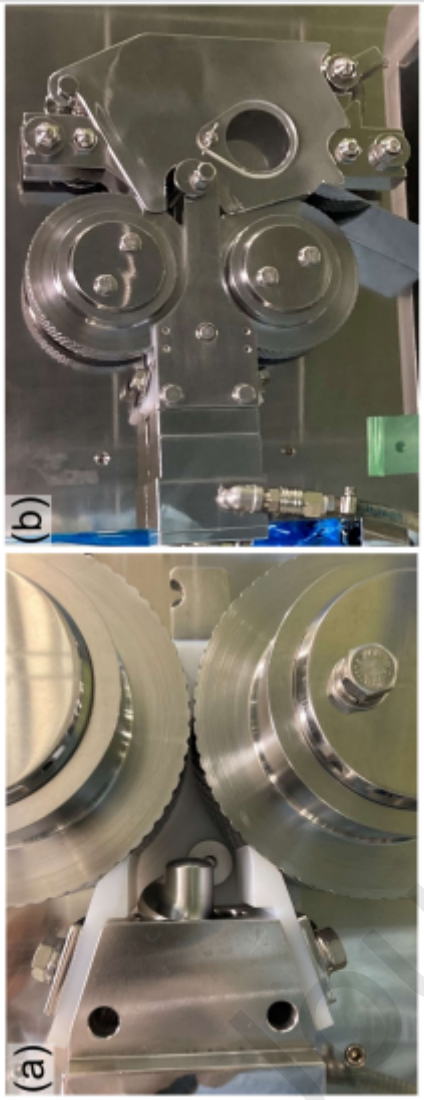
☒ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Kanta Sato reports a relationship with DAIICHI SANKYO COMPANY, LIMITED that includes: employment. Shuichi Tanabe reports a relationship with DAIICHI SANKYO COMPANY, LIMITED that includes: employment. Keita Yaginuma reports a relationship with DAIICHI SANKYO COMPANY, LIMITED that includes: employment. 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.

