Batch to batch consistency via multiparametric SPES principal component analysis (SPES-PCA)

INTRODUCTION

Modern industry research and QC of products based on micron and submicron particles must ensure a documented batch to batch consistency for the whole manufacturing processes. Standardized measurements of hundreds or thousands of production batches could be needed to guarantee reliability of the processes and the production consistency with expected product chemical-physical properties. Complex heterogeneous formulations require access to many variables to ensure an adequate overview. This large number of data and variables require adequate statistical approaches to reduce the risk of not being able to grasp the key relevant differences between samples.

Principal Component Analysis (PCA) is a mathematical and statistical method suitable in exploratory data analysis by reducing the dimensionality of multiparametric data and datasets, i.e., reducing the number of significant variables for a given system without losing data significance. PCA is often used in pattern recognition for image analysis, so it is perfect for being applied to quickly recognizing differences between EOS CLOUDS. The EOS Classizer™ software SPES-PCA add-on provides easy to use and effective PCA tool that automatically computes the first principal components of an experimental SPES dataset; the tool allows the user to label data, which is very useful to create a training set for e.g. machine learning techniques. Machine learning is divided into three categories: supervised, unsupervised, and semi-supervised. In the first case, the entire initial dataset is labelled, and the PCA algorithm finds the areas of classification of the measures. A further measure can be characterized by looking its position. Unsupervised and Semi-supervised Learning cases are not covered in this application note.

Supervised Machine Learning may be divided in two classes. **Classification**: the labels are discrete and finite, e.g. considering two possible labels such as True/False, Yes/No, OK/Not OK, or Good/Bad. **Regression**: labels retrieved are continuous and related to non-discrete variables, such as temperature or viscosity. Several algorithms and computation methods could be used in supervised machine learning processes. The choice depends on dimension, operator experience and so on, but for relatively small dataset, i.e., a hundred measures, K-nearest neighbour is a suitable choice for most the cases.

SPES-PCA software add-on relies on this latter PCA approach to classify EOS clouds and samples. Application examples are discussed in this note.

PARTICLE ANALYSIS METHOD

Among the several methods currently adopted, optical ones have unique advantages, and therefore, have brought light scattering into the forefront of analytical methods in many scientific and industrial applications. Unfortunately, the number of parameters typically affecting the scattering properties of a given particle is such that the basic measure of the scattering power (or even the power removal from a light beam -extinction- from one particle) is far from being enough to recover something more than a rough estimate of its size. Things change appreciably when considering a collection of many scatterers, with the immediate drawback of introducing the need for mathematical inversion and ill-posed problems to interpret experimental real data.

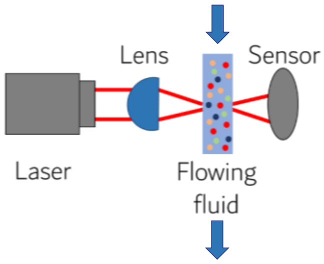
EOS Classizer™ ONE particle analyser is based on patented Single Particle Extinction and Scattering (SPES) method. It introduces a step forward in the way light scattering is exploited for single particle characterization.

EOS Classizer™ ONE – front view

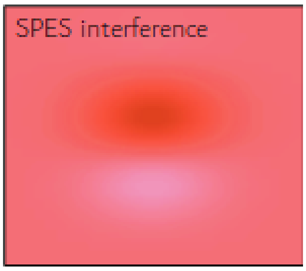
EOS Classizer™ ONE provides data that go beyond the traditionally optical approaches. EOS Classizer™ ONE discriminates, counts, and analyses single particles through their optical properties. It retrieves to the user several pieces of information such as: particle size distribution of the single observed populations, absolute and relative numerical concentrations, particle stability, information on optical particle structure and oversize. Classizer™ ONE works offline and online/real-time, enabling to verify consistency of intermediate and final formulations with target QbD, SbD, and Quality Control target expectations.

