Article:
Zhang, Y, Liu, JJ, Zhang, L et al. (2 more authors) (2016) Particle Shape Characterisation and Classification using Automated Microscopy and Shape Descriptors in Batch Manufacture of Particulate Solids. Particuology, 24. pp. 61-68. ISSN 1674-2001

https://doi.org/10.1016/j.partic.2014.12.012

Particle Shape Characterisation and Classification using Automated Microscopy and Shape Descriptors in Batch Manufacture of Particulate Solids

Yang Zhang[1], Jing Jing Liu[1], Lei Zhang[2], Jorge Calderon De Anda[3] and Xue Z. Wang[1,3]*

[1]School of Chemistry and Chemical Engineering, South China University of Technology, 381 Wushan Road, Guangzhou City, China 510641

[2]School of Bioscience and Bioengineering, South China University of Technology, Guangzhou Higher Education Mega Centre, Panyu District,Guangzhou, China,510006

[3]Institute of Particle Science and Engineering, School of Chemical and Process Engineering, University of Leeds, Leeds LS2 9JT, UK

* Prof. Xue Z. Wang (corresponding author), School of Chemistry and Chemical Engineering, South China University of Technology, Guangzhou, China 510641 & Institute of Particle Science and Engineering, School of Chemistry and Chemical Engineering, University of Leeds, Leeds LS2 9JT, UK. Email: xuezhongwang@scut.edu.cn, x.z.wang@leeds.ac.uk, Tel: 0086 20 8711 4000.

Abstract

It is known that size alone, as often defined as the volume equivalent diameter, is not sufficient for characterizing many particulate products. The shape of crystalline products can be as important as size in many applications. Traditionally particulate shape is often defined by some simple descriptors such as the maximum length and aspect ratio. Although these descriptors are intuitive, they result in loss of some information of the original shape. This paper presents a method to use principal component analysis (PCA) to derive simple latent shape descriptors from microscope images of particulate products made in batch processes, and the use of the descriptors for identification of batch to batch variations. Data from batch runs of both a laboratory crystalliser and an industrial crystallisation reactor are analysed using the approach. Qualitative and quantitative comparison with the use of traditional shape descriptors that have physical meanings and Fourier shape descriptors is also made.

Keywords: batch to batch variation, classification, principal components analysis, shape descriptors

1 Introduction

Quality assurance of particulate product manufacturing is moving towards using on-line, in-line and at-line process analytical technology (PAT) as a common practice for rapid and detailed characterization of chemical and physical properties of particles in samples. One of the properties for particulate products is shape, not only because it is now recognized that size alone as often defined as volume equivalent diameter is over simplified and sometimes misleading, but also because shape can affect product performance properties such as bioavailability of drug particles. Optical microscopy has been proved to be one of the most effective techniques to determine particle shape and size off-line or on-line (Brittain, 2001; Calderon De Anda, Wang, Lai, & Roberts, 2005; Calderon De Anda, Wang, Lai, Roberts, et al., 2005; Calderon De Anda, Wang, & Roberts, 2005; Larsen, Rawlings, & Ferrier, 2006; Larsen, Rawlings, & Ferrier, 2007; Wang, Calderon De Anda, & Roberts, 2007; Wang, Roberts, & Ma, 2007). The images can be analyzed using image analysis algorithms (Zhang & Lu, 2004) to obtain particle shape information, which is often defined by some descriptors that have physical meanings such as aspect ratio.

Alternative shape description techniques have been investigated in the last two decades, for example, using spectral or functional mathematical approaches such as the Fourier descriptors (Bernard-Michel, Rohani, Pons, Vivier, & Hundal, 1997; Calderon De Anda, Wang, Lai, & Roberts, 2005; Hentschel & Page, 2003; Jarvis, Jefferson, & Parsons, 2005; N. Laitinen, Antikainen, Rantanen, & Yliruusi, 2004; N Laitinen, Osmo, & Yliruusi, 2003; Podczeck, 1997; Pons, Vivier, Delcour, Authelin, & Pailleres-Hubert, 2002; Pourghahramani & Forssberg, 2005; Raj & Cannon, 1999). The main goals are to express or describe as accurate as possible the arbitrary and complex shapes of objects particularly for the purpose of classification, characterization, automated recognition and shape reconstruction. This type of

shape descriptors enables reconstruction of the original shape and hence retains high-frequency components of the shape boundaries. For practical use, they do not require a priori knowledge or a shape to be selected as a reference. Nevertheless, they lack of physical meanings of the particles described and therefore they can be considered as "latent". However, research has not yet been stressed for developing techniques that can methodologically arise appropriate shape descriptors for a given sample or series of samples immediately leading to obtain morphological information to improve manufacturing. The shapes of particles of different materials produced in laboratorial and industrial scales are often regarded as being arbitrary. Despite this degree of arbitrariness, previous research has demonstrated that the ratios of various particle size measurements can be used to derive shape descriptors of different orders (Hentschel & Page, 2003; Pourghahramani & Forssberg, 2005). However, the challenge lies in determining the combinations of size measurements that best provide the relevant morphological information of a sample or some samples.

In industrial manufacturing of particulate materials, techniques for particle shape description based on physical morphological measurements, shape factors or aspect ratios have been preferred over other shape description techniques. Some of the reasons for this tendency include the scarcity of possibilities to describe and quantify the shape of particles with indices that can retain morphological information and at the same time provide a meaningful shape indication for a direct physical understanding of the particle-process relationship. A major drawback for physical descriptors is the loss of morphological information, and hence the original shape cannot be reconstructed from this type of descriptors. Furthermore, their selection is traditionally made through a priori knowledge of particle shapes in the samples concerned, for example, selecting a convexity descriptor if the sample contains agglomerates. Nevertheless, such kind of knowledge may be difficult to obtain due to the variety of arbitrary shapes in samples. In this work, we investigate the use of principal component analysis to process a set of new shape descriptors calculated from the combination of different physical

size measurements of the particle contours. Concretely, the raw data of numerous physical shape descriptors derived from the combination of physical size measurements such as Length, Width and Maximum Distance has been reduced to a lower-dimensionality data using principal components analysis; the reduced lower-dimensionality data of physical shape descriptors are to be used for effective classification of polymorphic crystals with comparable results using latent Fourier descriptors; the raw data of numerous physical descriptors are used to identify batch to batch variations based on classification results of the reduced lower-dimensionality data. The paper is organized as follows: the shape descriptors concerned for characterization and classification of polymorphic crystals are introduced briefly in Section 2; then a standard sample of 55 imaged crystals is adopted for the comparable test of the proposed approach in Section 3; the addressed approach is further applied to two industrial examples for interpretable classification of various batches with different operation conditions in Section 4 and Section 5, respectively; and finally, some conclusions are drawn in Section 6.

