Minimal discrete matches for target grain size distributions

Lorne Arnold

Abstract

Soil is fundamentally a discrete material that is, nevertheless, commonly modeled as a continuum in part because of the computational expense of large-scale discrete element models (DEMs). Even at lab specimen scales, DEM's computational cost may be substantial depending on the grain sizes being modeled. Despite these limitations, discrete models have proven useful in furthering our understanding of soil mechanics because they can spontaneously replicate realistic soil behavior as an emergent macro-scale property from a collection of particles following relatively simple interaction rules. This makes DEM an attractive tool for a multi-scale modeling approach where the constitutive behavior of a representative volume element (RVE) is characterized with a discrete model and applied at a larger scale through a continuum model. The ability of the RVE to represent a soil depends strongly on an appropriate grain size distribution match. However, in order to achieve computationally feasible models, even at lab scales, DEM simulations often use larger minimum particle sizes and more uniform distributions than their intended targets. Intuitively, discrete matches of different grain size distributions (GSDs) will require vastly different numbers of particles. But the precise relationship between GSD characteristics and the number of particles needed to match the distribution (and by extension the associated computational cost associated) is not intuitive. In this paper, we present the minimal discrete match (MDM) concept. The minimal discrete match is the smallest set of discrete particles needed to match a given GSD. We present a method for determining the MDM for any GSD and discuss strategies for finding the smallest MDM within a set of tolerances on the GSD. A mapping of USCS classification and MDM reveals a broad distribution of computational cost over several orders of magnitude for granular soils.

Keywords

Discrete soil description, analytical solutions, open-source algorithms, computational cost

Introduction

Soil is a fundamentally discrete granular material whose complex mechanical behavior emerges from the numerous interactions between its constituent parts. The discrete element method (DEM), introduced by Cundall and Strack (1979), has proven to be a powerful tool in exploring the inter-particle interactions of granular assemblies. With DEM, a vast parametric space exists where grain sizes, shapes, contact models, etc. can be systematically varied and their influences on macro-scale assembly behavior can be quantified.

These models pose several challenges that must be managed in order for their benefits to be realized. Due to the number of elements involved in typical DEM simulations, both the computational resources needed to run them and the ability to characterize and interpret them are non-trivial even at lab specimen scales. Over the years, the number of particles used in DEM simulations has increased, but not proportionally to the reduction in costs of computational resources. O'Sullivan (2014) showed that the order of magnitude of average number of particles in DEM simulations increased substantially from 1998 to 2014, from approx. $\mathcal{O}(10^3)$ to $\mathcal{O}(10^5)$, but fell far short of the $\mathcal{O}(10^7)$ predicted by Cundall (2001) for "easy" geomechanics problems by 2011. More recent publications in DEM have shown that orders ranging from $\mathcal{O}(10^6)$ (Sufian et al. (2021), Dong, Yan, and Cui (2022)) to $\mathcal{O}(10^8)$ (Miyai et al. (2019), Fang et al. (2021), Zhang et al. (2024)) are possible, but rare and require high performance computing resources.

Because of the computational cost, DEM simulations whose particles approach realistic grain sizes generally model at the representative volume (RV) element scale. Several nuances to the RV concept exist, but broadly speaking, the RV represents the smallest volume of material to exhibit statistically consistent macro-scale behavior as its source material (Gitman, Askes, and Sluys 2007). In physical laboratory experiments and discrete numerical experiments, ensuring study samples are sufficiently large to behave as an RV is critical to the broader applicability of the results.

The RV size depends on several factors including the grain size distribution (GSD), sample density, loading conditions, grain shape and arrangement, and the behavior of interest. Studies have tied RV to sample height to maximum particle diameter ratios with values between 15 and 20 needed to achieve RV (Cantor and Ovalle 2025). Additionally, there is evidence that the average particle size (Zeraati-Shamsabadi and Sadrekarimi 2025) and overall polydispersity (Bandera et al. 2021) also influence RV size. Intuitively, with increasing GSD breadth, the number of particles needed in a DEM simulation of granular assemblies will also increase. Often, matching the true GSD of interest (e.g., from a physical soil sample) is sacrificed for computational efficiency by truncating or upscaling the distribution.

While the insights we gain from DEM simulations with upscaled or truncated GSDs are valuable, they come with acknowledged limitations on their broader applicability to natural soil fabrics. It becomes an important experimental design consideration, then, to understand the relative computational costs associated with different grain size distributions. For very specific GSDs,

formulas can be used to calculate the number of particles needed. For example, Tyler and Wheatcraft (1992) provide such a formula for self-similar fractal distributions. However, they note that the formula results are sensitive to the choice of representative radius for a given range. Further, they concluded that self-similar fractal distribution was a good model for pore structure, but not necessarily grain size. For general GSDs, particularly when a specific physical soil is being modeled, an ideal approach would be an analytical solution to the minimum number of particles required to reproduce the GSD.

This paper introduces such an approach through the minimal discrete match (MDM) concept. The minimal packing set is the smallest set of discrete particles needed to match a given grain size distribution by mass. Note that the MDM is based on mass-volume relationships only, not mechanical behavior. Therefore, it represents an important component to determining RVE and relative computational cost, but it is not equivalent to an RVE.

