

Supervised Learning for Accurate Mesoscale Simulations of Suspension Flow in Wall-bounded Geometries

Erika I. Barcelos,^{1,2} Shaghayegh Khani,¹ Mônica F. Naccache,² and Joao Maia^{*1}

¹Department of Macromolecular Science and Engineering, Case Western Reserve University, 2100 Adelbert Rd, Cleveland, OH, USA

²Department of Mechanical Engineering, PUC - Rio, 225 Marquês de São Vicente, Gávea, Rio de Janeiro, RJ, Brazil

(*Electronic mail: joao.maia@case.edu)

(Dated: 24 March 2022)

Herein we have employed a supervised learning approach combined with Core-Modified Dissipative Particle Dynamics Simulations (CM-DPD) in order to develop and design a reliable physics-based computational model that will be used in studying confined flow of suspensions. CM-DPD was recently developed and has shown promising performance in capturing rheological behavior of colloidal suspensions; however, the model becomes problematic when the flow of the material is confined between two walls. Wall-penetration by the particles is an unphysical phenomenon that occurs in coarse-grained simulations such as Dissipative Particle Dynamics (DPD) that mostly rely on soft inter-particle interactions. Different solutions to this problem have been proposed in the literature; however, no reports have been given on how to deal with walls using CM-DPD. Due to complexity of interactions and system parameters, designing a realistic simulation model is not a trivial task. Therefore, in this work we have trained a Random Forest (RF) for predicting wall penetration as we vary input parameters such as interaction potentials, flow rate, volume fraction of colloidal particles, confinement ratio and etc. The RF predictions were compared against simulation tests and a sufficiently high accuracy and low errors were obtained. This study shows the viability and potentiality of ML combined with DPD to perform parametric studies in complex fluids.

I. INTRODUCTION

Studying physical properties and transport phenomena in particulate flows is a multiscale problem encompassing a fundamental understanding of the connection between molecular physical phenomena and fluid macroscopic behavior. To bridge the gap between microscopic and macroscopic properties, mesoscale modeling approaches can be successfully applied in modeling fluid flows.

As a mesoscale particle-based method, Dissipative Particle Dynamics has been extremely successful in modeling soft matter and has successfully captured physics of complex fluids¹⁻⁷.

By utilizing soft potentials, DPD has shown extremely versatile and flexible in the modeling of different soft matter systems including but not limited to polymers^{1,5,8-13}, gels^{7,14,15}, and suspensions^{2-4,16}. In most industrial, biomedical, pharmaceutical applications and many biological systems such as blood flow, fluid is flown through a confined geometry. DPD despite its advantages could be associated with artifacts when modelling wall-bounded systems such as density oscillations at the wall^{17,18}. This matter has been a subject of study and many solutions have been developed for parametrizing DPD to avoid numerical artifacts¹⁷⁻²².

Different boundary conditions have been proposed for realistic modeling of fluid motion in between the walls^{21,23-26}. However, traditionally, the walls are made out of frozen particles which are still free to interact with the fluid particles. This approach is associated with challenges which are mainly due to the soft inter-particle interactions. Penetration of fluid particles in the walls is inevitable and avoiding penetration and controlling density fluctuations in the vicinity of the walls is an extremely difficult task which requires extensive theoretic

cal and parametric studies.

Preventing wall penetration in DPD-based walls was first addressed nearly three decades ago by²⁷. Posterior to this initial study, Pivkin and co-workers proposed a series of relevant studies targeting the nonphysical phenomena taking place at the wall^{17,19,20}. In those studies, several alternative strategies were introduced, including testing different boundary conditions, modifying wall forces, controlling density fluctuations and evaluating different wall densities and repulsion interactions. Additionally, walls adopting different geometries have been addressed in the past years²¹. In this particular work, the walls were not predefined, instead, local fluid particles were made in a way that they were able to detect the wall on-the-fly based on neighbouring particles. Previously, we explored²² the effect of fluid-wall interactions and wall density on controlling wall penetration and density fluctuations as these two parameters have been shown to have the most impact in determining a realistic physics-based simulation set up.

In the current study we expand upon our previous investigations of wall-bounded flows and we would like to propose a model for simulating flow of colloidal suspensions in a confined geometry. For this purpose, we have employed the framework proposed by Whittle and Travis²⁸ named Core-Modified Dissipative Particle Dynamics CM-DPD, which represents colloidal particles with a rigid core and a soft hydrodynamic shell. This method has been evaluated and expanded by our group in multiple studies^{2,3,7,16} and has been found to be very advantageous and promising in capturing the full spectrum of suspension rheology with much lower computational cost compared to models based on standard DPD simulations.

Reports of DPD applied to complex fluids in wall geometries can be found in the literature^{27,29-31}. In suspensions, however, the number of works is not as extensive, maybe be-

cause an additional complexity appears in setting the interactions since there are parameters at the particle level, such as rigidity and concentration, that may also play an unknown role in particle penetration. In other words, setting the wall-particle interactions might not be as trivial as in the case of pure solvent and wall system.

Adjusting colloid and solvent interactions in suspensions in a way that it prevents nonphysical effects is a key aspect in creating a successful simulating system for studying fluid properties. Since there are other parameters, in addition to wall repulsion, which play a role in particle penetration a systematic study about the individual and collective effect of those variables is essential to a better understanding of the problem and to the creation of a more reliable system.

Understanding the extension to which particle parameters and flow properties impact particle migration and wall penetration in DPD-based suspensions has not been addressed yet even though it is essential in the establishment of a geometry that can be posteriorly used to study the physical properties of a system. Adjusting colloidal and solvent interactions to avoid penetration is a key aspect in creating a successful simulating geometry for a reproducible study.

Exploring the essential relationships and associations between the parameters affecting penetration is a challenge in two aspects. First, there might exist interactions between the parameters that influence the response variable, thus, not only individual effects must be taken into account but also possible interactions between them. Secondly, particle penetration potentially varies in a non-linear way depending on the levels of the parameters adopted. Therefore, developing a strategy to quantify penetration would be an extremely complex task.

Due to the vast variety of data and the great number of Machine Learning (ML) algorithms, today it is possible to solve an enormous range of problems. Therefore, Machine Learning algorithms can be used as a powerful tool for understanding correlations between different parameters and hidden patterns within data through statistical analysis and decoupling of combined effect. Machine Learning, in a time efficient manner, provides a systematic understanding of individual effects and combined responses especially in cases when the response variable depends in an unknown and complex way on the original variables. There has been a plethora of interest in using Machine Learning in material science for making fast predictions about materials' behavior as introduced in recent review studies³²⁻³⁵. ML has also been combined with DPD simulations for modelling different fluid properties³⁶⁻⁴⁰.

As described earlier in the Manuscript, accurate definition of system parameters in confined flow of fluids in DPD simulations on its own is a difficult task and is still a subject of research. CM-DPD simulations of colloidal suspension flow between walls introduces more complexities in wall description. In particular, wall-particle and wall-solvent interactions, concentration of colloids, rigidity of the particles and flow rate are the main deterministic factors that give rise to the artifacts caused due to presence of the walls. In a highly concentrated system, for example, the overall motion of particles will be slowed down, since there is a higher number of colloidal particles that have real masses. More rigid particles are likely to

have an influence in penetration, since they are more viscous and more repulsive than soft ones. Flow rate and confinement ratio may also play a role in particle distribution in migration, and when all these factors are present, quantifying penetration in order to control the magnitude of the interactions become a very challenging task.

To the best of our knowledge, there has not been a report of accurate modeling CM-DPD suspensions in a wall-bounded flow which systematically studies the effect of simulation parameters on wall-penetration. It is, however, an exceptionally important study at the core development of a trustworthy simulated system. Therefore, the objective of this work is to understand how individual and collective effects in terms of particle and flow characteristics will influence wall penetration by solvent and colloidal particles. Intuitively, solvent penetration relies uniquely on wall-solvent interaction parameters²², however we intend to investigate if different properties of colloidal particles influence how solvent particles migrate towards the walls. Quantifying colloidal penetration and understanding how colloidal parameters affect penetration is also a key goal of this study. Developing a mathematical expression enabling quantification of penetration for both colloid and solvent would be a challenge. Therefore, we propose a machine learning approach to make predictions about penetration considering a series of input parameters.

Herein, a supervised learning approach is applied to predict solvent and colloid penetration in a pressure driven flow. In order to understand how particle rigidity, concentration, body force, confinement ratio and interaction parameters influence solvent penetration, and most importantly, to which extent colloidal penetration is affected when those variables are tuned. This work is organized as follows: Initially, the computational method behind the simulated system as well as the machine learning architecture used are introduced. Details of the chosen algorithm as well as the performance metrics used to evaluate the model accuracy are also highlighted. Next, the results of the training process are presented for colloid and solvent penetration. The choice of the best hyperparameters for each model was also performed and a systematic discussion of the results is carried out.

