Powder Technology 294 (2016) 134–145

Contents lists available at ScienceDirect

Powder Technology

journal homepage: www.elsevier.com/locate/powtec

Comparing particle size distributions of an arbitrary shape

Otto Scheibelhofer [a], Maximilian O. Besenhard [b], Michael Piller [b], Johannes G. Khinast [a,b,*]

[a] Institute for Process and Particle Engineering, Graz University of Technology, Graz, Austria
[b] Research Center Pharmaceutical Engineering GmbH, Graz, Austria

ARTICLE INFO

Article history:
Received 25 November 2015
Received in revised form 5 February 2016
Accepted 13 February 2016
Available online 16 February 2016

Keywords:
Particle size distribution
Statistics
Process analytical technology
Crystal growth

ABSTRACT

Since particle size distribution can provide crucial information regarding process and product quality, comparing one distribution with another and with modeled distributions is often necessary. Since only mean and median distribution values are often used for that purpose, important information may be missing. As such, methods are required that take into account the entire distribution and do not require modeling the distribution.

The χ^2-homogeneity test is a nonparametric homogeneity test based on the observed frequencies in the descriptive classes. In this work, we examined if it could be used for comparing particle size distributions and for large particle numbers. In conclusion, we demonstrate how this statistical test was successfully applied in several cases.

1. Introduction

Controlling particle size distribution (PSD) is essential for many industrial processes. Particle sizes can vary vastly, from rock debris in geological sciences to components in the millimeter range in the construction industry and crystals in the μm range (and below) in pharmaceutical applications [1–5]. Particle size often determines crucial quality attributes that relate to the processability of intermediates (e.g., handling and storing of cement, blending of active pharmaceutical ingredients (APIs) and excipients) and the quality of the final product (e.g., the strength of concrete or the bioavailability of dosage forms) [6–11].

Determining a particle size distribution typically involves sampling from a process stream. Sampling involves a mass reduction, and in that regard correct sampling of granular materials is challenging [12,13]. Since positioning of the sampling tool defines the part of process stream that is sampled and analyzed, it has to be performed meticulously. Although sampling is generally performed in a non-destructive matter, sampling of the entire particle population is rarely achieved due to the huge number of particles and data processing involved [14].

The samples' particle sizes are determined via methods that are suitable for the size range in question [15]. There are several approaches to defining a single characteristic number or size of particles that represents their size, i.e., called equivalent diameters. Many different equivalent diameters exist, e.g., the diameter of sphere having the same volume, surface area, sedimentation velocity and many more. The individual size data for a number of particles are then grouped into bins of a certain size range [16–18] called a histogram that determines a PSD [19,20].

Since the particle size is a continuous variable, the probability of finding a particle with a certain diameter is zero. However, there is a finite probability of finding a particle within a certain size range. Although PSDs are regarded as either continuous or discrete distributions (i.e., histograms), they are only truly continuous in simulated and modeled cases. Sampled PSDs always originate from a finite number of particles, leading to a discrete distribution. As a result, experimentally obtained particle sizes are typically split into size classes covering a certain range. The number of particles found within this range, compared to the total number of particles sampled, is the best guess with regard to the probability of finding a particle within this range [21,22].

A sample determines the number of particles in certain size classes, i.e., a PSD. A PSD can also be viewed as a multivariate description of the sample. However, since multivariate representations can be difficult to handle, simplifications have to be made [23]. Often, only a few relevant (integral) parameters (or moments), such as the mean, median diameter and the “broadness” of a PSD, are reported. However, this suffices only if PSDs have a known and common shape [19,24].

Various analytical distributions are used to approximate PSD shapes, e.g., normal distribution, log-normal distribution and others. PSDs however can take increasingly complex shapes, either due to combinations of several particle species, each having their own distribution, or due to sampling effects (e.g., a size cutoff in the measurement principle). Such complex PSDs are hard to describe by an analytical formula, and often only a visual representation of the PSD is informative [22,25].

Often, PSDs are compared with each other, e.g., for quality assurance purposes or if samples are drawn from the same population to ensure that sampling is representative and to understand the dependence of PSD on the sampling location. Theoretical and experimental

* Corresponding author at: Institute for Process and Particle Engineering, Graz University of Technology, Graz, Austria.
E-mail address: khinast@tugraz.at (J.G. Khinast).

http://dx.doi.org/10.1016/j.powtec.2016.02.028
0032-5910/

(i.e., discrete and continuous) PSDs have to be compared [26,27]. Comparing distributions only based on integral parameters may not reveal differences between them (see illustration in Fig. 1.). Although a visual comparison can be more informative, it does not have a comparative quantity. Another problem associated with comparing PSDs is that they may be obtained using different measurement techniques [28], in which case elaborate knowledge of the measurement process and particle shape is required.

As such, a method of comparing particle size distributions that cannot be suitably represented via an analytical function, or whose complexity cannot be grasped by summary parameters, would be advantageous. Hence, a method for comparison that takes into account the entire distribution rather than the summary statistics only is beneficial. For these purposes, the χ^2-homogeneity test can be used. This test is extensively applied in the psychological and medical sciences [29] or used for comparing cascade impactor profiles [30,31].

We applied the χ^2-homogeneity test for comparing particle size distributions of an arbitrary shape. Thus, there is no necessity in calculating or selecting summary parameters, but the whole distribution is considered. Furthermore, this test allows to quantify the similarity (or difference) of PSDs, including consideration of the sample size.

