Journal of Science and Technology in Civil Engineering, HUCE, 2026, 20 (3): 127–145

PREDICTING THE RELATIVE SHEAR STORAGE MODULUS OF HIGHLY FILLED BIDISPERSE PARTICULATE SUSPENSIONS BY MACHINE LEARNING AND CONSTRAINED SYMBOLIC REGRESSION

Vu Thai Son[a], Sengaloun Keoalounxay[b], Nguyen Ngoc Tuyen[a], Tran Viet Hung[a,*]

[a]*Faculty of Bridge and Road Engineering, Hanoi University of Civil Engineering, 55 Giai Phong road, Bach Mai district, Hanoi, Vietnam*
[b]*Department of Building Technology, Faculty of Architecture, National University of Laos, Dongdok Campus, Vientiane, Lao PDR*

Article history:
Received 28/8/2026, Revised 16/9/2026, Accepted 18/9/2026, Available online 25/9/2026, Version of Record 25/9/2026

Abstract

Highly filled suspensions of rigid spheres in a yield stress fluid have a relative shear storage modulus that is governed not only by the solid volume fraction but also by the grading of the particles. Classical homogenization schemes contain no maximum packing fraction and therefore cannot reproduce the rapid growth of the modulus that is observed as the solid skeleton approaches its densest state, whereas conventional machine learning delivers accurate predictions in the form of a trained model rather than an explicit expression that a designer can use. This paper compares six machine learning models and two symbolic regression variants on a published set of 70 data points covering seven mixtures of spherical particles in an inverse emulsion, one monodisperse and six bidisperse, with size ratios from 1 to 8.25 and fine fractions from 0 to 0.5; no new measurements were made. The set comprises 63 measurements of $G'(\varphi)/G'(0)$ and seven normalization points at $\varphi = 0$, where the ratio equals unity by definition, and every model is scored on the 63 measurements. The maximum packing fraction, computed from a compressible packing model, is used as an input feature, and predictive ability is assessed with a leave-one-mixture-out protocol that never places a mixture in both the training and the test set. Over the six bidisperse mixtures the Gaussian process model gives a macro-average mean squared error in logarithmic space of 0.0205 against 0.1164 for the one-parameter power law, and its pooled error rises by a factor of 6.4 when the maximum packing fraction is removed from the feature set. The constrained symbolic regression variant returns a power law with no additional numerical constant fitted in the final exponent and reaches a log-space coefficient of determination of 0.950 over the 63 measurements; under the same leave-one-mixture-out protocol, however, the search procedure that produces it is, for the seed-selection rules tested, less accurate than the one-parameter power law. On 16 measurements of Table 3 of [1] at eight gradings that entered no fit, the expression gives a mean squared error of 0.0308 against 0.0081 for the fixed exponent $\sqrt{2}$, so the fixed exponent is the better closed form for these suspensions. Constraining the functional form is therefore necessary but not sufficient for generalization to new gradings.

Keywords: yield stress fluid; relative shear storage modulus; machine learning; symbolic regression; genetic programming; maximum packing fraction; homogenization; bidisperse suspension.

1. Introduction

Mortar, concrete and particle-reinforced cementitious materials are inclusion–matrix composites of stiff, rigid grains in a much softer continuous binder. The solid volume fraction of an ordinary

*Corresponding author. *E-mail address:* hungtv@huce.edu.vn (Hung, T. V.)
https://doi.org/10.31814/stce.huce2026-20(3)-07 p-ISSN 1859-2996; e-ISSN 2734-9268
© 2026 The Author(s). Published by Hanoi University of Civil Engineering. Licensed under CC BY-NC-ND 4.0.

mortar or concrete typically exceeds one half, so the material is a nearly jammed granular skeleton whose voids are filled with binder, not the dilute dispersion assumed in micromechanics. Its effective elastic modulus governs the deformation of the structural element and is highly sensitive to the grading of the solid phase: two mixtures at the same solid volume fraction but different particle size distributions can differ in stiffness by close to an order of magnitude. The present study analyzes a model suspension of rigid spheres in a yield stress fluid, not a cementitious material, and makes no claim beyond that system. The quantity analyzed is the relative shear storage modulus $G'(\varphi)/G'(0)$ of such a suspension, measured in the linear viscoelastic regime; it plays the role that the effective elastic modulus plays for a composite, but it is a shear modulus normalized by that of the matrix, not a Young's modulus.

The classical route to such a prediction is analytical homogenization. The schemes in current use descend from the ellipsoidal inclusion solution and the mean-field, variational and self-consistent estimates built upon it [2–5]. All collapse to the dilute result of Einstein [6] as the volume fraction tends to zero, and none contains a maximum packing fraction: their predictions stay finite below unity, whereas experiments on concentrated suspensions show the modulus growing rapidly as the solid fraction approaches the densest packing the grading permits [1, 7].

Suspension rheology does contain a maximum packing fraction. Krieger and Dougherty [8] proposed the reference expression of the field, in which the relative property diverges as the reduced solid fraction approaches unity with an exponent fixed by the maximum packing fraction, so that the whole influence of the grading is carried by a single geometric parameter. Packing models for granular mixtures describe the loosening and wall effects between coarse and fine grains by two coefficients depending on the size ratio [9, 10], a geometric-interaction correction supplying those used here [11]. A homogenization framework for suspensions in yield stress fluids [12] and a refined morphological representative pattern approach for highly filled composites [13] complete this line.

Machine learning offers a third route: data-driven models predict concrete and construction-material properties with accuracies competitive with empirical design formulas [14, 15]. Their limitation is not accuracy but form: the procedure returns a trained model rather than an expression that can be written down, checked against a limiting case or embedded in a design code. Symbolic regression addresses this. Implemented by genetic programming on expression trees [16], it searches functional forms and parameters simultaneously and returns a closed-form expression, an approach that has recovered physical laws from raw data [17], structure-property relations in materials science [18] and explicit formulas in civil engineering [19, 20].

Two questions remain open. The first concerns physical admissibility: an unconstrained search optimizes a global error metric, and an expression can reach a very low error while violating the limits any relation for the relative modulus must satisfy, namely a relative modulus of unity at zero solid fraction and a divergence at the packing limit. An averaged error cannot detect either, since both occur at the edges of the domain where data are scarce. The second concerns how predictive ability is measured: when a data set is built from a few mixtures, each sampled at several solid fractions, a random split places points of every mixture in both training and test set, so the reported error measures interpolation rather than the ability to predict a new grading. The present paper therefore contributes a symbolic regression variant in which the functional form is constrained a priori, so that both limits hold by construction and only the exponent is searched; an explicit expression for that exponent in terms of the size ratio and the fine fraction, with no additional numerical constant fitted in the final exponent; a quantification of the gap between within-mixture interpolation and generalization to an unseen mixture, obtained with a leave-one-mixture-out protocol applied to every model; and an

external test of the resulting expression on 16 measurements of [1] that entered no fit. This work adds no new measurements: it reuses the data set of [1], the relative shear storage modulus of bidisperse suspensions of rigid spheres in a yield stress fluid, in which the modulus at a given solid fraction varies by more than a factor of eight between the extreme gradings. Fig. 1 outlines the study: Section 2 presents the experimental dataset and maximum packing fraction, Section 3 the analytical reference models and machine learning framework, Section 4 symbolic regression by genetic programming, Section 5 the results and discussion and Section 6 the conclusions.

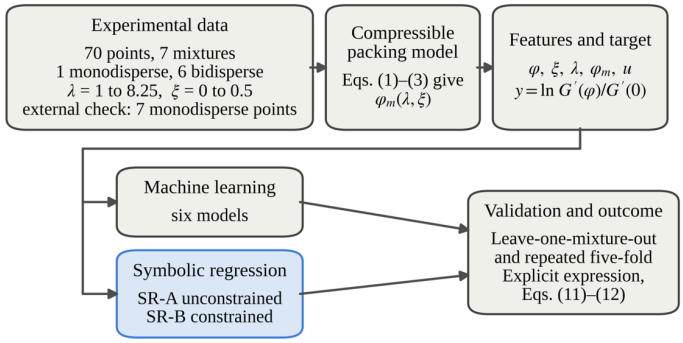


Figure 1. Outline of the study

2. Experimental dataset and maximum packing fraction

2.1. Model system and experimental data

The analysis rests on measurements published in [1], performed on a model system of rigid spheres dispersed in a yield stress fluid, an inverse water-in-oil emulsion, not on a real cementitious material. The particles are polystyrene beads of 80 and 315 μm for the size ratio 3.94 and glass beads of 40 and 330 μm for the size ratio 8.25. The present work adds no new measurements: every experimental value quoted below is taken from [1], the contribution claimed here being one of modeling and analysis of published data.