In order to characterize the MDM, descriptions of grain size distributions and granular samples are presented as mathematical sets and the rules describing several relationships between the sets are defined. Using these definitions, algorithms are presented for both an approximate and a rigorous solution for the MDM. The results show a broad range of MDM magnitudes over several GSDs of interest in geotechnical engineering. The MDM algorithms are implemented in an open-source Python module, available on GitHub.

Discrete definitions

Identifying the minimal discrete match requires rigorous mathematical definitions for particulate samples, S, and grain size distributions, G. Conceptually, each of these are collections of items (i.e., sets) with specific restrictions. The definitions are designed to capture the physical reality underlying a mechanical grain size analysis.

Sample definition

A sample, S is an indexed set describing its particles as pairs of size, X_S, and quantity, Q_S. X_S is an ordered (i.e., each entry is larger than the previous) set of unique values of positive real numbers. Q_s is an unordered set of positive integers related to X_S by a shared indexing set for the sample, I_S.

$$S = \{(X_S, Q_S) \in I_S\} \tag{1}$$

where:

$$X_S = \{x_i : i \in I_S\} \text{ with } x_1 < x_2 < \cdots < x_{n_S} \tag{2}$$

$$Q_S = \{q_i : i \in I_S\} \tag{3}$$

$$X_S \subset \mathbb{R}^+ \text{ and } Q_S \subset \mathbb{Z}^+ \tag{4}$$

Grain size distribution definition

A grain size distribution, G, has a similar structure. It is and indexed set describing size boundaries, X_G, and masses, M_G. Like X_S, X_G is an ordered set of unique values, however, it's domain also includes zero (representing the pan in a typical sieve analysis). Like Q_S, M_G shares an index (I_G) with the size set, but it's domain is not restricted to integers, only to non-negative real numbers. The values in M_G represent the masses retained on a sieve with opening size X_G.

$$G = \{(X_G, M_G) \in I_G\} \tag{5}$$

where:

$$X_G = \{x_j : j \in I_G\} \text{ where } x_1 < x_2 < \cdots < x_{n_G} \tag{6}$$

$$M_G = \{m_j : j \in I_G\} \tag{7}$$

$$X_G \subset \mathbb{R}^{non-neg} \; ; \; M_G \subset \mathbb{R}^{non-neg} \tag{8}$$

Comparison definition

The relationship between S and G cad be defined in terms of how G *describes* S. In the context of DEM modeling, this may seem backward because typically in DEM modeling, a target GSD is defined first and a sample generated to match it. On the other hand, a discrete sample (either physical or numerical) exists on its own, whereas a GSD only has meaning when interpreted as a description of a sample. Therefore the relationship between S and G will be defined in terms of G's description of S. In a later section, the topic of finding an instance of S that *matches* a target G will be addressed.

The relationship between S and G is formalized with the following conditions:

Condition 1: G is complete if and only if the final entry in M_G is zero:

$$G \text{ is complete } \Leftrightarrow m_{n_G} = 0 \tag{9}$$

This condition ensures that x_{nG} provides an upper bound to the sizes described by G. This is analogous to the limitation of scope in ASTM D6913 to grain sizes passing a 75-mm sieve, ensuring the cumulative GSD always reaches 100 percent at a known size (D18 Committee, n.d.).

Condition 2: G describes S (denoted $G \longrightarrow S$) if and only if Cond1 is met and the its smallest size is smaller than any size in S and its largest size is larger than any size in S:

$$G \longrightarrow S \Leftrightarrow \text{ Cond1} \wedge \min(X_G) < \min(X_S) \wedge \max(X_G) \geq \max(X_S) \tag{10}$$

Condition 3: G describes S articulately if Cond2 is met and for all consecutive pairs of sizes in X_S there exists at least one size in X_G between them.

$$G \overset{\text{articulately}}{\longrightarrow} S \Leftrightarrow \text{Cond2} \wedge \forall (x_i, x_{i+1}) \in X_S \quad \exists\, x_j \in X_G \text{ such that } x_i < x_j \leq x_{i+1} \tag{11}$$
(11)

Cond3 is rarely (if ever) met in physical scenarios because it requires that only a single size be present between any two adjacent sieves. The final MDM solution will not be constrained by Cond3, but it is helpful for formulating the solution.

Condition 4: G describes S accurately if Cond2 is met and the combined masses of all the particles with sizes between every pair of sizes in X_G is equal to the retained mass on the lower of the size pairs in G:

$$G \overset{\text{accurately}}{\longrightarrow} S \Leftrightarrow \text{ Cond2} \wedge \sum_{\substack{i \in I_S \\ x_j < x_i \leq x_{j+1}}} q_i \cdot f(x_i) = m_j \tag{12}$$

where $f(x)$ is a scaling function that converts size to mass.

Cond4 represents what is generally meant by a "match" between a sample and a grain size distribution. Note that Cond3 is not required for Cond4 to be satisfied.

Error: The error for a match between G and S is simply a measure of how far Cond4 is from being met:

$$\text{Error}(G, S) = m_j \quad - \sum_{\substack{i \in I_S \\ x_j < x_i \leq x_{j+1}}} q_i \cdot f(x_i) \tag{13}$$

Physical interpretation

Some assumptions are required in order for S and G to be interpreted as a physical sample and sieve analysis (or a numerical version of the same). First, the sizes in X_G are assumed to represent smallest enclosing diameter of whatever shapes the particles in S are. Second, as indicated in Equation 12, a size in X_S is considered between two sizes in X_G if it is larger than the x_j and smaller than or equal to x_{j+1}. In other words, particles pass through openings equal to their own size. Finally, the particles in S are assumed to be of similar enough shape that a single scaling function, $f(x)$, can represent the size-mass relationship for all of S. For spherical particles with of constant density (ρ), this is a straightforward definition:

$$f(x) = \frac{\rho\pi}{6}x^3 \tag{14}$$

Note, however, that the use of this particular scaling function is not required in general. It is only necessary that $f(x)$ be some mapping of $x \mapsto m$. It would be possible to combine sets of differently shaped particles of different densities with respective mapping functions, but this paper will limit itself to assuming uniform, spherical particles.

