



3D Particle Shape Modelling and Optimization through Proper Orthogonal Decomposition Application to Railway Ballast

Noura Ouhbi, Charles Voivret, Guillaume Perrin, Jean-Noël Roux

► To cite this version:

Noura Ouhbi, Charles Voivret, Guillaume Perrin, Jean-Noël Roux. 3D Particle Shape Modelling and Optimization through Proper Orthogonal Decomposition Application to Railway Ballast. Granular Matter, 2017, 10.1007/s10035-017-0771-0 . hal-01699213

HAL Id: hal-01699213

<https://hal.science/hal-01699213>

Submitted on 2 Feb 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

3D Particle Shape Modelling and Optimization through Proper Orthogonal Decomposition

Application to Railway Ballast

Noura Ouhbi · Charles Voivret · Guillaume Perrin · Jean-Noël Roux

Received: date / Accepted: date

Abstract Based on Proper Orthogonal Decomposition (POD), a new method is presented in order to statistically characterize arbitrary particle shapes using an optimal choice of shape functions identified on a set of 1000 digitized railway ballast particles obtained through 3D Scan. The coefficients of the POD expansion enable a description of ballast grains with varying levels of accuracy. On exploiting the knowledge of their statistical distribution we are able, implementing an appropriate Multivariate Kernel Density Estimation (Multivariate KDE) method, to generate irregular particles with similar morphological features. The description and generation methods are validated by comparing statistical distributions of basic characteristics: surface area, volume, average radius, elongation, flatness, and aspect ratio. Using suitable geometric descriptors defining local curvatures, we identify which surface points might be regarded as forming faces. This shows that the proposed particle generation method is well suited for irregularly shaped granular materials, as a first geometric definition step, before numerical simulations of their collective mechanical properties are carried out by a Discrete Element code dealing with

polyhedral shapes. We illustrate this process with the simple case of the assembling of a granular pack from a loose configuration, by one-dimensional compression, using different levels of accuracy in the representation of grain shape.

Keywords Railway Ballast · Shape · Characterization · Modelling · Proper Orthogonal Decomposition · Generation · Stochastic

1 Introduction

Understanding the mechanical behaviour of the ballasted track [1] is a major issue for railway companies. As with other granular materials, numerical simulations by the Discrete Element Method (DEM) [2,3] are a particularly valuable tool, thanks to its unique ability to control independently the micromechanical parameters (e.g. friction or restitution coefficients), and to the insights it provides on granular materials at different scales.

An important question, though, is the modeling of grain shape and of its influence on mechanical properties. As shown in Fig. 1, the ballast grains, with sizes of the order of a few centimeters, assume rather different, irregular shapes. They typically exhibit angular shapes and rough surfaces composed of large quasi planar faces limited by sharp edges. The issue of how to characterize the size and shape of solid particles is currently being investigated in many different fields of applied physical sciences and engineering, including chemical, atmospheric, pharmaceutical and soil sciences [4–8]. A number of numerical studies [9–16], as well as laboratory experiments [13], have shown the strong influence of grain shape on granular material behaviour, in both quasistatic and inertial flow regimes.

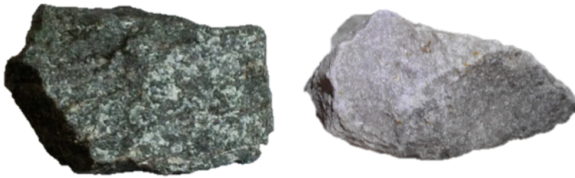
N.Ouhbi · C.Voivret
Department of Lines, Tracks & Environment
Engineering & Projects, SNCF-Reseau
Saint-Denis, France Tel.: +33-45-678910
E-mail: ouhbi.noura@gmail.com

N.Ouhbi · J.-N.Roux
Université Paris Est
Laboratoire Navier (UMR 8205), CNRS, ENPC, IFSTTAR
Marne-la-Vallée, France
E-mail: jean-noel.roux@ifsttar.fr

G. Perrin
CEA/DAM/DIF
Arpajon, France



(a)



(b)

Fig. 1 (a) Railway Ballast. (b) Two ballast grains on a larger scale.

The principal goal of this paper is to provide a quantitative method to model the shape of ballast grains, a necessary step in a research effort based on DEM, towards a quantitative and relevant evaluation of the sensitivity of ballast performance to shape variability. In this paper, we propose a novel statistical method to achieve an optimal and exhaustive particle shape characterization. The input quality is rendered relevant by the millimetric accuracy of 3D scans of real ballast grains. Our approach enables a DEM modelling of ballast grains as polyhedra with, in theory, an arbitrarily high accuracy.

In order to characterize particle shape, various shape descriptors of different levels of complexity have been introduced. As for railway ballast, we can find several shape indicators in the European and French norms (BS EN 13450, 2003). They are categorized in 1D, 2D and 3D descriptors, and are based on simple formulas combining some basic properties such as surface area, volume, or diameter. Such simple descriptors (e.g., elongation or roundness), while providing global shape information, fail to capture the complex morphology of ballast grains (faces, edges, angularity, etc.). More advanced shape descriptors have been introduced in the last years in order to both quantify and represent the realistic geometric characteristics of granular particles before incorporating them in DEM simulations [17–24].

Most of these approaches are based on computing statistical properties of different shape features, and using them to describe the shape of a set of particles.

Sophisticated three-dimensional shape characterizations recently used include 3D extensions of Fourier descriptors [25, 26], or spherical harmonics (SH) series (which appropriately generalize Fourier series in a spherical topology) [27–31]. They were implemented for shape characterization as well as generation of grain shapes. As explained further in this paper, such approaches, however powerful they may be, prove not convenient for highly angular shapes, as in the case of ballast grains.

Thus, the method we propose is based on the *Proper Orthogonal Decomposition* (POD) technique [32]. Historically introduced in studies in turbulence as an efficient method to identify and extract coherent structures of a flow [33], POD can be regarded as a way to substitute the Fourier decomposition [34] when the flow directions cannot be assumed uniform or periodic. POD is also known as the Karhunen-Loève Decomposition [35, 36] or Hotelling analysis [37], and may be viewed as a variant of Principal Component Analysis [38], or of Singular Value Decomposition [39]. It is being used through a broad range of applications (e.g., image processing, signal analysis, data compression and optimal control [40]), for which the common objective is to extract the dominant features out of a set of data, and to build a reduced model. As shown in [22], this method is also efficient for particle shape classification.

Sec. 2 presents the POD method and identifies its ingredients on a set of scanned ballast grains. Using statistical distributions provided by POD, the numerical generation of sets of equivalent grains is made possible as described in Sec. 3, on resorting to a stochastic method known as the *Multivariate Kernel Density Estimation* (Multivariate KDE). The method is validated by comparing statistical distributions of some basic shape descriptors (e.g., surface area, elongation or flatness) between the original grain and the numerically generated grain sets. More advanced geometric descriptors are discussed in Sec. 3.3, such as curvature and large faces (identified as low curvature subsets of the grain surfaces), allowing both more physical characterizations and more stringent tests of the modeling of grain morphology. Finally, Sec. 4 reports on a preliminary numerical simulation of some mechanical properties of the model material.

2 Grain shape characterization

2.1 Data Acquisition and Pre-processing

We start from a dataset (provided by SNCF) with $n = 1000$ point clouds obtained by 3D digitization (3D scan) of n real ballast grain surfaces, each one comprising about 300000 points. Fig. 2 shows 2 examples of these point clouds. To increase POD efficiency we have to

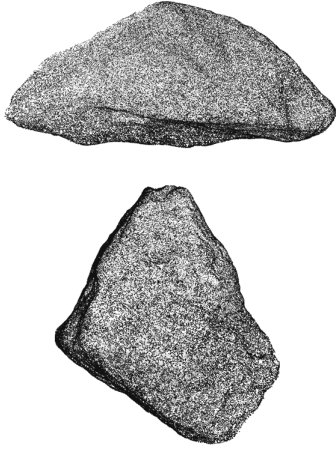


Fig. 2 Scanned ballast grains.

maximize the resemblance between the grains. After cleaning the point clouds from possible outliers, the centre of mass of all grains is moved to the origin of coordinates and the grains are rotated so that their principal axes of inertia coincide with a common fixed frame. The resulting grains (see Fig. 3), are represented by the values of their radii, computed, from their centre of mass to their surface, along $d = 800$ isotropically distributed directions, that are common to all the grains (using the point repulsion algorithm defined in Ref. [41]). This relies on the assumption, reasonable for ballast, of star-like grain shapes and uniquely defined radii in each direction. Denoting as r_{ij} the radius of grain i along direction j , we thus define a matrix $\mathcal{A}$, each line of which containing the d radii of one of the n grains, along each one of the d directions. Each one of the d columns of $\mathcal{A}$ contains the n radii of the grains as measured in one direction.