II. SIMULATION BACKGROUND

A. Dissipative Particle Dynamics

Dissipative Particle Dynamics⁴¹ is a particle-based simulation approach popularly employed to model complex fluids. One of the main advantages of DPD is the flexibility and versatility in modeling a large variety of structures. DPD particles in its essence are not actual real particles, but points in the space having an interaction range defined as r_c , representing the maximum distance in which particle interactions are active and it is commonly set to $r_c = 1$. In DPD, the Newton equation of motion Eq.(1) is solved for pairwise interaction between DPD particles according to three forces:

$$\mathbf{F}_i = \sum_{j \neq i} \mathbf{F}_{ij}^C + \mathbf{F}_{ij}^D + \mathbf{F}_{ij}^R \quad (1)$$

$$\mathbf{F}_{ij}^C = a_{ij} w_{ij}^C(r_{ij}) \mathbf{e}_{ij} \quad (2)$$

$$\mathbf{F}_{ij}^R = \frac{\sigma \omega_{ij}^R(r_{ij}) \theta_{ij} \mathbf{e}_{ij}}{\sqrt{dt}} \quad (3)$$

$$\mathbf{F}_{ij}^D = -\gamma \omega_{ij}^D(r_{ij}) (\mathbf{e}_{ij} \cdot \mathbf{v}_{ij}) \mathbf{e}_{ij} \quad (4)$$

The conservative force $\mathbf{F}_{ij}^C$, described in Eq.(2) expresses the repulsive character of the DPD particles and it governs the interactions between the components in the system. The strength of the interactions is controlled by the magnitude of the repulsive parameter, a_{ij} . $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ and $|\mathbf{r}_{ij}| = |\mathbf{r}_i - \mathbf{r}_j|$ which expresses the calculated distance between the particles. The unit vector $\mathbf{e}_{ij}$ is given by $\mathbf{e}_{ij} = \mathbf{r}_{ij}/r_{ij}$ and w_{ij}^C is a weight function.

$$w_{ij}^C(r_{ij}) = \begin{cases} \left(1 - \frac{r_{ij}}{r_c}\right) & r_{ij} < r_c \\ 0 & r_{ij} \geq r_c \end{cases} \quad (5)$$

Groot and Warren⁴², in their pioneer work in DPD, employed an equation of state to map the repulsion parameters to the compressibility of different systems. In order to keep the water compressibility, the repulsion parameter between DPD particles have to be set at $a=25k_B T$ when system density ρ_s is constant at 3.

The random force, $\mathbf{F}_{ij}^R$ in Eq.(3), represents the thermal fluctuations and works as a heat source in system. To compensate the energy added by this force a dissipative force is also introduced. The dissipative force, $\mathbf{F}_{ij}^D$ Eq.(4), expresses the viscous forces and it controls the relative velocities of the particles. Additionally, it removes the energy added by the random force. In the equations, γ_{ij} and σ_{ij} represent the dissipative and random coefficients, $\mathbf{v}_{ij}$ is the relative velocity of the pair of particles i and j , $\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j$ and θ_{ij} is the white noise that introduces randomness in the system. The random and dissipative force combined form a canonical ensemble NVT and are related by the fluctuation-dissipation theorem as describe by Espanol⁴³, which states that the following relationships must be valid in order to preserve the correct thermodynamics:

$$\omega^D(r_{ij}) = [\omega^R(r_{ij})]^2 \quad (6)$$

$$\sigma^2 = 2\gamma k_B T \quad (7)$$

ω is the weight function, σ and γ are, respectively, the random and dissipative coefficients.

Modified Velocity Verlet Algorithm⁴² was used to integrate the equation of motion and calculate and update the particles positions and velocities.

B. Core-Modified Dissipative Particle Dynamics

As an alternative approach to the traditional freezing method to model suspensions, CM-DPD was introduced in 2010 by Whittle and Travis²⁸. This method consists in representing the colloidal particles as a rigid core with an attached outer dissipative coating; this is added to make the particles more closely related to real physical particles. In contrast to traditional DPD, in which the interactions are soft and center-to-center, in CM-DPD they are semi-hard and surface-to-surface. As a consequence, the traditional r_c , is replaced by a h_{ij} term, given by $h_{ij} = r_{ij} - R_i - R_j$, which is the surface-surface distance equivalent version accounting for colloid interactions. R_m and R_n are the radii of the pair of particles and the weight function takes the form:

$$\omega_{ij} = \begin{cases} \left(1 - \frac{h_{ij}}{r_c}\right) & h_{ij} < r_c \\ 0 & h_{ij} \geq r_c \end{cases} \quad (8)$$

There are three different type of interactions present in the system: DPD-DPD, which represents solvent-solvent and solvent-wall interactions, DPD-colloid, given by solvent-colloid and wall-colloid interactions and finally, colloid-colloid. In the first case the forces taking place are the regular ones for DPD interactions (Eq.1). In the second case, the interaction when the distances between the center of the DPD particle and the core surface of the colloidal particle is approximately zero the repulsion forces are maximum and it vanishes when the distance between them are beyond $1h_{ij}$. Fig. 1a illustrates the scenario where the interaction is maximum and Fig. 1b minimum. In the colloid-colloid case, the forces are maximum when the cores are about to form a contact (Fig. 1c) and vanishes as a function of the ranges in which the colloidal forces are applied.

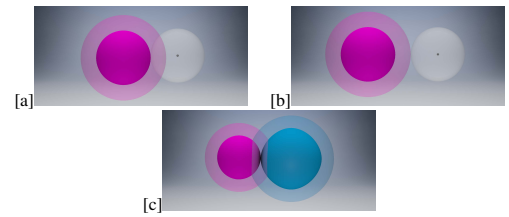


FIG. 1. Interactions between DPD-colloid and colloid-colloid particles. White beads represent DPD particles and pink and blue spheres constitute the colloidal particles having a size ratio difference of 1.4. Interaction between a DPD bead and a colloid particle when the repulsion is maximum (a) and minimum (b). (c) Maximum interaction between two colloidal particles.

The physical phenomena in the colloidal world is driven by colloidal forces⁴⁴. In a simulated environment, not all of them can be evaluated. Instead, the most significant ones for the phenomena of interest are usually included in the system. Hydrodynamic interactions, as one of the most relevant forces in

colloidal mixtures, are typically considered⁴⁴. In short, it describes the flow disturbance promoted by the presence of the particles. In CM-DPD the short-range hydrodynamic interactions are given as:

$$\mathbf{F}_{ij}^H = -f_{ij}^H (\mathbf{e}_{ij} \cdot \mathbf{v}_{ij}) \mathbf{e}_{ij} \quad (9)$$

where f_{ij}^H is the pair-drag term⁴⁵

$$f_{ij}^H = \begin{cases} \frac{(3\pi\eta_0 R^2)}{2h_{ij}} & h_{ij} > \delta \\ \frac{(3\pi\eta_0 R^2)}{2\delta} & 0 < h_{ij} \leq \delta \end{cases} \quad (10)$$

The hydrodynamic force is much stronger at very close separation distances, and although it is always active in the system it becomes very weak at $h_{ij} > 1r_c$. For this reason, in this model it is accounted only short-range hydrodynamic force, i.e. lubrication. At contact point ($h_{ij} = 0$) the force diverges to infinity, thus, a δ term is added to truncate the force at $\delta = 0.000001$.

To replicate the repulsive character of the interactions and avoid cores overlapping, a core force is also introduced in the system (Eq.11). In opposition to the hydrodynamic force, this force is active in a very narrow window, limited to when the particles are in near contact. The force has been slightly modified by² and accordingly, it is the version adopted in this work. In⁴⁶ the core force was mapped to physical units by the shear modulus; the softest particles ($f_{ij}^{core} = 100$) correspond to 400MPa and the most rigid ones ($f_{ij}^{core} = 25,000$) to 100GPa³.

$$\mathbf{F}_{ij}^{core} = \begin{cases} f_{ij}^{core} \mathbf{e}_{ij} & h_{ij} < 0.02 \\ 0 & h_{ij} \geq 0.02 \end{cases} \quad (11)$$

C. Predictive Modeling

Machine learning consists of data-driven approaches to support the decision-making process in a recurrent problem. Basically, a model can learn and understand how to solve complex problems from data reflecting past experiences [42]. Essentially, an input matrix containing the features and their associated values is fed in a pre-determined machine learning algorithm and the model will make predictions from it. To compute how accurate the predictions are a performance metric is used, in which the differences between true values and predictions are computed. At first the predictive power will be poor and the outputs will be the result of a random guess, but as the model keeps receiving new data it starts learning from it and eventually it will improve the output prediction.

Depending on the problem to be solved and the type of data different ML approaches can be used. Supervised learning, which corresponds to the class of algorithms with the goal of making predictions, can be classified into two tasks according to desired predicted output. In classification tasks, the algorithm predicts a categorical variable from the independent

variables while in regression tasks the response is a continuous value. Several ML models can be applied to a specific problem and the choice of the most appropriate one depends, essentially, on the type of data and type of problem.

Decision learning tree is one of the Machine Learning algorithms that enables developing predictive models via data observation. Herein, we train a Random Forest (RF), which is a learning method that is mostly used for classification and regression constituted of ensembles of decision trees. RF are used to predict solvent and wall penetration value from a collection of input parameters. A detailed description of the method is given in following.