Grain sizes evaluated

This paper will analyze several grain size distributions in the range of interest in geotechnical engineering (although the methods are applicable over a wider range). Specifically, 2750 randomized GSDs based on normal distributions across the sieve set specified in ASTM D6913 (14 opening sizes from 0.075-mm to 75-mm) are analyzed. The distributions were generated by adding random noise to normal mass distributions around a "central" size. The central size itself was randomly selected from the midele half of the sieves used. Between 5 and 14 consecutive sieves were used in each GSD. For classification purposes, articles smaller than 0.075 are assumed to be low plasticity silt, regardless of their size.

Table 1 summarizes the GSDs used in this study using several traditional characteristics including coefficient of curvature, C_c, and coefficient of uniformity, C_u. The grain size index, I_{GS}, as defined by Erguler (2016) and a new parameter, the curvature index, I_C, are also presented. The curvature index is conceptually similar to the grain size index in that it is a ratio of areas related to the cumulative distribution function (in percent passing, log size space) of the GSD. I_C normalizes the area under the GSD curve to the trapezoidal area bounded by the percent passing the smallest size above the pan (x_{min+1}) and the largest size (x_{max}). Figure 1 shows a small subset (10 percent) of the GSDs considered and illustrates the definition of I_C on an example GSD.

Some traditional grain size distribution characteristics

Table 1: Grain size distribution parameters in this study

Parameter	Value(s) in study
Instances of G	2750
Sieves in G	5 to 14
x_{min+1}	0.075 to 25 mm
x_{max}	9.5 to 75 mm
Percent fines	0.4% to 34%
C_c	0.2 to 25.5
C_u	1.2 to 214.5
I_{GS}	0.03 to 0.60
I_C	0.29 to 1.57

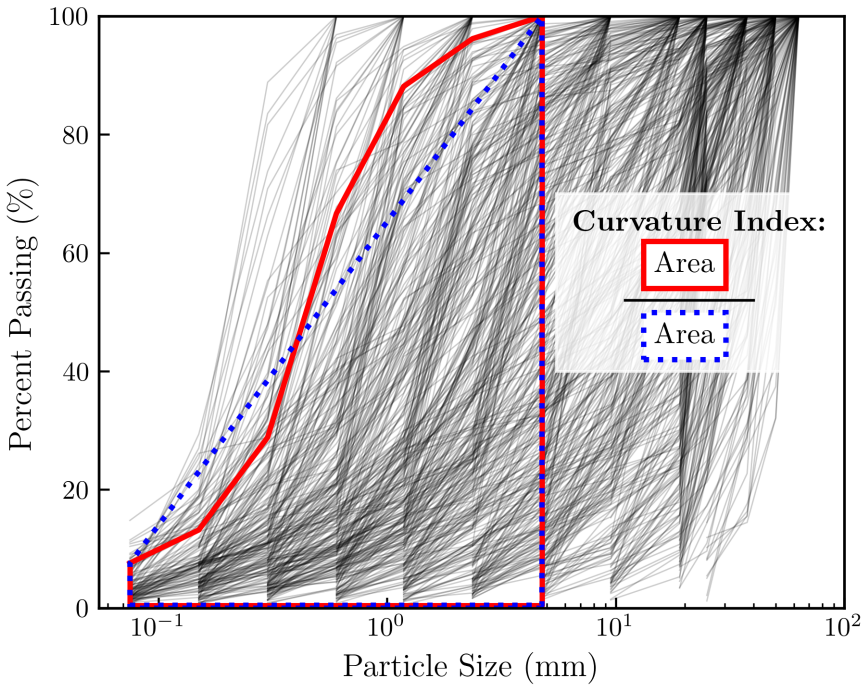

Figure 1: A representative subset (~10%) of the GSDs evaluated. The area ratio defining the curvature index is shown. The curve in the highlighted example has a curvature index of 1.09.

Minimal discrete match solution

For any given sample, creating a grain size distributionthat accurately describes it is no more complicated than following the numerical equivalent to a sieve analysis. The reverse, however, finding a solution for S that is accurately described by some target G, is non-trivial. One of the reasons is that unlike a grain size analysis, the sizes of the solution (X_S) are bounded, but not explicitly defined. Additionally, the goal for this procedure is not only to find any S that is accurately described by G, but the minimal packing set S_{min} that is accurately described by G.

As a starting point, assuming that G is an articulate description of S limits the number of sizes in S to $n_G - 1$. Temporarily setting aside the minimization goal, any valid sizes may be selected for X_S. From G, an indexed set of mass ratios, Φ, can be defined to describe the masses of each size range relative to the mass retained on the second to largest sieve (recall that no mass is retained on the largest sieve):

$$\Phi = \left\{ \frac{m_j}{m_{n_G-1}} : j \in I_G - 1 \right\}; \text{ with elements } \phi_j \tag{15}$$

The length of Φ is one less than I_G, making it equal in length to I_S because of Cond3.