We then define the average grain by its radii:

$$\bar{r}_j = \frac{1}{n} \sum_{k=1}^n r_{kj} \quad (1 \leq j \leq d). \quad (1)$$

The average grain, as shown in Fig. 4, is close to an ellipsoid with large ($\bar{L}$), intermediate ($\bar{I}$) and small ($\bar{S}$)

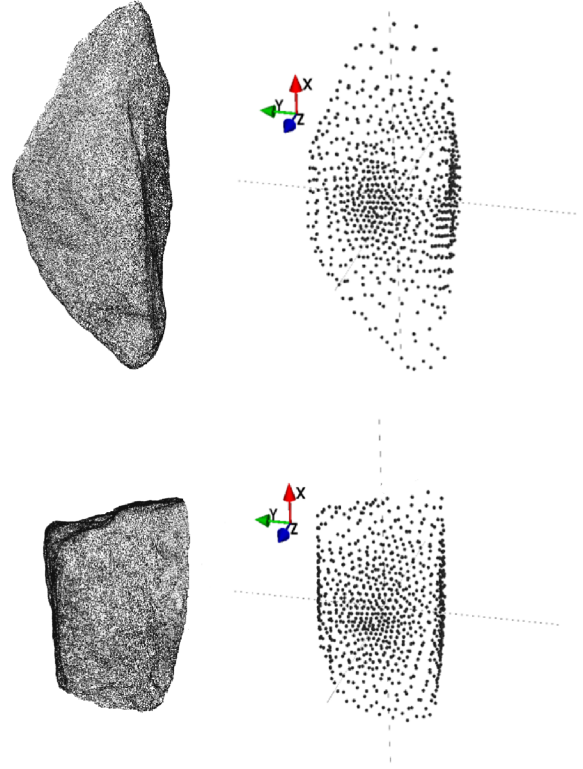


Fig. 3 Pre-processed grains: effects of reorientation (left), of description with 800 directions (right).

lengths along principal axes given by:

$$\bar{L} \simeq 29.6 \text{ mm}; \quad \bar{I} \simeq 25.7 \text{ mm}; \quad \bar{S} \simeq 21.6 \text{ mm}. \quad (2)$$

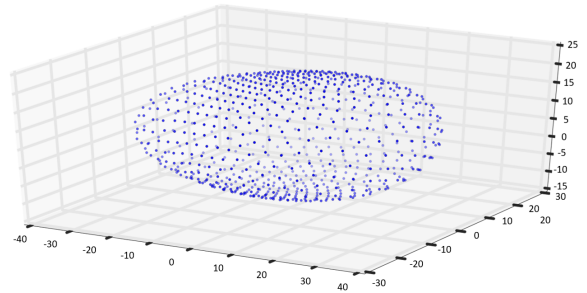


Fig. 4 Average grain shape.

2.2 POD applied to particle shape

To organize morphological information more efficiently, it is natural to work on the difference between each grain and the average grain, using

$$\Delta r_{ij} = r_{ij} - \bar{r}_j. \quad (3)$$

The radius differences written in (3) define matrix $\mathbf{A}$, and its line vectors $\Delta \mathbf{r}_i$ for $1 \leq i \leq n$, by subtracting vector $\bar{\mathbf{r}}$ from each line of $\mathcal{A}$. Let us, for any matrix $\mathbf{a}$ denote its Frobenius norm as $\|\mathbf{a}\|$: $\|\mathbf{a}\|^2 = \sum_i \sum_j a_{ij}^2$. $\|\bar{\mathbf{r}}\|$ being the norm of d -dimensional vector $\bar{\mathbf{r}}$, one has by definition:

$$\|\mathcal{A}\|^2 = n\|\bar{\mathbf{r}}\|^2 + \|\mathbf{A}\|^2, \quad (4)$$

and thus the importance of the variation from grain to grain, relatively to the average shape, may be expressed by the following ratio:

$$\delta_1 = \frac{\|\mathbf{A}\|}{n^{1/2}\|\bar{\mathbf{r}}\|} \simeq 20\% \quad (5)$$

Applying the POD to the data contained in $\mathbf{A}$ consists in carrying out a singular value decomposition of this matrix, i.e.,

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot {}^T\mathbf{V} \quad (6)$$

in which $\mathbf{U}$ and $\mathbf{V}$ are orthogonal square matrices, of respective order n and d , and $\mathbf{\Sigma}$ is a $n \times d$ matrix with all elements equal to zero except the first r ones on the diagonal, σ_k , $1 \leq k \leq r$, which are positive and classified in decreasing order, r being the rank of $\mathbf{A}$. Equivalently, matrix ${}^T\mathbf{A} \cdot \mathbf{A}$, which is square symmetric of order d and contains the correlation coefficients of the column vectors of $\mathbf{A}$, has to be diagonalised. Its $d \geq r$ eigenvalues λ_j are $\lambda_j = \sigma_j^2$ for $1 \leq j \leq r$, $\lambda_j = 0$ for $j \geq r + 1$, and, written in decreasing order from the upper left corner, define diagonal matrix $\mathbf{D}$. Matrix $\mathbf{V}$, as defined in (6), contains the coordinates of the corresponding (normalized) eigenvectors, such that

$${}^T\mathbf{A} \cdot \mathbf{A} = \mathbf{V} \cdot \mathbf{D} \cdot {}^T\mathbf{V}. \quad (7)$$

As to matrix $\mathbf{U}$, it is easily deduced from equality

$$\mathbf{U} \cdot \mathbf{\Sigma} = \mathbf{A} \cdot \mathbf{V}, \quad (8)$$

on exploiting the simple diagonal structure of $\mathbf{\Sigma}$.

The main interest of the POD in the context of data modeling is its ability to define a set of approximations $\mathbf{A}^{(k)}$ for $\mathbf{A}$, each one being optimal in the sense that $\mathbf{A}^{(k)}$ minimizes $\|\mathbf{A} - \mathbf{a}\|^2$ among all matrices $\mathbf{a}$ of rank k .

$\mathbf{A}^{(k)}$ is obtained on setting all singular values σ_l to zero for $l \geq k + 1$ in SVD decomposition (6), thereby

replacing $\mathbf{\Sigma}$ by $\mathbf{\Sigma}^{(k)}$, containing only k non-vanishing diagonal elements:

$$\mathbf{A}^{(k)} = \mathbf{U} \cdot \mathbf{\Sigma}^{(k)} \cdot {}^T\mathbf{V}. \quad (9)$$

A relative error in approximating $\mathbf{A}$ by $\mathbf{A}^{(k)}$ might be defined as:

$$\epsilon(k) = \frac{\|\mathbf{A} - \mathbf{A}^{(k)}\|^2}{\|\mathbf{A}\|^2} = \frac{\sum_{l=k+1}^r \lambda_l}{\sum_{l=1}^r \lambda_l}. \quad (10)$$

By construction, it decreases as k increases and is equal to 0 for $k = r$ (with $r = d$ in the present case, because $800 = d < n = 1000$). Conversely, the difference $E(k) = 1 - \epsilon(k)$, to which we refer as the information content of the approximate morphological description at order k , grows from zero to one. Returning to the vectors (or sets of radius differences) $\Delta \mathbf{r}_i$, with j coordinates that describe the shapes of the n grains, relation 6 expresses them as linear combinations, with coefficients $U_{il}\sigma_l$, of a set of basis vectors, hereafter referred to as *shape functions* or modes, $\mathbf{S}_l$, defined by the columns of matrix $\mathbf{V}$, $S_{lj} = V_{jl}$ (the eigenvectors of ${}^T\mathbf{A} \cdot \mathbf{A}$). Upon truncating the decomposition at order k one only uses the k first functions, and characterise the shape of the grains by the values taken by k coefficients.

2.3 Results and discussion

Fig. 5 shows the error defined in (10) as a function of the number k of basis functions. This error measures a relative shape deviation, averaged over all grains, between their real and their approximated shapes (truncated at the order k). Fig. 5 shows that the method is successful in extracting the important information: keeping only the 20 first modes instead of 800 (2.5% of the total number), the error remains below 10%. This performance can be increased as we have larger sets of particles. For more details, see [42].

It is interesting to test the sensitivity of the set of basis functions to the initial number of scanned grains. We reproduced the procedure using randomly chosen subsets of the grain population, and compared the resulting covariance matrices using the method of Ref. [43]. The outcome of this study [44] is that the differences become negligible as soon as a population of 400 ballast grains is used in the statistical analysis.