SPES TECHNOLOGY IN A NUTSHELL

The patented Single Particle Extinction and Scattering (SPES) method is based on a self-reference interferometric measurement of the scattered wavefront in the forward direction by a single illuminated particle.

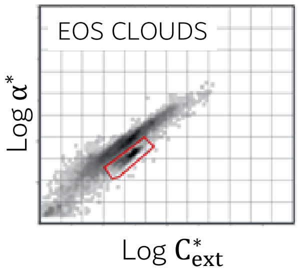


Particles are driven by a laminar fluid flow (liquid or gas depending on the application/CLASSIZER™ version) through the waist region of a tightly focused laser beam.



The intense transmitted beam interferes with the faint scattered wavefront in the far field, thus superimposing the two waves with the same curvature. This causes the interference pattern to exhibit intensity modulations on the spatial scale of the beam itself.

Two scattering features are sampled to follow the evolution of the intensity modulations during the passage of each single particle through the beam: i) the beam attenuation, given by the particle which removes a small fraction of the incoming power; ii) the fringes given by the partially constructive interference, which depth is proportional to the amplitude of the complex forward scattered field S(0). These two features are directly related to the real and imaginary components of S(0), $\Re e\, S(0)$ and $\Im m\, S(0)$, accordingly to the Optical Theorem [H. C. van de Hulst, Light Scattering by Small Particles, 1981]. On the other side they are also related to **the Extinction Cross Section $C^*_{ext} = \frac{k^2}{4\pi}\Re e\, S(0)$ and the Polarizability $\alpha^* = k^3\,\alpha = \Im m\, S(0)$**, where $k = 2\pi n/\lambda$ is the wave number in the medium with refractive index n at wavelength λ. Whatever the choice, two independent parameters **are thus retrieved for each detected, validated, and counted particle** thanks to a robust Pulse Shape Analysis scheme and proprietary algorithms. No need is there of adopting ill-posed problems, like inversion or deconvolution. At will, other optical parameters could be alternatively retrieved, eg. particle optical thickness ϱ.



In a few minutes SPES/ CLASSIZER™ ONE creates the unique **EOS CLOUDS**: a 2D histogram which is the optical fingerprint of the sample. Heterogeneous samples produce different clouds for each particle population simultaneously. They can be selected, analyzed, and compared. Statistical approaches as PCA are possible for extracting additional valuable information typically inaccessible with current instrumentation.

Added-value information is provided thanks to **SPES** and **EOS Classizer™ ONE** unique data and analysis libraries:

- **Optical Classification,** Absolute **Particle Size Distribution, Numerical Concentration** of each single population irrespectively of polydispersity/composition.

- Quality Control of particle **porosity, wetting**, **aspect ratio**, **payload**, **impurities**, **scraps**, and **shelf-life without intermediate steps** (purification/filtration).

- Measurement of **particle behavior and formulation stability** directly in real **heterogeneous non-filtered target biological, industrial, or environmental fluids**.

- Hi-Resolution **Continuous Flow Analysis**, also coupling SPES information with other analytical devices as CF3 separators, small chemical reactors, and pilot line.

- Statistical approaches as **Oversize Measure** and **PCA** for Hi-Quality Batch-2-Batch analysis and out-of-specifics identification in product formulation and production.

EOS Classizer™ ONE, based on patented SPES method, is the ideal solution for improving formulations and for verifying product consistency with the target Quality-by-Design final expectations.

Depending on the system configuration and sample, EOS Classizer™ ONE covers a dynamic range of 0.1 – 20 µm, concentration range of 1E5-1E7 ptc/mL @ 0.5-5ccm. External sample manager and autosampler are available.

This document presents representative examples of applications of EOS Classizer™ ONE and does not cover all the cases where the patented SPES method solves the particle identification, classification, and characterisation of challenges in heterogeneous samples and complex liquids. EOS software release SW1.4.44 is used for the data analysis and generation of the figures.