The mechanical response was characterized by small-amplitude oscillatory shear, a regime in which the suspension deforms reversibly; G' is the shear storage modulus of the linear regime, called elastic modulus in [1]; it is not a Young's modulus. The normalized ratio $G'(\varphi)/G'(0)$ is called the relative shear storage modulus throughout, or relative modulus for short. Each modulus was normalized by that of the particle-free emulsion measured under identical conditions, giving the dimensionless ratio $G'(\varphi)/G'(0)$. That normalization is what makes the data comparable across mixtures: the absolute modulus of the emulsion depends on its composition and preparation, which vary from batch to batch, whereas the ratio is a purely microstructural quantity governed by the particle volume fraction and the grading of the particle phase.

Two descriptors characterize that grading: the size ratio $\lambda = d_c/d_f \geq 1$ between the coarse and the fine particle diameter, equal to unity for a monodisperse mixture, and the fine fraction ξ, the volume fraction of fines within the solid skeleton, so that $\xi = 0$ and $\xi = 1$ both describe monodisperse assemblies. The dataset comprises seven mixtures, one monodisperse and six bidisperse, and contains 70 points: 63 measurements and 7 normalization points at $\varphi = 0$, where the ratio equals unity by definition. The particle volume fraction spans 0 to 0.68 and the relative modulus spans 1 to 106.25. All values were typed from Table 1 of [1] and checked twice against the published table; no figure was digitized. Table 3 of [1] gives 22 further values at $\lambda = 3.94$, $\varphi = 0.4$ and 0.5 and $\xi = 0, 0.1, \ldots, 1.0$; they were overlooked when the data were extracted. They are reconstructed in Table 4 and used in

Section 5.7 as an external test: the 16 values at $\xi = 0.2$ and 0.4 to 1.0 enter no fit, the 4 values at $\xi = 0.1$ and 0.3 are second measurements, at $\varphi = 0.4$ and 0.5, of the gradings of mixtures M2 and M3, and the 2 values at $\xi = 0$ are measurements on the 315 μm polystyrene beads alone, the particles of the monodisperse mixture M1, and therefore repeat measurements of M1. These 6 values add no new grading and are excluded from the main analysis; Section 5.7 reports the effect of treating them as independent observations. The seven monodisperse points of [7] are held out as an external check; they enter no fit and no model selection step, and serve in Section 5 to test a formula already fixed. This check is not fully independent, because the packing fraction $\varphi_m = 0.57$ used in [1] and here is the value measured in [7].

Such a system separates variables entangled in a real mortar: matrix and particles are prepared independently, the modulus contrast is fixed, the particles are rigid, spherical, non-colloidal and sieved to known diameters, and there is neither hydration nor an interfacial transition zone. The price is a restricted scope: transfer to cement paste, mortar or concrete, where grains are angular, gradings continuous and the matrix chemically active, is extrapolation rather than demonstration, a limitation revisited in Section 5.6.

2.2. *Compressible packing model for bidisperse mixtures*

The maximum packing fraction φ_m, at which the granular skeleton locks and the modulus diverges, is computed from the compressible packing model of de Larrard [9], built on the linear packing density model of [10], with the geometric-interaction coefficients of [11]. Two mechanisms reduce the packing efficiency of a binary mixture. In the loosening effect, fines too large to fit inside the voids between the coarse grains push those grains apart and dilate the coarse skeleton. In the wall effect, a coarse grain acts as a rigid wall against which the surrounding fines cannot pack as densely as in the bulk, leaving a shell of excess porosity around it. They are quantified by the loosening coefficient $a(\lambda)$ and the wall-effect coefficient $b(\lambda)$, functions of the size ratio alone,

$$a(\lambda) = \left[1 - \left(1 - \frac{1}{\lambda}\right)^{1.13}\right]^{0.57} \quad ; \quad b(\lambda) = \left[1 - \left(1 - \frac{1}{\lambda}\right)^{1.79}\right]^{0.82} \tag{1}$$

Both tend to unity as $\lambda \to 1$ and to zero as $\lambda \to \infty$, where the fines fill the coarse voids without disturbing them. Depending on which class carries the load, the mixture is coarse-dominant, the fines filling the voids of a coarse skeleton, or fine-dominant, the coarse grains being isolated inclusions in a skeleton of fines. Each state yields its own threshold,

$$\begin{aligned} \varphi_m^c &= \frac{\varphi_{cm}}{1 - \xi\left[1 - a(\lambda)\frac{\varphi_{cm}}{\varphi_{fm}}\right]} \\ \varphi_m^f &= \frac{\varphi_{fm}}{1 - (1-\xi)\left[1 - \varphi_{fm} + b(\lambda)\left(\varphi_{fm} - \frac{\varphi_{fm}}{\varphi_{cm}}\right)\right]} \end{aligned} \tag{2}$$

and, since the skeleton locks as soon as either class saturates, the mixture threshold is the smaller of the two,

$$\varphi_m = \min(\varphi_m^c, \varphi_m^f) \tag{3}$$

where φ_{cm} and φ_{fm} are the residual packing fractions of the coarse and the fine class; following [1], both are set equal, $\varphi_{cm} = \varphi_{fm} = 0.57$, measured on the monodisperse assembly. Eqs. (1)–(3) use only ξ and λ, so φ_m of a held-out or a new mixture is obtained without any response data. The largest discrepancy between the φ_m of Eqs. (1)–(3) and the φ_m reported in [1] is then 0.0041, below 0.6%,

as Table 1 shows. This is not an independent validation, since φ_{cm} is itself taken from mixture M1. Sections 5.1–5.6 use the φ_m reported in [1]; substituting those from Eq. (3) changes the results only marginally and favorably, the error of the proposed expression falling from 0.0580 to 0.0536 and that of the $\sqrt{2}$ power law from 0.0823 to 0.0734.

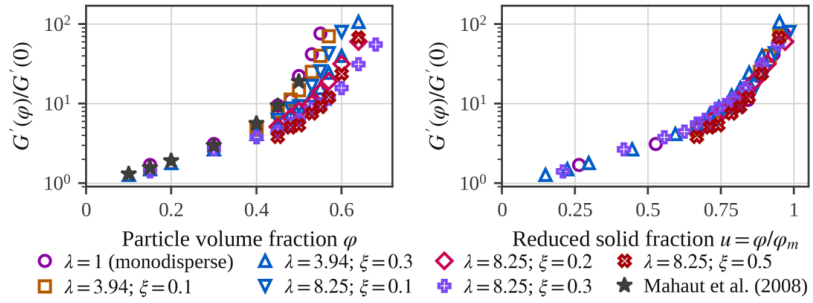
One property matters later: with the two residual packing fractions equal, the model returns the monodisperse value in each degenerate limit, $\varphi_m(\xi \to 0) = \varphi_{cm}$, $\varphi_m(\xi \to 1) = \varphi_{fm}$ and $\varphi_m(\lambda \to 1) = \varphi_{cm}$, so a mixture that ceases to be bidisperse is assigned a single-size packing fraction.

2.3. *Distribution of the data set*

Table 1 summarizes the seven mixtures: size ratio, fine fraction, the two estimates of φ_m, number of points, number of measurements at $\varphi > 0$ and largest φ.

Table 1. Seven particle mixtures of the experimental data set

Mixture	M1	M2	M3	M4	M5	M6	M7
Size ratio λ	1.00	3.94	3.94	8.25	8.25	8.25	8.25
Fine fraction ξ	0.0	0.1	0.3	0.1	0.2	0.3	0.5
φ_m from [1]	0.570	0.600	0.674	0.610	0.660	0.720	0.675
φ_m from Eq. (3)	0.570	0.601	0.674	0.612	0.660	0.716	0.675
Number of points	9	8	14	8	9	13	9
Measurements ($\varphi > 0$)	8	7	13	7	8	12	8
Largest φ	0.55	0.57	0.64	0.60	0.64	0.68	0.64



(a) Versus the particle volume fraction, together with the data of [7]

(b) Versus the reduced solid fraction

Figure 2. Experimental data set, 63 measurements (the normalization points at $\varphi = 0$ are not shown)

Fig. 2 displays the data in two ways. Panel (a) plots the relative modulus against the particle volume fraction, together with the monodisperse points of [7]; the seven mixtures form distinct curves, and at a given φ those with the larger λ and ξ give the lower modulus, a better graded skeleton being further from its packing limit. At $\varphi = 0.55$ the spread exceeds a factor of eight, so the volume fraction alone is not a sufficient descriptor. Panel (b) plots them against the reduced solid fraction $u = \varphi/\varphi_m$, and the seven clusters collapse onto a narrow band, most of the grading dependence being absorbed by φ_m. This collapse is a result of [1], not a finding of the present work; it is reproduced here because it defines the reference for the models of Section 3, and Section 5.1 quantifies how much predictive power comes from φ_m.