In order to prove the efficiency of our method, we compare its results with those of the well known method shape characterisation method relying on an expansion in Spherical Harmonics (SH). Using the same number of

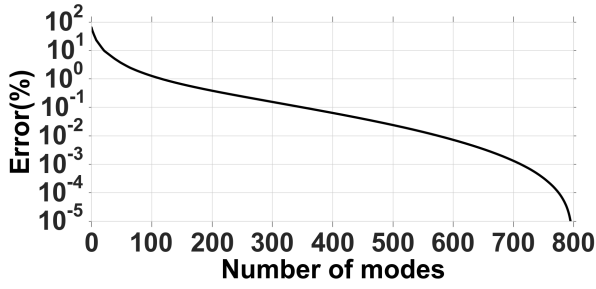


Fig. 5 Error $\epsilon(k)$, as defined in Eq. 10, versus number k of shape functions.

coefficients, $N_c = 800$, corresponding to 100% of the information within the POD approach, we reconstructed our dataset with different shape precision levels and compared with POD results. Fig. 6 shows an example of a grain reconstructed with 800 coefficients using POD and using SH, and one can notice that for a complex shaped particle as ballast, a larger SH basis is required: SH expansions with 800 coefficients fail to construct a grain similar to the scanned one. On trying different basis sizes, we found that the complete precision (100% of the shape information) is approached with 3000-3500 SH coefficients. Fig. 7, plotting SH and POD errors versus the number of coefficients, clearly shows that the POD approach is, by far, the more efficient one.

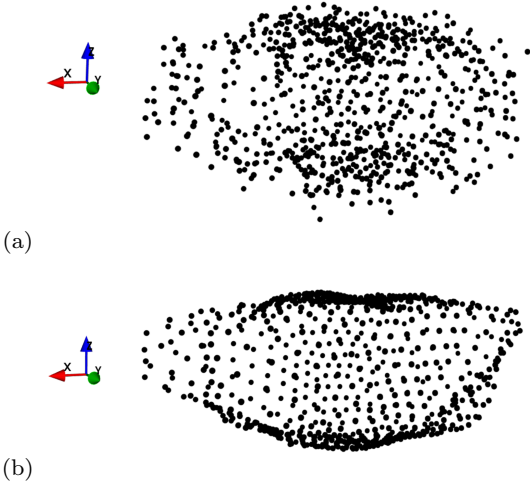


Fig. 6 A grain reconstructed using (SH) analysis (a) and POD procedure (b).

As in the SH analysis, the POD modes of increasing order tend to characterise details of smaller scale, from global shape to surface texture and roughness details. This is apparent in Fig. 8, showing the first mode, containing more than 50% of the morphological information ($E(1) > 0.5$), to have roughly ellipsoidal shape,

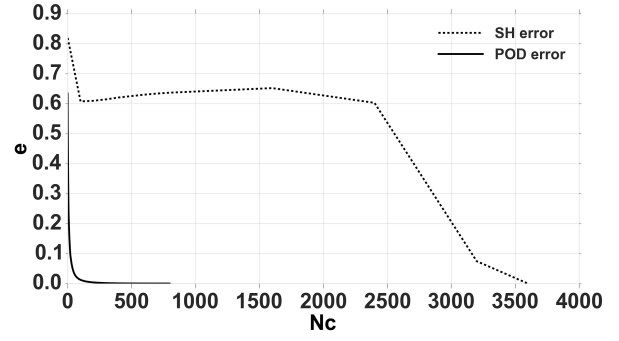


Fig. 7 Error $\epsilon(k)$, as defined in Eq. 10 for POD and SH, versus number N_c of coefficients.

while the following ones, with positive and negative lobes, are associated with regions of higher curvatures. The first mode (Fig. 8a) is quite similar to the average

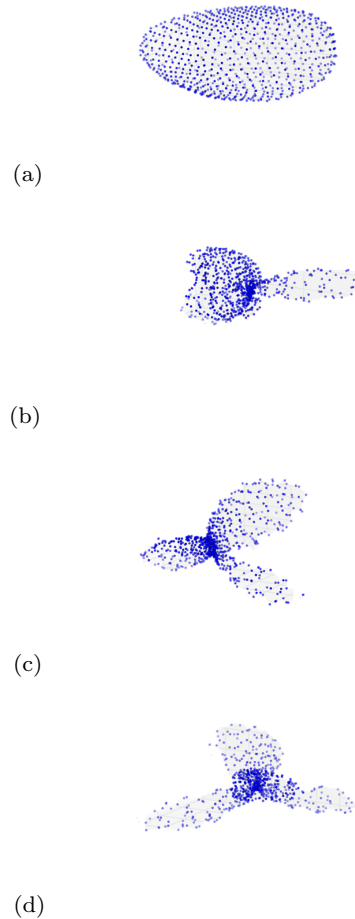


Fig. 8 First shape functions. From top to bottom: mode 1 to mode 4.)

grain (Fig. 4), albeit slightly more elongated in shape

DE COMBIEN ???? Consequently, the coefficient of the first mode is, to a large extent, characteristic of the grain size, but also of its shape, which is correlated to it.

Fig. 9 illustrates how the shape of one grain is gradually retrieved by an increasing number of modes used to reconstruct the shape. Table 1 lists the number of shape functions necessary to achieve the different levels of accuracy illustrated on this figure, as well as in the measurements of characteristic morphological data in the following.

Table 1 Number of modes corresponding to different error levels. The error decreases to zero for 800 modes.

Error	20%	15 %	10%	5%	1%	0.1%
Mode number	11	16	22	40	117	350

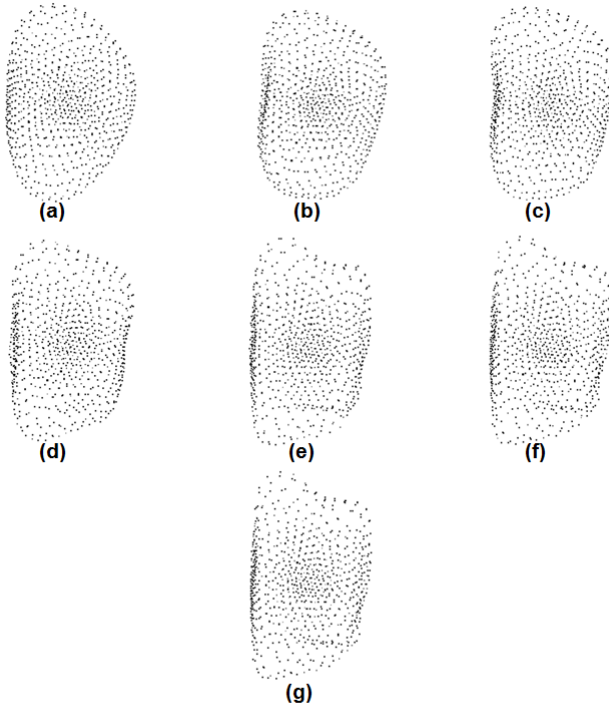


Fig. 9 One grain, approximated with the different levels of approximations of Tab. 1, improving from (a) to (f), compared to exact shape (g).

To quantify the evolution of shape descriptors with shape accuracy, we computed surface areas $\mathcal{S}_i$, volumes Ω_i , and average radii $\bar{R}_i$, $1 \leq i \leq n$, defined as follows, from radius values interpolated along directions distributed over the unit sphere such that both angles

θ and ϕ vary by constant steps, $\Delta\theta$ and $\Delta\phi$.

$$\begin{aligned}\mathcal{S}_i &= \Delta\theta\Delta\phi \sum_{j=1}^d r_{ij}^2(\theta_j, \phi_j) \sin \phi_j \\ \Omega_i &= \frac{\Delta\theta\Delta\phi}{3} \sum_{j=1}^d r_{ij}^3(\theta_j, \phi_j) \sin \phi_j \\ \bar{R}_i &= \frac{\Delta\theta\Delta\phi}{\mathcal{S}_i} \sum_{j=1}^d r_{ij}^3(\theta_j, \phi_j) \sin \phi_j = \frac{3\Omega_i}{\mathcal{S}_i}.\end{aligned}\quad (11)$$

Other morphological featured are also investigated. Identifying three linear dimensions of a grain, as length (largest dimension L), width (intermediate dimension I) and thickness (smallest dimension S), as in Ref. [17] (which involves embedding the smallest projected surface of the grain onto a plane within a minimum rectangle), one defines the grain *flatness* as $F = S/I$, its *elongation* as $E = I/L$ and its *aspect ratio* as $a = EF = S/L$. The shape of mode 1 is such that the larger grains (with large positive coefficient C_1) have larger flatness and aspect ratio than the samller ones (those with negative C_1 of larger absolute value).