For a general introduction to SPES data with standard samples, as polystyrene spheres, please refer to the Application Note AN001/2021, available for free online at EOS website: www.eosinstruments.com/publications/

EOS SPES PCA - a brief intro

The EOS SPES-PCA tool calculates the Principal Components (PC) of a SPES dataset. Each Principal Component is a linear combination of uncorrelated variables, reduced via an orthogonal linear transformation of the initial variables of the SPES dataset: PC1 is the first component found by the PCA algorithm, along the maximum variance of the dataset. Analysis may be limited to PC1, which typically classifies the samples considering their particle concentration differences. It is possible to iterate the argument with subsequent Principal Components until the desired accuracy or level of analysis detail is achieved. E.g., PC2 is the second component along the maximum variance and typically differentiates data basing on differences of an optical property, e.g. as the refractive index.

The result of the PCA is a basis of eigenvectors common to all the dataset, each of which has associated an eigenvalue PD for each single SPES data. A single 2D EOS CLOUDS measure is thus describable by a 2D vector sum:

$$M_i = \sum_j PC_{ij} PD_j$$

where each iterative sum corresponds to adding a new level of detail, finer than the previous one, to the SPES histogram. EOS PCA tool focusses the analysis to the first three components which are typically suitable for most of the exploratory data and batch-2-batch analyses:

$$M_i = PC_{i1}PD_1 + PC_{i2}PD_2 + PC_{i3}PD_3$$

By comparing the PDs between the samples, it is possible to classify samples on the basis of many variables, e.g. different absolute numerical concentrations, secondary particle populations as aggregates or contaminants, relative numerical concentration between each population, and different tails in particle size distributions.

APPLICATION EXAMPLES

In this methodological application note of SPES-PCA tool we present two representative scenarios where PCA and supervised learning are useful for sample QC classification and impurity or secondary population detection:

1) Batch to batch sample quality control
2) New batches compared to a reference dataset

1) Batch to batch sample quality control

It is important to standardize industrial processes to produce consistently high-quality finished products and/or identify batches which do not comply with production standards, previous batches, or design formulations. In this example, a PCA test is performed on batches of PS particle suspensions to identify unwanted contaminations. Notwithstanding its fainty presence is almost negligible in the EOS CLOUDS and in the particle size distribution, the contamination may ruin and spoil of the production batch.

EOS CLOUDS are 120x100 8-bit histograms with the greyscale normalized to numerical particle concentration. The EOS SPES-PCA tool reduces all these variables to three components (the eigenvectors, PD) and plots and compares the values (the eigenvalues, PC) of each measure.

As a training set, twenty-five samples of 750nm polystyrene particles with a 5E6 concentration are prepared and measured. We label these as “good” measures. Then ten samples of the same kind plus an impurity of silicon oil are prepared and measured. We label these as “bad” measures.

Looking at a representative EOS CLOUDS of a “good” sample (Figure 1) and a “bad” (Figure 2) one, few differences can be observed. The impurities are polydisperse and about less than 10% of the whole numerical concentration. Also, from a study of the whole SPES CLOUDS, the software calculates for both the groups the same refractive index of 1.61 with only a 2% shift in average diameter. Even the similarity of the two histograms is >99%. Moreover, comparing manually one-by-one hundreds of data sets so similar basing on the EOS CLOUDS or particle size distribution would be very time-demanding, not efficient, and with a low level of precision.