D. Random Forests

Random forests RF, introduced in 1995² and extended in 2001 by⁴⁷ is a very popular machine learning model able to solve regression and classification problems. As an ensemble method, RF combine decisions trees that work as weak learners and return predictions based on the average results returned by the individual trees. The fundamental concept is that these uncorrelated trees when combined will most frequently outperform the individual components and will yield more accurate predictions.

RF are essentially a modification of the bagging technique⁴⁸, which is particularly useful in reducing the model variance and therefore, it is very effective in preventing overfitting. The overall idea behind bagging is that by combining multiple and uncorrelated predictors a more stable and assertive prediction can be obtained. Bagging is often used in decision trees, that are considered high variance algorithms.

Decision trees are the constituents of RF. The fundamental idea behind them is that if a model can learn certain rules from the training data it will be able to make predictions on new and unseen data. In short, they are built from the root node, corresponding to the attribute that best separates the observations. From the root nodes the tree is split in smaller subsets (nodes) until it reaches a decision node. An important step in building a decision tree is selecting the attribute to be used in the root node as well as in the branches. A random selection would yield bad accuracy results, thus, a common criterion for splitting the trees is the Gini index and Entropy.

The most used performance metrics to evaluate how accurate the model predictions are, which, in regression, means how close the predictions made by the model are compared to the simulated results are mean squared error MSE, root mean squared error RMSE and R^2 .

$$MSE = \frac{1}{n} \sum_{i=1}^m (y - y_i)^2 \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^m (y - y_i)} \quad (13)$$

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (14)$$

y is the actual value obtained by the simulations and y_i is the model predicted output. SS_{res} is the sum of squared residuals and SS_{total} the total sum of squares.

III. SIMULATION DETAILS

The simulation box was built by placing the walls at the edges of the computational domain. The number of solvent DPD particles, colloids, and wall DPD particles were calculated to ensure a constant density in the entire system of approximately $\rho_s = 3$. The box size used was regulated according to the desired confinement ratio. Once the y -direction was fixed, the other two dimensions were set in a way that the in all the cases the box volume remained practically unchanged. Periodic boundary conditions were applied in the x and z directions and no wall boundary condition was employed.

The wall design was inspired by the previous works^{17,19}, in which a double layer of symmetrically spaced DPD particles were frozen to represent it. These particles are not allowed to move neither interact among themselves, and therefore are excluded from the calculations. However, they are free to interact with the remaining particles in the system.

Colloidal particles were built as Core-Modified particles having a size ratio difference of 1.4 and masses calculated as $\rho_s \frac{4}{3} \pi R^3$, where d_s is the system density. A bimodal suspension was used and to achieve the desired global concentration the number of particles was calculated in a way that:

$$VF_{global} = VF_{type1} + VF_{type2} \quad (15)$$

given

$$VF_{type1} = VF_{type2} \quad (16)$$

VF_{type1} and VF_{type2} are, respectively, the volume fraction of colloidal particle type 1 and type 2. The solvent assumes the water compressibility, and hence the a_{ij} between solvent DPD particles remained fixed at $a_{ij}=25$. For DPD-colloid interactions, the a_{ij} used was 100^6 . The dissipative γ and random σ coefficients adopted are 50 and 10^6 . It is important to emphasize that the a_{ij} that was varied, as shown in Table I, is the one between wall/solvent and wall/colloids. In this work, these two wall interactions varied by the same proportion. The time step used was 5×10^{-5} and the simulations ran for 500,000 time steps, after an equilibrium step of 100,000 time steps. Five input variables were used with varying levels, as showed in Table II. Combining all of them resulted in the total of 558 simulations. An uniform body force is applied to all solvent particles which is equivalent to a pressure drop in a Poiseuille flow.

The box dimensions differs according to the confinement ratio, however an approximate volume of $21,600r_c$ was employed in all cases. The number of colloidal and solvent particles respective to each VF is displaced in Table II.

Wall penetration is calculated as:

a_{ij}	Body Force	Rigidity	y/D	VF
25	0.1	100	5	0.08
50	1	1000	7.5	0.28
75	10	25,000	10	0.48
100			20	0.58

TABLE I. Parameters with the respective levels adopted in the simulations. a_{ij} refers to the interactions between wall/DPD and wall/colloids. y/D represents the confinement ratio, y is the box width and D is DPD particle diameter. Rigidity is set by the strength of the core force f_{core} in colloid-colloid interactions.

VF	Number of Colloidal Particles	Number of Solvent Particles
0.08	93	19872
0.28	327	15552
0.48	562	11232
0.58	679	9072

TABLE II. Total number of colloidal and solvent particles with the corresponding VF for $y/D=10$. In the other confinement ratios the number is slightly different considering that the box volume is not precisely the same.

$$Penetration = \frac{d_n}{d_{100\%}} \quad (17)$$

where

$$d_{n_{dpd}} = \frac{N_{dpd}}{V_{layer}} \quad (18)$$

$$d_{n_{col}} = \frac{N_{col}M_{col1} + N_{col}M_{col2}}{V_{layer}} \quad (19)$$

V_{layer} is the volume of the layer expressed as $V_{layer} = V_{box}/n_{layers}$ and n_{layer} is the number of layer in which the box is divided, being the first and last layer of the channel corresponding to the wall. $d_{100\%}$ corresponds to the layer density considering 100% of penetration. N_{dpd} is the number of DPD particles; for simplification, since the DPD mass is one, the mass term is not included in the equation. N_{col} is the number of colloids an M_{col} is the colloid mass. In order to compute penetration, the densities were calculated in the wall layers and penetration was calculated according to Eq. (17).

A schematic representation of the simulation box with an example of a case when interactions are not adjusted correctly leading to penetration is illustrated in Fig. 2.

For predictive modeling, the data was randomly split in 70% for training and 20% for testing. The data was normalized to avoid scaling problems using MinMaxScaler class in sklearn library and log transformations were also applied to the output columns. The model performance was measured by calculating MSE and RMSE and R^2 . The predictive power of several ML models was compared and RF was the one yielding the highest accuracy and performance, which explains our choice for this specific model.

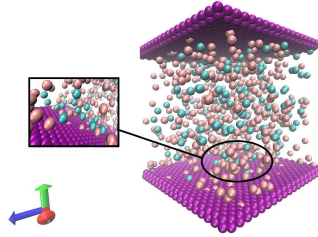


FIG. 2. Illustration of a scenario when the simulation parameters are not adjusted appropriately resulting in wall penetration. In the Figure, blue and pink particles represent the two type of colloidal particles in the system. For a clearer visualization DPD particles are not represented. $f_B=1$, $VF=0.48$ and $rigidity=25,000$.

IV. RESULTS

A. EXPLORATORY AND STATISTICAL ANALYSIS

As previously stated, solvent particles represented as DPD beads, are, in reality, single points in space with a r_c interaction range. Hence, there is no direct particle parameter that could be affecting particle penetration. Yet, the presence of colloidal particles in the system is likely to influence the way solvent particles move across the channel and migrate towards the walls. In this case, although there is no direct relationship, colloidal particles properties may impact solvent penetration in different levels.

Fig. 3 illustrate penetration of colloidal particle at two different body forces, all rigidities and volume fractions evaluated. For simplicity, only colloidal penetration is shown here; a similar behavior for solvent penetration was also observed.

VF.pdf

FIG. 3. Mean colloidal penetration shown for different VF and particle rigidities. The right figure corresponds to a $f_B = 0.1$ and left $f_B = 1$. The data points are averages over all repulsion and confinement ratios for the parameter at interest. That explain the large error bars. For this analysis, only the mean values should be, therefore, considered for comparison.

It can be clearly observed that colloidal penetration is strongly dependent on the VF. Colloidal particles, as opposite to DPD particles have masses and sizes and the force necessary to keep them away from the walls has to be greater than in fluid DPD particles. Particle rigidity also affect penetration, specially evidenced at a stronger volume fraction. Rigid particles are more viscous and the energy necessary to avoid their penetration has to be higher comparative to the soft ones. In terms of body force comparison, it seems that the change in magnitude is not significantly affecting particle penetration.

A second analysis was carried out varying, this time, the repulsion between wall and colloidal particles and the results are displayed in Fig. 4. As expected, an increasing in the repulsion(a_{ij}) has a significant effect on particle penetration. Regardless the particle rigidity of body force employed, when a strong interaction is applied a low number of particles will migrate towards the walls. Stronger repulsion combined with small rigidities leads to an nonexistent particle penetration.

Figures/aij.pdf

FIG. 4. Mean colloidal penetration shown for different a_{ij} at different rigidities. The right figure corresponds to a $f_B = 0.1$ and left $f_B = 1$. As in the previous Figure, the data points are averages over all VF and confinement ratios for the parameter at interest, explaining the large error bars. Only the mean values should be considered for comparison.

In order to complement and extend this analyses to all parameters and levels for the two different cases, a correlation analysis can be performed to gain more insights about the relationship between input parameters and final outcomes (penetration).