Probability density functions (PDF) of all these descriptors computed for reconstructed particles with different levels of shape precision are compared with those of the scanned set. Fig. 10 plots the probability density functions (PDF) of $\mathcal{S}$, Ω , $\bar{R}$. The ability of the sequence of shape approximation to describe those distributions is more quantitatively appreciated on evaluating the Kolmogorov-Smirnov (KS) likelihood parameter (which quantifies the probability, assuming equidistribution of two sets of data, that the difference in their empirical distributions is as large as the one that is actually measured). Likelihood values for all 6 descriptors are very good, showing that the distributions of morphological features are well reproduced, even with more than 10% error on shape (Tab. 2).

Table 2 Kolmogorov-Smirnov test results, at different levels of approximation k , for the distributions of volume Ω , surface area $\mathcal{S}$, average radius $\bar{R}$, flatness F , elongation E and aspect ratio a .

k	11	16	22	40	117	350
KS (Ω)	0.866	0.879	0.984	0.997	0.997	0.997
KS ($\mathcal{S}$)	0.975	0.984	0.997	0.999	0.999	0.999
KS ($\bar{R}$)	0.725	0.737	0.764	0.785	0.849	0.947
KS (F)	0.926	0.943	0.965	0.973	0.987	0.998
KS (E)	0.934	0.941	0.945	0.981	0.986	0.996
KS (a)	0.887	0.888	0.919	0.928	0.965	0.983

The POD modeling scheme thus enables the control of shape precision, but also significant data reduction,

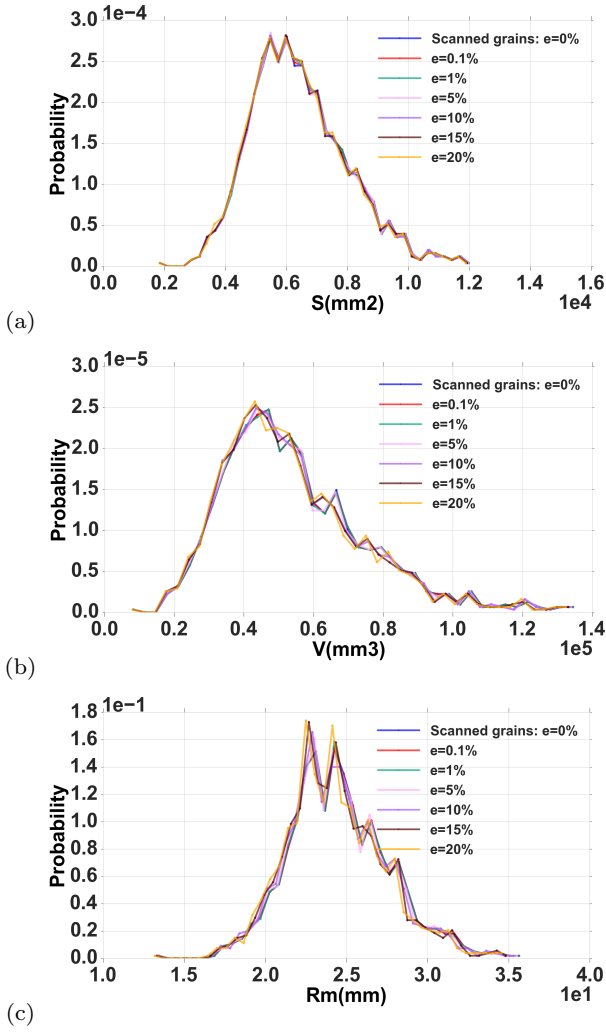


Fig. 10 Convergence study for distributions of morphological properties (a) surface area, (b) volume, (c) average radius.

as we noticed the very good agreement of shape approximation with truncated information when it comes to simple shape descriptors. This raises the question of comparing descriptors that are representative of a complex shape (faces, edges, angularity etc.). In other words, if the method shows significant reduction, then we should be able to generate grains that are equivalent to the scanned set (in terms of shape), using only a reduced number of modes. According to the obtained KS results, we are able to generate equivalent grains using only 117 modes instead of 800 modes, since it represents 99% of all information. This will be the objective of Sec. 3.

3 Generating new grains

3.1 Procedure

We now exploit the POD procedure output statistics in order to generate distinct and representative 3D ballast point clouds. The distributions of some of the coefficients of the expansion of grain shapes as series of basic shape functions are shown in Fig. 11. By construction, these coefficients are uncorrelated [45] and of zero mean, while their variances add up to $\|\mathbf{A}\|^2/n$. In order to generate grains with properties similar to

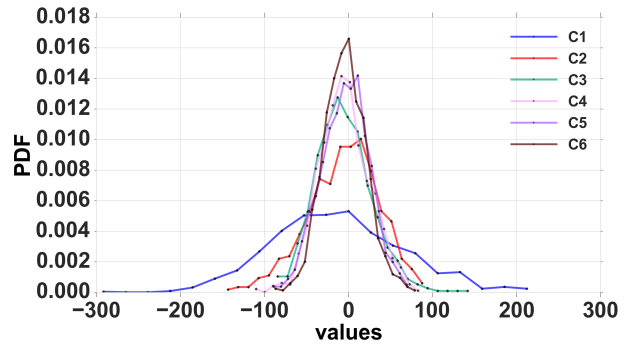


Fig. 11 Distributions of the first POD coefficients.

the scanned ballast grains from which the basic data set was obtained, one should reproduce those distributions, as well as the dependencies between the different coefficients. From a stochastic point of view, generating statistically equivalent particles means to estimate the multivariate probability density functions (PDF) representing our experimental data. To do so, the parametric approaches propose to search this PDF in an algebraic class of distributions, which are parametrized by a relatively small number of quantities. However, when the dependence structure associated with the unknown PDF is complex, the definition of a relevant parametric class can become very difficult. For instance, assuming for the PDF of the grain coefficients simple representations (such as the multidimensional Gaussian distribution) can lead to the generation of non physical shapes, once the generated coefficients have been applied to the modes.

To circumvent this problem, nonparametric approaches, such as the multivariate Kernel Density Estimation, are generally preferred. This method consists in approximating the PDF of the coefficients as a sum of multidimensional Gaussian PDFs, which are centred at each available value of these coefficients. In that case, the dependence structure that is present in the available data is directly integrated in the modeling of the PDF. How-

ever, to correctly concentrate the new generated coefficients on their regions of high probability, and therefore obtain realistic and representative grain shapes, the covariance of these Gaussian PDFs has to be carefully chosen. This is particularly important when the number of available data is not very high compared to the number of coefficients. More details about the method and the algorithm associated with this covariance optimization are presented and discussed in another publication [46]. The method enables generating as many grains as wanted in some seconds. To be able to compare correctly PDFs with the scanned set, 1000 new ballast grains are thus generated, using 100 shape functions – a limitation motivated by the observation made in Sec. 2 (see Fig. 5) that modes of higher order only contain about 1% of the total information.

3.2 Representativity: simple shape descriptors.

Fig. 12 shows some of these simulated grains. Beyond

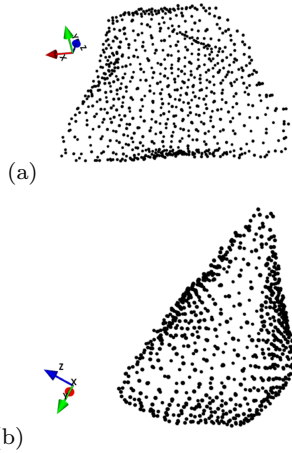


Fig. 12 Two generated ballast grains.

their realistic visual appearance, their representativity is tested with the distribution of the morphologic parameters measured on real grains: volume, surface area, average radius, flatness, elongation, and aspect ratio. The distribution of the three latter quantities is shown in Fig. 13, showing good coincidence between original and simulated grain data.

As a more decisive check, the Kolmogorov-Smirnov test values are above 0.7 for all six quantities Ω , $\mathcal{S}$, $\bar{R}$, F , E and a .