2 Shape Descriptors for Classification

Shape descriptors are to describe the crystals individually so as to distinguish them in a mathematical way. An automated vision system for analysis of the size and shape of particles, the Pharma Vision System 830 (PVS 830) from Malvern Instruments (Malvern-Instruments-Ltd) is used in this study to measure the physical size dimensions of the particles. The samples are homogeneously dispersed onto a glass slide using a special sample preparation device based on pressurized air. Thus particle orientations on the glass slide can be assumed to be statistically random, which is beneficial for the following imaged-based measurement. The glass slide is placed under a CCD camera that scans the sample for particle image acquisition. Particles are automatically segmented from their background by embedded computer software for obtaining the individual particle images and a variety of imaged-based

measurements according to the inherent configuration of the system. Typical particle shape and size measurements as well as their physical meanings are shown in Table

Table 1

2.1 Physical Descriptors

Basic shape measurements from the PVS 830 contains 3 descriptors, namely, Roundness, Contour/Area and Convexity, whilst the majority of other measurements are associated to particle size, i.e., Mean Diameter, Diameter, Length, Width, Maximum Distance, Area and Volume. Given that the particle size can be expressed in various ways according to the morphological dimensions of the particles such as length, breadth, volume or maximum distance, a large set of physical shape descriptors with various orders can be created by the arithmetic combination of the size measurements. Each of such combinations is sensitive to a specific attribute of shape according to the size measurements selected. Based on this principle, the size measurements obtained by the PVS 830 were combined as follows to create 85 new shape descriptors (Hentschel & Page, 2003):

$$\{\frac{V}{L^3};\ \frac{A}{L^2};\ \frac{L_1}{L_2};\ \frac{L_1L_2}{L_3^2};\ \frac{L_1L_2}{L_3L_4};\ \frac{V}{L_1L_2L_3};\ \frac{A}{L_1L_2}\}, \quad (1)$$

where V represents the particle volume calculated from the area and mean diameter, A is the particle area and L represents the different particle size measurement, i.e., Length, Width, Diameter, Mean Diameter and Maximum Distance. In combination with Roundness, Contour/Area and Convexity, a total of 88 physical shape descriptors were derived for each particle measured.

2.2 Latent Physical Descriptors

Physical shape descriptors deduced from the combinations of basic measurements directly could be numerous. Such numerous physical shape descriptors are in fact highly correlated.

The correlated high-dimensionality data can be reduced to a lower-dimensionality data without much loss of information using principal components analysis (PCA).

PCA is an established methodology well documented in text books therefore will only be briefly described here. PCA uses a linear mathematical algorithm to derive a new set of variables, called principal components (PCs), from the original measured variables in such a way that the new set of variables are no longer correlated. The new set of variables, the principal components (PCs), each is the linear combination of the original variables, are latent variables. Often the first few PCs can capture the majority variation in the data, so only the first few PCs are needed in further processing of the data, such as in classification.

Every PC is the linear addition of the original variables. The principal components are also called scores, while the weight that quantifies the linear contribution of each original variable to a PC is often known as loading. It can be interesting to find the relative contribution of the original variables to a specific principal component. This can be analyzed by plotting the loadings. Since the original variables are scaled before PCA is applied, the larger the loading, the more contribution that original variable makes.

The obtained lower-dimensionality data after PCA analysis can be regarded as a kind of latent physical descriptors since they correspond to the new coordinates of the space spanned by the original physical shape descriptors. There were various attempts to build the missing link between the original data and the reduced lower-dimensionality data so as to identify the cause of abnormity based on the reduced lower-dimensionality data in literature (Kourti & MacGregor, 1996; Miller, Swanson, & Heckler, 1998). However, these results were mainly instructive without intuitive interpretation and guaranteed generality. Here, PCA is also applied to reduce the dimensionality of the original highly-correlated physical descriptors for the classification based on the reduced lower-dimensionality data while the contribution of every original physical descriptor for the resulting classification of two clusters is estimated

based on the original data and the classification result, which is calculated by the following equation:

$$\varphi_i = \frac{\left|\overline{C_i^1} - \overline{C_i^2}\right|}{\max_j d_{ij}}, \quad i = 1,2,\cdots,n, \tag{2}$$

where φ_i is the estimated contribution of the i th physical descriptor; $\overline{C_i^1}$ and $\overline{C_i^2}$ are the mean values of the i th physical descriptor for the first classified cluster and the second classified cluster, respectively; d_{ij} is the i th physical descriptor for the j th particle in the sample while j is the index of the individual particle in the sample. The estimated contribution can be intuitively interpreted as the original physical descriptor contributes most to the resulting classification. The calculated values of φ_i can be normalized to obtain a contribution fraction of each variable in the dataset.

2.3 Latent Fourier Descriptors

A further comparison for shape classification was carried out using latent Fourier descriptors. Fourier descriptors for particle shape have been the subject of several previous studies (Zhang & Lu, 2004). Unlike physical shape factors, Fourier descriptors can be considered as a kind of latent descriptors, i.e., they do not provide a direct physical meaning of the described object although they do provide means to retain the morphological information and the capability to reconstruct the shape. In order to obtain the descriptors, a chain code was applied to obtain the coordinates of the points (x_t, y_t) composing the shape boundary to derive the shape signatures. In previous studies on classification of particles produced in crystallization processes, we have explained the details of the procedure to calculate the latent Fourier descriptors from the contour coordinate points (Calderon De Anda, Wang, Lai, & Roberts, 2005). Here, we focus on comparing the performance of the latent Fourier descriptors for classification of polymorphic crystals against the same task using physical shape descriptors.

2.4 Particle Classification

Polymorphic crystals can be classified according to their respective shape descriptors by means of unsupervised classification methods such as Kohonen neural networks and Fuzzy C-means clustering. Here, fuzzy C-means clustering is used for the classification task because the differences of imaged polymorphic crystals are sometimes ambiguous and thus the grade of membership returned by fuzzy C-means clustering provides additional insightful information for the classification results. As a method of clustering which allows one piece of data to belong to two or more clusters, fuzzy C-means is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m}\left\|x_i - c_j\right\|^2, \quad m \geq 1, \tag{3}$$

where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the ith of d-dimensional measured data, c_j is the d-dimension centre of the cluster, and $\|*\|$ is any norm expressing the similarity between any measured data and the centre. Fuzzy partitioning is carried out through an iterative optimization of the objective function with the update of membership u_{ij} and the cluster centres c_j by:

$$u_{ij} = 1 \Big/ \sum_{k=1}^{C} \left(\frac{\left\|x_i - c_j\right\|}{\left\|x_i - c_k\right\|}\right)^{\frac{2}{m-1}}, \quad c_j = \sum_{i=1}^{N} u_{ij}^{m} \cdot x_i \Big/ \sum_{i=1}^{N} u_{ij}^{m}. \tag{4}$$

This iteration is to stop when $\max_{ij}\left\{\left|u_{ij}^{(k+1)} - u_{ij}^{(k)}\right|\right\} < \varepsilon$, where ε is a termination criterion between 0 and 1, whereas k are the iteration steps. This procedure converges to a local minimum or a saddle point of the object function (Xie & Beni, 1991).