Correlations and statistical analysis are extremely useful when one wants to investigate associations and dependencies between variables. Spearman correlation [46] represented in Eq. 20 is a bivariate analysis based on creating ranks in the variables and measuring the strength of the monotonic relationship between pairs of them. In the Spearman correlation, as opposite to the traditional Pearson correlation, the association does not need to be linear neither normally distributed, as it expresses only the direction of the relationship. The correlations coefficients for all the pairs of variables is illustrated as a heatmap in Fig. 5.

correlations.pdf

FIG. 5. Heatmap showing the Spearman correlation coefficients obtained for pairs of variables

$$r_s = 1 - \frac{6 \sum D^2}{n(n^2 - 1)} \quad (20)$$

Where D^2 is the difference between ranks of the variables and n the total number of observations.

It can be observed that while a moderate negative correlation exists between a_{ij} and colloid penetration in the solvent case it is negatively strong. The closer the coefficient values are to +1 or -1, the stronger the correlation. That means that the higher the a_{ij} the weaker the penetration. Volume fraction is moderately correlated with penetration for both solvent and colloids, being stronger in the colloidal case. Particle rigidity, on the other hand, only correlates moderately with colloidal penetration and does not have a very strong effect in solvent penetration. Confinement and body forces do not show a significant correlation with neither of the response variables.

Metric	Solvent		Colloid	
	Train	Test	Train	Test
MSE	0.05	0.18	0.12	0.34
RMSE	0.22	0.43	0.35	0.35
r^2	0.97	0.92	0.95	0.89

TABLE III. Model performance obtained for the train and test sets in each estimator

B. MACHINE LEARNING MODELING

Tuning the parameters to obtain a desired penetration is not a trivial task since several factors may play a role in particle penetration. ML offers a very good alternative to track and predict penetration given a set of initial parameters. It learns the non-linear dependencies between the variables and how they operate together to return the final predictions. Initially the predictions are not accurate but as more data is fed in the model, the performance improves gradually and eventually the model become highly performant. In addition to adding more data, there are alternative techniques to improve a model overall accuracy. One of the most important and popular is hyperparameters tuning, which is based on selecting the best parameters that optimize a given model response. A key parameter in random forest is the number of trees, especially if one considers the computational cost associated in the modeling step. Usually, the higher the number of trees the better the model performance, but when a threshold is reached the addition of more trees has no further influence on the model predictive power.

The results of the training process for both cases with the optimized hyperparameters selection is depicted in Table III. For a better visualization Fig. 6 illustrates the predicted versus true values for the train and test sets. The errors found are considerably low for both cases, and the fit of the data to the model is pretty satisfactory. Some data dispersion is seen in the two cases, being more evident in the colloid penetration model, where the performance was slightly worse. r^2 for the training set was above 0.95 and for the test set 0.89, which indicate that the results are valid although a slightly overfitting is taking place. That is not the case for solvent penetration, where not overfitting is observed as the r^2 as well as the errors between train and test sets were not very high.

Tree based methods calculates internally the feature importance, which expresses the magnitude of the contribution of a variable in predicting the output. The main idea behind the concept is that a variable that affects significantly the error when permuted is assumed to be important. Consequently, features that do not influence the error are seen as not relevant for making predictions. RF assigns a score to each feature based on how big the errors in the predictions are. Fig. 7 shows the ranking of the top features for solvent and colloids returned by the model.

In the solvent case, a_{ij} is the most useful attribute in predicting penetration. This result is expected once solvent particles do not have any parameters other than the repulsive interactions that can affect penetration directly. Colloid volume fraction is the second most relevant variable in predicting solvent

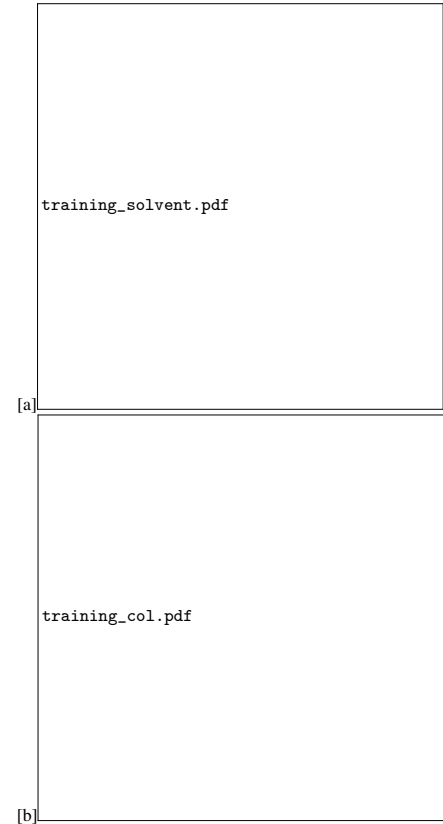


FIG. 6. Predicted outputs versus ground truth. a) Solvent, b) Colloid. Simulation values are plotted against the model predictions. For a perfect fit ($r^2=1$ and $MSE=0$) all points should lie on the line $x=y$. The variance was low for both cases, indicating the quality of the predictions. The units of x and y axis are a result of the predictions on the data that was normalized and the response (penetration) log transformed. penetration predictions.

penetration. By having a rigid core, colloidal particles may push the solvent ones towards the wall leading to an increase in penetration, as previously highlighted. In colloidal penetration, more parameters come to play. The concentration of particles in the system is the most important attribute determining penetration. As more particles are added the repulsive interaction needs to be higher to compensate for the increase in particle number and to prevent their migration inside the wall. VF is followed by a_{ij} and rigidity.

Body force does not seem to be a relevant attribute in predicting penetration, not being the case for confinement. From the correlations analysis previously discussed (Fig. 5), it was seen that there is no monotonic relationship between confine-

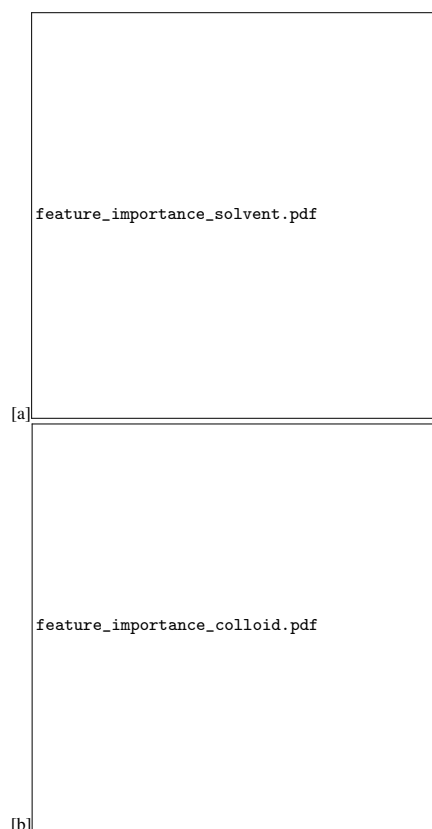


FIG. 7. Feature importance ranked by the Random Forest model. a) Solvent Model, b) Colloid Model.

ment and penetration, and yet, herein it is clear that confinement does play a role in predicting penetration. The effects of confinement in a system depends on many factors, such as particle volume fraction, confinement ratios, particle rigidity, shape, etc.^{49–51}. Particle aggregation might be taking place specially in highly concentrated systems and in some confinement ratios; therefore the agglomerated structures formation will depend on many factors and those might be affecting indirectly the model predictive power. A more in depth study on the confinement effect on the physics of the system is necessary; this will be addressed in a future work.

V. CONCLUDING REMARKS

Setting up the appropriate parameters in a computational system is a crucial step in the development of a successful

simulated system. Particularly, in DPD based models, preventing wall penetration from all the components present in the system is a fundamental step towards reproducing real physical phenomena in confined geometries. The traditional methodologies adopted to prevent DPD penetration are well studied and understood. In the presence of a second component, such as colloidal particles, additional phenomenological effects might take place that can impact the way particles penetrate the walls. Understanding the factors associated and tune them in order to control penetration is a challenge task due to the complex relationships between the parameters. Statistical analysis and ML represents a very powerful approach to deal with those challenges, providing a deeper understanding of the data and the relationships between the input parameters. With ML it is possible to make penetration predictions based on the selection of numerical values of the input parameters. Using RF, we were able to develop successfully a highly performant predictive model displaying low errors and low overfitting. Additionally, it was possible to understand the most relevant features in predicting outputs. This study showed the enormous potentiality of data-driven models and ML tools to systematically study complex systems. In effect, the model developed in this study will be employed to set up properly our wall-solvent and wall-particle interactions in the subsequent works.

VI. ACKNOWLEDGMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 - 1910/2016, Conselho Nacional de Desenvolvimento Científico e Tecnológico CNPq 307100/2017-0 and Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro - FAPERJ E-26/202.835/2017. The authors of this work also thank the National Science Foundation financial support through grant CBET 1703919. This work made use of the High Performance Computing Resource in the Core Facility for Advanced Research Computing at Case Western Reserve University.

¹F. Paiva, A. Boromand, J. Maia, A. Secchi, V. Calado, and S. Khani, "Interfacial aggregation of Janus rods in binary polymer blends and their effect on phase separation," *Journal of Chemical Physics* **151** (2019).

²S. Jamali, M. Yamanoi, and J. Maia, "Bridging the gap between microstructure and macroscopic behavior of monodisperse and bimodal colloidal suspensions," *Soft Matter* **9**, 1506–1515 (2013).