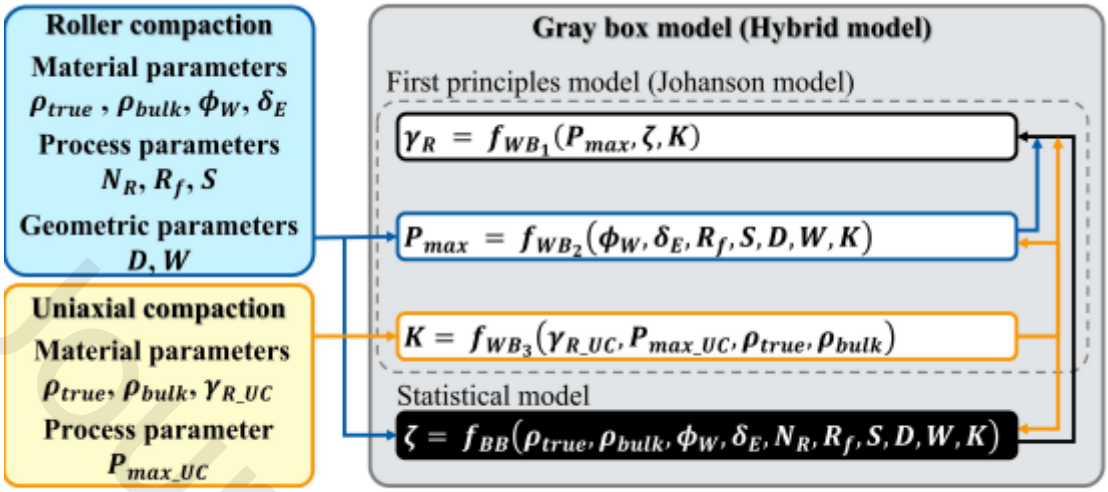
Journal Pre-proofs



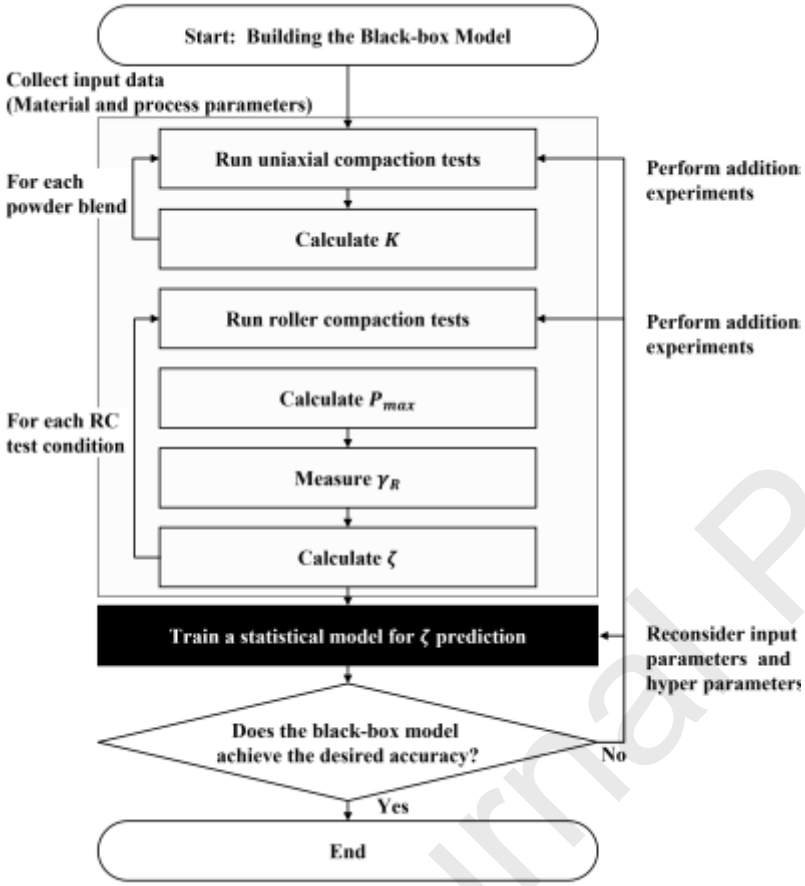
Journal Pre-proofs



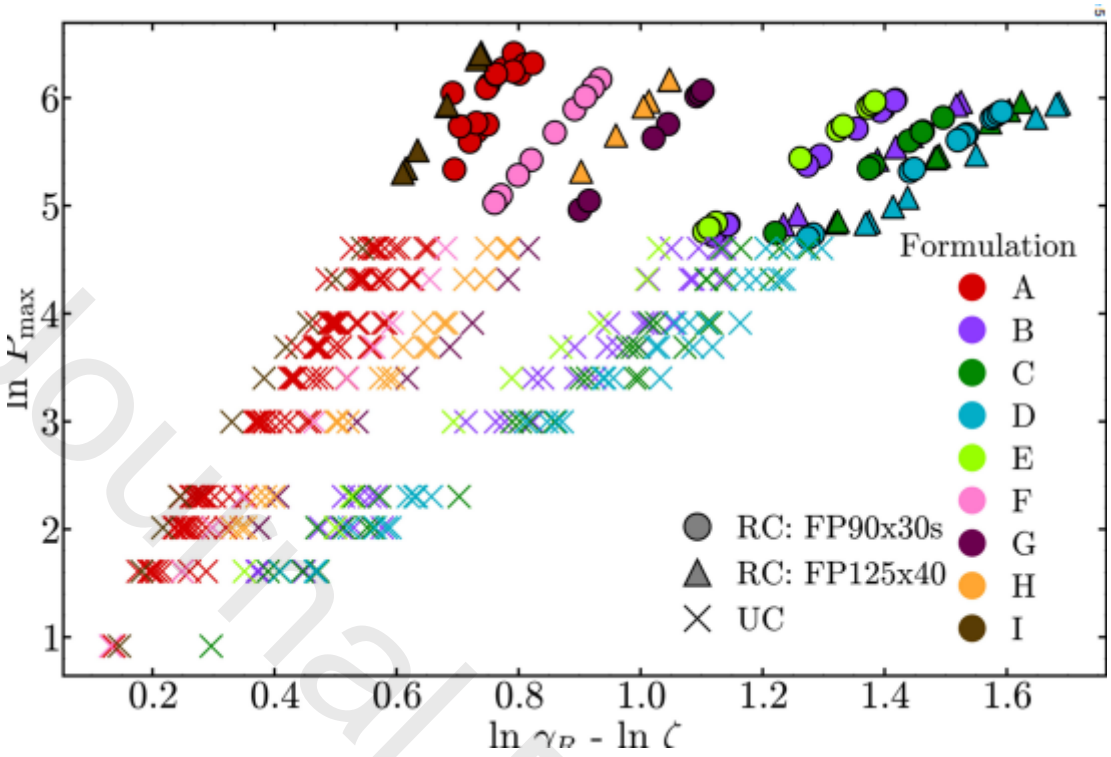