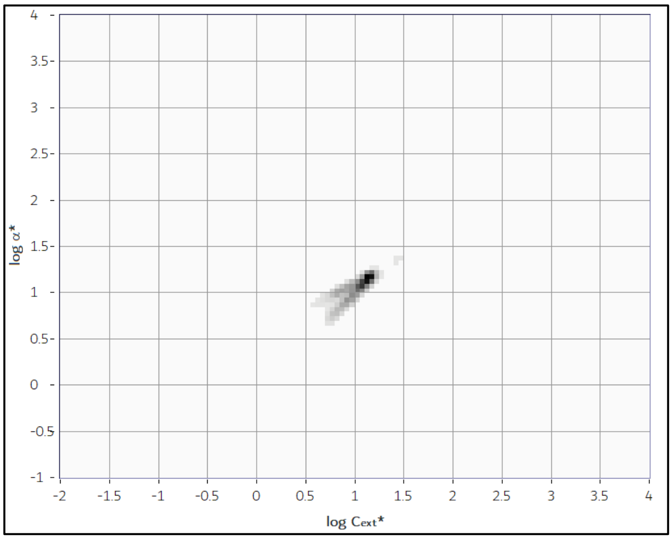


Figure 1 EOS CLOUDS experimental data for a “good” sample: 750nm polystyrene particles at 5E6 ptc/mL in filtered milliQ-grade water.

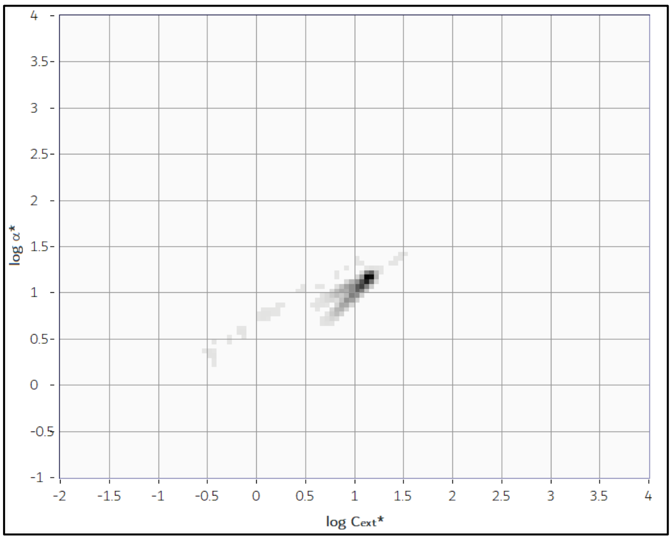


Figure 2 EOS CLOUDS experimental data for a "bad" sample: 750nm polystyrene particles at 5E6 ptc/mL dispersed in filtered milliQ-grade water with a fainty contamination of silicon oil.

The EOS SPES-PCA add-on performs quick statistical analysis of the data and provides the needed classification in term of batch-2-batch optical properties differences.

The two groups of SPES data are merged and loaded in the EOS software PCA add-on as a single dataset. Running the PCA analysis, the algorithm automatically finds the principal directions (PDs), or eigenvectors, of the selected dataset. PD1 indicates the most common population, which is the monodispersed polystyrene (Figure 3). All the eigenvalues of this first matrix are positive, since the raw EOS clouds cannot return negative numbers. In the remote case the algorithm retrieves negative values for PD1, the M_i sign is compensated by a negative eigenvalue.

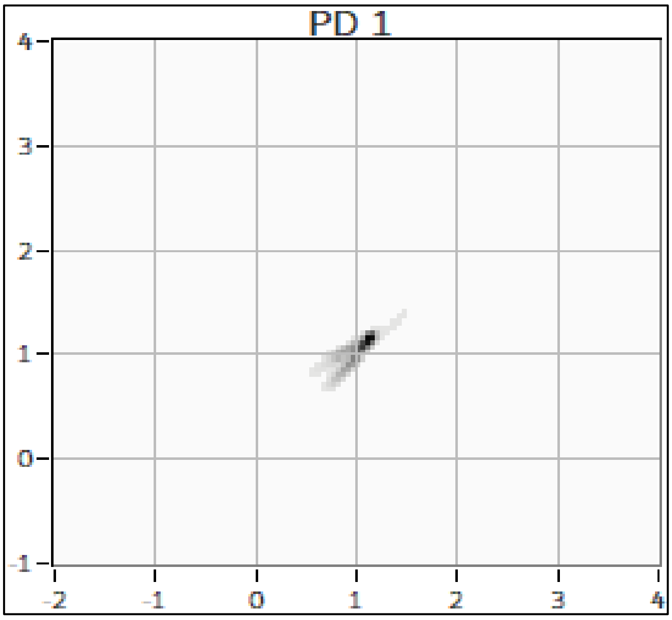


Figure 3 PD1 corresponds to *the main population*. In this example, the PS population is isolated and represented as the main one in the matrix.