3. Analytical reference models and machine learning framework

3.1. Analytical reference models

Three analytical expressions serve as quantitative benchmarks. The first is the dilute result of Einstein [6], the leading correction to the modulus of a matrix filled with a vanishingly small volume fraction of rigid spheres,

$$\frac{G'(\varphi)}{G'(0)} = 1 + 2.5\varphi + O(\varphi^2) \tag{4}$$

The coefficient 2.5 is exact, but Eq. (4) holds only as $\varphi \to 0$, where the perturbation produced by one inclusion is not felt by any other. Between 50 and 70% solid fraction the stiffening is instead controlled by near-contact geometry, so a first-order expansion carries almost none of the relevant physics; Eq. (4) is included precisely to show how badly a dilute expansion fails there.

The second benchmark is the Krieger–Dougherty relation [8], evaluated with the maximum packing fraction φ_m of Table 1,

$$\frac{G'(\varphi)}{G'(0)} = \left(1 - \frac{\varphi}{\varphi_m}\right)^{-2.5\varphi_m} \tag{5}$$

The exponent $2.5\varphi_m$ is not adjusted to the data; it is chosen so that the expansion of Eq. (5) about $\varphi = 0$ matches the dilute result to first order. That is the physical logic of the form — a divergence at the packing limit grafted onto the correct dilute slope — and it leaves no free coefficient: once φ_m is known the curve is completely determined.

The third benchmark relaxes that constraint: the one-parameter power law

$$\frac{G'(\varphi)}{G'(0)} = (1 - u)^{-n}; \quad u = \frac{\varphi}{\varphi_m} \tag{6}$$

in which u is the reduced solid fraction and n is a single exponent fitted globally, that is, one constant shared by all seven mixtures — unlike Eq. (5), whose exponent varies with φ_m from one mixture to the next. This form was proposed in [1] with $n = 1.43$.

The classical homogenization estimates contain no packing threshold but are finite for every $\varphi < 1$, so they can be evaluated at every data point. For rigid spheres in an incompressible matrix the Mori–Tanaka estimate [3] coincides with the Hashin–Shtrikman lower bound [4], $G'(\varphi)/G'(0) = (2 + 3\varphi)/(2 - 2\varphi)$, which [1] plots against its data; it is used here as a fourth benchmark. The self-consistent estimate [5] diverges at a threshold independent of the grading and is not used. Eqs. (5) and (6) take φ_m as an explicit input.

3.2. Input features, target variable and error metrics

All models are trained on a logarithmic target, because $G'(\varphi)/G'(0)$ spans more than two orders of magnitude and its error is multiplicative rather than additive, so a squared error in linear space would be dominated by the few largest points. The target variable and the associated divergence variable are

$$y = \ln \frac{G'(\varphi)}{G'(0)}; \quad L = -\ln\left(1 - \frac{\varphi}{\varphi_m}\right) \tag{7}$$

By Eq. (6), $y = nL$ is exactly linear whenever n is constant — the structure the constrained symbolic regression variant of Section 4 exploits.

The feature set comprises the particle volume fraction φ, the fine fraction ξ, the size ratio λ, the maximum packing fraction φ_m from Table 1, and the reduced solid fraction u. The machine learning

models use λ; symbolic regression uses $1/\lambda$ instead, to avoid the division operator; the constrained variant uses only ξ, $1-\xi$, $1/\lambda$ and φ_m, with φ and u entering solely through L. Apart from these feature-encoding differences, the two model classes share the same data set and the same validation protocols.

Accuracy is reported through the mean squared error in logarithmic space, the log-space coefficient of determination $R^2_{\ln}$ and the mean relative error,

$$\mathrm{MSE}_{\ln} = \frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2; \quad R^2 = 1-\frac{\sum_i(y_i-\hat{y}_i)^2}{\sum_i(y_i-\bar{y})^2}; \quad \mathrm{MRE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\hat{G}_i-G_i}{G_i}\right| \tag{8}$$

where y_i is the measured logarithmic target at point i, $\hat{y}_i$ the corresponding prediction, $\bar{y}$ the mean of the measured values, N the number of points scored, and G_i and $\hat{G}_i$ the measured and predicted relative moduli. The first is the quantity minimized during training. All scores are computed on the 63 measurements with $\varphi > 0$; the seven normalization points are used in training but never scored.

3.3. Regression models

Six regression models were trained on the features of Section 3.2, chosen for the contrast between their inductive biases so that the resulting ranking is interpretable rather than a leaderboard. The random forest [21] and the extremely randomized trees [22] use 500 trees each, averaging piecewise-constant partitions that cannot extrapolate beyond the training targets. Gradient boosting [23] uses 400 trees of a maximum depth of 3 and a learning rate of 0.05, sharing that bias but fitting residuals sequentially. Support vector regression [24] uses a radial basis function kernel, a penalty $C = 50$ and an ε-insensitive margin $\varepsilon = 0.02$, ignoring residuals below the margin. The multilayer perceptron has two hidden layers of 32 units, a smooth global parameterization with no built-in locality. Gaussian process regression [25] uses an anisotropic radial basis kernel plus a white-noise term, a smooth interpolator that adapts one length scale per feature and can therefore learn which inputs matter.

All six are implemented in scikit-learn [26] with fixed settings and fixed random seeds. No hyper-parameter search was performed: no setting was chosen by comparing validation or test scores, so the numbers reported correspond to a single configuration fixed before any scoring. The kernel parameters of the Gaussian process are not tuned in this sense; they are estimated as part of fitting the model on each training fold, as described below.

The tree ensembles use 500 trees with a minimum leaf size of one; gradient boosting 400 trees of depth 3 with learning rate 0.05; support vector regression an RBF kernel with $C = 50$, $\varepsilon = 0.02$ and the scale heuristic for γ; the multilayer perceptron two hidden layers of 32 ReLU units trained by Adam [27] with an L_2 penalty of 10^{-3}; and the Gaussian process a constant kernel times an anisotropic RBF kernel (one length scale per feature) plus a white-noise kernel. Its parameters start from a constant of 1, unit length scales and a noise level of 10^{-3}, are bounded to $[10^{-5}, 10^{5}]$ (the scikit-learn defaults), and are re-estimated in every training fold by maximizing the log marginal likelihood with L-BFGS-B [28] from the initial point and from three further starts drawn log-uniformly within the bounds; the fitted values are listed per fold in Supplementary Table S3. Support vector regression, the perceptron and the Gaussian process are preceded by a standard scaler fitted inside each training fold, so no test information enters the preprocessing. The models are those of scikit-learn 1.8.0 under Python 3.11, NumPy 2.4 [29] and SciPy 1.17 [30]; the seed 20260825 is used throughout (plus the running index of the 30 folds in the repeated random split), and the code and data are provided as a reproducibility package.

3.4. Validation protocols

Two protocols are used: a five-fold random split of the 70 points, repeated six times; and leave-one-mixture-out (LOMO) validation, in which one whole mixture is held out and the models are trained on the remaining six.

Because each mixture contributes only 8 to 14 points, random splitting necessarily places points of the same mixture in both the training and the test set, so it measures interpolation along a curve that is already known; LOMO measures what actually matters in practice, prediction for a grading that was never seen. The monodisperse mixture M1 is the boundary extrapolation fold, its λ and ξ lying at the edge of the design domain, while the six bidisperse folds are pooled as the interior-mixture generalization column of Table 2.

Under LOMO the error is scored on the measurements of the held-out mixture only and reported per mixture, pooled over the 55 bidisperse measurements, and as an equal-weight macro-average over the six bidisperse mixtures with a 95% percentile interval from 20 000 bootstrap resamples of the six mixtures, an exploratory measure given the small number of clusters. The power-law exponent of Eq. (6) is refitted inside each fold. The random split is scored on the $\varphi > 0$ points of each test fold; each repeat scores every measurement once, and the results are reported as the mean and standard deviation over the six repeats.

4. Symbolic regression by genetic programming

4.1. Tree representation and safe operator set

In a conventional regression the functional form is fixed in advance and only its coefficients are free. Symbolic regression instead treats the algebraic expression itself as a variable of the problem: candidate formulas are represented as expression trees and evolved by genetic programming [16], so that the functional form and the numerical constants are optimized simultaneously. The engine used here is a custom implementation supplied in the reproducibility package. The method has recovered physical laws from measurements [17, 31] and property models in materials science [18].

Here a tree carries an operator at every internal node and an input feature or a random constant at every leaf. An initial population of random trees is evolved by tournament selection, subtree crossover and subtree mutation, with the constants of the leading individuals periodically refined by nonlinear least squares.