Journal Pre-proofs



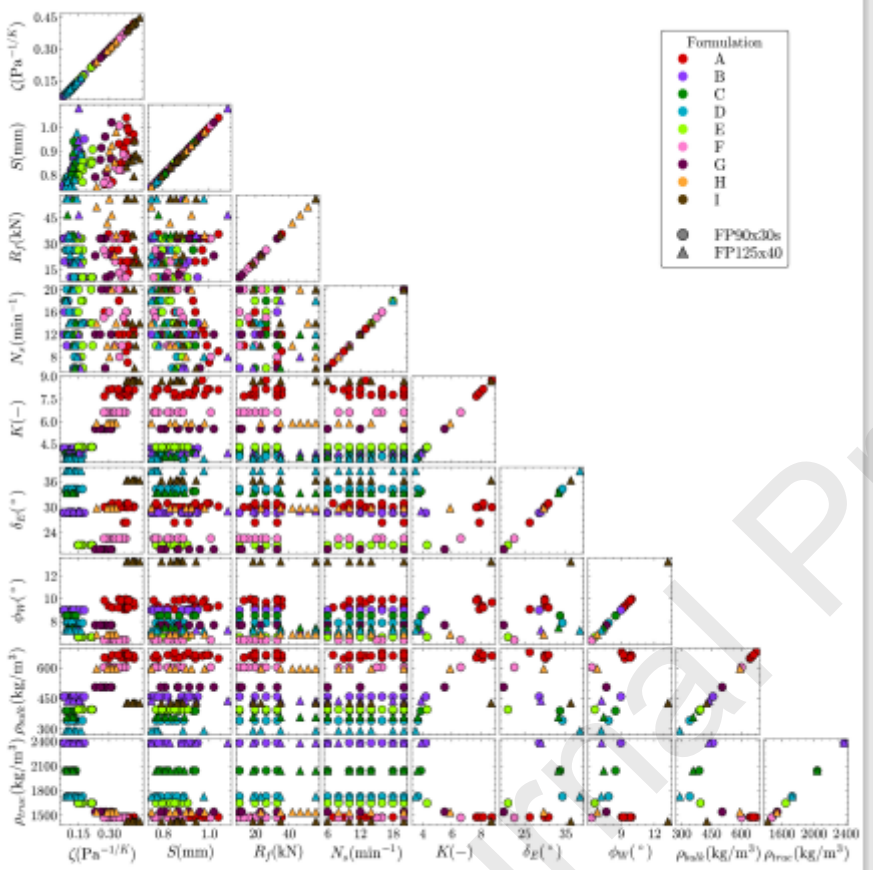
Journal Pre-proofs



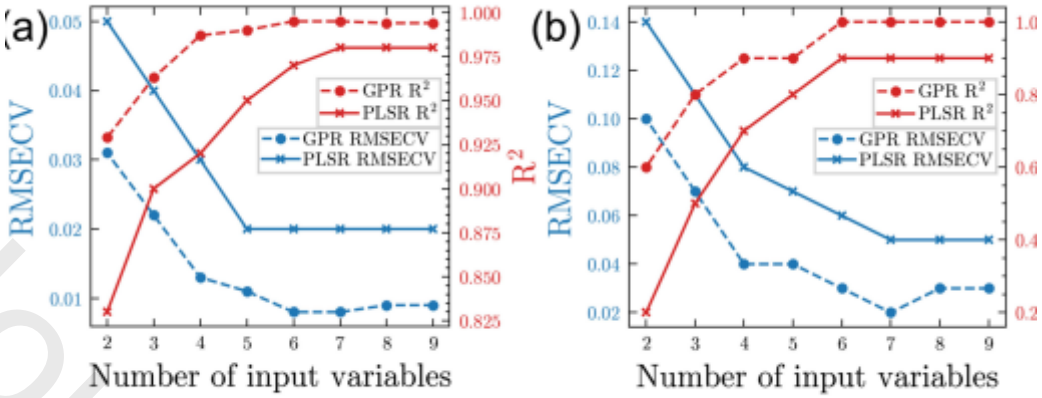
Journal Pre-proofs



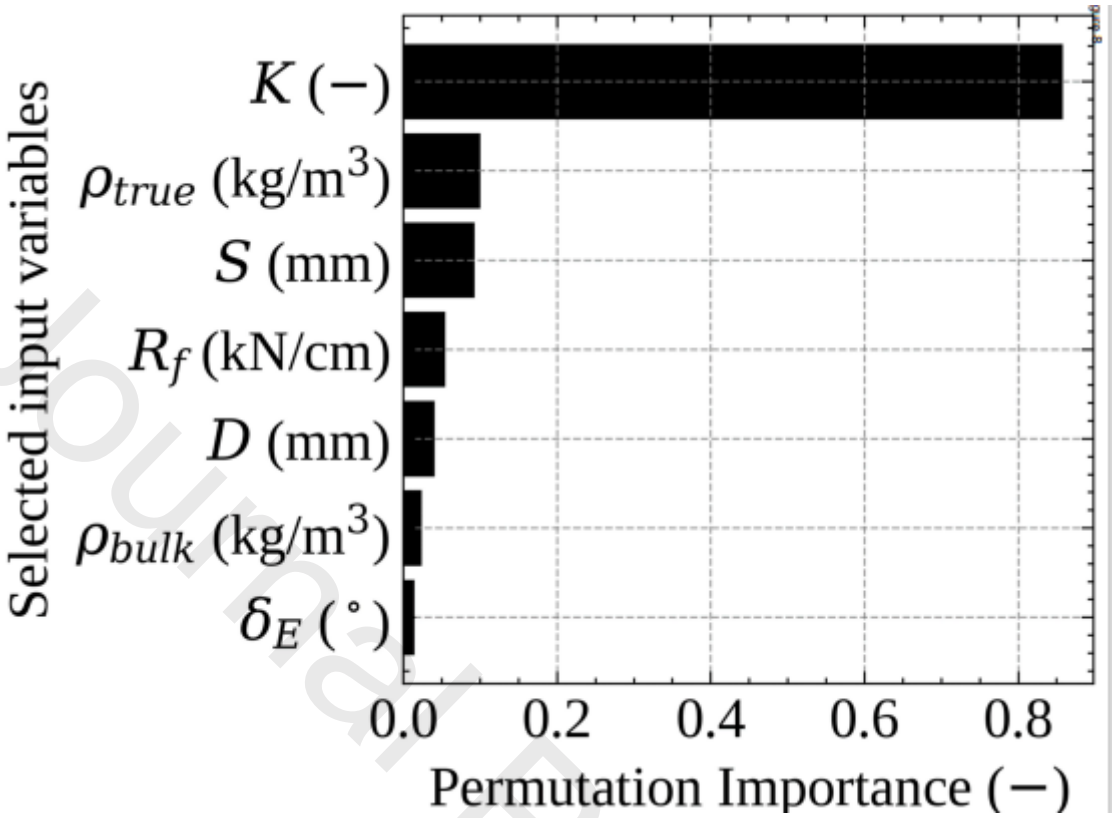