3 Classification Results of a Test Sample

In order to test the performance of various descriptors for shape classification using Fuzzy C-means clustering, a standard sample of 55 imaged crystals with known morphology was selected for the test. The 55 imaged crystals obtained from the PVS 830 are shown in Figure 1, where both prismatic and needle-shaped particles are contained. For the classification of the selected 55 imaged crystals, three cases are to be compared:

- Using 64 latent Fourier descriptors, the 55 imaged crystals are to be classified into two clusters using Fuzzy C-means clustering.
- Using normalized 88 physical descriptors, the 55 imaged crystals are also to be classified into two clusters using Fuzzy C-means clustering.
- Using 3 latent physical descriptors derived from PCA with the overall 90% contribution, the 55 imaged crystals are to be classified into two clusters using Fuzzy C-means clustering as well.

Figure 1

The classification results of the above three cases are listed in Table 2 and it turns out that the classification using only 3 latent physical descriptors is similar to the classification using normalized 88 physical descriptors and also effective compared to the benchmark of the classification using 64 latent Fourier descriptors since the boundary between prismatic and needle shape is quite ambiguous for some imaged crystals. The classification results demonstrate that the reduced lower-dimensionality data using PCA maintain most of morphological information of these 55 imaged crystals. The contribution of every original physical descriptor for the resulting classification using 3 latent physical descriptors, which is shown in Figure 2 with corresponding grade of membership, is estimated according to (2) and the hierarchical list of the first 10 physical descriptors relative to their contribution order is shown in Table 3. It can be seen that roundness is the most important physical descriptor that

contributes most for the resulting classification of these 55 imaged crystals, which is consistent to the intuitive observation of these 55 imaged crystals.

Table 2

Figure 2

Table 3

4 Application to a Real Process of Laboratory Scale --- Crystallization Samples of Different Polymorphic Forms of L-glutamic Acid

L-glutamic acid is known to have two polymorphic forms, α and β, when it crystallizes from aqueous solutions (Calderon De Anda, Wang, Lai, Roberts, et al., 2005; Kitamura, 1989). Each of the polymorphs exhibits a characteristic crystal habit, the alpha form being prismatic and the beta form being needle-shaped. Although the beta form is the thermodynamically stable form, the crystallization of the alpha form may be achieved by rapid cooling of aqueous solutions of L-glutamic acid. The alpha form can be kept indefinitely if the crystals in this form are separated from the mother liquor. Without separation, transformation is observed into the beta form with dissolution of alpha crystals and growth of beta crystals.

This example studies batch to batch variations related to the cooling rate through characterizing and classifying polymorphic crystals sampled from three different laboratorial crystallization batches of L-glutamic acid using latent physical descriptors and Fuzzy C-means clustering:

- Two samples of crystals, LGA A01 and LGA A02, were obtained from two batch crystallization runs carried out by cooling in a laboratory 20L scale. The solution was cooled from 70^{o}C down to 10^{o}C using a cooling rate of 0.5 ^{o}C/min.

- One sample, LGAAB001, was obtained from a third crystallization run. For the crystallization experiment, the solution was also cooled from 70^{o}C down to 10^{o}C, in this case with a cooling rate of 0.2 ^{o}C/min in the same reactor.

For each experiment, the crystals obtained at the end of the experiment were filtered and dried. The dried particles were characterized by image analysis using the PVS 830 instrument. We investigated the possibility to identify the sample for the batch with slower cooling rate and the main effect of the cooling rate on the morphology of the particles in the two batches through descriptor selection. Example frames obtained by the PVS 830 for each sample are shown in Figure 3. The raw data of 88 physical descriptors have been reduced to a lower-dimensionality data using PCA for the following classification. The classification result is shown in Figure 4, where all the crystals in these three batches are classified into two clusters. The contribution of every original physical descriptor is estimated according to the classification result and the hierarchical list of the first 10 physical descriptors is shown in Table 4. It is found that the roundness descriptor also provides the largest difference in shape between the batches due to the change of the cooling rate. At the same time, once the descriptor Roundness was pointed out, its distribution for each sample shows the evident difference of the sample in the batch with slower cooling rate compared to the other two samples, as shown in Figure 5. In a new examination of the two samples under microscope, it was found that for the batch with faster cooling rate most of the crystals were of prismatic form suggesting the predominance of the α form, whereas for the slower cooling rate, a mixture of prismatic and needle shapes were observed providing indication of the existence of the two polymorphic forms.

Figure 3 Figure 4 Table 4 Figure 5

5 Application to an Industrial Scale Process - Crystallization Batches Produced in a Pilot Plant

Crystallization is an important operation in the production of various fine chemicals including pharmaceutical products. During the process, crystals are nucleated and grown from a solution beyond the equilibrium concentration, i.e., a supersaturated solution. The degree of supersaturation, S, at which the formed crystals are exposed, constitutes the main driving force for the growth of their individual external crystal faces. Due to differences in surface chemistry, different faces normally present different growth rates and the face with the slowest growth rate critically defines the morphology of the produced particles. For these reasons, supersaturation is an essential parameter in crystallization for manipulating particulate shape and size. Particularly, it is desirable to control supersaturation in order to produce crystals with desired and homogeneous physical shape properties.

This example compares the samples obtained from two different crystallization batches of L-glutamic acid produced in an industrial-scale pilot plant at Switzerland (Al-Ghafran, et al., 2005). The two batches were carried out in a 250L batch reactor under supersaturation control, one batch with supersaturation, S=1.1, and the other with S=1.2. The samples were obtained towards the end of the run from the reactor using a long sample stick to scoop up the crystals from the solution. The crystals were filtered and dried. The two batches produced the stable polymorph, β-form, hence the characteristic needle shape of the polymorph was observed in both runs. Nevertheless, it is of interest to know the morphological differences between the needles as a consequence of the supersaturation effect on the growth of the crystal faces, which may be difficult to detect visually from visual inspection of the samples. Snapshots of the two samples obtained by the PVS 830 are shown in Figure 6. The two samples were analyzed using image analysis software of the PVS 830. The samples of S=1.1, lgaS1101, contained 7511 particles measured. The sample of S=1.2, lgaS12, contained 20000 particles

measured. The later sample was subdivided into three sub-samples in order to make similar the number of particles among the samples for comparison, i.e., the new samples are lgaS1201 (6500 particles), lgaS1202 (6500 particles) and lgaS1203 (7000 particles), respectively.