Similarly, a set of volume ratios, Z can be defined to describe the volume per particle relative to the largest particle:

$$Z = \left\{ \frac{f(x_{nS})}{f(x_i)} : i \in I_S \right\}; \text{ with elements } \zeta_i \tag{16}$$

Note that Z is referred to as a volume ratio to avoid confusion with Φ despite the fact that the mapping function $f(x)$ converts size to mass. Interpreting Z as a volume ratio is valid under a constant density assumption.

Because they are ratios relative to the largest considered size, $\phi_{nS} = \zeta_{nS} = 1$. Since Φ describes relative masses of each size and Z describes the relative volume (and mass) per particle of each size, their product will describe a relative quantity ratio, K:

$$K = \{\phi_i \times \zeta_i : i \in I_S\}; \text{ with elements } \kappa_i \tag{17}$$

The quantity ratio K is the relative number of particles of each of the assumed sizes in X_S needed to satisfy Cond4. It is directly proportional to the key target parameter, Q_S. Being the product of ϕ_{nS} and ζ_{nS}, κ_{nS} will be an integer equal to 1, making it the smallest allowable candidate for q_{nS}. Unfortunately, with the exception of κ_{nS}, the entries in K are not guaranteed to be integers. Rounding each element in K to the nearest integer provides a first-order, approximate solution for the minimal discrete match.

$$S_{approx} = \{(X, \text{int}(K)) \in I_S\} \approx S_{min} \tag{18}$$

Depending on the application and acceptable error tolerance, S_{approx} may be a suitable substitute for S_{min}. The error for the approximate solution that assumed fixed sizes can be reduced using the following fixed size (FS) algorithm:

1. Find K using Equation 17.
2. Use int(K) to approximate $Q_{S(0)}$ and build $S_{(0)} = \{X_S, Q_{S(0)}\}$.
3. Calculate the error, E, between G and the current approximation for S with Equation 13.
4. If E exceeds the desired tolerance, iteratively approximate larger estimates for Q as int($i \times K$) with i being integer values (starting with $i = 2$).
5. Repeat Steps 3 and 4 until the desired tolerance is reached.

This approach can reduce the match error to an arbitrarily low value, but the total number of particles grows by a factor of $|Q_{S(0)}|$ with each iteration.

But since the sizes in X_S are only bounded by X_G and not explicitly defined, the opportunity exists to find a better selection of entries in X_S that will minimize Q_S. The best allowable sizes for X_S can be found using the following spanned integer (SI) algorithm:

1) Find the quantity ratios associated with the minimum ($-$) and maximum ($+$) allowable sizes in x_1 through x_{nS-1} (i.e. K_- and K_+).
2) Check whether an integer is spanned between each entry in K_- and K_+. An integer is spanned between two consecutive numbers a and b if $\lceil a \rceil \leq \lfloor b \rfloor$ (i.e., a is rounded to the next highest integer and b is rounded to the next lowest integer).
3) If a spanned integer (SI) does not exist for each entry, repeat Steps 1 and 2 with incrementally larger sizes for x_{nS}.
4) If, after reaching the maximum allowable particle size for x_{nS}, an SI does not exist between each entry in K_- and K_+, perform Step 4 from the FS algorithm and repeat Steps 1 to 3 until the span between each entry contains an integer.
5) When an integer quantity exists between each entry in K_- and K_+, these integers can be used to populate the quantity set Q_{SI}.
6) Invert the mass scaling function $f(x)$ on the ratio Φ / Q_{SI} to identify the compatible sizes for the spanned integer quantities:

$$X_{SI} = f^{-1}\left(\frac{\Phi}{Q_{SI}}\right) \tag{19}$$

Because Q_{SI} was generated using the allowable sizes of X_S and the mass ratios of the target G, the resulting X_{SI} satisfies Cond4 with Q_{SI} as the smallest possible quantity of particles, or the minimal discrete match:

$$MDM = S_{min} = \{X_{SI}, Q_{SI}\} \tag{20}$$

The total number of particles in the MDM is the sum of the quantities in Q_{SI}:

$$N_{MDM} = \sum_{i \in I_S} q_{SI,i} \tag{21}$$

The effectiveness of the spanned integer algorithm is shown in the rapid convergence to numeric zero error compared to a fixed-size approximation in Figure 2. The figure shows both approaches attempting to find a suitable sample for each of the GSDs shown in Figure 1 with an allowable total error tolerance of 1 percent. In nearly all cases (over 98% for the GSDs in this study), the spanned integer approach requires no quantity-increasing iteration (i.e., Step 4 is skipped). Where iteration is needed, it converges to numerical zero within one or two iterations. The fixed size approach converges with between 1 and 15 iterations and retains measurable error. This difference is significant, not because the spanned integer algorithm itself is computationally expensive, but because each iteration represents an increase in total particles in the final discrete sample.