³S. Jamali, A. Boromand, N. Wagner, and J. Maia, "Microstructure and rheology of soft to rigid shear-thickening colloidal suspensions," *Journal of Rheology* **59**, 1377–1395 (2015).

⁴M. Wang, S. Jamali, and J. F. Brady, "A hydrodynamic model for discontinuous shear-thickening in dense suspensions," *Journal of Rheology* **64**, 379–394 (2020).

⁵S. Khani, S. Jamali, A. Boromand, M. J. Hore, and J. Maia, "Polymer-mediated nanorod self-assembly predicted by dissipative particle dynamics simulations," *Soft Matter* **11**, 6881–6892 (2015).

⁶A. Boromand, S. Jamali, and J. M. Maia, "Viscosity measurement techniques in Dissipative Particle Dynamics," *Computer Physics Communications* **196**, 149–160 (2015).

⁷A. Boromand, S. Jamali, and J. M. Maia, "Structural fingerprints of yielding mechanisms in attractive colloidal gels," *Soft Matter* **13**, 458–473 (2017).

- ⁸A. G. Schlijper, P. J. Hoogerbrugge, and C. W. Manke, "Computer simulation of dilute polymer solutions with the dissipative particle dynamics method," *Journal of Rheology* **39**, 567–579 (1995).
- ⁹N. A. Spenley, "Scaling laws for polymers in dissipative particle dynamics," *Europhysics Letters (EPL)* **49**, 534–540 (2000).
- ¹⁰F. Lahmar and B. Rousseau, "Influence of the adjustable parameters of the dpd on the global and local dynamics of a polymer melt," *Polymer* **48**, 3584–3592 (2007).
- ¹¹M. M. Nardai and G. Zifferer, "Simulation of dilute solutions of linear and star-branched polymers by dissipative particle dynamics," *The Journal of Chemical Physics* **131**, 124903 (2009).
- ¹²T. Zhao, X. Wang, L. Jiang, and R. G. Larson, "Dissipative particle dynamics simulation of dilute polymer solutions—Inertial effects and hydrodynamic interactions," *Journal of Rheology* **58**, 1039–1058 (2014).
- ¹³F. L. Paiva, M. J. A. Hore, A. Secchi, V. Calado, J. Maia, and S. Khani, "Dynamic interfacial trapping of janus nanorod aggregates," *Langmuir* **36**, 4184–4193 (2020).
- ¹⁴X. Yong, O. Kuksenok, K. Matyjaszewski, and A. C. Balazs, "Harnessing interfacially-active nanorods to regenerate severed polymer gels," *Nano Letters* **13**, 6269–6274 (2013).
- ¹⁵S. Biswas, A. Singh, A. Beziau, T. Kowalewski, K. Matyjaszewski, and A. C. Balazs, "Modeling the formation of layered, amphiphilic gels," *Polymer* **111**, 214–221 (2017).
- ¹⁶A. Boromand, S. Jamali, B. Grove, and J. M. Maia, "A generalized frictional and hydrodynamic model of the dynamics and structure of dense colloidal suspensions," *Journal of Rheology* **62**, 905–918 (2018).
- ¹⁷I. V. Pivkin and G. E. Karniadakis, "Controlling density fluctuations in wall-bounded dissipative particle dynamics systems," *Physical Review Letters* **96**, 1–4 (2006).
- ¹⁸A. Mehboudi and M. S. Saidi, "Physically based wall boundary condition for dissipative particle dynamics," *Microfluidics and Nanofluidics* **17**, 181–198 (2014).
- ¹⁹I. V. Pivkin and G. E. Karniadakis, "A new method to impose no-slip boundary conditions in dissipative particle dynamics," *Journal of Computational Physics* **207**, 114–128 (2005).
- ²⁰D. A. Fedosov, I. V. Pivkin, and G. E. Karniadakis, "Velocity limit in DPD simulations of wall-bounded flows," *Journal of Computational Physics* **227**, 2540–2559 (2008).
- ²¹Z. Li, X. Bian, Y. H. Tang, and G. E. Karniadakis, "A dissipative particle dynamics method for arbitrarily complex geometries," *Journal of Computational Physics* **355**, 534–547 (2018).
- ²²I. Barcelos, S. Khani, A. Boromand, L. F. Vieira, J. A. Lee, J. Peet, M. F. Naccache, and J. Maia, "Controlling particle penetration and depletion at the wall using Dissipative Particle Dynamics," *Computer Physics Communications* **258**, 513–526 (2021), arXiv:1711.03136.
- ²³D. Duong-Hong, N. Phan-Thien, and X. Fan, "An implementation of no-slip boundary conditions in DPD," *Computational Mechanics* **35**, 24–29 (2004).
- ²⁴D. Zhang, Q. Shangguan, and Y. Wang, "An easy-to-use boundary condition in dissipative particle dynamics system," *Computers and Fluids* **166**, 117–122 (2018).
- ²⁵S. K. Ranjith, B. Patnaik, and S. Vedantam, "No-slip boundary condition in finite-size dissipative particle dynamics," *Journal of Computational Physics* **232**, 174–188 (2013).
- ²⁶D. Visser, H. Hoefslot, and P. Jedema, "Comprehensive boundary method for solid walls in dissipative particle dynamics," *Journal of Computational Physics* **205**, 626–639 (2005).
- ²⁷Y. Kong, C. W. Manke, W. G. Madden, and A. G. Schlijper, "Simulation of a confined polymer in solution using the dissipative particle dynamics method," *International Journal of Thermophysics* **15**, 1093–1101 (1994).
- ²⁸M. Whittle and K. P. Travis, "Dynamic simulations of colloids by core-modified dissipative particle dynamics," *Journal of Chemical Physics* **132** (2010).
- ²⁹P. Malfreyt and D. J. Tildesley, "Dissipative particle dynamics simulations of grafted polymer chains between two walls," *Langmuir* **16**, 4732–4740 (2000).
- ³⁰X. Fan, N. N. Phan-Thien, N. T. Yong, X. Wu, and D. Xu, "Microchannel flow of a macromolecular suspension," *Physics of Fluids* **15**, 11–21 (2003).
- ³¹X. Fan, N. Phan-Thien, S. Chen, X. Wu, and T. Y. Ng, "Simulating flow of DNA suspension using dissipative particle dynamics," *Physics of Fluids* **18** (2006).
- ³²J. Wei, X. Chu, X.-Y. Sun, K. Xu, H.-X. Deng, J. Chen, Z. Wei, and M. Lei, "Machine learning in materials science," *InfoMat* **1**, 338–358 (2019), <https://onlinelibrary.wiley.com/doi/pdf/10.1002/inf2.12028>.
- ³³S. M. Moosavi, K. M. Jablonka, and B. Smit, "The Role of Machine Learning in the Understanding and Design of Materials," (2020), 10.1021/jacs.0c09105.
- ³⁴G. Pilania, C. Wang, X. Jiang, S. Rajasekaran, and R. Ramprasad, "Accelerating materials property predictions using machine learning," (2013).
- ³⁵K. T. Butler and W. Daniel, "Review Machine learning for molecular and materials science," *Nature* (2018), 10.1038/s41586-018-0337-2.
- ³⁶T. Inokuchi, N. Li, K. Morohoshi, and N. Arai, "Multiscale prediction of functional self-assembled materials using machine learning: High-performance surfactant molecules," *Nanoscale* **10**, 16013–16021 (2018).
- ³⁷S. Chen and X. Yong, "Janus Nanoparticles Enable Entropy-Driven Mixing of Bicomponent Hydrogels," *Langmuir* **35**, 14840–14848 (2019).
- ³⁸T. Inokuchi, R. Okamoto, and N. Arai, "Predicting molecular ordering in a binary liquid crystal using machine learning," *Liquid Crystals* **47**, 438–448 (2020).
- ³⁹L. Zhao, Z. Li, Z. Wang, B. Caswell, J. Ouyang, and G. E. Karniadakis, "Active- and transfer-learning applied to microscale-macroscale coupling to simulate viscoelastic flows," *Journal of Computational Physics* **427**, 110069 (2021), arXiv:2005.04382.
- ⁴⁰Y. Wang, J. Ouyang, and X. Wang, "Machine learning of lubrication correction based on gpr for the coupled dpd-dem simulation of colloidal suspensions," *Soft Matter*, – (2021).
- ⁴¹P. J. Hoogerbrugge and J. M. Koelman, "Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics," *Epl* **19**, 155–160 (1992).
- ⁴²R. D. Groot and P. B. Warren, "Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation," *The Journal of Chemical Physics* **107**, 4423–4435 (1997).
- ⁴³P. Espanol and P. Warren, "Statistical mechanics of dissipative particle dynamics," *Epl* **30**, 191–196 (1995).
- ⁴⁴J. Mewis and N. J. Wagner, *Colloidal Suspension Rheology*, Cambridge Series in Chemical Engineering (Cambridge University Press, 2011).
- ⁴⁵R. Ball and J. Melrose, "Lubrication breakdown in hydrodynamic simulations of concentrated colloids," *Advances in Colloid and Interface Science* **59**, 19–30 (1995).
- ⁴⁶A. Ghoufi and P. Malfreyt, "Coarse grained simulations of the electrolytes at the water–air interface from many body dissipative particle dynamics," *Journal of Chemical Theory and Computation* **8**, 787–791 (2012).
- ⁴⁷L. Breiman, "Random forests," *Machine Learning* **45**, 5–32 (2001).
- ⁴⁸L. E. O. Breiman, "Bagging Predictors," **140**, 123–140 (1996).
- ⁴⁹I. Cohen, T. G. Mason, and D. A. Weitz, "Shear-induced configurations of confined colloidal suspensions," *Physical Review Letters* **93**, 046001–1 (2004).
- ⁵⁰S. Gallier, E. Lemaire, L. Lobry, and F. Peters, "Effect of confinement in wall-bounded non-colloidal suspensions," *Journal of Fluid Mechanics* **799**, 100–127 (2016).
- ⁵¹M. Ramaswamy, N. Y. Lin, B. D. Leahy, C. Ness, A. M. Fiore, J. W. Swan, and I. Cohen, "How confinement-induced structures alter the contribution of hydrodynamic and short-ranged repulsion forces to the viscosity of colloidal suspensions," *Physical Review X* **7**, 1–14 (2017).
- ⁵²M. Revenga, I. Zúñiga, P. Español, and I. Pagonabarraga, "Boundary Models in DPD," *International Journal of Modern Physics C* **09**, 1319–1328 (1998).
- ⁵³J. M. Koelman and P. J. Hoogerbrugge, "Dynamic simulations of hard-sphere suspensions under steady shear," *Epl* **21**, 363–368 (1993).
- ⁵⁴E. S. Boek, P. V. Coveney, and H. N. Lekkerkerker, "Computer simulation of rheological phenomena in dense colloidal suspensions with dissipative particle dynamics," *Journal of Physics Condensed Matter* **8**, 9509–9512 (1996).
- ⁵⁵E. S. Boek, P. V. Coveney, H. Cross, and M. Road, "Simulating the rheology of dense colloidal suspensions using dissipative particle dynamics," **55**, 3124–3133 (1996).
- ⁵⁶Ö. E. Aw, I. A. Aey, and Ü. Y. A. By, ATT Bell Laboratoires.
- ⁵⁷T. Mitchell, *Machine Learning*, McGraw-Hill Series in Computer Science (1997).

