Powder Technology 412 (2022) 118003

Contents lists available at ScienceDirect

Powder Technology

journal homepage: www.journals.elsevier.com/powder-technology

Sums of amplitudes analysis – A new non-parametric classification method for time series deviation evaluation in pharmaceutical processes

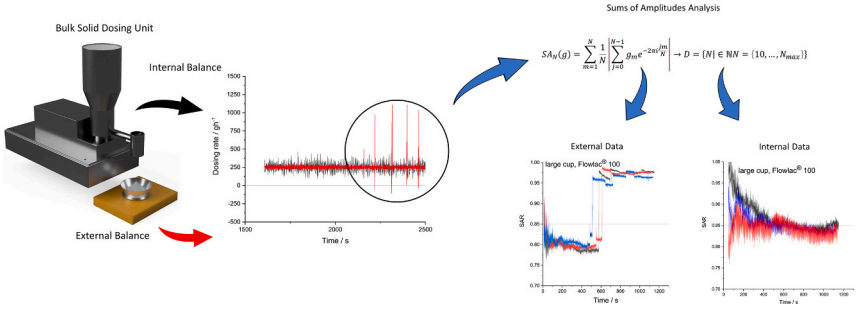
Stefan Klinken [a,*], Julian Quodbach [a,b]

[a] Institute of Pharmaceutics and Biopharmaceutics, Heinrich Heine University, Duesseldorf, Germany
[b] Department of Pharmaceutics, Utrecht Institute of Pharmaceutical Sciences, Utrecht University, Utrecht, the Netherlands

HIGHLIGHTS

- Quantification of the systematic information in pharmaceutics related time series.
- Description of artefacts in powder dosing evaluations.
- Fourier based non-parametric time series evaluation.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Fourier transform
Time series
Data analysis
Continuous manufacturing
Powder feeding
Bulk solid dosing

ABSTRACT

The evaluation of various types of data and the understanding of their properties is the backbone of scientific work. Time series data are abundant in the investigation of manufacturing processes. Since they result from measurements of parameters over time, time series have distinctive statistical characteristics. Focused on bulk solid dosing, the authors illustrate the potential for misinterpretations if these characteristics are ignored in data evaluation. The sums of amplitudes analysis (SAA) is developed, based on discrete Fourier transformations as a novel non-parametric statistical method for description of time series. Subsequently, the authors use the SAA to investigate bulk solid dosing data. In this process, artifications of data of external catch scales, commonly used for the evaluation of bulk solid dosing, are found, and discussed.

1. Introduction

Data acquisition and evaluation is of utmost importance in modern pharmaceutical manufacturing processes to obtain information about critical attributes of product and process. The increasing use of process-analytical-technologies (PAT) to specify the critical quality attributes (CQA) of the produced goods results in an increase of the quantity and complexity of the data. This underlines the importance of suitable data processing [1,2]. The collection of data can be done by various measurement methods which are frequently subdivided in on-line, off-line and in-line tools. While in-line and on-line tools measures the material inside the process or a bypass, for off-line measurements the removal of the material from the process is required. Therefore, the first two are capable of ensuring real-time monitoring of processes.

* Corresponding author.
E-mail address: Stefan.Klinken@hhu.de (S. Klinken).

https://doi.org/10.1016/j.powtec.2022.118003
Received 15 August 2022; Received in revised form 6 October 2022; Accepted 7 October 2022
Available online 13 October 2022

Properties of the data depend on the type of the applied method and on the investigated process. Many data are generated in replicates without a determined order. They are mathematically described by realizations of a random variable. The exact values of these realizations depend on the parameters of the respective random distribution function. Following the central limit theorem, these functions in many cases approximate a normal distribution. Thus, the parameters of the random variables are the expectation of the distribution μ and the variance σ^2, while the realizations are fully independent of one another. The preceding realizations of the random variable does not influence the results of the subsequent measurements. This is the most common type of data and nearly all widely used statistical measures and tests require these properties, for example the standard deviation or arithmetic mean as robust measures of the data and their confidence intervals, as well as basic ANOVA and *t*-tests. Various applied examples of the use of these measures and tests can be found in literature. They range from drug content data of tablets over rheology parameters of gels to particle size distributions of bulk solids [3–5].