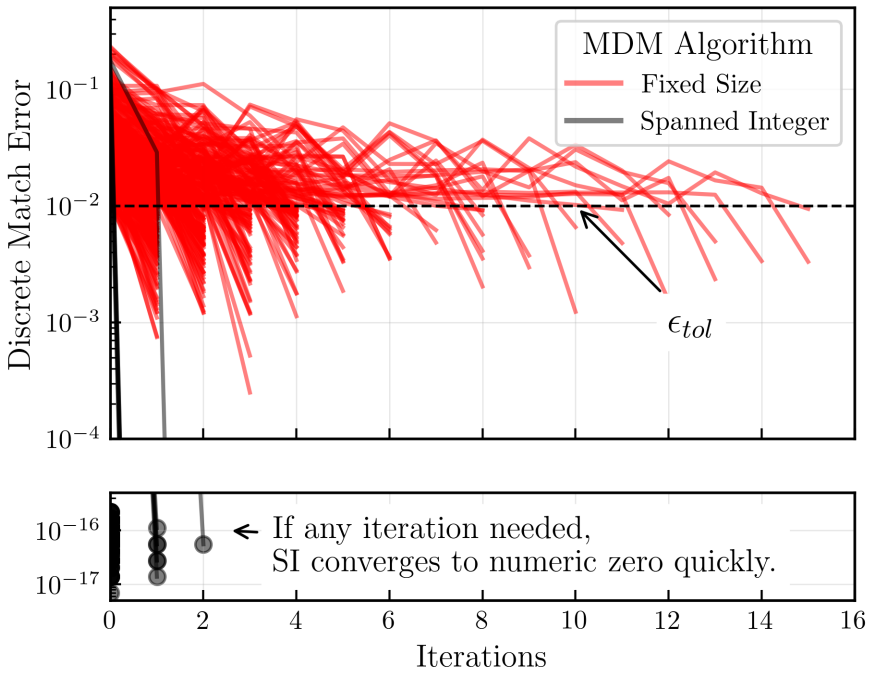

Figure 2: Packing Algorithm Convergence

If the fixed size approach, with its many iterations produce larger solutions for a "minimal" set than a different algorithm, can its solutions even be considered valid? They are, in fact,

valid solutions, but to an over-constrained version of the problem at hand. This differently constrained solution serves at least two potential purposes. First, there are scenarios in which very specific, pre-determined sizes are required. For example, in space-filling solutions, such as presented by Botet, Kwok, and Cabane (2021), particle size is constrained by 3-D geometry rather than the 1-D bounds of a mass distribution-filling problems. Second, comparing the "minimal" solution sizes from the fixed size and spanned integer algorithms provides a measure of how sensitive the minimal discrete match problem is to the selection of sizes. The SI algorithm finds the ideal sizes where the relative particle quantities fall neatly into place to match the target GSD. Even over the fairly small available size ranges, the FS algorithm performance shows that deviation from these ideal sizes can carry a significant penalty.

The results in the rest of this paper use the spanned integer algorithm.

Results

The minimal discrete matches for the grain size distributions described in Table 1 range in magnitude from $\mathcal{O}(10^1)$ to $\mathcal{O}(10^{10})$ particles. The results are presented in terms of ϕ_1, ζ_1, I_C, and USCS classification. As intuition suggests, increasing the percent mass of the smallest particle sizes or the ratio of the largest to the smallest particle sizes both tend to increase N_{MDM}. Figure 3 illustrates a scattered linear relationship between ϕ_1 (subscript 1 indicates the smallest particle size) and N_{MDM} in log-log space for constant volume ratio, ζ_1. For $\phi_1 \approx 0.4$, $\log_{10}\zeta_1$ correlates roughly 1:1 with N_{MDM}. The color bands of $\log_{10}\zeta_1$ indicate distinct combinations of min and max sieve sizes in the target G.

A similar pattern exists between N_{MDM} , ζ_1, and I_C as shown in Figure 4. At a given volume ratio, an increase in curvature index tends to result in an increased N_{MDM}. For $I_C \approx 0.7$, $\log_{10}\zeta_1$ provides a very rough 1:1 correlation with N_{MDM}. Curves with high I_C tend to involve larger relative masses at lower sizes than those with low I_C, which is conceptually similar to ϕ_1, but averaged over the entire distribution.

The grain size parameters C_c, C_u, and I_{GS} were found to not correlate well with N_{MDM}. However, I_{GS} is helpful for visualizing the breadth of correlation between N_{MDM} and USCS classifications. Figure 5 shows the range of N_{MDM} associated with the granular USCS classifications in this study. The lower boundary of the data results indicate that all but the broadest distributions have instances that can be matched with N_{MDM} magnitudes of $\mathcal{O}(10^2)$ or lower. Yet even the classifications with the smallest observed matches also have numerous instances requiring magnitudes from $\mathcal{O}(10^6)$ to $\mathcal{O}(10^9)$.

Relative to "poorly graded gravel", "well-graded gravel" tends to have larger N_{MDM}. Where "with sand" is applicable to the gravel group name, this tendancy and lower limit of N_{MDM} remains the same, but the upper range of N_{MDM} is higher. Instances where "with silt and sand" are applicable to the gravel group name were rare in the generated GSDs and had very high values of N_{MDM}. The rarity and large N_{MDM} for these gradations are due to strict

requirements for these group names, which only a small range of values across a broad range of particle sizes can satisfy.

Relative to "poorly graded sand", "well-graded sand" tends to have a similar (and quite broad) range of N_{MDM}. The addition of "with gravel" or "with silt" makes very little difference in the range of N_{MDM} at similar ranges of I_{GS} (although a small increase in lower bound is apparent in "well-graded sand with gravel". Even "silty sand" appears to share orders of magnitude with "poorly graded sand" at comparable ranges of I_{GS}. Similar to the trends for gravels, sands whose group names include both silt and gravel were rare in the generated GSDs and had high values of N_{MDM}.