Figure 6

The four sample datasets were analyzed using PCA for the classification based on the lower-dimensionality data. The classification result is shown in Figure 7. Table 5 displays the top 10 descriptors that are indicated as the most variable. The output indicated that the main difference in particle shape between the two batches lies in the roundness of the crystals. The supersaturation of S=1.1 led to the production of longer needles reflecting by lower values of roundness. Figure 8 displays the distribution of roundness for the four sample datasets analyzed, where it is clearly observed that the single sample distribution for S=1.1 corresponds to particles with lower roundness compared to the fairly similar three sample datasets for S = 1.2. Reviewing the hierarchical list of all the physical descriptors, it was found that the descriptor that appeared to have the least difference between the batches is the ratio (Length/MaximumDistance) that appeared at the bottom of the hierarchical list, which is not shown in Table 5. It is not surprising for such result because of the close values of these two size measurements for the prismatic and needle-shaped crystals.

Figure 7 Table 5 Figure 8

6 Final Remarks

The investigation on particle shape descriptors is not for the purpose of reducing computational time, since with the current technology of microelectronics, computation time

and memory are never a burden in particle characterization anymore. Instead, the motivation for investigating new particle shape descriptors is that simple and small number of descriptors are preferred as long as they can capture the same amount of information about the shape of the particles as more complicated ones such as Fourier descriptors, because the former are more intuitive and interpretable than the later.

. In this work, we have studied an approach to achieve interpretable classification of polymorphic organic crystals from samples or series of samples for identification of batch to batch variations. The approach is based on the derivation of new shape descriptors produced from multiple combinations of physical size measurements and the application of principal components analysis to reduce the dimensionality of the obtained data. The contribution of every original physical descriptor for the classification result using the reduced lower-dimensionality data is estimated intuitively. A standard sample of 55 crystals has been adopted for testing the consistency of the approach. The application of the method has been further demonstrated in two practical examples of crystallization processes at laboratory and industrial scale for assisting in discovering sample variations and the main affects of operating conditions on the batches. It was concluded that both physical descriptors derived from the combination of size measurements and the latent Fourier descriptors derived from boundary information of imaged crystals provide high sensitivity and promising means to classify large samples with arbitrary shapes while the advantages of the first type over the latent Fourier descriptors being the ability to associate batch to batch variations on physical shape properties with operational conditions.

Some other imaging instruments can provide more descriptors such as smoothness, compactness, ellipticity, rectangularity, polygon order, interior angle, fiber curl, aspect ratio, surface uniformity, as well as texture information such as roughness. These descriptors can often be correlated which can lead to poor classification result. . The method described in this

paper using principal component analysis can also be useful to process the data to remove the correlations prior to classification of particle shape.

The work described in this paper is based on 2D imaging. Clearly more accurate particle shape information can be captured by 3D imaging such as the 3D process imaging work described in literature(Wang, Roberts, et al., 2007). Due to the particular shapes investigated in this work, 2D imaging does not cause large errors. There is also research work going on that is aimed at developing 3D microscopes. For 3D images, the method investigated in this paper, the principal component analysis can still be used.

Acknowledgements

Financial support from the China Thousand Talents Scheme is acknowledged. The work has also benefited from early research supported by UK Engineering and Physical Sciences Research Council (EPSRC) for the projects of Shape (EP/C009541) and Stereo Vision (EP/E045707), as well as from Malvern Instruments Ltd (project name Vision) .

References

Al-Ghafran, M., Andrews, J., Dallin, P., Gibson, N., Grieve, B., Hall, A., Khan, S., Ma, C., Mahmud, T., Morris, J., Ozkan, L., Penchev, R., Price, C., & Roberts, K. (2005). Crystallisation of organic compounds at supersaturation close loop control in a 250 litre industrial pilot-scale batch reactor. Proceedings of the 7th World Congress of Chemical Engineering, Glasgow, UK.

Bernard-Michel, B., Rohani, S., Pons, M. N., Vivier, H., & Hundal, H. S. (1997). Classification of crystal shape using Fourier descriptors and mathematical morphology. Particle & Particle Systems Characterization, 14, 193-200.

Brittain, H. G. (2001). What is the "correct" method to use for particle-size determination? Pharmaceutical Technology North America, 25, 96-97.

Calderon De Anda, J., Wang, X. Z., Lai, X., & Roberts, K. J. (2005). Classifying organic crystals via in-process image analysis and the use of monitoring charts to follow polymorphic and morphological changes. Journal of Process Control, 15, 785-797.

Calderon De Anda, J., Wang, X. Z., Lai, X., Roberts, K. J., Jennings, K. H., Wilkinson, M. J., Watson, D., & Roberts, D. (2005). Real-time product morphology monitoring in crystallization using imaging technique. AIChE Journal, 51, 1406-1414.

Calderon De Anda, J., Wang, X. Z., & Roberts, K. J. (2005). Multi-scale segmentation image analysis for the in-process monitoring of particle shape with batch crystallisers. Chemical Engineering Science, 60, 1053-1065.

Hentschel, M. L., & Page, N. W. (2003). Selection of descriptors for particle shape characterization. Particle & Particle Systems Characterization, 20, 25-38.

Jarvis, P., Jefferson, B., & Parsons, S. (2005). Measuring floc structural characteristics. Reviews in Environmental Science and Biotechnology, 4, 1 - 18.

Kitamura, M. (1989). Polymorphism in the Crystallization of L-Glutamic Acid. Journal of Crystal Growth, 96, 541-546.

Kourti, T., & MacGregor, J. F. (1996). Multivariate SPC methods for process and product monitoring. Journal of Quality Technology, 28, 409-428.

Laitinen, N., Antikainen, O., Rantanen, J., & Yliruusi, J. (2004). New perspectives for visual characterization of pharmaceutical solids. Journal of Pharmaceutical Sciences, 93, 165-176.

Laitinen, N., Osmo, A., & Yliruusi, J. (2003). Characterization of particle sizes in bulk pharmaceutical solids using digital image information. AAPS PharmSciTech, 4, 383 - 391.

Larsen, P. A., Rawlings, J. B., & Ferrier, N. J. (2006). An algorithm for analyzing noisy, in situ images of high-aspect-ratio crystals to monitor particle size distribution. Chemical Engineering Science, 61, 5236-5248.