The operator set was deliberately restricted to addition, subtraction, multiplication, squaring and the square root of the absolute value; there is no division. Protected division, which returns a fixed value whenever the denominator falls below a threshold, hides a singularity instead of removing it. A trial run made this concrete by returning $\xi/(\lambda - 3.94)$, which divides by zero for the two mixtures with $\lambda = 3.94$. Taking the square root of the absolute value keeps that operator total on the real line without introducing a discontinuity, so every expression the search can generate is finite and continuous. This restriction is a deliberate injection of prior knowledge into the search.

4.2. Fitness function, complexity penalty and hyper-parameters

Individuals are ranked for selection by a fitness that combines accuracy with a penalty on size,

$$F = \mathrm{MSE}_{\ln} + \alpha C(T) \tag{9}$$

where $\mathrm{MSE}_{\ln}$ is the mean squared error in logarithmic space defined in Eq. (8), to which the bound penalty of Section 4.4 is added for the constrained variant, $C(T)$ is the number of nodes of the tree T, and α is the parsimony coefficient, set here to $\alpha = 0.004$, with tournament size 7. Trees were

generated to a maximum depth of 4 in the constrained variant and 5 in the unconstrained variant, offspring exceeding that depth by two levels or 40 nodes being discarded. Each leave-one-mixture-out fold was run with 2000 individuals over 70 generations with four random seeds, rerun with eight seeds for the seed spread reported below, and the final search on all 70 points with 3000 individuals over 120 generations with eight random seeds. Leaf constants are drawn uniformly from −3 to 3 (probability 0.25 per leaf) and refined by nonlinear least squares for the 25 best individuals every 10 generations; offspring are produced by subtree crossover (60%), subtree mutation (25%) and constant mutation (15%) with one elite; the generation budget is fixed, without early stopping, and the seed whose retained individual has the smallest training error is reported. Within a run, the individual retained as the result is the one with the smallest training error met during the run, the training error being the $\mathrm{MSE}_{\ln}$ over the training points plus, for the constrained variant, the bound penalty of Section 4.4; the parsimony term of Eq. (9) acts only through tournament selection. The reported expression of Eq. (11) has 7 nodes (the unconstrained expression of Eq. (13) has 16); the packing model behind φ_m carries four calibrated coefficients (Eq. (1)) and the measured $\varphi_{cm} = 0.57$.

With only 70 data points a tree of even moderate depth can memorize the measurements, so the penalty of Eq. (9) is what keeps the returned expression short enough to be read and reported.

All random seeds were fixed, so the searches are exactly reproducible, but reproducibility is not uniqueness. The eight final runs of the constrained variant returned expressions whose $\mathrm{MSE}_{\ln}$ over the 63 measurements ranges from 0.0580 to 0.0694 (0.052 to 0.062 on the 70-point training error), so Eq. (11) is the best member of a spread of comparable candidates rather than a unique global optimum. The spread is larger under leave-one-mixture-out: across eight seeds the held-out error of a single bidisperse fold varies by a factor of 1.5 to 34 for the constrained variant and 1.4 to 7.0 for the unconstrained one, and the training error used to select the seed is a poor guide to the held-out error (Spearman correlation between −0.69 and 0.64 across the bidisperse folds). Supplementary Tables S2a and S2b summarize the seed spread, and the reproducibility package lists every seed of every fold.

4.3. Unconstrained variant (SR-A)

The unconstrained variant searches directly for $y = g(\varphi, \xi, 1/\lambda, \varphi_m, u)$, where g is any expression the algorithm can assemble from the operator set within the depth limit. This buys maximum freedom, and with it the lowest fitting error of any expression considered in this work.

The cost is that nothing in the search enforces the two physical limits that any model of a filled suspension must satisfy: the relative modulus must equal one when the volume fraction vanishes, and it must diverge as the packing threshold is approached. Both are statements about the endpoints of the curve, whereas the fitness of Eq. (9) is a global average. A formula can therefore violate either limit, or both, while scoring extremely well, and no value of $\mathrm{MSE}_{\ln}$ alone can reveal the violation. Section 5 shows that this variant does violate both limits.

4.4. Constrained variant (SR-B)

The constrained variant imposes the functional form of Eq. (6) in advance and searches only for its exponent,

$$y = n(\xi, 1 - \xi, 1/\lambda, \varphi_m) \cdot L \tag{10}$$

with L given by Eq. (7). This construction satisfies both physical limits automatically, for any positive exponent whatsoever. As $\varphi \to 0$ the divergence variable $L \to 0$, so $y \to 0$ and $G'(\varphi)/G'(0) \to 1$; as $\varphi \to \varphi_m$ the variable $L \to \infty$, so the predicted modulus diverges. Both limits follow from the form itself, not from the particular exponent the search returns.

The exponent is searched over four features only, namely the fine fraction ξ, its complement $1-\xi$, the inverse size ratio $1/\lambda$ and the maximum packing fraction φ_m. The volume fraction and the reduced fraction are excluded from the exponent and enter the prediction solely through L, so that the exponent is by construction a property of the grading alone. An additional penalty keeps the exponent inside the interval from 0.2 to 10: it adds 20 times the mean violation of that interval over the training points to the error. It is therefore a soft penalty evaluated at the sampled gradings only, not a hard constraint on the continuous (ξ, λ) domain, and a candidate can satisfy it at the training gradings and leave the interval between them. This happens once among the 64 constrained candidates of this study (eight seeds in each of the seven folds and in the final search), evaluated on a dense grid $\xi \in [0, 1]$, $\lambda \in [1, 8.25]$: the expression selected for the fold that holds out M6 vanishes along a curve near $\xi = 0.84$ to 0.93 (Supplementary Table S2c). For the reported expression the bound holds analytically: Eq. (11) is concave in ξ and lies between $\min(\sqrt{1 + 1/\lambda}, 1 + \sqrt{1/\lambda}) > 1$ and 2 for every $\xi \in [0, 1]$ and $\lambda \geq 1$.

The two variants answer different questions. SR-A tests how well an unconstrained search can fit these measurements; SR-B tests whether constraining the functional form is enough to make the resulting expression generalize to a grading it has never seen.

5. Results and discussion

5.1. Role of the maximum packing fraction

Figure 2(b) showed the seven clusters of Figure 2(a) collapsing onto a narrow band in u: the grading effect is largely absorbed by φ_m, a result of [1], not of the present work.

Over the 63 measurements, Eq. (6) with $n = \sqrt{2}$ gives $\mathrm{MSE}_{\ln} = 0.0823$ and $R^2_{\ln} = 0.929$; a one-parameter fit returns $n = 1.407$ and 0.0821, only 0.2% better, so $\sqrt{2} \approx 1.414$ is a convenient rounding with no theoretical basis, and the $n = 1.43$ of [1] gives 0.0839. Eq. (5), whose exponent varies from 1.425 to 1.800 with φ_m, gives 0.2163, so a constant exponent fits better than one varying with φ_m.

The monodisperse data of [7], which entered no fit, give an external check: Eq. (6) with $n = \sqrt{2}$ and $\varphi_m = 0.57$ gives $\mathrm{MSE}_{\ln} = 0.0006$ and a mean relative error of 2.2% over seven points, below the 4.6% to 23% standard deviation of [7]. The check covers only a monodisperse system with $\varphi \leq 0.50$, and $\varphi_m = 0.57$ is the value measured in [7] itself (Section 2.1).

5.2. Model ranking under leave-one-mixture-out validation

Table 2 ranks the models by increasing macro-average interior-mixture error over the six bidisperse folds (55 measurements) and reports the boundary extrapolation error on the monodisperse mixture M1 (8 measurements). Every row is out of sample; per-mixture $\mathrm{MSE}_{\ln}$, mean relative error and $R^2_{\ln}$ of every model are listed in Supplementary Tables S1a to S1c.

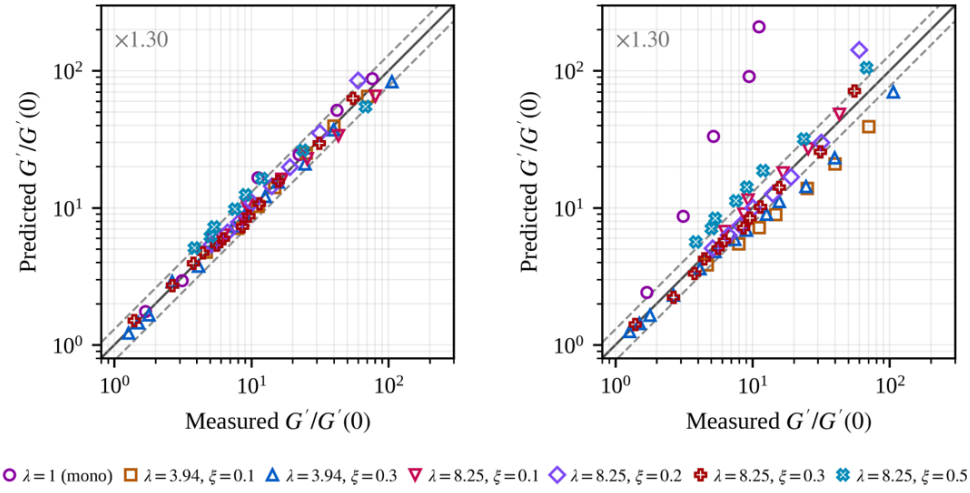
The Gaussian process gives the smallest interior-mixture error, 0.0188, a mean relative error of 10.4% over the 55 bidisperse measurements, with per-mixture errors from 0.0044 to 0.0664. Four models fall between 0.048 and 0.057, below the 0.1107 of the one-parameter power law of Eq. (6); support vector regression and the multilayer perceptron are worse than that law.