PD2 shows positive (grey) and negative (red) values, typically corresponding to secondary populations and / or minor differences in the whole dataset (Figure 4).

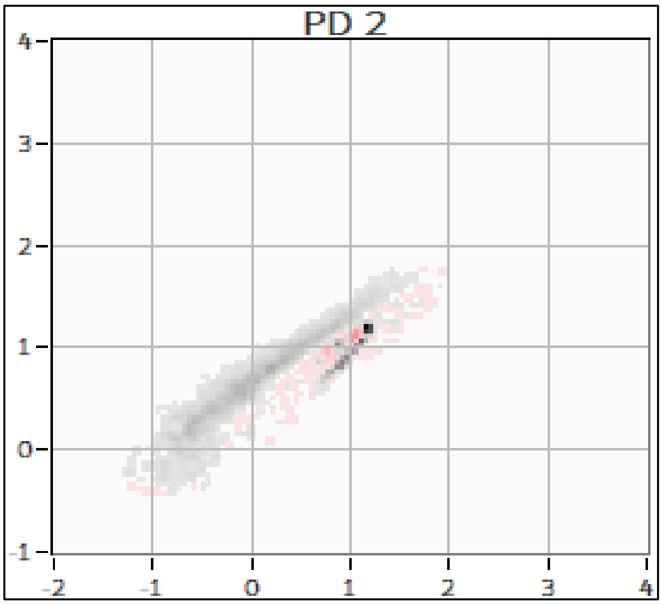


Figure 4 PD2 indicates, in this case, the emulsion plus some other mixed minor differences between PS batches.

Note. In PD2 it is clear the presence of a secondary particle population at lower refractive index respect to the main one (silicon oil contamination). Minor differences between PS batches are also considered and enlightened in PD2.

PD3 provides another info (Figure 5). Except for some points close to the position of the main population, most of the values are uncorrelated and may be related to a non-uniform non-constant background, i.e., water impurities not correlated with the two main populations observed.

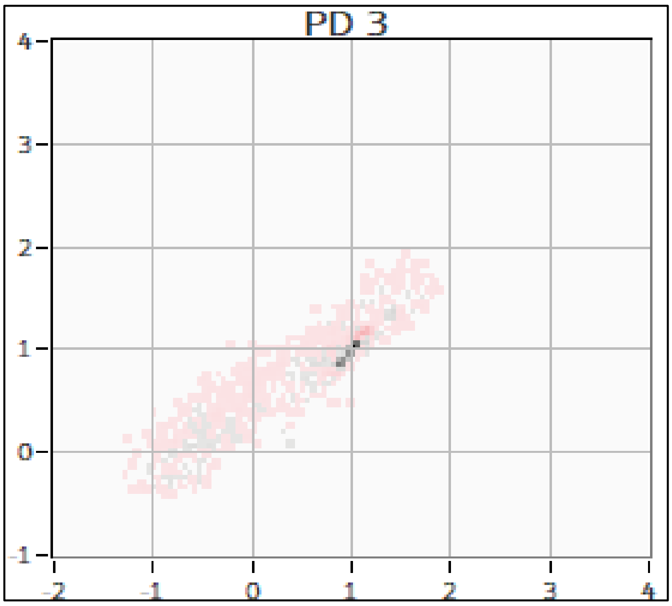


Figure 5 PD3 could be related to negligible differences and/or to the background impurities.

Once the autovector is defined, the software automatically computes and plots the eigenvalues PC1, PC2, and PC3 of each data in the dataset, reducing the batch-2-batch comparison to three numbers per measure.