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

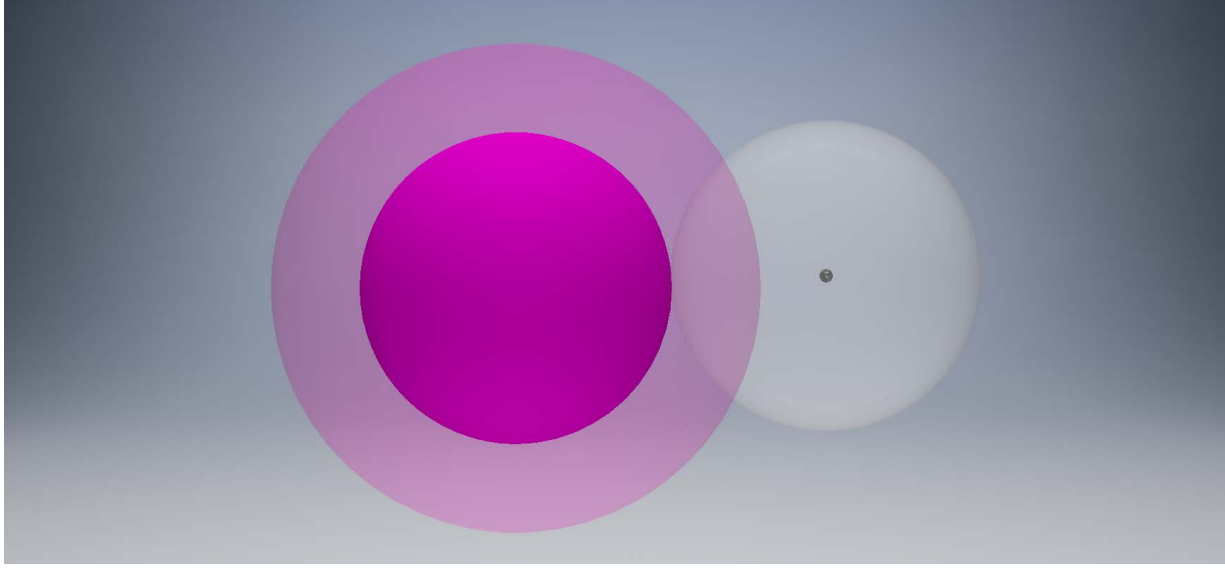
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0086759

Accepted to Phys. Fluids 10.1063/5.0086759

⁵⁸C. Wissler, "The spearman correlation formula," Science **22**, 309–311 (1905).

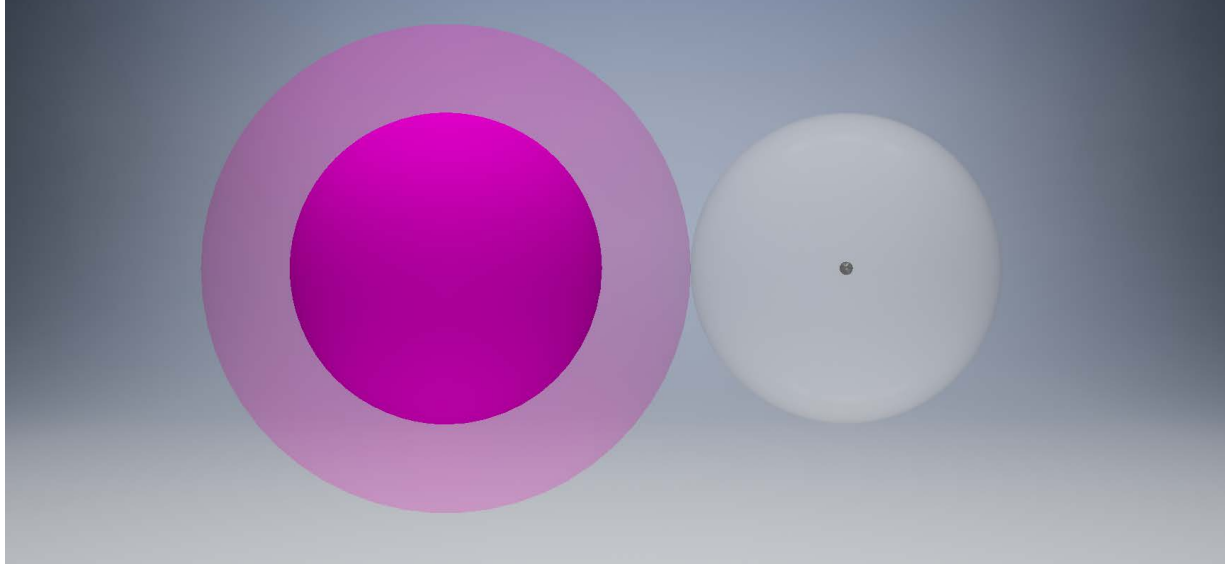
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0086759



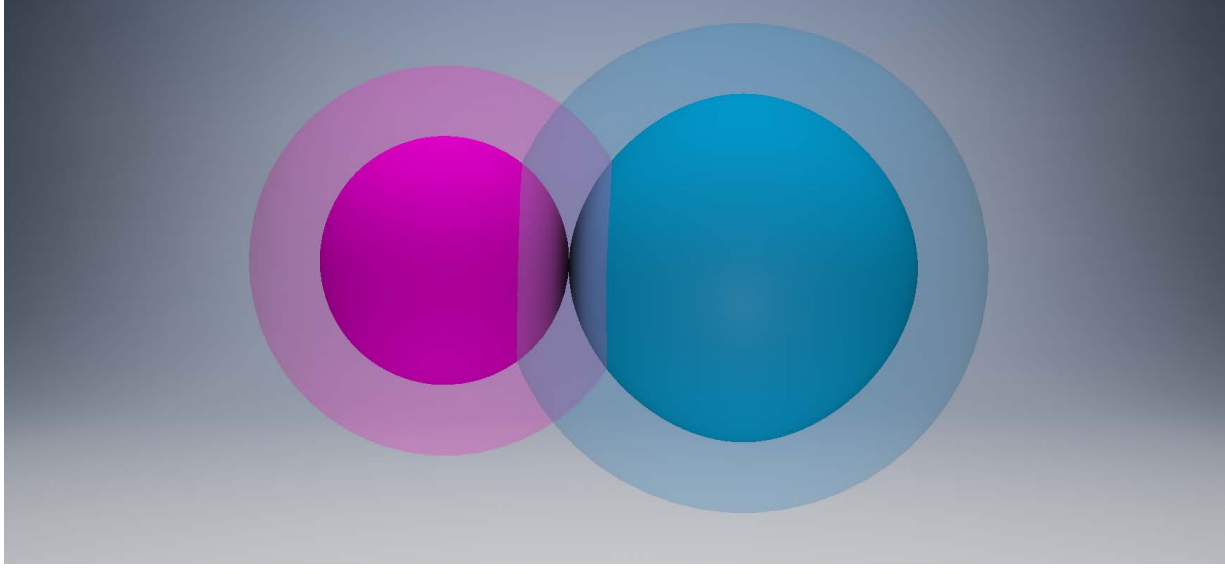
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0086759