On the boundary fold M1 the tree ensembles and the power law keep small errors, from 0.0161 to 0.0303, because the u of M1 still lies inside the training range, whereas both symbolic regression variants blow up, to 5.36 and 8.95. For the constrained variant the fold holding out M1 returns an exponent of 2.90 instead of the 1.37 obtained when M1 is fitted alone: the functional constraint of Eq. (10) is not sufficient for extrapolation. Fig. 3 compares predictions with measurements for the Gaussian process and for the constrained search, both under leave-one-mixture-out. The two symbolic regression rows of Table 2 use the selection rule of Section 4.2, the best of four seeds by training error

in each fold. Under other rules the pooled interior-mixture error of the constrained search is 0.1873 for the best of eight seeds and 0.1671 for the median seed, and a single seed used in every fold gives 0.080 to 0.795; seven of the eight single seeds are less accurate than the one-parameter power law (0.1107), one is more accurate. The statement that the search is less accurate than the power law therefore holds for both training-error rules tested and for the median seed, but not for every seed (Supplementary Table S2b).

Table 2. Mean squared error in logarithmic space under leave-one-mixture-out validation on the 63 measurements; the interior-mixture generalization columns pool and macro-average the six bidisperse mixtures, the boundary extrapolation column is the monodisperse mixture; the interval is a mixture-level percentile bootstrap

Model	Interior-mixture generalization, pooled (55)	Interior-mixture generalization, macro-average [95% CI]	Boundary extrapolation, M1 (8)
Gaussian process	0.0188	0.0205 [0.0078, 0.0400]	0.0355
Extra trees	0.0479	0.0466 [0.0131, 0.0871]	0.0203
Random forest	0.0572	0.0531 [0.0223, 0.0851]	0.0161
Gradient boosting	0.0560	0.0541 [0.0277, 0.0843]	0.0170
Unconstrained symbolic regression (SR-A)	0.0571	0.0575 [0.0229, 0.1062]	5.3556
One-parameter power law, Eq. (6)	0.1107	0.1164 [0.0414, 0.2247]	0.0303
Support vector regression	0.1360	0.1337 [0.0609, 0.2256]	0.6052
Constrained symbolic regression (SR-B), Eq. (10)	0.1355	0.1559 [0.0800, 0.2407]	8.9541
Krieger–Dougherty, Eq. (5)	0.2427	0.2715 [0.1038, 0.4393]	0.0348
Multilayer perceptron	0.4453	0.4509 [0.1812, 0.8310]	0.3277
Mori–Tanaka / Hashin–Shtrikman	1.8909	2.0325 [1.4481, 2.6378]	2.6441
Einstein, Eq. (4)	3.4069	3.6087 [2.8463, 4.4010]	3.9726



(a) Gaussian process under leave-one-mixture-out validation

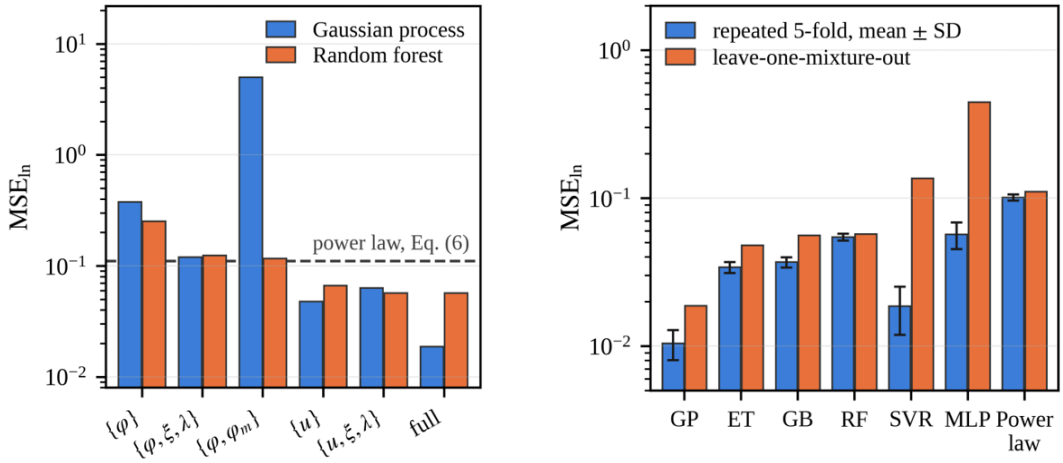
(b) Constrained symbolic regression search under leave-one-mixture-out; four points lie outside the axes

Figure 3. Predicted versus measured relative modulus for the 63 measurements (normalization points at $\varphi = 0$ not shown). Dashed lines bound the factor-of-1.30 band; symbols as in Fig. 2

5.3. Feature ablation and sensitivity to the validation protocol

Feature ablation isolates the role of φ_m with six feature sets: φ alone; φ, ξ, λ; φ, φ_m; u alone; u, ξ, λ; and the full set. Under leave-one-mixture-out, pooled over the 55 bidisperse measurements, the Gaussian process gives 0.3770, 0.1200, 5.0146, 0.0478, 0.0632 and 0.0188, and the random forest 0.2532, 0.1245, 0.1170, 0.0663, 0.0569 and 0.0572. Withholding φ_m (set φ, ξ, λ) multiplies the pooled Gaussian process error by 6.4 and brings both models back to the level of the one-parameter power law; the single feature u already gives 0.0478 and 0.0663. The set φ, φ_m fails for the Gaussian process because the held-out mixture has a φ_m never seen in training and the anisotropic kernel extrapolates poorly in that direction, whereas the random forest treats φ_m as a split variable. The advantage of machine learning comes mainly from φ_m, not from model flexibility.

The validation protocol matters as much. Under the repeated five-fold random split the Gaussian process reaches $\mathrm{MSE_{ln}} = 0.0115 \pm 0.0022$ and the random forest 0.0508±0.0023 (mean ± standard deviation over the six repeats), with R^2_{ln} of 0.990±0.002 and 0.956±0.002 over the 63 measurements (Supplementary Table S4a). Scored on the same 55 bidisperse measurements as the pooled column of Table 2 (Supplementary Table S4b), leave-one-mixture-out is higher by factors of 1.79, 1.40 and 1.51 for the Gaussian process, extra trees and gradient boosting, but only 1.05 for the random forest. With 8 to 14 points per mixture a random split places the same mixture on both sides of the partition, so a reported R^2_{ln} depends on the protocol and flatters the model. Fig. 4 makes both effects visible: withholding φ_m lifts the ablation bars above the power-law reference, and the error of most models rises under leave-one-mixture-out, that of the random forest changing little.



(a) Factorial feature ablation (pooled over the 55 bidisperse measurements) under leave-one-mixture-out validation

(b) Effect of the validation protocol on the 55 bidisperse measurements: repeated five-fold random split (mean ± standard deviation over six repeats) versus leave-one-mixture-out

Figure 4. Role of the maximum packing fraction

5.4. The explicit closed-form expression

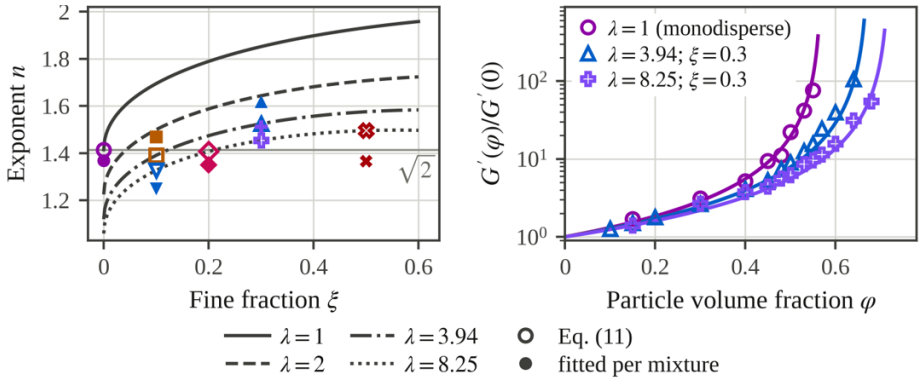
Across the eight final runs on all 70 points, the constrained variant's best expression contains no additional numerical constant fitted in the final exponent,

$$n(\xi, \lambda) = \sqrt{1 - \xi + \frac{1}{\lambda}} + \sqrt{\xi} \tag{11}$$

and the prediction formula is

$$\frac{G'(\varphi)}{G'(0)} = \left(1 - \frac{\varphi}{\varphi_m}\right)^{-n(\xi,\lambda)} \tag{12}$$

with φ_m from Table 1; Table 3 collects the errors of the closed-form expressions fitted on the whole data set. In Eq. (12) φ_m enters only through u, not through the exponent — a property of this expression, not a constraint of the algorithm. The eight final runs span $\mathrm{MSE}_{\ln} = 0.0580$ to 0.0694 over the 63 measurements (Section 4.2). For the monodisperse mixture Eq. (11) gives $n = \sqrt{2} = 1.414$, close to the 1.407 of Section 5.1. Fig. 5(a) compares the seven values of Eq. (11), 1.327 to 1.524, with the exponents fitted mixture by mixture, 1.249 to 1.618; both cluster around $\sqrt{2}$. Fig. 5(b) compares Eq. (12) with three mixtures.