In Figure *6*, the data are divided into two groups. In this case, the "good" measures are green, while the red ones are bad. The user can change the colors based on personal preferences. Green data on the left indicates that good measures have a higher component of polystyrene particles (the first principal dimension, PD1), while they have a negligible or even negative component of the second principal dimensions, that is the emulsion and a few particles. On the other hand, the bad samples have a positive value for the second component, as expected.

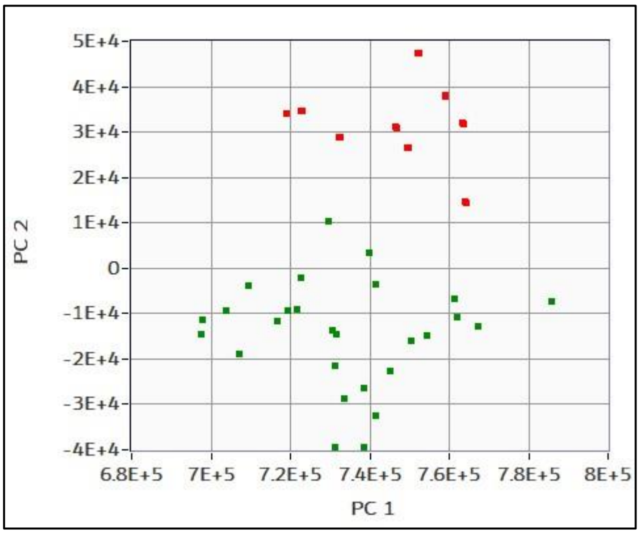


Figure 6 Graphical representation of PC1 to PC2 for each measure. The eigenvalues, and thus point locations, are automatically calculated by the PCA algorithm without relying on information provided by the user. The labels inputted are used to color the points and help the user in the data interpretation. In this case two separate clouds of data results, red one corresponding to "good" sample and green corresponding to "bad" samples. The separation is on the horizontal axis (PC1) which corresponds to the numerical concentration and on the vertical axis (PC2) which corresponds to optical differences between the measures.

In Figure 7, data are not divided into two groups. This could be interpreted as a different uncorrelated background in the data. Thus, all the measures have a small quantity of generic non-repeatable impurities.

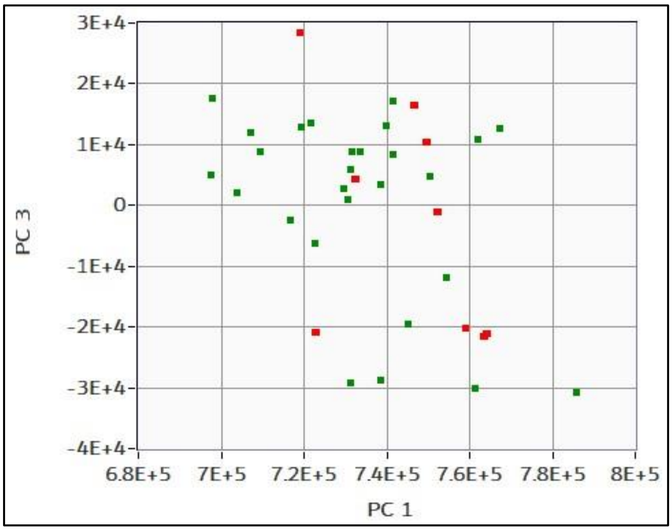


Figure 7 Graphical representation of PC1 to PC3 for each measure. The eigenvalues, and thus point locations, are automatically calculated by the PCA algorithm without relying on information provided by the user. The labels inputted are used to color the points and help the user in the data interpretation. In this case two separate clouds of data result: the red one corresponding to "good" sample and the green one corresponding to "bad" samples. There seems to be no clear separation between the two cases.

PCA method easily differentiates and classifies the EOS CLOUDS. Secondary populations or a difference in the shape of the main population may be evidenced and then further investigated.