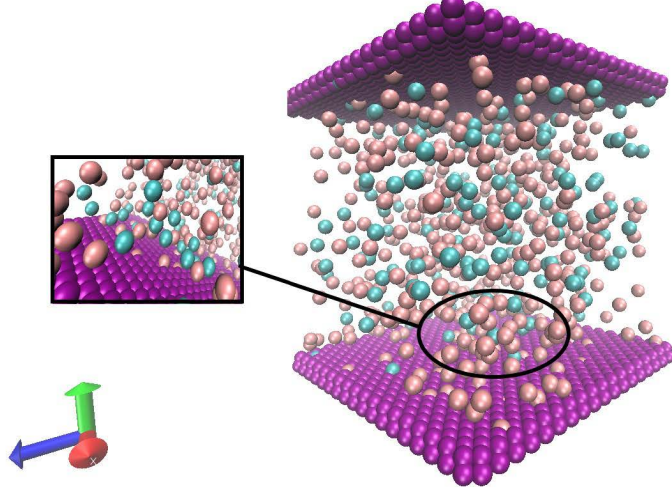
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0086759



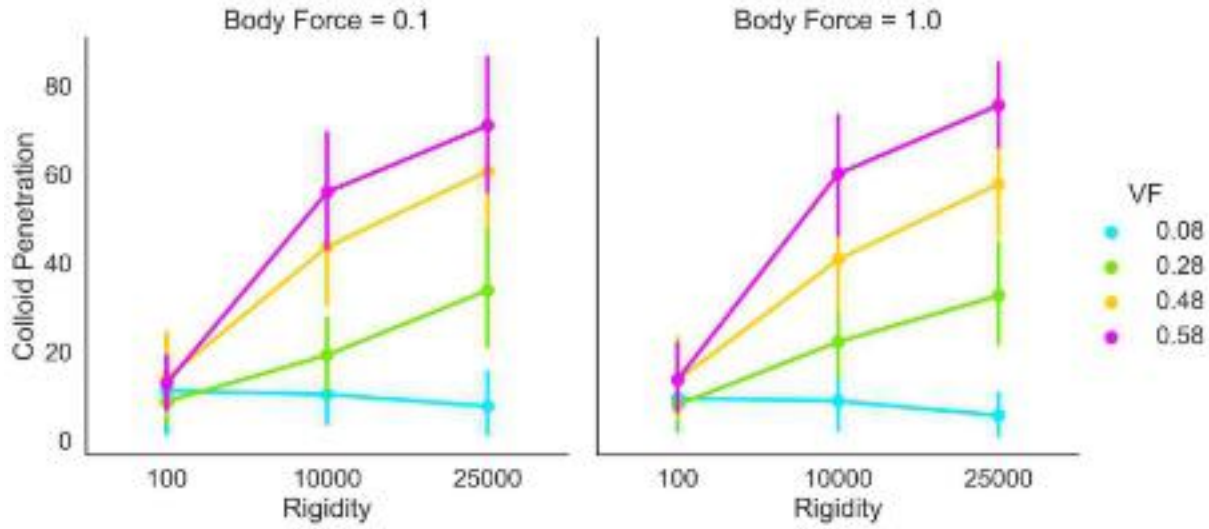
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0086759



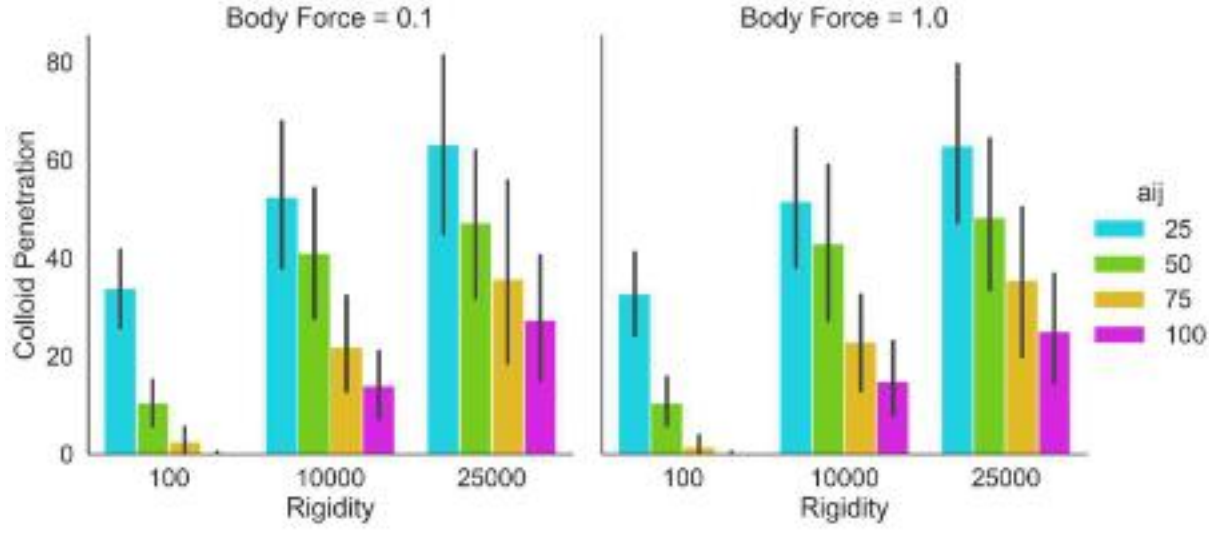
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0086759



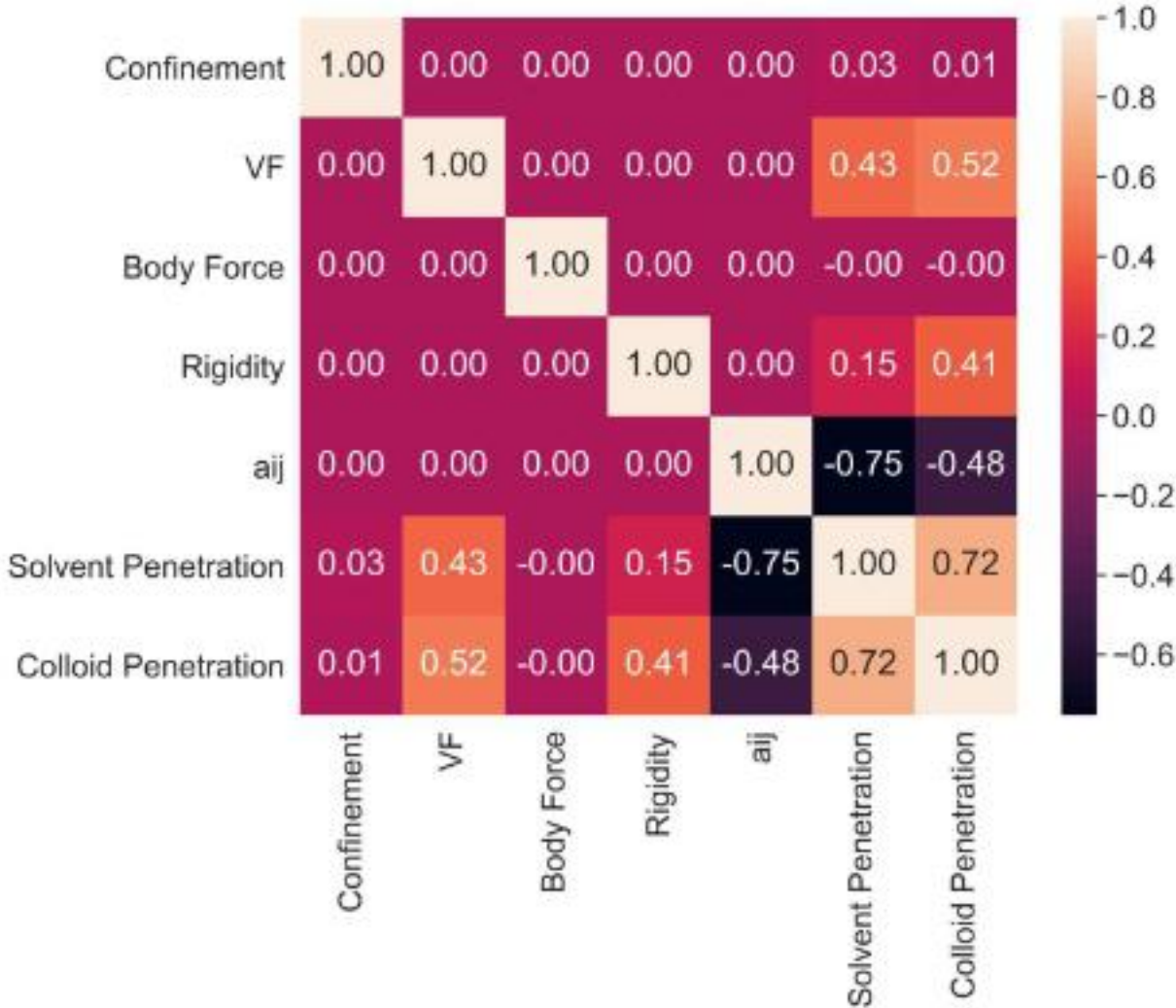
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0086759