(a) Exponent given by Eq. (11) compared with the exponent fitted mixture by mixture

(b) Eq. (12) compared with three representative mixtures

Figure 5. The closed-form expression of Eqs. (11)–(12)

Table 3. Error of the closed-form expressions over the 63 measurements

Expression	Numerical constants fitted	$\mathrm{MSE}_{\ln}$	$R^2_{\ln}$	MRE (%)
Krieger–Dougherty, Eq. (5)	0	0.2163	0.814	47.7
Power law with $n = \sqrt{2}$, Eq. (6)	0	0.0823	0.929	21.2
One-parameter power law, Eq. (6)	1	0.0821	0.929	21.1
Constrained symbolic regression, Eqs. (11)–(12)	0	0.0580	0.950	19.1
Unconstrained symbolic regression (SR-A)	1	0.0295	0.975	14.2
Mori–Tanaka / Hashin–Shtrikman	0	1.9865	−0.708	59.3

Eqs. (11)–(12) give $\mathrm{MSE}_{\ln} = 0.0580$ and $R^2_{\ln} = 0.950$, about 30% below the $\sqrt{2}$ power law and 27% of the Krieger–Dougherty error. Three qualifications follow. First, 0.0580 is an in-sample error, not a prediction error as in Section 5.2. Second, a paired mixture-level percentile bootstrap, in which the seven mixtures are resampled with replacement 20 000 times and the difference between the pooled $\mathrm{MSE}_{\ln}$ of the $\sqrt{2}$ power law and that of Eq. (12) is recomputed on each resample, gives a 95% interval from −0.0118 to +0.0607, which contains zero (Supplementary Table S6); Eq. (11) is better on four mixtures, equal on M1 and worse on two (two-sided sign test, $p = 0.69$). With seven clusters, one of them tied by construction, the interval is exploratory: it shows that the in-sample gain is not established, not that the two expressions are equivalent. Third, under leave-one-mixture-out the

variant that produced Eq. (11) gives 0.1355, above the 0.1107 of the power law; that row of Table 2 evaluates the search procedure, not Eq. (11) itself.

5.5. *Physical bounds and asymptotic limits*

The unconstrained variant returns, in the logarithmic target of Eq. (7),

$$y = (u^4 + u)^2 + \frac{\sqrt{\xi}}{\lambda} + 1.627\varphi \tag{13}$$

Eq. (13) attains $\mathrm{MSE}_{\ln} = 0.0295$, $R^2_{\ln} = 0.975$ and a mean relative error of 14.2%, the best fit of all, yet it violates both physical bounds: at $\varphi = 0$ it gives $G'(0)/G'(0) = \exp(\sqrt{\xi}/\lambda)$, exceeding unity by up to 14.9% unless $\xi = 0$, and at $\varphi = \varphi_m$ it stays finite instead of diverging. The Gaussian process violates both bounds too, returning 0.966 to 1.061 at $\varphi = 0$ and a finite value at the threshold.

The structure of Eq. (10) does not encode the degenerate limits, so Eq. (11) does not satisfy them. As $\xi \to 0$, $\xi \to 1$ or $\lambda \to 1$ the exponent should return to about 1.37 to 1.41, but Eq. (11) gives $n = 1.06$ at $\lambda = 8.25$ with $\xi = 0$, and $n = 1.85$ at $\lambda \to 1$ with $\xi = 0.3$. The packing model of Eq. (3), by contrast, satisfies them. This explains the extrapolation failure of Section 5.2 and is the central argument of the paper: constraining the functional form is necessary but not sufficient.

5.6. *Range of applicability and limitations*

The factor-of-1.30 band, $0.77 \le G'_{\text{predicted}}/G'_{\text{measured}} \le 1.30$, contains 57 of the 63 measurements for the Gaussian process under leave-one-mixture-out and 53 of 63 for Eq. (12) fitted on all data. Fig. 6 shows the logarithmic error of four models against u: trendless for $u < 0.6$, then growing toward the packing limit. For $u > 0.9$ the ratio of prediction to data for Eq. (12) spans 0.79 to 2.92 over 12 points with a median of 0.97, so Eq. (12) should not be used close to the threshold.

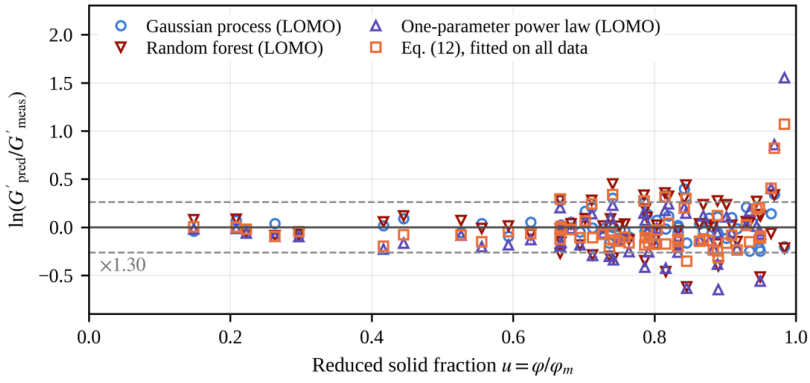


Figure 6. Logarithmic error of four models against the reduced solid fraction for the 63 measurements (normalization points at $\varphi = 0$ not shown); dashed lines bound the factor-of-1.30 band

A first-order expansion of Eq. (12) gives n/φ_m from 2.02 to 2.48, so it does not exactly recover the Einstein coefficient of 2.5. The data set contains only bidisperse spheres with size ratios up to 8.25, so Eq. (11) is unvalidated outside that range, and the functional form was searched on these same 70 points. No direct comparison is made with [13], which needs per-mixture calibration of a morphological parameter; Eq. (12) needs only λ, ξ and φ_m. The packing model and the regression models used here have already been applied to cementitious materials, for other properties. Compressible packing models guide the mix design of ultra high performance concrete [32], and data-driven models predict strength and workability from mix proportions [33]. The Gaussian process exists only as a trained model, Eqs. (11)–(12) as a closed-form expression.

5.7. *External test on the unused gradings of [1]*

Table 3 of [1] reports the relative modulus at $\lambda = 3.94$ for $\varphi = 0.4$ and 0.5 at eleven fine fractions from 0 to 1; Table 4 reconstructs it. The gradings $\xi = 0.1$ and 0.3 are those of mixtures M2 and M3 and are excluded as repeat measurements, and the values at $\xi = 0$, the 315 μm polystyrene beads alone, are excluded as repeat measurements of M1. The repeats indicate the scatter of the data: at $\xi = 0.1$ Table 3 gives 4.68 and 15.16 where Table 1 gives 4.68 and 14.80, and at $\xi = 0.3$ it gives 3.87 and 9.05 where Table 1 gives 4.13 and 9.01. At $\xi = 0$ Table 3 gives 5.60 and 21.01 where mixture M1 gives 5.19 and 22.28. The remaining 16 values at eight gradings entered no fit and no selection step. The two values at $\xi = 1.0$, the 80 μm beads alone, are retained although they are monodisperse, because the 80 μm beads are never measured alone in Table 1. Their φ_m comes from Eqs. (1)–(3), which use only ξ and λ, since [1] gives no packing fraction for these gradings. The machine learning models are trained once on the 70 points of Table 1, the closed-form expressions are those of Table 3, and every model is scored on the same 16 measurements. Table 5 gives the results, split into the gradings $\xi = 0.2$ to 0.5, which lie between or just beyond the training gradings $\xi = 0.1$ and 0.3, and the gradings $\xi = 0.6$ to 1.0, which lie outside the training range of ξ.