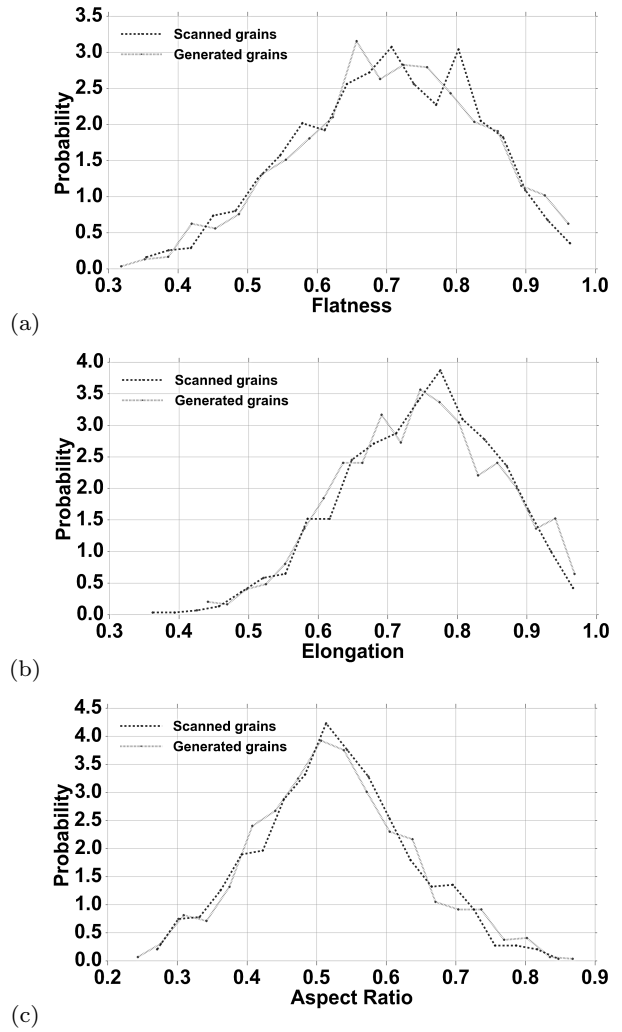


Fig. 13 Distributions of (a) flatness, (b) elongation and (c) aspect ratios in simulated grains, compared to the real ones.

3.3 Representativity: advanced shape characterization.

While distributions of simple shape parameters appear to validate our model, one essential motivation of the study was the modeling of angular and faceted grains. To assess the performance of the method in this respect, we first have to explain how to identify normal vectors, faces and edges on grains represented by point clouds.

3.3.1 Normal vectors and surface variation at vertices

We implement the method described in Ref. [41]. After a selection of a local neighborhood N_p of each point p (9 neighbours per vertex using the KDtree algorithm from Scikit-learn [47]), an eigenanalysis of the covariance matrix of N_p is carried out. This amounts to ap-

plying PCA to matrix C defined as follows:

$$C = \begin{bmatrix} \mathbf{p}_{i_1} - \bar{\mathbf{p}} \\ \dots \\ \mathbf{p}_{i_k} - \bar{\mathbf{p}} \end{bmatrix}^T \cdot \begin{bmatrix} \mathbf{p}_{i_1} - \bar{\mathbf{p}} \\ \dots \\ \mathbf{p}_{i_k} - \bar{\mathbf{p}} \end{bmatrix}, i_j \in N_p \quad (12)$$

where $\bar{\mathbf{p}}$ is the centroid of the neighbours $\mathbf{p}_{i_j}$ of the sample point $\mathbf{p}$. This 3x3 matrix is used to solve the eigenvector problem:

$$C \cdot \mathbf{v}_l = \lambda \cdot \mathbf{v}_l, l \in \{0, 1, 2\}. \quad (13)$$

λ_l measure the variation of $\mathbf{p}_i, i \in N_p$, along the direction of the corresponding eigenvectors $\mathbf{v}_l$. Labelling eigenvalues such that $\lambda_0 \leq \lambda_1 \leq \lambda_2$, the plane:

$$T(\mathbf{x}) : (\mathbf{x} - \mathbf{p}) \cdot \mathbf{v}_0 = 0 \quad (14)$$

through $\mathbf{p}$ minimizes the sum of squared distances to the neighbours of $\mathbf{p}$. Thus $\mathbf{v}_0$ approximates the surface normal at $\mathbf{p}$. Consequently, λ_0 quantitatively describes the variation along the surface normal. We define the surface variation $\sigma_m(\mathbf{p})$ at point $\mathbf{p}$ in a neighbourhood of size m by:

$$\sigma_m(\mathbf{p}) = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} \quad (15)$$

A small value of the surface variation at $\mathbf{p}$ signals a well identified apparent tangential plane on the scale of a neighbourhood of m points about point $\mathbf{p}$. Large values of $\sigma_m(\mathbf{p})$, conversely, correspond to large surface curvatures. Given a choice of m (in practice we set $m = 9$), to each vertex is attributed a normal vector and a value of surface variation. Fig. 14 visualizes the normal vectors computed at each one of the 800 points used to represent one ballast grain, while the surface variation values are shown in Fig. 15.

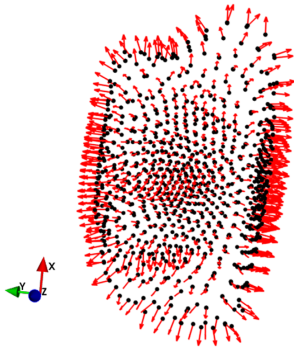


Fig. 14 Normal vectors at each vertex of one grain.

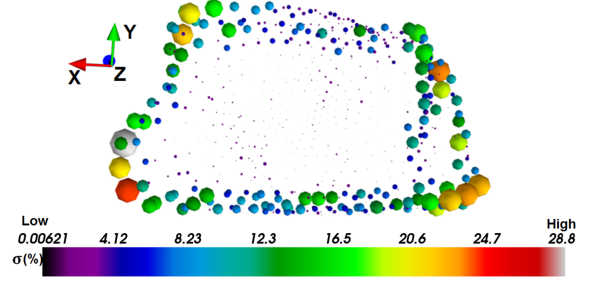


Fig. 15 Surface variation at each vertex of one grain, with values encoded as dot size and color.

3.3.2 Clustering procedure, identification of faces and edges.

The procedure for face detection is based on a clustering algorithm called DBSCAN [47]. A clustering is defining groups of objects (points) sharing the same chosen property. In this case, we chose to define a cluster as a group of neighbouring points with normal vectors satisfying the following condition:

$$\| \vec{n}_i - \vec{n}_j \|^2 = \sum_{k=1}^3 (n_i^k - n_j^k)^2 \leq \epsilon^2 \quad (1 \leq i, j \leq d) \quad (16)$$

The chosen value of ϵ has a significant impact on how the points will be distributed in clusters. An optimal choice, minimizing the sensitivity of the obtained set of clusters to variations of ϵ in some range about the set value, is identified on implementing the HDBSCAN algorithm [48]. The best value is $\epsilon = 0.116$, which corresponds to a solid angle of 0.17 steradians.

The clustering procedure attributes cluster labels to the vertices of each grain. All vertices belonging to the same cluster abide by the criterion defined in Eq. 16. Some points are not labelled, for lack of neighbours sharing the same normal direction within the set tolerance. In order to interpret the geometrical meaning of these clusters we classify them as follows.

F Large faces, clusters with more than 50 vertices: see e.g. large faces 1, 2, 3 and 4 in Fig. 16.

sF Small faces, clusters with less than 50 vertices.

E Edge vertices, defined as non-labelled ones with a high surface variation (we chose the threshold of 5%, the standard deviation of the surface variation distribution of non-labelled points, whose mean is 0.5%)(Fig. 16).

R Remaining points (Fig. 16), which belong to none of the previous categories.

The result of such a clustering procedure enables the detection of faces and edges, as shown in Fig. 16. Fig. 17 also illustrates this classification, showing edges and remaining points U, i.e., all points that do not belong to large or small faces.

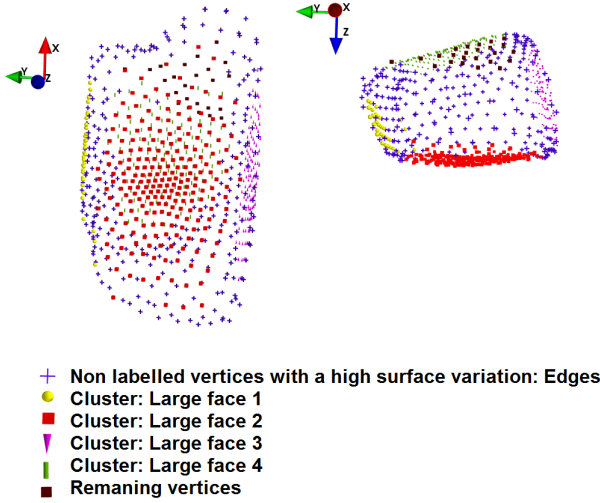


Fig. 16 Clustering of normal directions for one grain (different colours represent different clusters).

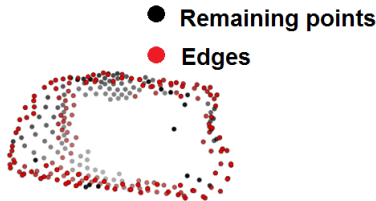


Fig. 17 Edges E and remaining points U on the surface of one grain.

The clustering method has shown its ability to detect faces and edges in the same way we would while looking at a ballast grain. To quantify the obtained results, Fig. 18 shows the distribution of the numbers of large faces observed on both scanned and numerically generated grains. Fig. 19 is a plot of distribution, among the population of grains, of the proportion ζ_{LF_i} of the d vertices belonging to large faces. We clearly observe a good agreement between scanned and generated distributions. The distributions of surface variation are shown in Figs. 20 and 21 for large faces and edges.