Larsen, P. A., Rawlings, J. B., & Ferrier, N. J. (2007). Model-based object recognition to measure crystal size and shape distributions from in situ video images. Chemical Engineering Science, 62, 1430-1441.

Malvern-Instruments-Ltd. http://www.malvern.co.uk.

Miller, P., Swanson, R. E., & Heckler, C. E. (1998). Contribution plots: a missing link in multivariate quality control. Applied Mathematics and Computer Science, 8, 775-792.

Podczeck, F. (1997). A shape factor to assess the shape of particles using image analysis. Powder Technology, 93, 47-53.

Pons, M. N., Vivier, H., Delcour, V., Authelin, J. R., & Pailleres-Hubert, L. (2002). Morphological analysis of pharmaceutical powders. Powder Technology, 128, 276-286.

Pourghahramani, P., & Forssberg, E. (2005). Review of applied particle shape descriptors and produced particle shapes in grinding environments. Part I: Particle shape descriptors. Mineral Processing and Extractive Metallurgy Review, 26, 145-166.

Raj, P. M., & Cannon, W. R. (1999). 2-D particle shape averaging and comparison using Fourier descriptors. Powder Technology, 104, 180-189.

Wang, X. Z., Calderon De Anda, J., & Roberts, K. J. (2007). Real-time measurement of the growth rates of individual crystal facets using imaging and image analysis: a feasibility study on needle-shaped crystals of L-glutamic acid. Chemical Engineering Research & Design, 85, 921-927.

Wang, X. Z., Roberts, K. J., & Ma, C. Y. (2007). Crystal growth measurement using 2D and 3D imaging and the persepctives for shape control. Chemical Engineering Science(2007), 65, 1173-1184.

Xie, X. L., & Beni, G. (1991). A validity measure for fuzzy clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13, 841-847.

Zhang, D., & Lu, G. (2004). Review of shape representation and description techniques. Pattern Recognition, 37, 1-19.

Table 1: Crystal morphological measurements by the PVS 830

Parameter	Meaning
Mean Diameter(μm)	The radius is measured in steps of 3° and the mean is calculated.
Diameter(μm)	The diameter of a circle with the same area.
Roundness (-)	Roundness is a measurement of the length to width relationship, with a value in the range (0 1), 1 for a perfect circle and close to 0 for a very narrow elongated object.
Maximum Distance (μm)	The maximum possible distance between two points in the contour of the shape.
Length (μm)	The maximum projected distance on the major axis.
Width (μm)	The maximum projected distance on the minor axis.
Area (μm^2)	The visual projected area.
Contour/Area (1/μm)	The particle perimeter divided by the area.
Convexity (-)	The convexity number illustrates particle roughness in a range (0 1), with low values for highly rough particles.
Volume (μm^3)	The volume calculated from the area and mean diameter.
Intensity (-)	The mean object intensity in the range (0 – 100).

Table 2: Classification results of 55 imaged crystals using various descriptors

Descriptors / Shapes	64 latent Fourier Descriptors	Normalized 88 latent physical descriptors	3 latent physical Descriptors
Prismatic shape	33 members: 1-21,23,26,28-30,35,36,38,40,45,48,52	25 members: 1-5,7-21,29,40,46,48	27 members: 1-5,7-23,29,40,46,48,52
Needle-shape	22 members: 22,24,25,27,31-34,37,39,41-44,46,47,49-51,53-55	30 members: 6,22,24-28,30-39,41-45,47,49-55	28 members: 6,24-28,30-39,41-45,47,49-51,53-55

Table 3: The order of contribution among 88 physical descriptors for the classified two forms of crystals

Order	Descriptor ID	Descriptor definition
1	D88	Roundness
2	D37	Diameter*Width/(MaximumDisdance)^2
3	D28	MeanDiameter*Width/(MaximumDistance)^2
4	D38	Diameter*Width/(Lengh)^2
5	D29	MeanDiameter*Width/(Length)^2
6	D3	Volume/(MaximumDistance)^3
7	D18	Width/MaximumDistance
8	D50	Width*Length/(MaximumDistance)^2
9	D19	Width/Length
10	D36	Diameter*Width/(MeanDiameter)^2

Table 4: The order of contribution among 88 physical descriptors for the three laboratorial batches

Order	Descriptor ID	Descriptor definition
1	D88	Roundness
2	D37	Diameter*Width/(MaximumDisdance)^2
3	D28	MeanDiameter*Width/(MaximumDistance)^2
4	D18	Width/MaximumDistance
5	D3	Volume/(MaximumDistance)^3
6	D50	Width*Length/(MaximumDistance)^2
7	D8	Area/(MaximumDistance)^2
8	D21	MeanDiameter*Diameter/(MaximumDistance)^2
9	D19	Width/Length
10	D36	Diameter*Width/(MeanDiameter)^2

Table 5: The order of contribution among 88 physical descriptors for the two Pilot Plant batches

Order	Descriptor ID	Descriptor definition
1	D88	Roundness
2	D37	Diameter*Width/(MaximumDistance)^2
3	D28	MeanDiameter*Width/(MaximumDistance)^2
4	D18	Width/MaximumDistance
5	D50	Width*Length/(MaximumDistance)^2
6	D3	Volume/(MaximumDistance)^3
7	D8	Area/(MaximumDistance)^2
8	D21	MeanDiameter*Diameter/(MaximumDistance)^2
9	D52	MeanDiameter*Diameter/(MaximumDiameter*Length)
10	D36	Diameter*Width/(MeanDiameter)^2