Table 4. Relative modulus of Table 3 of [1] at $\lambda = 3.94$, reconstructed, with φ_m from Eq. (3) and the use made of each grading

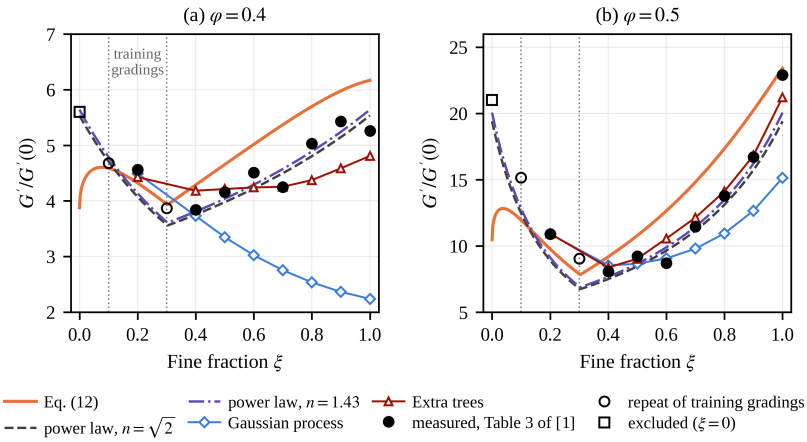
ξ	$G'/G'(0)$ at $\varphi = 0.4$	$G'/G'(0)$ at $\varphi = 0.5$	φ_m from Eq. (3)	Use
0.0	5.60	21.01	0.570	Coarse beads alone (315 μm); repeat of M1, excluded
0.1	4.68	15.16	0.601	Grading of M2; repeat, excluded
0.2	4.56	10.90	0.635	External test
0.3	3.87	9.05	0.674	Grading of M3; repeat, excluded
0.4	3.84	8.07	0.658	External test
0.5	4.16	9.23	0.642	External test
0.6	4.51	8.71	0.626	External test
0.7	4.25	11.45	0.611	External test
0.8	5.03	13.78	0.597	External test
0.9	5.43	16.71	0.583	External test
1.0	5.26	22.91	0.570	External test; fine beads alone (80 μm)

Eq. (12) gives $\mathrm{MSE}_{\ln} = 0.0308$ and a mean relative error of 17.0%, against 0.0081 and 7.3% for Eq. (6) with $n = \sqrt{2}$ and 0.0062 and 6.1% for the exponent $n = 1.43$ of [1]. Fig. 7 shows why: at $\lambda = 3.94$ Eq. (11) gives exponents from 1.47 to 1.58 for the tested gradings, above 1.43, so Eq. (12) over-predicts at every tested grading except $\xi = 0.2$, and the over-prediction grows with u: it is largest at $\varphi = 0.5$ for $\xi = 0.6$ to 0.9. The dependence of the exponent on ξ that the search extracted from the two gradings at $\lambda = 3.94$, $n = 1.390$ at $\xi = 0.1$ and 1.524 at $\xi = 0.3$, is not supported by the eight further gradings; a constant exponent with φ_m from Eq. (3) describes the dependence on ξ better. Restricted to the 14 bidisperse values, the ranking is unchanged: Eq. (12) gives 0.0333 and 18.0% against 0.0071 and 6.9% for $n = \sqrt{2}$. The one-parameter fit of Table 3, $n = 1.407$, gives 0.0093, Eq. (5) 0.0245, and Eq. (13) 0.0176 with 0.0067 on the first group and 0.0242 on the second, so the unconstrained expression also degrades outside the training range of ξ. The other seven final runs of the unconstrained search give 0.0055 to 0.0127 on the 16 measurements, five of them below the 0.0081 of the $\sqrt{2}$ power law. None of the seven contains ξ or λ: the grading enters only through u and φ_m, so they return unity at $\varphi = 0$, but like Eq. (13) none diverges at $\varphi = \varphi_m$. The eight final runs of the constrained search give 0.0115 to 0.2899, all above 0.0081 (Supplementary Table S7).

Among the machine learning models the extra trees give 0.0073, the random forest 0.0080 and gradient boosting 0.0130, close to the power laws. The Gaussian process, best under leave-one-mixture-out, gives 0.1640: 0.0092 on the gradings $\xi = 0.2$ to 0.5 but 0.2569 on $\xi = 0.6$ to 1.0, because outside the training range of ξ it reverts toward its prior mean, whereas the tree ensembles and the power law hold. Support vector regression and the multilayer perceptron are worse still. The Mori–Tanaka / Hashin–Shtrikman estimate gives 0.9686 and Eq. (4) 1.8085, finite baselines that lie below all 16 measurements.

Table 5. External test on the 16 measurements of Table 3 of [1]: mean squared error in logarithmic space over the 16 measurements and over the two groups of gradings, and mean relative error; φ_m from Eq. (3); machine learning models trained on the 70 points of Table 1 only

Model	$\mathrm{MSE_{ln}}$ (16)	MRE (%)	$\mathrm{MSE_{ln}}$, $\xi = 0.2$–0.5	$\mathrm{MSE_{ln}}$, $\xi = 0.6$–1.0
Power law with $n = 1.43$ of [1], Eq. (6)	0.0062	6.1	0.0081	0.0050
Extra trees	0.0073	6.3	0.0017	0.0107
Random forest	0.0080	7.1	0.0040	0.0104
Power law with $n = \sqrt{2}$, Eq. (6)	0.0081	7.3	0.0115	0.0060
One-parameter power law, Eq. (6), $n = 1.407$	0.0093	7.9	0.0134	0.0068
Gradient boosting	0.0130	9.7	0.0107	0.0143
Unconstrained symbolic regression (SR-A), Eq. (13)	0.0176	11.4	0.0067	0.0242
Krieger–Dougherty, Eq. (5)	0.0245	14.1	0.0286	0.0220
Constrained symbolic regression, Eqs. (11)–(12)	0.0308	17.0	0.0134	0.0412
Multilayer perceptron	0.1141	35.2	0.0699	0.1406
Gaussian process	0.1640	22.8	0.0092	0.2569
Support vector regression	0.3098	60.4	0.0240	0.4813
Mori–Tanaka / Hashin–Shtrikman	0.9686	55.5	0.5912	1.1951
Einstein, Eq. (4)	1.8085	68.2	1.2924	2.1182



(a) $\varphi = 0.4$ (b) $\varphi = 0.5$

Figure 7. External test on Table 3 of [1] ($\lambda = 3.94$; φ_m from Eq. (3)); machine learning models trained on the 70 points of Table 1): relative modulus against the fine fraction at (a) $\varphi = 0.4$ and (b) $\varphi = 0.5$; dotted lines mark the training gradings

The six excluded values were then treated as independent observations (Supplementary Tables S5a to S5c). Adding the four repeats to M2 and M3 (74 points, 67 measurements) moves the pooled

interior-mixture error of the Gaussian process from 0.0188 to 0.0177 and that of the one-parameter power law from 0.1107 to 0.1061, leaves the ranking of Table 2 unchanged, and moves the refitted exponent from 1.407 to 1.411. Retrained on the enlarged set, the Gaussian process gives 0.1321 on the 16 external measurements instead of 0.1640 and the extra trees 0.0083 instead of 0.0073, while the closed forms with a fixed exponent are unaffected. The constrained search run on the enlarged set again selects Eq. (11) as its best final expression, and its leave-one-mixture-out error, 0.2135 over 59 bidisperse measurements, stays above that of the power law. Adding the two $\xi = 0$ values as further points of M1 gives the same picture. Scored instead as external points, the two $\xi = 0$ values lower the error of the $\sqrt{2}$ power law on the extended set of 18 to 0.0075 but raise that of Eq. (12) to 0.0618, because at $\xi = 0$ and $\lambda = 3.94$ Eq. (11) gives $n = 1.12$ instead of a monodisperse value near $\sqrt{2}$: the degenerate-limit failure of Section 5.5. The repeated pairs also set a floor: their log differences give a replicate variance of about 0.0012 per measurement (0.0006 from the four bidisperse repeats), so the best external errors, 0.006 to 0.008, lie five to seven times above the scatter of the data.

The result is negative: Eq. (12), which fits the training gradings best, is inferior to the fixed exponent on gradings it has not seen; the dependence of the exponent on ξ that the search extracted from two gradings at $\lambda = 3.94$ is not supported by eight further gradings. For these suspensions the recommended closed form is Eq. (6) with a fixed exponent and φ_m from Eqs. (1)–(3).

6. Conclusions

The maximum packing fraction explains most of the variation of the relative shear storage modulus across the seven mixtures. Once the solid fraction is rescaled into the reduced variable $u = \varphi/\varphi_m$, they collapse onto one curve that a power law of $1 - u$ reproduces with an exponent of $\sqrt{2}$, a convenient rounding of the fitted value 1.407 and close to the exponent 1.43 proposed in [1]. With no fitted constant it attains $R^2_{\ln} = 0.929$ over the 63 measurements.