As the population of size N usually cannot be analyzed, samples of a sample size n are drawn. Each element of the sample must be drawn randomly and independently from the same population distribution, which means that the probability for any element to be included in the sample is the same. In reality, these conditions are only partially fulfilled [6,7].Therefore, the sampling procedure is particularly crucial as it determines the quality of the drawn sample [8].

In contrast, time series data are of a defined order. They can be produced by discrete measurement of a certain parameter over the change in time. While time series can be defined differently, we will use the here presented definition in the scope of this article. The intervals between the individual measurement points are usually equal, which means that the number of measurement points per time can be specified as the sample rate. Due to the sampling procedure and structure of the time series data, their properties differ from classic samples. For instance, different measurements of time series data can show high collinearity [9], which can result in deviations of the measured distribution from a normal distribution to varying extent. This will be explained in more detail later. Furthermore, the individual measured values of some time series do not all relate to the same population. This can be the case if the process is not in a steady state but has different states divided by changing points, such as pressure data of rotary tablet presses or plug flow dominated extrusion processes. As the normal distribution and the independency of the data is important for many statistical methods, they are, in fact, often not suitable for common time series evaluation. In contrast to this, the use of classical statistical parameters like the variance or confidence intervals is widespread for pharmaceutical time series evaluations [10–12]. Due to the importance of time series data for the increasingly used analytical technologies and modern process control systems, the understanding of this kind of data is of high interest for research and pharmaceutical manufacturing. The classification of time series is well described in literature of data science. Mostly, multivariate statistical methods, autocorrelation or artificial intelligence are used in the approaches of time series evaluation in signal processing [13].

The data of the continuous dosing of powders and granules are an example for time series data. Bulk solid dosing is a standard procedure in many pharmaceutical processes, among them twin-screw granulation and extrusion [14]. Different devices are available to ensure continuous feeding of solid material into the process. Subgroups of the devices can be defined by the mechanism of transportation of the bulk solid material from the hopper to the downstream process. Vibratory tray feeders are commonly used in instrumental powder analytics like laser diffraction analysis or in continuous milling processes. The second large group of bulk solid dosing units is comprised of screw feeders. Co-rotating twin-screw devices are often used in pharmaceutical continuous granulation or extrusion processes. These devices can be further divided in volumetric and gravimetric dosing units. The first work at a constant setpoint of the vibration intensity or screw speed while the second adjust the setpoint to prevent and compensate deviations in the bulk solid dosing process. This is done by weighting the whole dosing unit and measuring the amount of powder dispensed at every point in the process [15]. Evaluations of bulk solid dosing units, considering the given dosing rate over the process time, the so called dosing curve, are widely published in the literature [11,12,16–33]. An external catch scale, on which the powder is dosed, is frequently used to avoid errors in the data by internal data preprocessing methods in the dosing units. Additionally, the maximum possible sample rate can be much higher in the external catch scale than in the dosing device. 16 of the 20 publications mentioned above as examples use an external catch scale set-up. 14 of them evaluate dosing data regarding measures of the spread, for example the variance or the residuals of a regression, assuming normal distribution or at least symmetric distribution of the data or deviations and disregard the effect of collinearity in the data. The quantitative impact of the artification of these methods can be high in the field of bulk solid dosing evaluations. However, the individual qualitative impact of these artifications depend on the respective study and gathered data. Nevertheless, the exploration of the influence of potential misconceptions and errors is of great interest, to improve the consistency of the applied methods. The call of Wagenmakers et al. to implement multiple statistical methods for data evaluation in a single study underlines the need for novel mathematical perspectives on data [34].