The excellent agreement between the distribution functions shows the success of the POD method to generate statistically representative ballast grains with the correct faceted, angular morphology. The method also

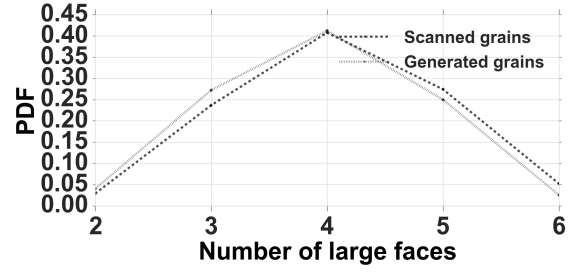


Fig. 18 Statistical distributions of numbers of large faces for scanned and generated grains.

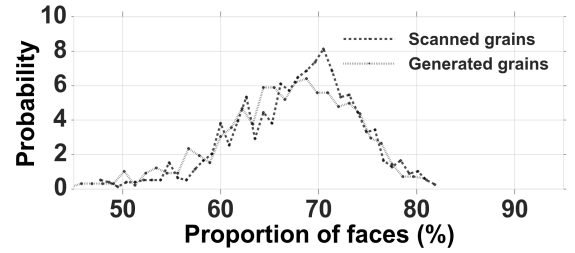


Fig. 19 Statistical distributions of the proportion ζ_{LF} for scanned and generated grains.

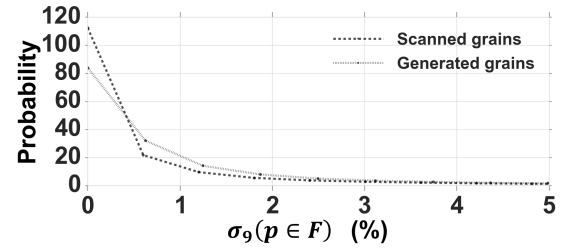


Fig. 20 Statistical distributions of surface variation of large faces for scanned and generated grains.

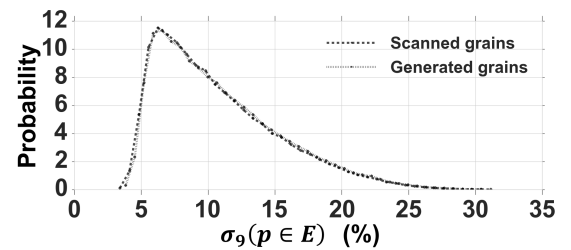


Fig. 21 Distribution of surface variation of edges for scanned and generated grains.

shows its ability to detect faces and edges of a particle even with a high level of shape irregularity.

Next we turn, in a preliminary study, to the scanned and generated grains mechanical properties.

4 Meshing and Numerical Simulations

A necessary step to carry out DEM simulations of the ballast grains as modelled via the POD approach is to mesh them, so that the point cloud is transformed into a polyhedron with a set of faces and edges. We may then implement the DEM numerical platform LMGC90 [49] to simulate mechanical tests.

The meshing procedure should be carried out in a way that modifies the least the morphology of the point cloud. Our practical procedure involves the three following steps. First, we resort to the α -shape method [50], involving a Delaunay triangulation [51], thereby building a mesh on the grain surface, as a set of triangles, whose number is chosen to produce the least error on shape. The second step is a *surface subdivision*, for which the “Loop S3 method” [52] is applied, within the Meshlab software [53], resulting in a more regular mesh with many faces. Finally, the third operation reduces the number of faces to the preset value *via* the “edge collapse” method [54].

Fig. 22 shows how these steps operate on one grain, transforming the initial point cloud into a polyhedron with a number of faces set to 100. Fig. 23 shows how the choice of a number of faces may affect the grain shape.

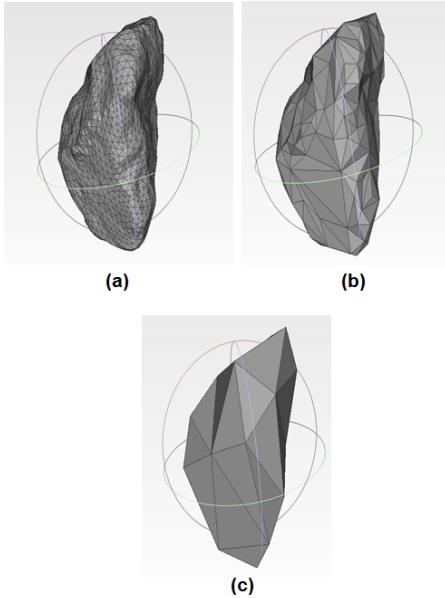


Fig. 23 One grain meshed, as in Fig. 22, using different numbers of faces: (a) 4000, (b) 400, and (c) 40.

In a first stage, the shapes are further simplified as convex polyhedra that correspond to the convex hull of our non-convex meshed polyhedra, which enables much faster computations in terms of contact detection. This

entails a loss of precision on shape, which needs to be quantified. In case a meshing is done with a high number of faces (50000), and thus a better quality, the loss of information (in the sense of the POD error) is about 20% (which would be reduced to less than 5% with non-convex polyhedra) (Fig. 24). In our case, we use only 100 faces, which induces a larger error.

The evolution of the error restricts the relevance of simulations with large numbers of modes. Here we simulate an assembling procedure via one-dimensional compression between parallel planes with periodic lateral boundary conditions, without gravity. It is carried out using 300 grains, with 100 faces and 3 different shape precision levels after convex meshing: 30% (1 mode), 80% (11 modes) and 100% of information (800 modes). The sample is first generated by depositing the spheres circumscribed to the polyhedron (Fig. 25) in a cell with a constant square section ($S = 30 \text{ cm} \times 30 \text{ cm}$). Three values of the inter-granular friction coefficient μ are compared: 0, 0.3 and 0.8. As a reference, a similar simulation is carried out with meshed spheres with 100 faces (and $\mu=0.3$). Contacts with the lower and upper plane walls are frictionless. A constant force of $F_0 = 500 \text{ N}$ is applied to the upper mobile plane until an equilibrium state is approached, in which the contacts resist further compression (Fig. 25). Figs. 26 to 29 show how the solid fraction and the coordination numbers corresponding to simple (vertex-face or edge-edge), double (edge-face) and triple (face-face) contacts respectively denoted as z_1 , z_2 , and z_3 evolve in time towards their final equilibrium values. We note that the force supported by the bottom fixed plane converges to F_0 after 3.8s. The ratio $K = \frac{E_c}{\sigma D^3}$ of the kinetic energy per grain E_c , to the characteristic energy scale σD^3 (σ denoting stress $\frac{F_0}{S}$, and D the average grain diameter) gradually decreases to $4 \cdot 10^{-6}$ in 10 s. Such time scales are related to the initial state density and to grain and upper wall inertia.

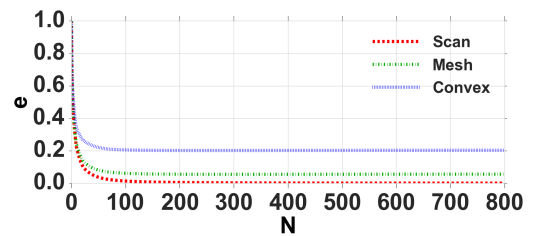


Fig. 24 Meshing effect on POD error.

The final, equilibrium values of solid fraction ρ , of coordination numbers z_1 , z_2 , z_3 are given in Tab. 3, along with those of x_0 , the proportion of “rattlers”

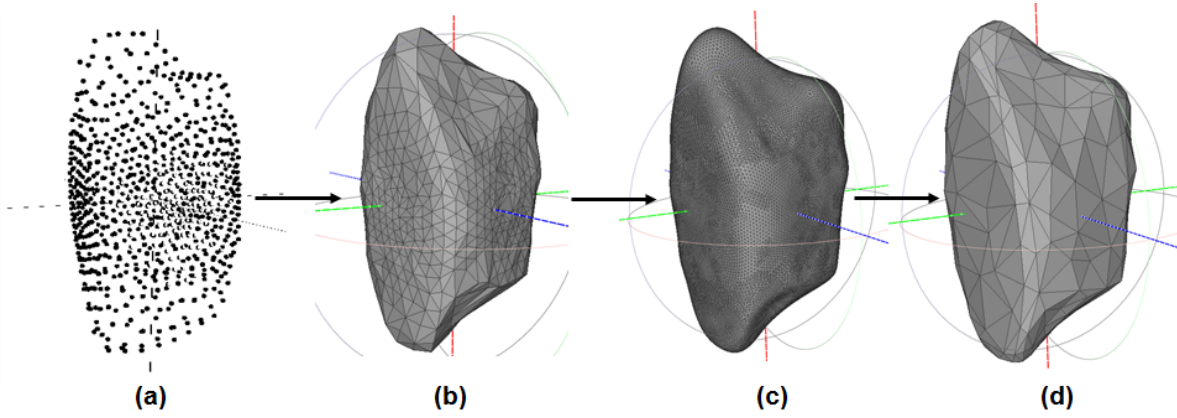


Fig. 22 Meshing steps: (a) point cloud; (b) surface construction; (c) surface subdivision; (d) mesh decimation.