2) New batches compared to a reference dataset

In this example, PCA automatically compares the data of novel measure to a reference dataset of "good" and "bad" samples. This statistical approach classifies if a novel sample is more comparable with a first class of measures (e.g. "good"), a second one (e.g. "bad), or none at all, reducing analysis-time and operator-dependent issue in data analysis. The EOS SPES-PCA software add-on standardizes the procedure via the machine-learning algorithm K-Nearest Neighbor (K-nn). The K-nn algorithm computes the first k closest elements of the training set for each measure and classify the whole dataset basing on distance.

In Figure 8, a batch of ten novel unknown measures are loaded and compared to the same directions PDs of the training set of the previous example. PCs of the novel samples (blue) are thus compared to the PCs of the reference dataset (red/green). It can be observed that six measures fall in the "good" green area, three fall in the "bad" red area, while the remaining three are far from both.

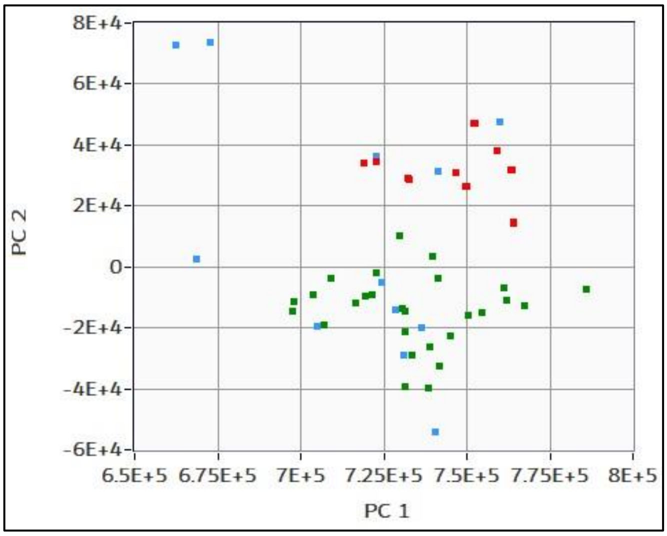


Figure 8 Graphical representation of PC1 to PC2 for each measure, both for the new batches (in blue) and for the training set (in red and green), as already seen in Figure 7. It is possible to classify the measures by calculating the distance between them and the nearest points in the graph.

Using the K-nn algorithm, the EOS software suggests a suitable value of K. If needed, the user may change it basing on its experience and needs. The lower the K, the higher the correlation required to evidence a novel sample as one of the two populations in the dataset. The software thus retrieves a list and a “good”/“bad”/“NA” label for each measure. “NA” corresponds to data non compatible with “good” or “bad” population for the K value used.

In this case example case, using a K of 5, the two blue measures in the top left corner of Figure 8 are valued as “bad”, the one in the center and the one in the bottom of the figure are valued “good”. Note. For the latter two points, a warning informs that that these two are far from the training set. The software provides the user with the name of the file corresponding to each point. A further analysis can be processed on single sample EOS CLOUDS.

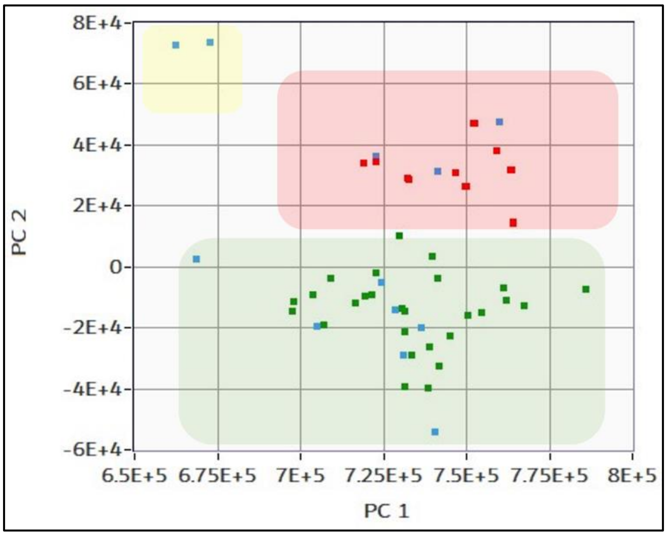


Figure 10 Graphical representation of the sample PCA classification done for the dataset and novel data plotted in Figure 8.