This study aims to identify and discuss problems and potential errors in pharmaceutical time series evaluation, highlighted by the example of bulk solid dosing processes. A new classification method based on a permutation in frequency domain is introduced to characterize the properties of the obtained data in dosing evaluation and to identify potential error sources. In addition, examples of time series of other process and parameters are discussed regarding their respective properties and classification.

2. Definitions and time series properties

2.1. Definitions

The properties of different types of time series are rarely considered in the evaluation of various data in the pharmaceutics focused literature. In bulk solid dosing investigations, Kruisz et al. demonstrated the effect of a change in the sample rate on the coefficient of variation of the time series data [35]. Nonetheless, previous studies in this field disregarded this effect in the statistical methods to describe the respective data [17,18,36,37]. The terminology of the respective variations from a target value or a steady state mean need to be more precise. For example, variations are mostly described by the term “fluctuations”, but differences in the various forms of deviations are not considered. In fact, Kruisz et al. [35] mention different causes for disturbances in bulk solid dosing data without specifying their properties.

To ensure consistent terminology and to heed the differences in mathematical properties, the authors propose a systematic classification system for time series in pharmaceutical processes.

Fundamental and derivative parameters: Considering the properties of different parameters that form a time series, subgroups of these can be defined. The magnitude of perfect fundamental parameters is independent of the sample rate. Derivative parameters on the other hand are calculated from fundamental ones as the mathematical derivative over the change of the x dimension, which is mostly the time, so that the change in the intervals of x changes the magnitude of derivative parameters. Examples for fundamental parameters are the die pressure or temperature data in extrusion or the punch force and in-die density in tableting processes. Derivative parameters can be found in the data of mass changes over time, for example in bulk solid dosing, punch or screw speed or in air velocity in fluid bed coating machines.

Homogeneous and heterogeneous deviations: Most time series in pharmaceutical processes scatter homogeneously around a global steady