The machine-learning study confirmed this from a different direction. Under leave-one-mixture-out validation, Gaussian process regression reached a macro-average $\mathrm{MSE}_{\ln}$ of 0.0205 against 0.1164 for the one-parameter power law, an apparently decisive margin; but removing the maximum packing fraction from the feature set multiplied its pooled error by a factor of 6.4. The advantage of machine learning therefore rests mainly on that single physically motivated feature rather than on model capacity: what the algorithm exploits is the packing information supplied to it, not a richer functional vocabulary.

Constrained symbolic regression returned Eq. (11), with no additional numerical constant fitted in the final exponent, giving $\mathrm{MSE}_{\ln} = 0.0580$, $R^2_{\ln} = 0.950$ and a mean relative error of 19.1% over the 63 measurements. The improvement must be stated with care: the 95% paired mixture-level bootstrap interval of the reduction relative to the $\sqrt{2}$ power law runs from −0.0118 to +0.0607, an exploratory interval over seven mixtures that contains zero. Under leave-one-mixture-out the same variant gives 0.1355 against 0.1107, so an advantage in fitting does not translate into an advantage in predicting a new grading. Constraining the functional form is necessary but not sufficient for generalization to new gradings: the search space of Eq. (10) does not encode the degenerate limits, and Eq. (11) misses them.

The external test on the 16 measurements of Table 3 of [1] settles the comparison: Eq. (12) gives 0.0308 against 0.0081 for the $\sqrt{2}$ power law and 0.0062 for the exponent 1.43 of [1], so the exponent expression does not transfer to gradings outside the training set, and the Gaussian process, best under leave-one-mixture-out, gives 0.1640 on the 16 measurements and 0.2569 outside the training range of the fine fraction, against 0.0073 and 0.0107 for the extra trees. This negative result is the main outcome of the study.

For practical use on suspensions of this kind, the fixed-exponent power law of Eq. (6) with φ_m from Eqs. (1)–(3) is the recommended closed form; Eq. (12) keeps 53 of the 63 measurements inside the factor-of-1.30 band but should not be used outside the gradings on which it was fitted nor close to the packing threshold, where the scatter grows sharply.

Two directions follow: adding soft penalty terms inside the genetic fitness function so that the exponent converges to the monodisperse value in the degenerate limits $\xi \to 0$, $\xi \to 1$ and $\lambda \to 1$, and validating the expression on other yield stress fluids, continuous gradings and non-spherical particles, the condition for any transfer to a real cementitious material.

Acknowledgment

The experimental data reanalyzed in this study were published in [1]; no new measurements were performed for the present work. The authors gratefully acknowledge the institutional support that made this research possible.

Declaration of generative AI and AI-assisted technologies

The author(s) declare that no generative AI or AI-assisted technologies were used in the preparation of this manuscript.

References

[1] Vu, T.-S., Ovarlez, G., Chateau, X. (2010). Macroscopic behavior of bidisperse suspensions of noncolloidal particles in yield stress fluids. *Journal of Rheology*, 54(4):815–833.

[2] Eshelby, J. D. (1957). The determination of the elastic field of an ellipsoidal inclusion, and related problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 241(1226):376–396.

[3] Mori, T., Tanaka, K. (1973). Average stress in matrix and average elastic energy of materials with misfitting inclusions. *Acta Metallurgica*, 21(5):571–574.

[4] Hashin, Z., Shtrikman, S. (1963). A variational approach to the theory of the elastic behaviour of multiphase materials. *Journal of the Mechanics and Physics of Solids*, 11(2):127–140.

[5] Hill, R. (1965). A self-consistent mechanics of composite materials. *Journal of the Mechanics and Physics of Solids*, 13(4):213–222.

[6] Einstein, A. (1906). Eine neue Bestimmung der Moleküldimensionen. *Annalen der Physik*, 324(2): 289–306.

[7] Mahaut, F., Chateau, X., Coussot, P., Ovarlez, G. (2008). Yield stress and elastic modulus of suspensions of noncolloidal particles in yield stress fluids. *Journal of Rheology*, 52(1):287–313.

[8] Krieger, I. M., Dougherty, T. J. (1959). A mechanism for non-Newtonian flow in suspensions of rigid spheres. *Transactions of the Society of Rheology*, 3(1):137–152.

[9] de Larrard, F. (1999). *Concrete mixture proportioning: A scientific approach*. E & FN Spon, London.

[10] Stovall, T., de Larrard, F., Buil, M. (1986). Linear packing density model of grain mixtures. *Powder Technology*, 48(1):1–12.

[11] Bournonville, B., Coussot, P., Chateau, X. (2004). Modification du modèle de Farris pour la prise en compte des interactions géométriques d'un mélange polydisperse de particules. *Rhéologie*, 7:1–8.

[12] Chateau, X., Ovarlez, G., Trung, K. L. (2008). Homogenization approach to the behavior of suspensions of noncolloidal particles in yield stress fluids. *Journal of Rheology*, 52(2):489–506.

[13] Vu, T.-S., Tran, B.-V., Nguyen, H.-Q., Chateau, X. (2023). A refined morphological representative pattern approach to the behavior of polydisperse highly-filled inclusion-matrix composites. *International Journal of Solids and Structures*, 270:112253.

[14] Ben Chaabene, W., Flah, M., Nehdi, M. L. (2020). Machine learning prediction of mechanical properties of concrete: Critical review. *Construction and Building Materials*, 260:119889.

[15] Le, B.-A., Vu, V.-H., Seo, S.-Y., Tran, B.-V., Nguyen-Sy, T., Le, M.-C., Vu, T.-S. (2022). Predicting the compressive strength and the effective porosity of pervious concrete using machine learning methods. *KSCE Journal of Civil Engineering*, 26(11):4664–4679.

[16] Koza, J. R. (1992). *Genetic programming: On the programming of computers by means of natural selection*. MIT Press, Cambridge, MA.

[17] Schmidt, M., Lipson, H. (2009). Distilling free-form natural laws from experimental data. *Science*, 324 (5923):81–85.
[18] Wang, Y., Wagner, N., Rondinelli, J. M. (2019). Symbolic regression in materials science. *MRS Communications*, 9(3):793–805.
[19] Le, B.-A., Tran, B.-V., Vu, T.-S., Vu, V.-H., Nguyen, V.-H. (2024). Predicting the compressive strength of pervious cement concrete based on fast genetic programming method. *Arabian Journal for Science and Engineering*, 49(4):5487–5504.
[20] Vu, V.-H., Le, B.-A., Tran, B.-V., Bui, T.-L., Nguyen, V.-H. (2025). A hybrid micromechanical-based symbolic regression model for transverse effective conductivity of high-contrast component composites. *Archive of Applied Mechanics*, 95(8):194.
[21] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.
[22] Geurts, P., Ernst, D., Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1):3–42.
[23] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232.
[24] Smola, A. J., Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14 (3):199–222.
[25] Rasmussen, C. E., Williams, C. K. I. (2006). *Gaussian processes for machine learning*. MIT Press, Cambridge, MA.
[26] Pedregosa, F., Varoquaux, G., Gramfort, A. et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
[27] Kingma, D. P., Ba, J. (2015). Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations (ICLR)*, San Diego, CA.
[28] Byrd, R. H., Lu, P., Nocedal, J., Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208.
[29] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825): 357–362.
[30] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, I., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., Vijaykumar, A., Bardelli, A. P., Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C. N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D. A., Hagen, D. R., Pasechnik, D. V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G. A., Ingold, G.-L., Allen, G. E., Lee, G. R., Audren, H., Probst, I., Dietrich, J. P., Silterra, J., Webber, J. T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J. L., de Miranda Cardoso, J. V., Reimer, J., Harrington, J., Rodríguez, J. L. C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N. J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P. A., Lee, P., McGibbon, R. T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T. J., Robitaille, T. P., Spura, T., Jones, T. R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y. O., Vázquez-Baeza, Y. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272.
[31] Udrescu, S.-M., Tegmark, M. (2020). AI Feynman: A physics-inspired method for symbolic regression. *Science Advances*, 6(16):eaay2631.
[32] Dong, P. S., Tuan, N. V., Thanh, L. T., Thang, N. C., Yang, K.-H. (2021). Mix design of high-volume fly ash ultra high performance concrete. *Journal of Science and Technology in Civil Engineering (STCE) - HUCE*, 15(4):197–208.
[33] Vu, V.-T., Dinh, Q.-T. (2025). Extreme gradient boosting model for forecasting slump and compressive strength of high-performance concrete. *Journal of Science and Technology in Civil Engineering (JSTCE) - HUCE*, 19(2):78–91.