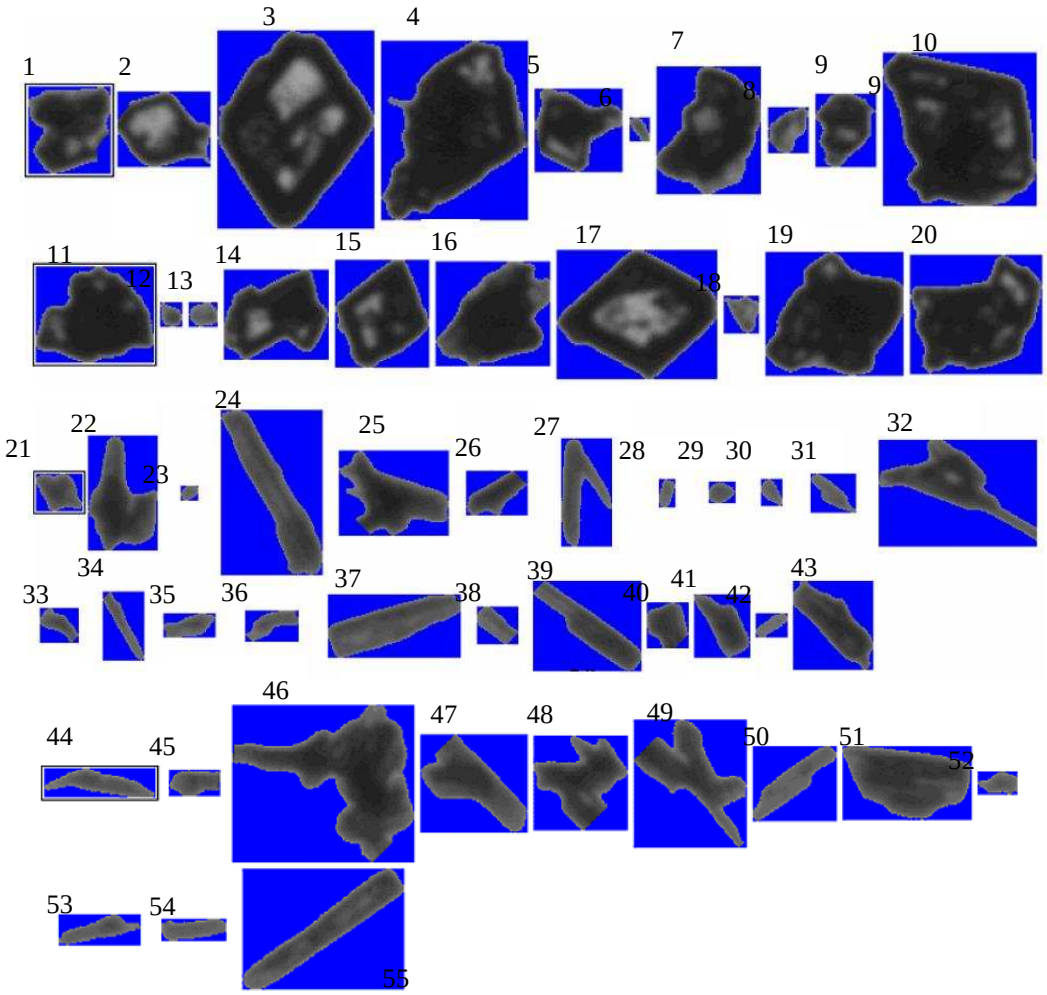


Fig. 1: Particle images and their numbering used for classification

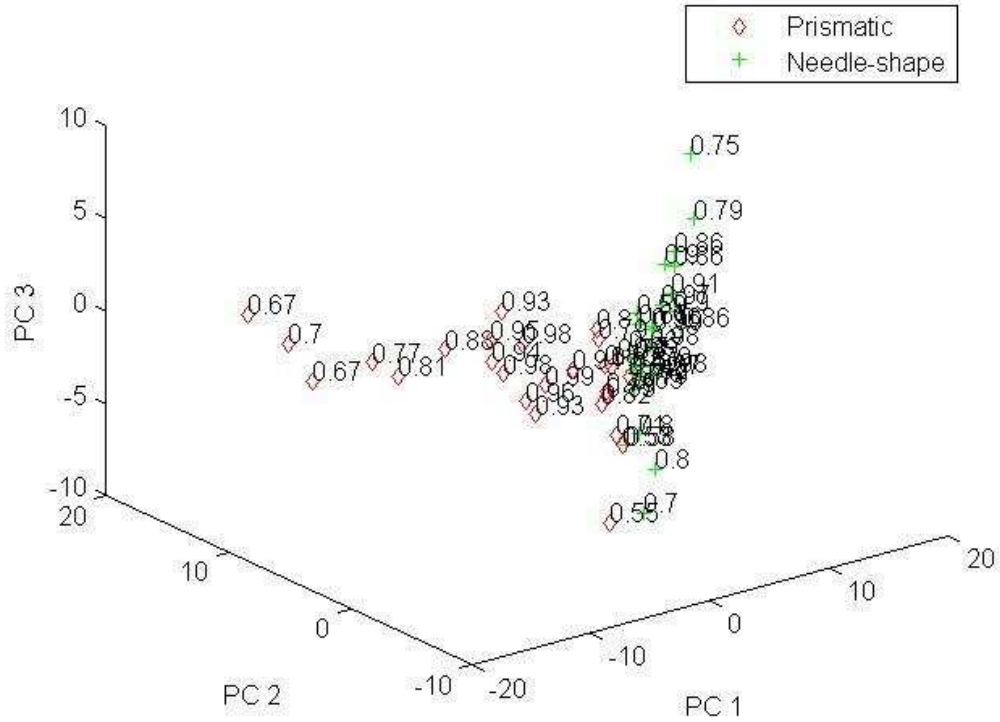
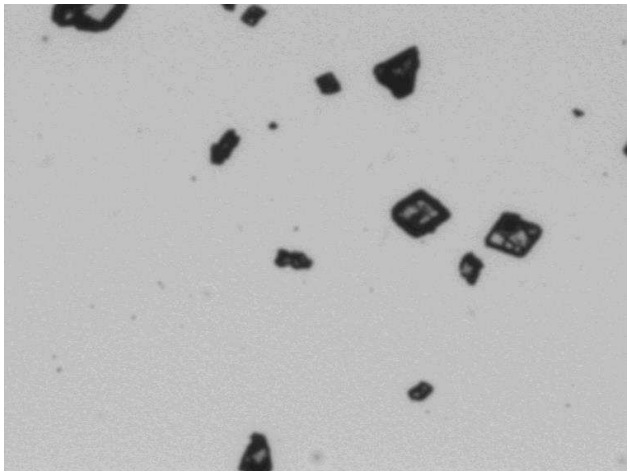
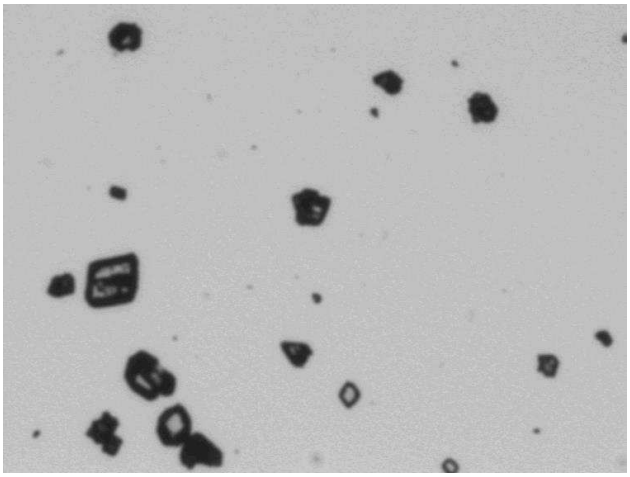