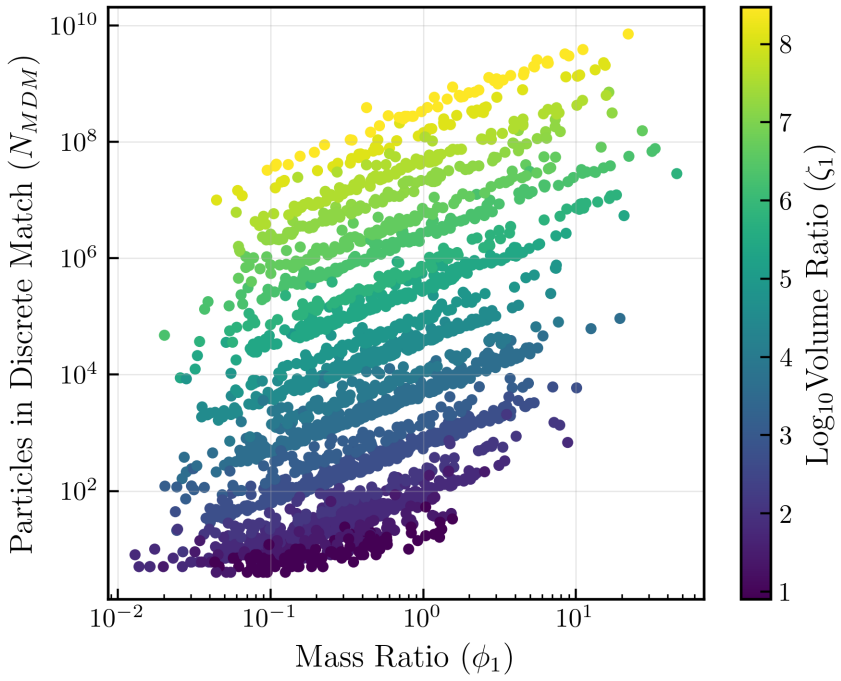

Figure 3: Total particles in MDM with with mass ratio. Color scale indicates volume ratio.

Discussion

Depending on the GSD, loading conditions, and behavior of interest, the minimum particle size expected to participate mechanically in the soil matrix may be substantially smaller than the largest particle size.

The spanned integer algorithm component of the approach presented here is effective in finding the MDM. However, some caution is warranted in using the full flexibility allowed by a GSD on the discrete particle sizes. While traditional sieve analyses, and the analogous mathematical

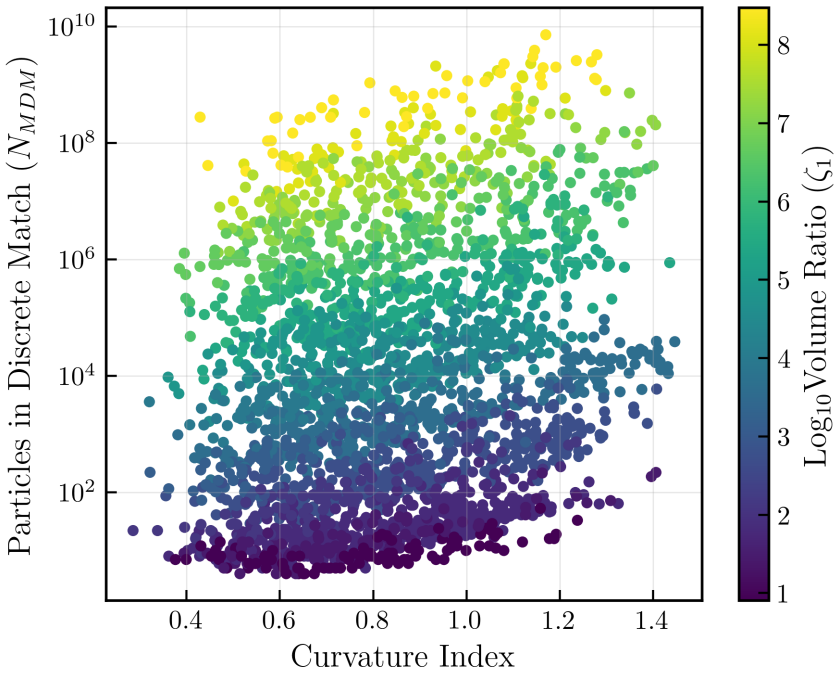

Figure 4: Total particles in MDM with with curvature index. Color scale indicates volume ratio

representation presented herein do not provide any additional constraints, natural soil grain distributions are known to be broadly distributed between sieves (Ghalib and Hryciw 1999). Modelers should inspect the MDM sizes of their target GSDs for results that would indicate abnormal underlying characteristics (e.g., repeated adjacent discrete sizes converging at the sieve boundary). Approximate solutions to the MDM, which give greater control over particle sizes, may be sufficient depending on the application and the desired rigor in the $S : G$ match.

It is important to understand that an MDM does not constitute a representative volume (RV). An MDM is defined purely in terms of discrete mass-volume relationships. The RV has additional requirements about the mechanical behavior. Nevertheless, the mechanical behavior of granular materials is strongly influenced by its GSD, so the discrete mass-volume relationship cannot be fully decoupled from the RV. In this sense, the MDM can be thought of as a building block and minimum bound on the particle count in RVE, which can be useful in assessing computational cost.