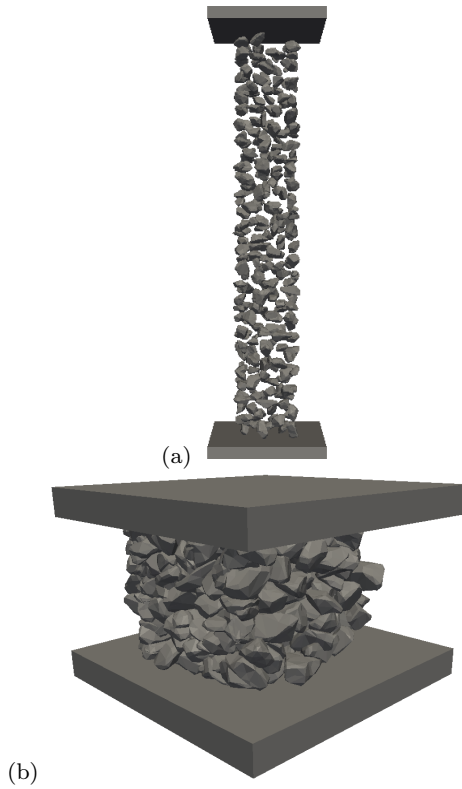


Fig. 25 Compression case: initial (a) and final (b) states for polyhedra.

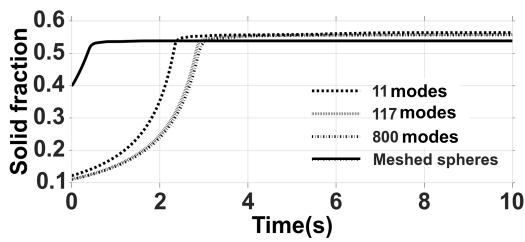


Fig. 26 Solid fraction vs time ($\mu=0.3$).

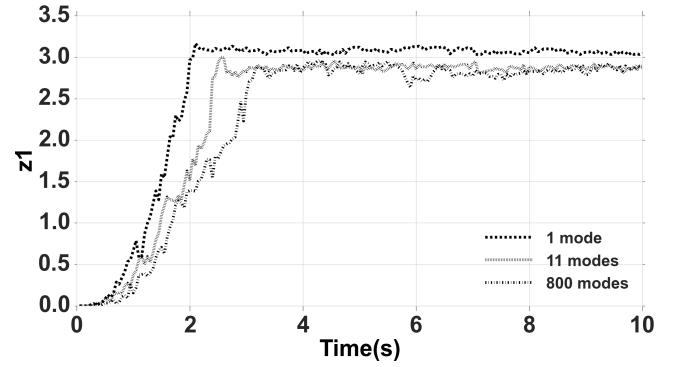


Fig. 27 z_1 vs time ($\mu=0.3$).

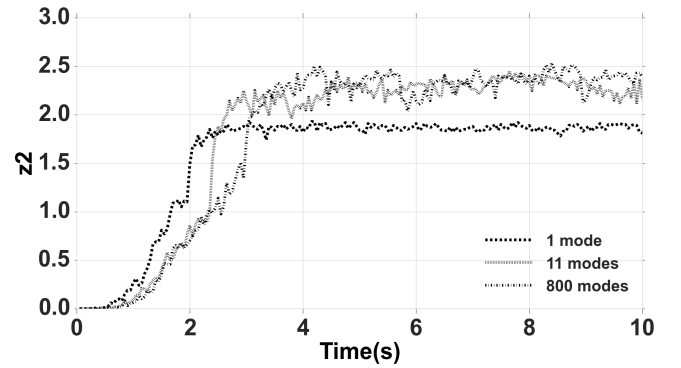


Fig. 28 z_2 vs time ($\mu=0.3$).

(grains carrying no force), for different numbers of modes and friction coefficients.

A slight increase of solid fraction ρ with the number of modes is observed, while z_1 decreases and z_2 and z_3 increase. This effect of a more accurate, and therefore more complex, shape description, could reflect some evolution in the accommodation of steric hindrance. This shape effect is stronger for low friction coefficients.

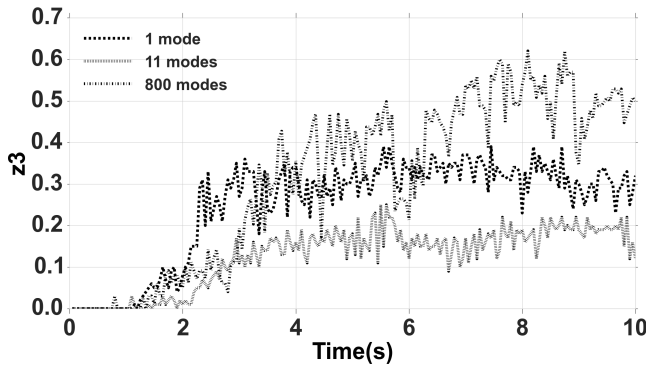


Fig. 29 z_3 vs time ($\mu=0.3$).

Table 3 Final state characteristics for different numbers of modes N and friction coefficients μ .

N	μ	ρ	z_1	z_2	z_3	x_0
11	0	0.690	4.360	3.766	0.320	0%
800	0	0.697	3.877	0.997	0.427	0%
1	0.3	0.590	3.046	0.764	0.213	7.6%
11	0.3	0.594	2.893	1.440	0.280	7.3%
800	0.3	0.602	2.874	1.593	0.332	5.1%
1	0.8	0.566	1.893	0.860	0.080	26.6%
11	0.8	0.567	1.886	0.980	0.080	26.3%
800	0.8	0.567	1.884	0.981	0.094	22.4%

A strong increase of rattler fraction x_0 should also be noted for growing μ , as in spherical bead assemblies.

Fig. 30 plots the distribution of connectivity Z (the fractions of grains having exactly Z contacting neighbours). A significant difference with spheres can be noted, while different shape models yield very similar results: increasing the precision on shape modelling changes the type of contacts without affecting contact numbers.

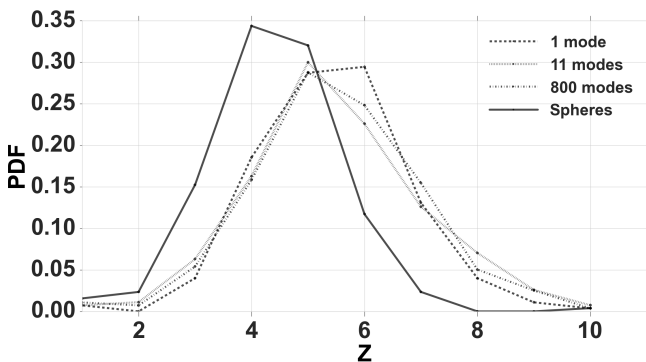


Fig. 30 Connectivity distribution ($\mu=0.3$).

The main outcome of this preliminary mechanical study is thus the observed influence of the accuracy with which the grain shape is modeled on the collective properties of static granular systems assembled by a

compression process and equilibrated under prescribed confining stresses. This influence is stronger for low friction coefficients, and tend to fade out for large μ values.

5 Conclusions.

We propose an innovative method to characterize particle shape, with ballast grains as an application. This method enables shape description with a controlled accuracy, and the generation of new sets of grains, that are morphologically equivalent to real ones, as regards simple geometric characteristics (volume, surface, elongation and aspect ratio) as well as suitably introduced more sophisticated descriptors, such as the number of large faces and the proportion of grain surface area they contain.

Some preliminary numerical simulations, carried out to quantify the impact of particle shape on the mechanical behaviour, reveal that a very significantly reduced number of parameters (11 shape functions) may be sufficient for our model, and also a dominance of friction over particle shape for high friction coefficients. Beyond the simulated assembling stage for a solid granular packing, we expect the shear response to exhibit a higher sensitivity to shape, as observed in [12] for polyhedra with varying numbers of faces inscribed within the same sphere.

Acknowledgements This work has been supported by the Lines, Track & Environment Department in SNCF-Réseau, Engineering & Projects.