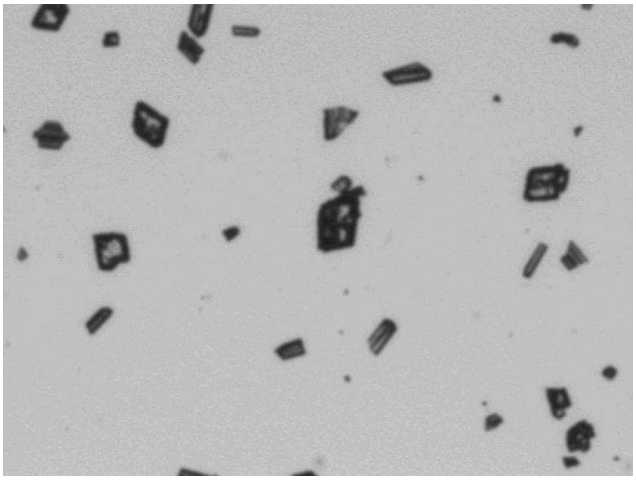
Fig.2: The classified two clusters using 3 latent physical descriptors with the grade of membership



(a)



(b)



(c)

Fig.3: Sample frames of the three sets: (a) LGA A01, (b) LGA A02 containing α-form particles, and (c) LGAAB001 containing both α-form and β-form of L-glutamic acid

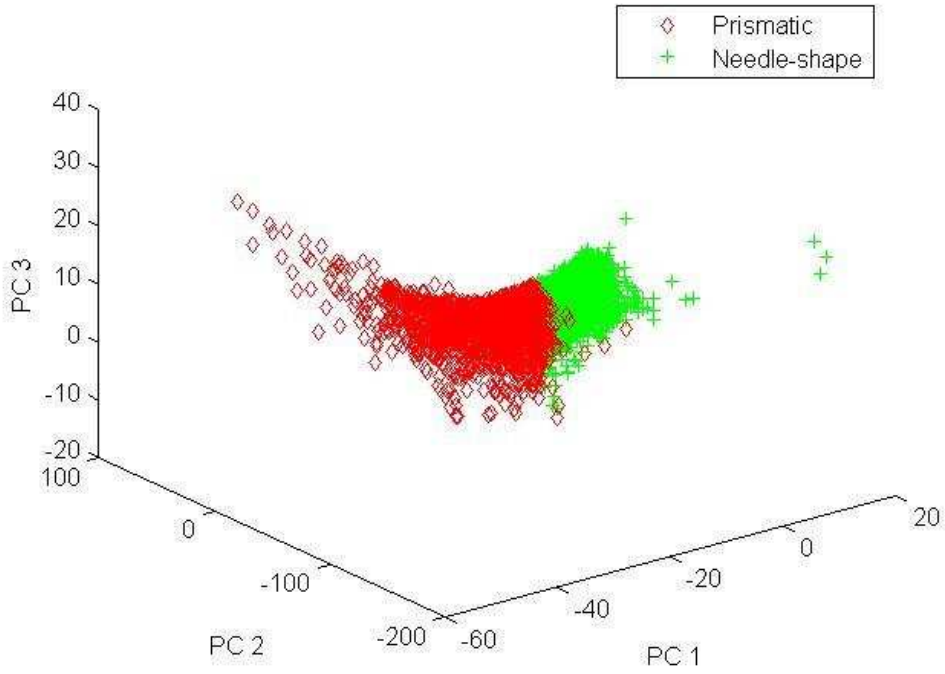


Fig. 4: The classified two clusters using 3 latent physical descriptors for the three laboratorial batches

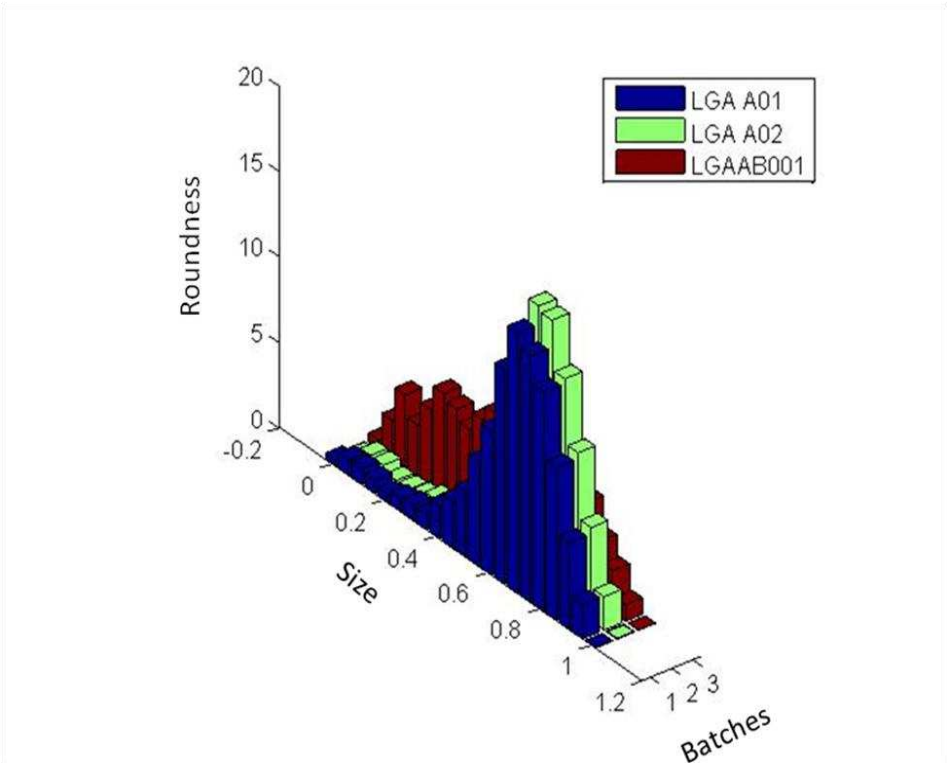
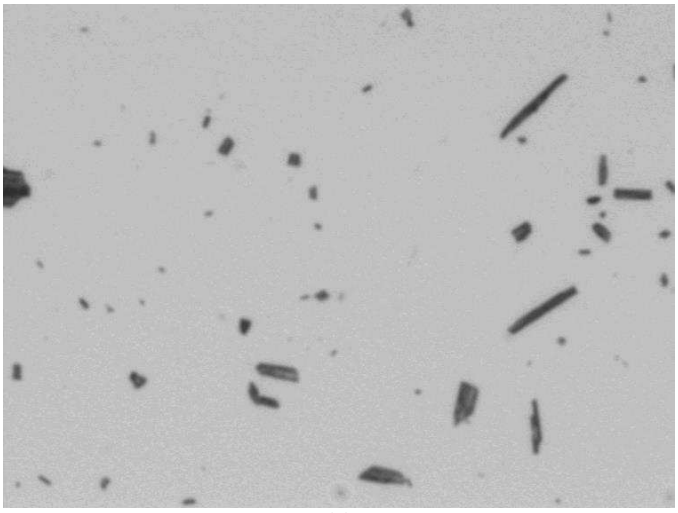
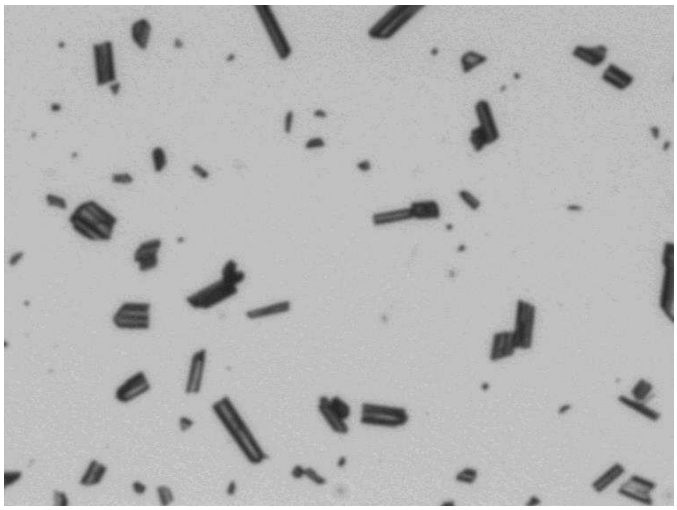