CONCLUSIONS

The capability of EOS SPES-PCA software add-on to perform a Principal Component Analysis on a measure dataset via supervised learning technique has been presented. SPES-PCA provides added value exploratory data analysis and sample classification based on SPES multiparametric data. It speeds up batch to batch processing in a Quality Control production line, detecting small differences between apparently very similar sets of measures. A quantitative PCA classification of data consistency via K-nn algorithm provides quick and effective batch analysis comparing data of novel batches with data libraries and dataset of reference samples.

RELEVANT PUBLICATIONS AND REFERENCES

Presentation of Single Particle Extinction and Scattering (SPES) method for particle analysis
AN001-2021 Analysis of Polymeric Particles via SPES Technology – a general introduction to SPES method

AN006-2021 Multiparametric Classification of Particles as a Pathway to Oversize Analysis in Complex Fluids via SPES Technology

Potenza MAC *et al.*, «Measuring the complex field scattered by single submicron particles », AIP Advances 5 (2015)

Example of CFA application of SPES technology
AN002-2021 Continuous SPES Flow Analysis CFA-SPES

Example of PCA application of SPES technology
AN005-2022 Multiparametric Principal Component Analysis of Heterogeneous Samples via SPES Technology

Classizer™ ONE with Sample Manager Autosampler
AN008-2022 Automatic Liquid Sample Management, Dilution, and System Cleaning with EOS Sample Manager

AN009-2022 Standardize SPES Operative Procedure of Liquid Samples Analysis via EOS Autosampler

Example of SPES application to aggregates
AN003-2021 Addressing the Issue of Particle Wetting and Clustering by means of SPES Technology

Potenza MAC *et al.*, «Single-Particle Extinction and Scattering Method …», ACS Earth Space Chem 15 (2017)

SPES application to non-spherical particles
AN004-2021 Addressing the Classification of Non-Spherical Particle by means of SPES Technology

Simonsen MF *et al.*, «Particle shape accounts for instrumental discrepancy in ice …», Clim. Past 14 (2018)

Example of SPES application to emulsions w/o payload in environmental waters
AN012-2021 Monitoring the Fate of a Lipid/ZnO Emulsion in Environmental Waters

Examples of SPES application to particle analysis and behavior characterization in biotech applications
AN007-2021 Quantitative Classification of Particles in Biological Liquids via SPES Technology

Sanvito T *et al.*, «Single particle extinction and scattering optical method unveils in real…", Nanomedicine 13 (2017)

Potenza MAC *et al.*, «Single particle optical extinction and scattering allows real time quantitative…», Sci Rep (2015)

Example of SPES application to oxide particles, abrasives, and industrial slurries w/o impurities
Potenza MAC *et al.*, «Optical characterization of particles for industries», KONA Powder and Particle 33 (2016)

AN002-2021 Analysis of Abrasives via SPES Technology

Example of SPES application to ecotoxicity analysis
Maiorana S *et al.*, «Phytotoxicity of wear debris from traditional and innovative brake pads», Env Int., 123 (2019)

Example of SPES application to aerosol analysis
Mariani F *et al.*, «Single Particle Extinction and Scattering allows novel optical …», J Nanopart Res 19 (2017)

Cremonesi L *et al.*, «Multiparametric optical characterization of airborne dust …. », Env Int 123 (2019)

Quantum Design EUROPE

Quantum Design GmbH
Breitwieserweg 9
D-64319 Pfungstadt

Please contact: Stefan Wittmer
+49 6157 80710-63, wittmer@qd-europe.com
Find your local contact at www.qd-europe.com