References

1. E.T. Selig and John M Waters. *Track geotechnology and substructure management*. Thomas Telford, 1994.
2. P.A. Cundall and O.D.L. Strack. A discrete numerical model for granular assemblies. *Geotechnique*, 29(1):47–65, 1979.
3. F. Radjaï and F. Dubois. *Editors. Discrete-element modeling of granular materials*. Wiley-Iste, 2011.
4. C.M. Riley, W.I. Rose, and G.J.S. Bluth. Quantitative shape measurements of distal volcanic ash. *Journal of Geophysical Research*, 108(10):ECV 8, 2003.
5. G. Vallebuona, K. Arburo, and A. Casali. A procedure to estimate weight particle distributions from area measurements. *Minerals engineering*, 16(4):323–329, 2003.
6. M.A. Taylor, E.J. Garboczi, S.T. Erdogan, and D.W. Fowler. Some properties of irregular 3 D particles. *Powder Technology*, 162(1):1–15, 2006.
7. M. Matzl and M. Schneebei. Stereological measurement of the specific surface area of seasonal snow types: Comparison to other methods, and implications for mm-scale vertical profiling. *Cold Regions Science and Technology*, 64(1):1–8, 2010.

8. G.H. Bagheri, C. Bonadonna, I. Manzella, P. Pontelandolfo, and P. Haas. Dedicated vertical wind tunnel for the study of sedimentation of non-spherical particles. *Review of Scientific Instruments*, 84(5):054501, 2013.
9. W. L. Lim. *Mechanics of railway ballast behaviour*. PhD thesis, University of Nottingham, 2004.
10. E. Azéma, Y. Descantes, N. Roquet, J.N. Roux, and F. Chevoir. Discrete simulation of dense flows of polyhedral grains down a rough inclined plane. *Physical Review E*, 86(3):031303, 2012.
11. E. Azéma. *Etude numérique des matériaux granulaires à grains polyédriques: rhéologie quasi-statique, dynamique vibratoire, application au procédé de bourrage du ballast*. PhD thesis, Université Montpellier 2, 2007.
12. E. Azéma, F. Radjai, and F. Dubois. Packings of irregular polyhedral particles: Strength, structure, and effects of angularity. *Physical Review E*, 87(6):062203, 2013.
13. J. C. Quezada, P. Breul, G. Saussine, and F. Radjai. Penetration test in coarse granular material using contact dynamics method. *Computers and Geotechnics*, 55:248–253, 2014.
14. B. Saint-Cyr, K. Szarf, C. Voivret, E. Azéma, V. Richefeu, J.Y. Delenne, G. Combe, C. Nouguier-Lehon, P. Villard, P. Sornay, et al. Particle shape dependence in 2D granular media. *Europhysics Letters*, 98(4):44008, 2012.
15. J. Yang and X.D. Luo. Exploring the relationship between critical state and particle shape for granular materials. *Journal of the Mechanics and Physics of Solids*, 84:196–213, 2015.
16. E. Azéma, F. Radjai, and G. Saussine. Quasistatic rheology, force transmission and fabric properties of a packing of irregular polyhedral particles. *Mechanics of Materials*, 41(6):729–741, 2009.
17. S.J. Blott and K. Pye. Particle shape: a review and new methods of characterization and classification. *Sedimentology*, 55(1):31–63, 2008.
18. K.C Williams, W. Chen, S. Weeger, and T.J. Donohue. Particle shape characterisation and its application to discrete element modelling. *Particuology*, 12:80–89, 2014.
19. D. Petrak, S. Dietrich, G. Eckardt, and M. Köhler. Two-dimensional particle shape analysis from chord measurements to increase accuracy of particle shape determination. *Powder Technology*, 284:25–31, 2015.
20. G.H. Bagheri, C. Bonadonna, I. Manzella, and P. Vonlanthen. On the characterization of size and shape of irregular particles. *Powder Technology*, 270:141–153, 2015.
21. X. Jia and E.J. Garboczi. Advances in shape measurement in the digital world. *Particuology*, 26:19–31, 2016.
22. Y. Zhang, J.J. Liu, L. Zhang, J.C. De Anda, and X.Z. Wang. Particle shape characterisation and classification using automated microscopy and shape descriptors in batch manufacture of particulate solids. *Particuology*, 24:61–68, 2016.
23. W. Yuan, K.S. Chin, M. Hua, G. Dong, and C. Wang. Shape classification of wear particles by image boundary analysis using machine learning algorithms. *Mechanical Systems and Signal Processing*, 72:346–358, 2016.
24. O. Scheibelhofer, M.O Besenhard, M. Piller, and J.G. Khinast. Comparing particle size distributions of an arbitrary shape. *Powder Technology*, 294:134–145, 2016.
25. G. Mollon and J. Zhao. Generating realistic 3d sand particles using Fourier descriptors. *Granular Matter*, 15(1):95–108, 2013.
26. G. Mollon and J. Zhao. 3d generation of realistic granular samples based on random fields theory and fourier shape descriptors. *Computer Methods in Applied Mechanics and Engineering*, 279:46–65, 2014.
27. E.J. Garboczi. Three-dimensional mathematical analysis of particle shape using x-ray tomography and spherical harmonics: Application to aggregates used in concrete. *Cement and concrete research*, 32(10):1621–1638, 2002.
28. M. Grigoriu, E. Garboczi, and C. Kafali. Spherical harmonic-based random fields for aggregates used in concrete. *Powder Technology*, 166(3):123–138, 2006.
29. V. Venkatraman, L. Sael, and D. Kihara. Potential for protein surface shape analysis using spherical harmonics and 3d zernike descriptors. *Cell biochemistry and biophysics*, 54(1-3):23–32, 2009.
30. B. Zhou and J. Wang. Random generation of natural sand assembly using micro x-ray tomography and spherical harmonics. *Géotechnique Letters*, 5(1):6–11, 2015.
31. B. Zhou, J. Wang, and B. Zhao. Micromorphology characterization and reconstruction of sand particles using micro x-ray tomography and spherical harmonics. *Engineering Geology*, 184:126–137, 2015.
32. volume 252. Elsevier, 2002.
33. R.J. Adrian. Stochastic estimation of the structure of turbulent fields. In J.P. Bonnet, editor, *Eddy structure identification*, pages 145–195. Springer, 1996.
34. R. Lasser. *Introduction to Fourier series*, volume 199. CRC Press, 1996.
35. K. Karhunen. *Zur spektraltheorie stochastischer prozesse*. Suomalainen tiedeakatemia, 1946.
36. M. Loève. *Probability Theory; Foundations, Random Sequences*. New York: D. Van Nostrand Company, 1955.
37. H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.
38. K. Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.
39. K. Baker. Singular value decomposition tutorial. *The Ohio State University*, 24, 2005.
40. Y. Yin, F. Liu, X. Zhou, and Q. Li. An efficient data compression model based on spatial clustering and principal component analysis in wireless sensor networks. *Sensors*, 15(8):19443–19465, 2015.
41. L.A. Albrecht. Curvature-based analysis of point clouds. 2014.
42. X. Liu, E.J. Garboczi, M. Grigoriu, Y. Lu, and S.T. Erdoğan. Spherical harmonic-based random fields based on real particle 3d data: improved numerical algorithm and quantitative comparison to real particles. *Powder Technology*, 207(1):78–86, 2011.
43. C. Garcia. A simple procedure for the comparison of covariance matrices. *BMC evolutionary biology*, 12(1):222, 2012.
44. N. Ouhbi. An accurate description of grain shape for the numerical study of the mechanical behaviour of railway ballast. 2017.
45. Wikipedia. Correlation and dependence — wikipedia, the free encyclopedia, 2017. [Online; accessed 10-April-2017].
46. G. Perrin, C. Soize, and N. Ouhbi. Data-driven kernel representations for sampling with an unknown block dependence structure under correlation constraints. *Journal of Computational Statistics and Data Analysis*, (submitted) 2016.
47. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

48. R.J.G.B. Campello, D. Moulavi, and J. Sander. Density-based clustering based on hierarchical density estimates. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 160–172. Springer, 2013.
49. F. Dubois, M. Jean, M. Renouf, R. Mozul, A. Martin, and M. Bagn  ris. Lmgc90. 10e colloque national en calcul des structures. *Giens, France*, 2011.
50. K. Fischer. Introduction to alpha shapes. *Department of Information and Computing Sciences, Faculty of Science, Utrecht University*, 17, 2000.
51. D.T. Lee and B.J. Schachter. Two algorithms for constructing a delaunay triangulation. *International Journal of Computer & Information Sciences*, 9(3):219–242, 1980.
52. S. Boy  , G. Guennebaud, and C. Schlick. Least squares subdivision surfaces. In *Computer Graphics Forum*, volume 29, pages 2021–2028. Wiley Online Library, 2010.
53. P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, and G. Ranzuglia. Meshlab: an open-source mesh processing tool. In *Eurographics Italian Chapter Conference*, volume 2008, pages 129–136, 2008.
54. L. Kobbelt, S. Campagna, and H.P. Seidel. A general framework for mesh decimation. In *Graphics interface*, volume 98, pages 43–50, 1998.