Journal Pre-proofs



Journal Pre-proofs

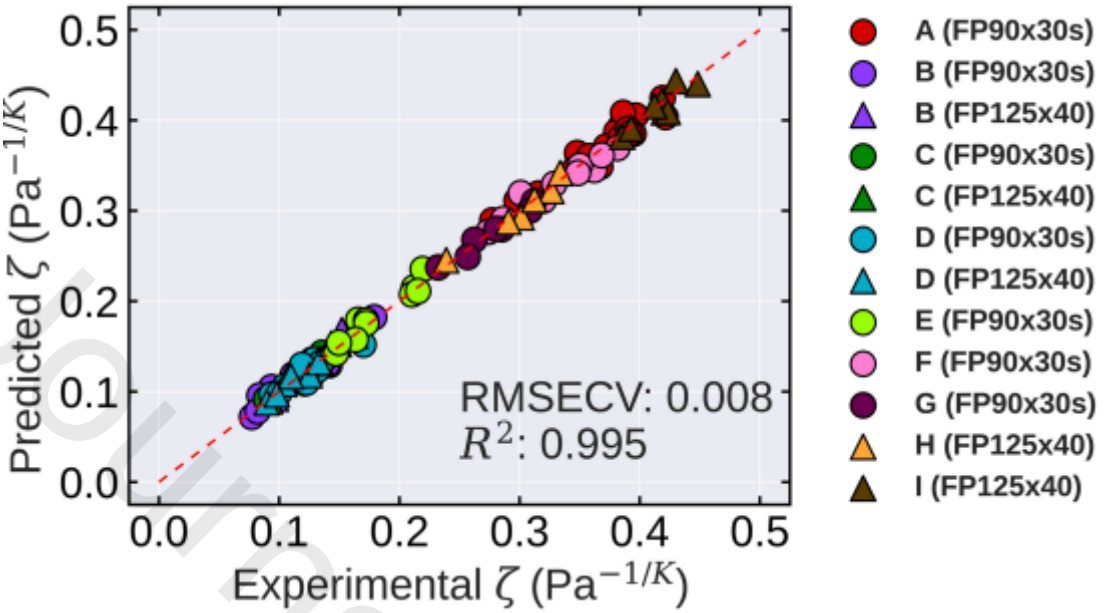


Journal Pre-proofs



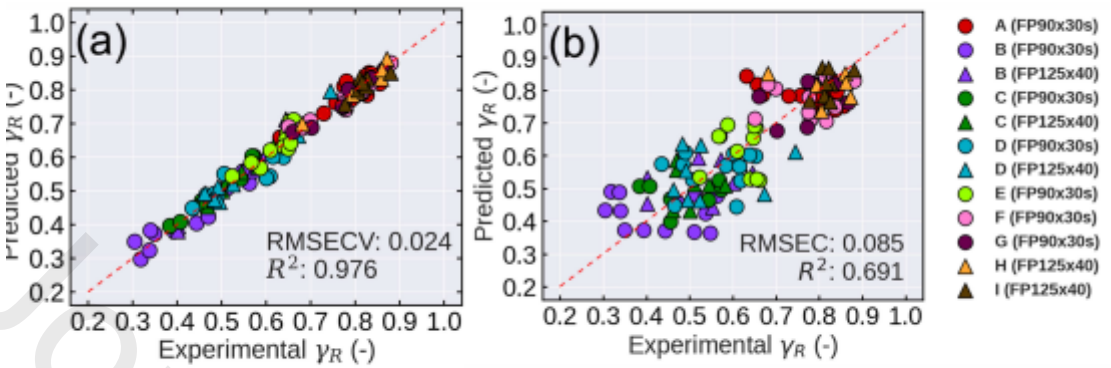
Journal Pre-proofs