2. Method

The χ^2-tests belong to a family of hypothesis tests, from which only the χ^2-homogeneity test will be discussed below. The χ^2-homogeneity test is a non-parametric test applied to 2 or more samples described by 2 or more categorical variables, to determine if they originate from the same population [32–35]. It is a hypothesis test whose null-hypothesis is: H_0: *All samples are drawn from populations that have the same proportions of observations between classes.* In other words, the populations from which the samples originate are homogeneous, i.e., their size distributions are identical. Thus, the χ^2-homogeneity test can determine if several samples of a limited size represent the same particle size distribution or if the difference between samples is too large to be explained by random sampling.

2.1. Prerequisites for the χ^2-homogeneity test

An agreement between the mathematical and experimental requirements for conducting the test will be established first.

The population is defined as the total number of all particles, i.e., in the investigated process stream, or all particles inside a hopper or a heap. During the process, samples are obtained containing a reduced (compared to the population) number of particles. Although the sampled PSDs should be representative of the population, the real PSD of the population can only be established when measuring all particles of a population. As such, sampled PSDs are inherently only an approximation and contain an error.

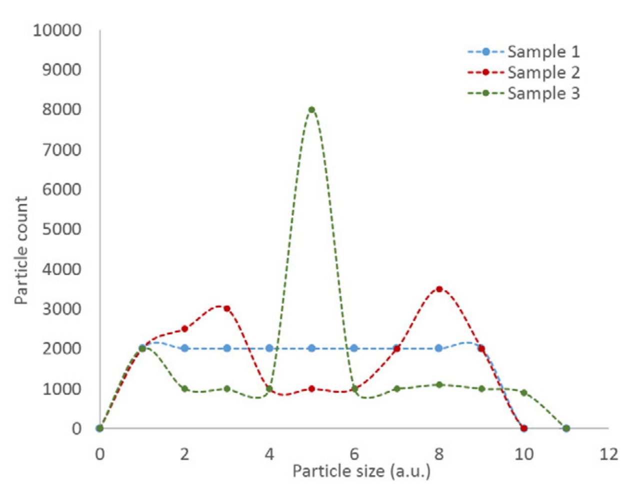

Fig. 1. Particle size distribution in three samples. Although their shapes are very different, they have the same summary statistics: $x_{10} = 1.4$, $x_{50} = 5.0$, and $x_{90} = 8.6$. The line connecting the actual data points is the interpolated spline setting of Excel (Microsoft Office 2013).

Correct sampling of process streams is an elaborate task, and numerous recommendations are available in the literature [36]. As a prerequisite for the χ^2-homogeneity test the number of independent samples must be equal to or greater than 2. Repeated measurements of the same sample are not considered independent. If there are 2 samples, one is compared to the other. When there are more than 2 samples, all of them are compared and a single deviating one will dominate the outcome of the test (i.e., no distinction is made between reference and test samples).

Suppose a number of M samples is measured, with sample m containing a number of N_m particles. The particle sizes are distributed in K different classes, where class k spans within a size region w_k, represented by a mean size value of x_k. This results in a particle count of n_{mk} for class k and sample m, and a total test size of $N = \sum_m N_m = \sum_m \sum_k n_{mk}$.

Observations (i.e., size measurements of single particles) should be unique and distinct. This is usually the case in particle size measurements. Every particle is only measured once, resulting in a single size value of this particle. In sieve analysis, a particle is found on top of a single non-passing mesh size. The size of the next particle measured does not relate to the previous one. Classes in PSDs are generally exclusive (i.e., not overlapping) and exhaustive (i.e., there is no particle that is not belonging to a class). The χ^2-test can be applied on any scale, even non-ordinate (i.e., classes do not have to have an order). Since classes are categories for describing samples, the specific size and order of classes is of no importance to the test, and there is a significant amount of freedom in choosing the classes. For example, certain size classes may be neglected during an analysis, since (irrelevant) foreign particles (dust) or measurement uncertainties can obliterate those classes. Furthermore, this enables the pooling of classes, as discussed below.

Since the χ^2-homogeneity test is based on the frequency of particles per sample and size class, the *absolute* number of particles is important. If q_{mk} is the relative distributional density of class k and sample m and if the total number of particles N_m in the sample m is known, the number of observations per class k can be calculated as $n_{mk} = q_{mk}\, w_k\, N_m$.

For every particle, the question if it belongs to class k can be answered modally. If a particle belonging to the population is measured (and hence is now part of sample m), the number of particles in class k increases by one with probability p_k, which is the (unknown) fraction of particles of class k in the total population. Thus, sampling N_m particles, the number of particles in class k of sample m follows a binomial distribution and the expected value is $\langle n_{mk} \rangle = p_k N_m$ with a variance of $\sigma_{n_{mk}}{}^2 = N_m p_k (1 - p_k)$. For large samples, the binomial distribution can be approximated via a normal distribution according to the Moivre–Laplace theorem with $\mathcal{N}(\langle n_{mk} \rangle, \sigma^2_{n_{mk}})$. In this notation, $\mathcal{N}(\mu, \sigma^2)$ indicates the normal distribution with mean μ and variance σ^2.

The sum of squares of a number of f independent variables (e.g., sample and particle size class when comparing PSDs, see Section 2.2.), all of which follow a standard normal distribution (i.e., $\mathcal{N}(0,1)$), is distributed as the $\chi^2_f(x)$-distribution. The parameter f is generally referred to as the degrees of freedom. For the purposes of the χ^2-homogeneity test, a test statistic, termed χ^2, is calculated from the experimental data, indicating the difference between the observed and expected particle numbers. The exact calculation will be shown in Section 2.2. This χ^2 test statistic is a sum of squared standard-normally distributed variables and hence follows the χ^2-distribution. If the test statistic is much higher than the expected χ^2_{crit}, which is determined analytically based on the assumption of random sampling, this excess cannot be explained only by sampling effects and it is