The behavior of interest being modeled may allow some portions of a realistic GSD to be truncated without loss of fidelity, allowing a partial decouple of MDM and RV. For example, studies have shown that for certain combinations of grain size distribution (GSD) and density, there may be significant portions of the soil mass that are not mechanically engaged with the soil matrix (Sufian et al. 2021). This may lead to thresholds of particle size that can be omitted from a given simulation without significant impact on the behavior being studied. In

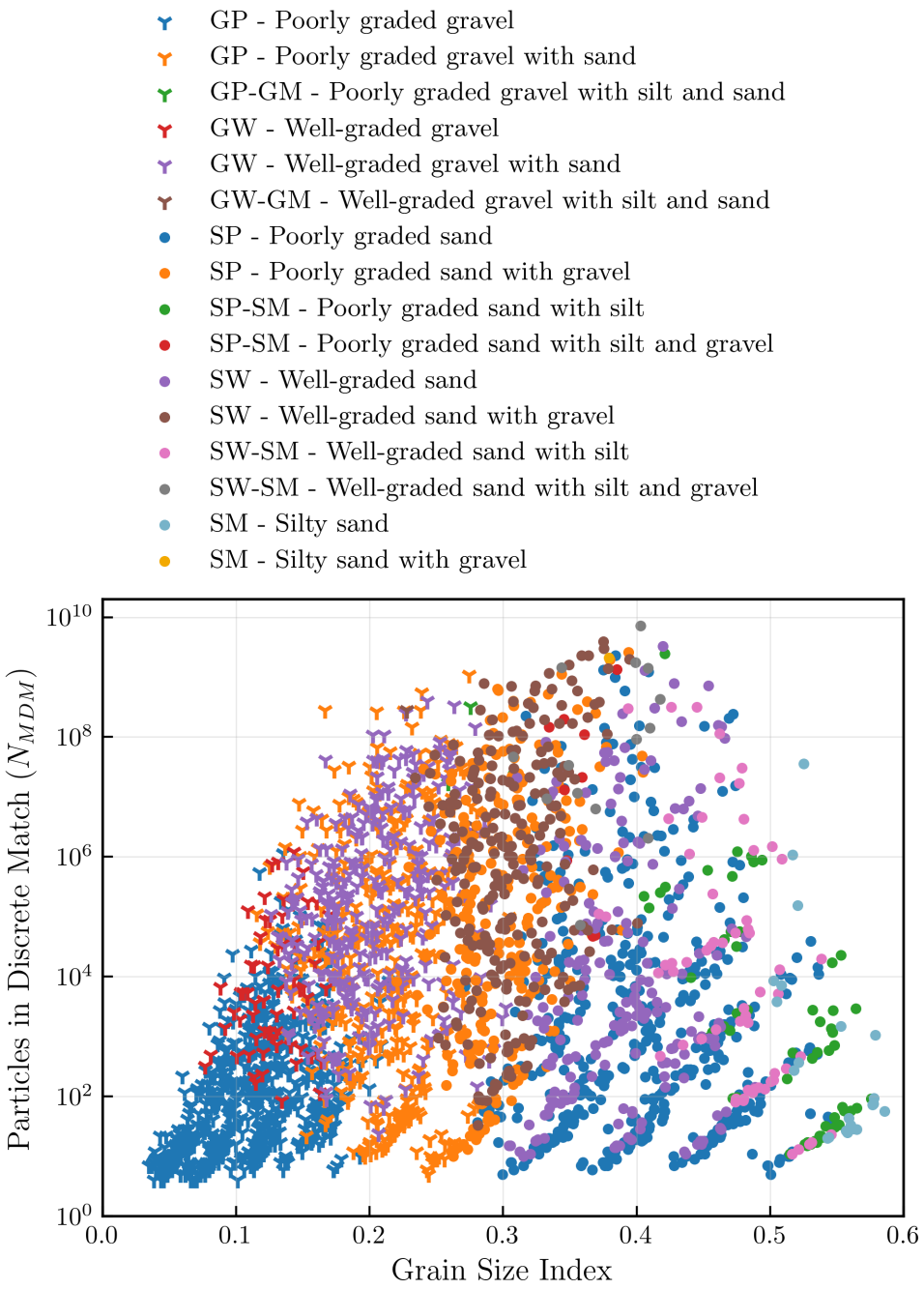

Figure 5: USCS classifications of the MPS data.

such cases, the MDM concept is still valuable for comparing relative computational effort with varying levels of truncation.

Conclusions

This paper described algorithms to 1) find a first-order approximation of a minimal discrete match (MDM) to general grain size distributions, 2) quantify the error in approximate minimal matches, and 3) find a rigorous solution for the MDM with a spanned integer approach. Depending on whether the modeling application calls for a rigorous match between a target GSD and its DEM representation, either a rigorous or approximate MEM approach may be appropriate.

The MDM is a valuable tool in quantifying relative computational cost and assessing feasibility of building DEM models of different grain size distributions. To be used effectively, the representative volume for the modeling application must be considered.

The MDM results for a large distribution of USCS show several trends: increasing the ratio of maximum to minimum particle size, the relative mass of the smallest particles, and the curvature index all increase the size of the MDM.