Journal Pre-proofs



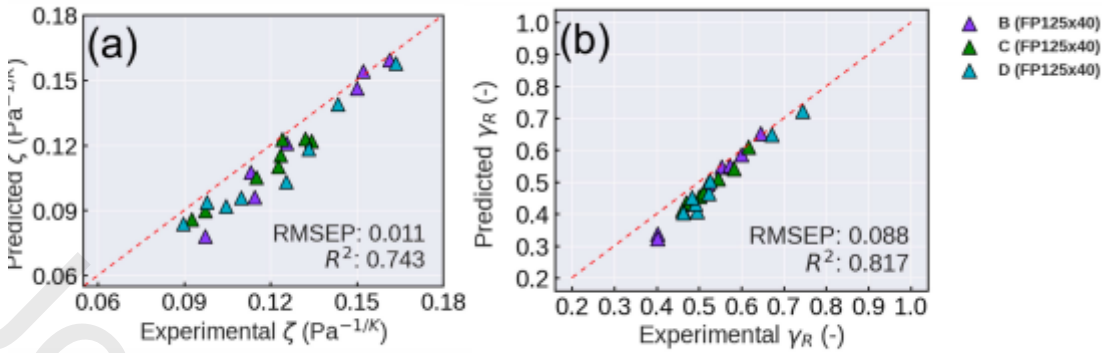
Journal Pre-proofs

Journal Pre-proofs

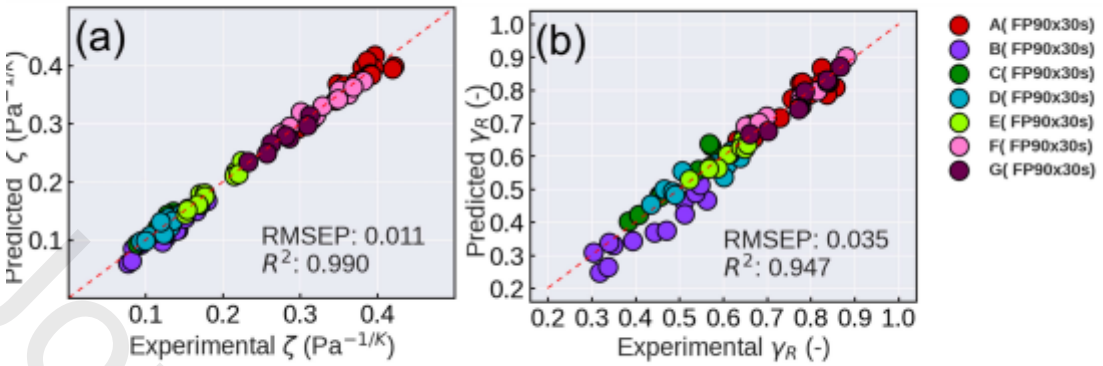


Journal Pre-proofs