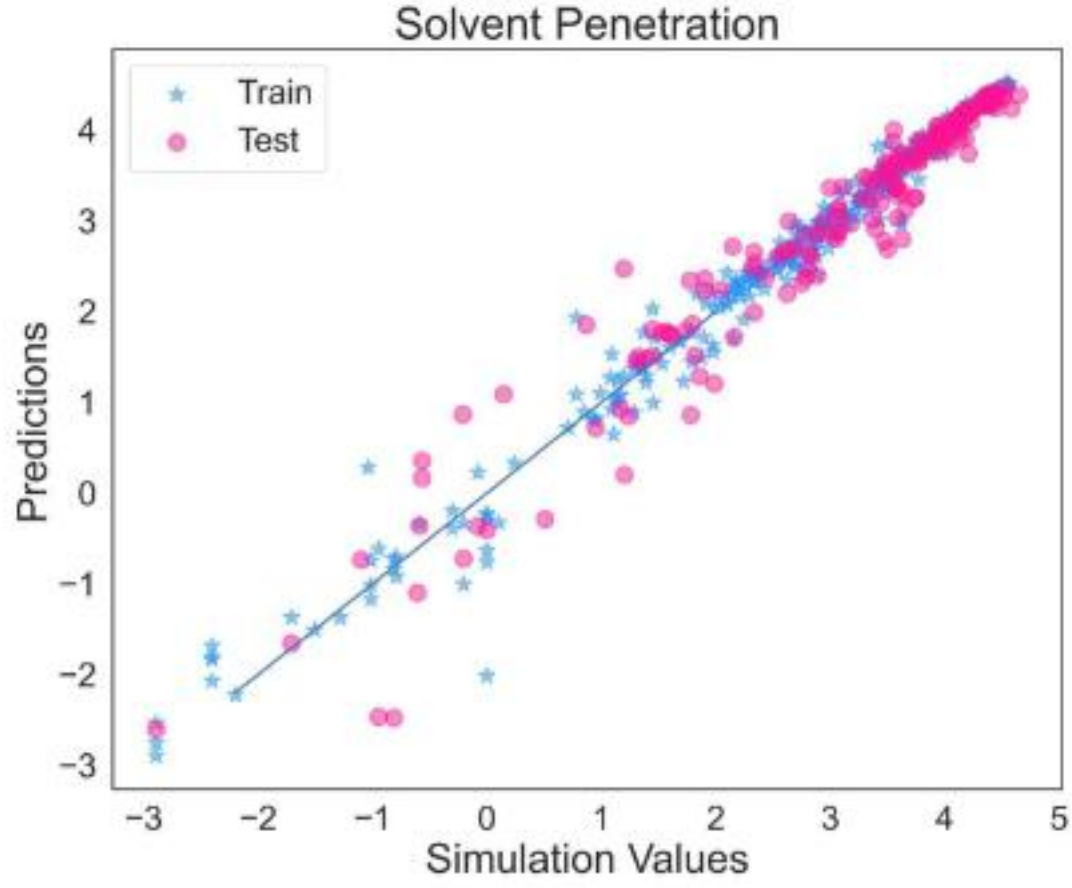
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0086759



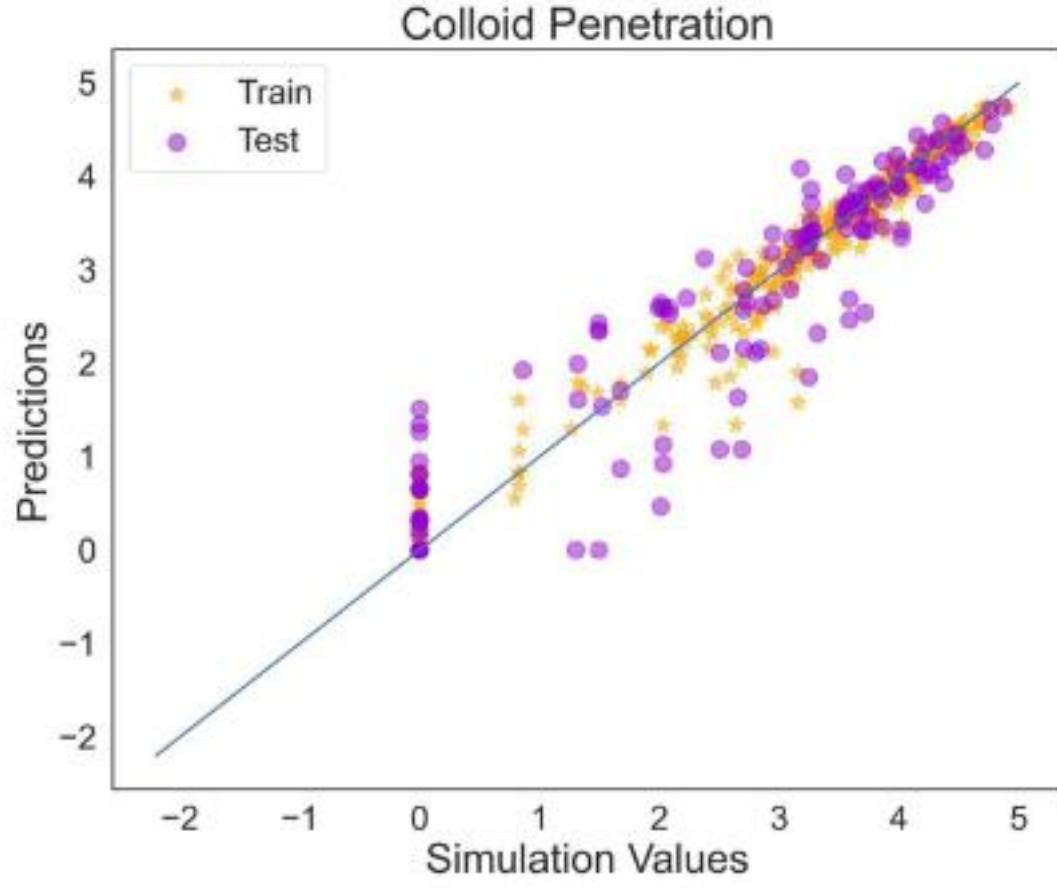
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0086759



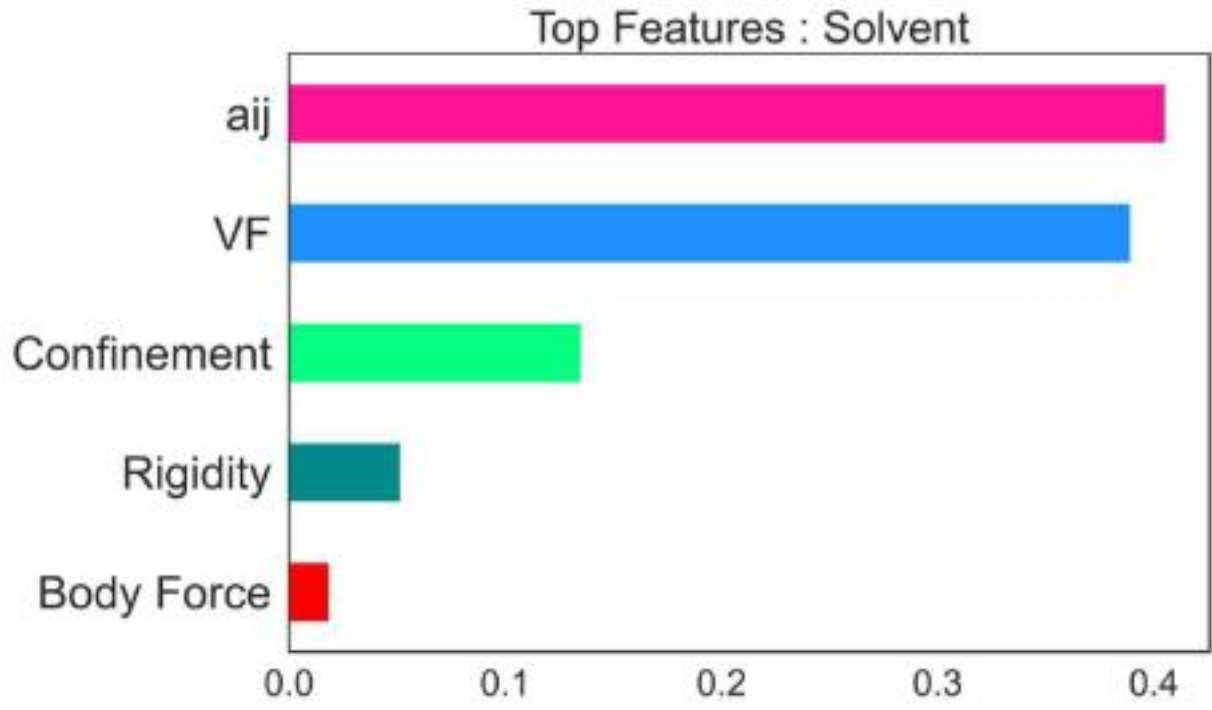
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0086759



This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0086759



This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0086759