Fig. 5: The distribution of the Roundness values among three batches



(a)



(b)

Fig. 6: Two samples of L-glutamic acid particles obtained from two different Pilot Plant batches corresponding to different set-points for supersaturation control: (a) S=1.1 and (b) S=1.2

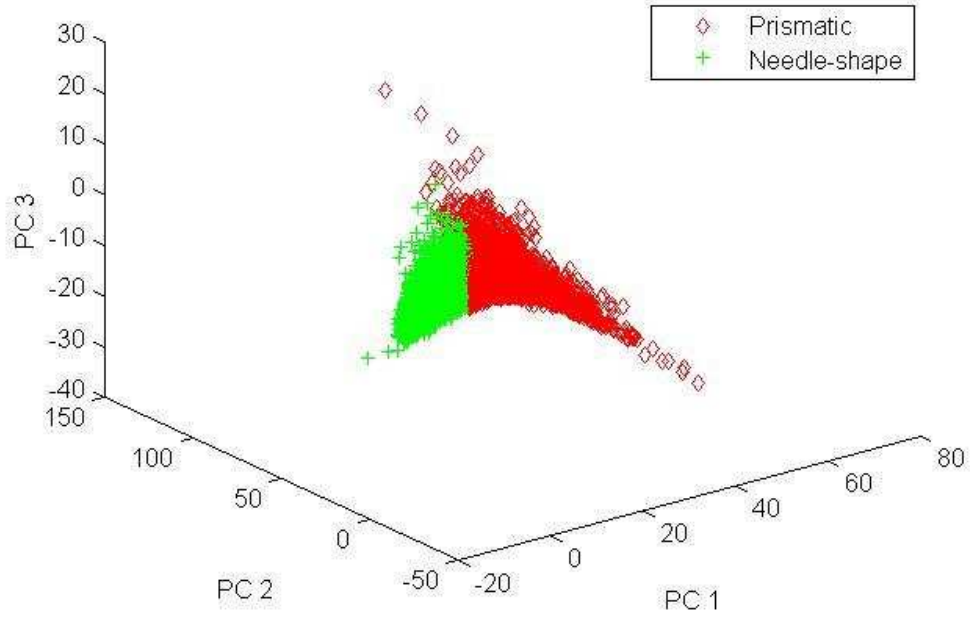


Fig. 7: The classified two clusters using 3 latent physical descriptors for the two Pilot Plant batches

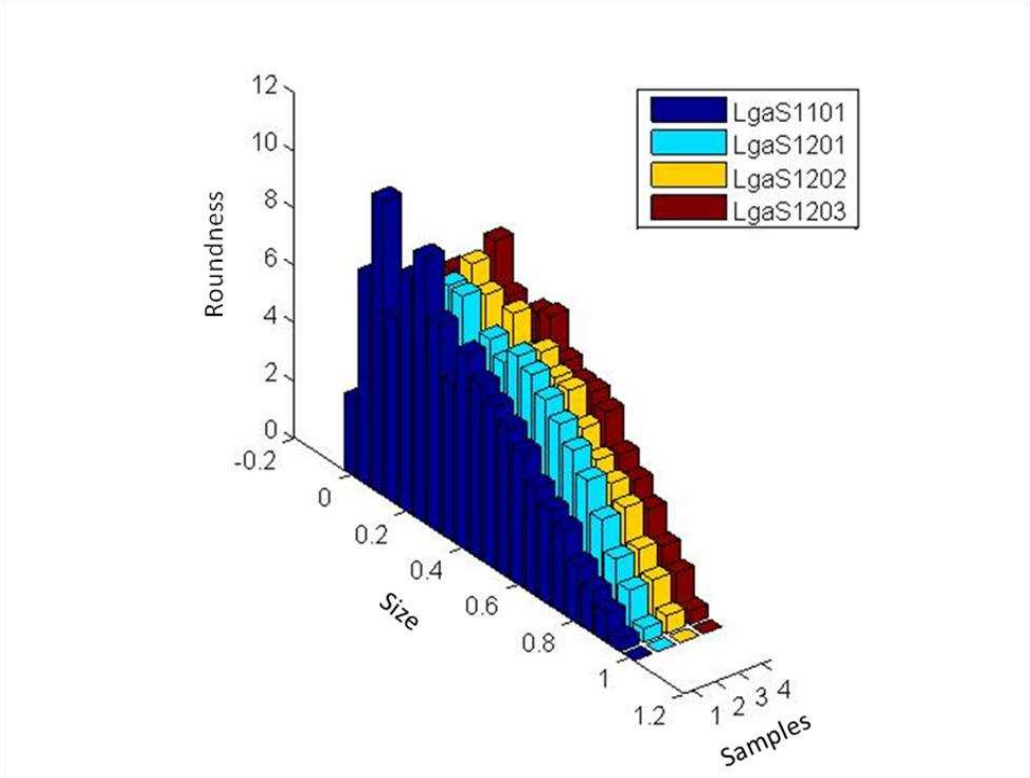


Fig. 8: The distribution of Roundness values among 4 samples