Journal Pre-proofs

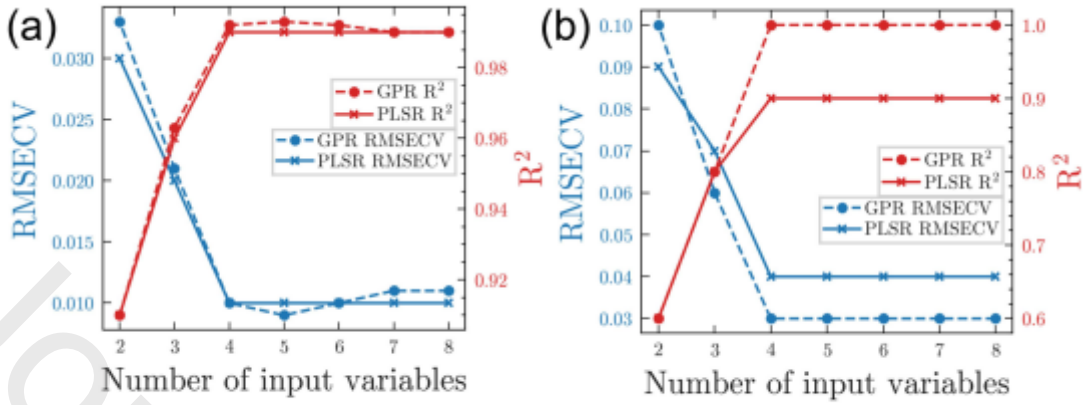


Journal Pre-proofs



Journal Pre-proofs

Journal Pre-proofs



Journal Pre-proofs

Journal Pre-proofs

Journal Pre-proofs