The distribution of MDM size across USCS classifications indicates that a wide range of realistic granular soils are feasible for DEM modeling. The distribution also indicates that neither USCS classification, curvature index, size, and mass ratios alone are sufficient to characterize MDM size. This indicates that MDM magnitudes for specific GSDs of interest should calculated rather than estimated based on index parameters.

Data and code availability

The Python code used to generate data, solve for minimal discrete matches, and plot figures are available on GitHub.

Bandera, Sara, Catherine O'Sullivan, Paul Tangney, and Stefano Angioletti-Uberti. 2021. "Coarse-Grained Molecular Dynamics Simulations of Clay Compression." *Computers and Geotechnics* 138 (October): 104333. https://doi.org/10.1016/j.compgeo.2021.104333.

Botet, Robert, Sylvie Kwok, and Bernard Cabane. 2021. "Filling Space with Polydisperse Spheres in a Non-Apollonian Way." *Journal of Physics A: Mathematical and Theoretical* 54 (19): 195201. https://doi.org/10.1088/1751-8121/abef81.

Cantor, David, and Carlos Ovalle. 2025. "Sample Size Effects on the Critical State Shear Strength of Granular Materials with Varied Gradation and the Role of Column-Like Local Structures." *Géotechnique* 75 (1): 29–40. https://doi.org/10.1680/jgeot.23.00032.

Cundall, P. A. 2001. "A Discontinuous Future for Numerical Modelling in Geomechanics?" *Proceedings of the Institution of Civil Engineers - Geotechnical Engineering* 149 (1): 41–47. https://doi.org/10.1680/geng.2001.149.1.41.

Cundall, P. A., and O. D. L. Strack. 1979. "A Discrete Numerical Model for Granular Assemblies." *Géotechnique* 29 (1): 47–65. https://doi.org/10.1680/geot.1979.29.1.47.

D18 Committee. n.d. "Test Methods for Particle-Size Distribution (Gradation) of Soils Using Sieve Analysis." https://doi.org/10.1520/d6913_d6913m-17.

Dong, Youkou, Dingtao Yan, and Lan Cui. 2022. "An Efficient Parallel Framework for the Discrete Element Method Using GPU." *Applied Sciences* 12 (6): 3107. https://doi.org/10.3390/app12063107.

Erguler, Zeynal Abiddin. 2016. "A Quantitative Method of Describing Grain Size Distribution of Soils and Some Examples for Its Applications." *Bulletin of Engineering Geology and the Environment* 75 (2): 807–19. https://doi.org/10.1007/s10064-015-0790-1.

Fang, Luning, Ruochun Zhang, Colin Vanden Heuvel, Radu Serban, and Dan Negrut. 2021. "Chrono::GPU: An Open-Source Simulation Package for Granular Dynamics Using the Discrete Element Method." *Processes* 9 (10): 1813. https://doi.org/10.3390/pr9101813.

Ghalib, Ali M., and Roman D. Hryciw. 1999. "Soil Particle Size Distribution by Mosaic Imaging and Watershed Analysis." *Journal of Computing in Civil Engineering* 13 (2): 80–87. https://doi.org/10.1061/(ASCE)0887-3801(1999)13:2(80).

Gitman, I. M., H. Askes, and L. J. Sluys. 2007. "Representative Volume: Existence and Size Determination." *Engineering Fracture Mechanics* 74 (16): 2518–34. https://doi.org/10.1016/j.engfracmech.2006.12.021.

Miyai, Shinichiro, Murino Kobayakawa, Takuya Tsuji, and Toshitsugu Tanaka. 2019. "Influence of Particle Size on Vertical Plate Penetration into Dense Cohesionless Granular Materials (Large-Scale DEM Simulation Using Real Particle Size)." *Granular Matter* 21 (4): 105. https://doi.org/10.1007/s10035-019-0961-z.

O'Sullivan, C. 2014. "Advancing Geomechanics Using DEM." In, edited by Kenichi Soga, Krishna Kumar, Giovanna Biscontin, and Matthew Kuo, 21–32. CRC Press. http://www.crcnetbase.com/doi/abs/10.1201/b17395-4.

Sufian, Adnan, Marion Artigaut, Thomas Shire, and Catherine O'Sullivan. 2021. "Influence of Fabric on Stress Distribution in Gap-Graded Soil." *Journal of Geotechnical and Geoenvironmental Engineering* 147 (5): 04021016. https://doi.org/10.1061/(ASCE)GT.1943-5606.0002487.

Tyler, Scott W., and Stephen W. Wheatcraft. 1992. "Fractal Scaling of Soil Particle-Size Distributions: Analysis and Limitations." *Soil Science Society of America Journal* 56 (2): 362–69. https://doi.org/10.2136/sssaj1992.03615995005600020005x.

Zeraati-Shamsabadi, Mohammad, and Abouzar Sadrekarimi. 2025. "A DEM Study on the Effects of Specimen and Particle Sizes on Direct Simple Shear Tests." *Granular Matter* 27 (2). https://doi.org/10.1007/s10035-025-01513-y.

Zhang, Ruochun, Bonaventura Tagliafierro, Colin Vanden Heuvel, Shlok Sabarwal, Luning Bakke, Yulong Yue, Xin Wei, Radu Serban, and Dan Negruţ. 2024. "Chrono DEM-Engine: A Discrete Element Method Dual-GPU Simulator with Customizable Contact Forces and Element Shape." *Computer Physics Communications* 300 (July): 109196. https://doi.org/10.1016/j.cpc.2024.109196.