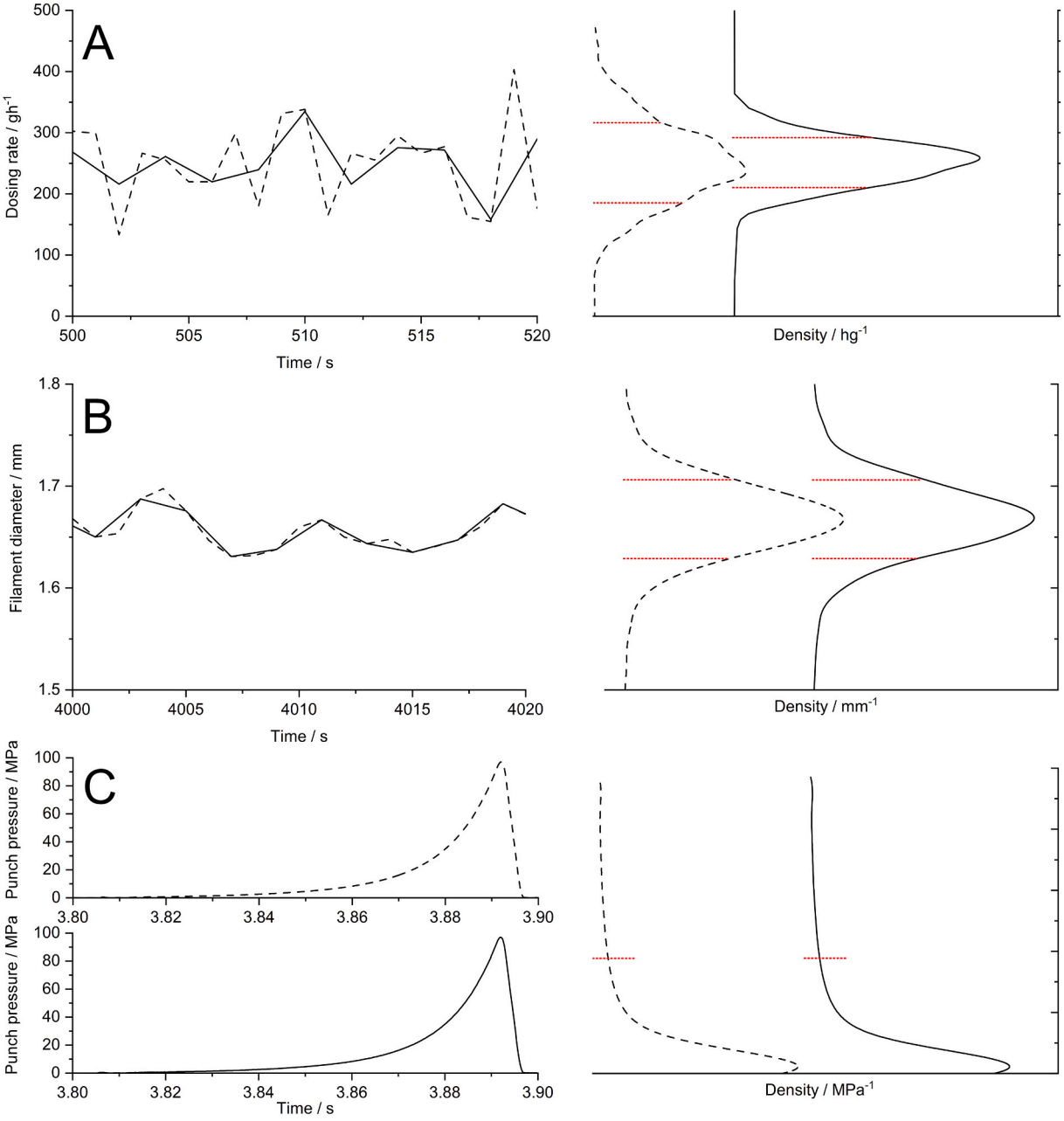


Fig. 1. Data in time domain (left) and kernel PDF (right) at different sample rates, red lines: $\bar{x} \pm s$, A: bulk solid dosing (dashed line: 1 Hz, solid line: 0.5 Hz), B: filament diameter (dashed line: 1 Hz, solid line: 0.5 Hz), C: upper punch pressure of a tableting cycle of Fujicalin® (dashed line: 10^4 Hz, solid line: $5 \bullet 10^3$ Hz). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

state mean or a target value. This results in a symmetric density function of their respective values. As the term "fluctuation" describes the continuous change between one level and another [38], we consider fluctuations to be homogeneous deviations. A different group of time series is global heterogeneous and is characterized by singular or repetitive events that cause asymmetric value distributions. Residence time distribution data or punch force data of tableting cycles are dominated by a heterogenous shape, while temperature profiles, turret speeds of rotary presses, change of mass over time in dosing processes, among others, show mainly homogeneous deviations.

Sinusoidal, singular, chaotic deviations: Sinusoidal deviations in time series show a defined Fourier spectrum. Purely sinusoidal deviations are described by just one sinusoidal function (Eq. (1)), while cumulated sinusoids are a cumulation of multiple sine waves.

$$y(t) = A sin(2\pi f t + \varphi) + C \quad (1)$$

Sinusoidal and cumulated sinusoidal deviations are both fluctuations as they are symmetric in their shape. Furthermore, they show high collinearity between the different measured values. This can be understood in the continuous nature of these data series. Examples can be found in the pattern of the response of controllers, like PID controllers, vibrations, or pressure fluctuations in extruders. Singular events are mostly asymmetric in shape. They are not continuously repetitive like sinusoidal deviations but can also be described by a defined Fourier spectrum. The force time data in tableting processes or the refilling process of gravimetric dosing units are of singular characteristics. Chaotic deviations can be symmetric or asymmetric. The magnitudes of time series data of pure chaotic deviations are realizations of a random variable of a certain probability distribution so that their Fourier spectra

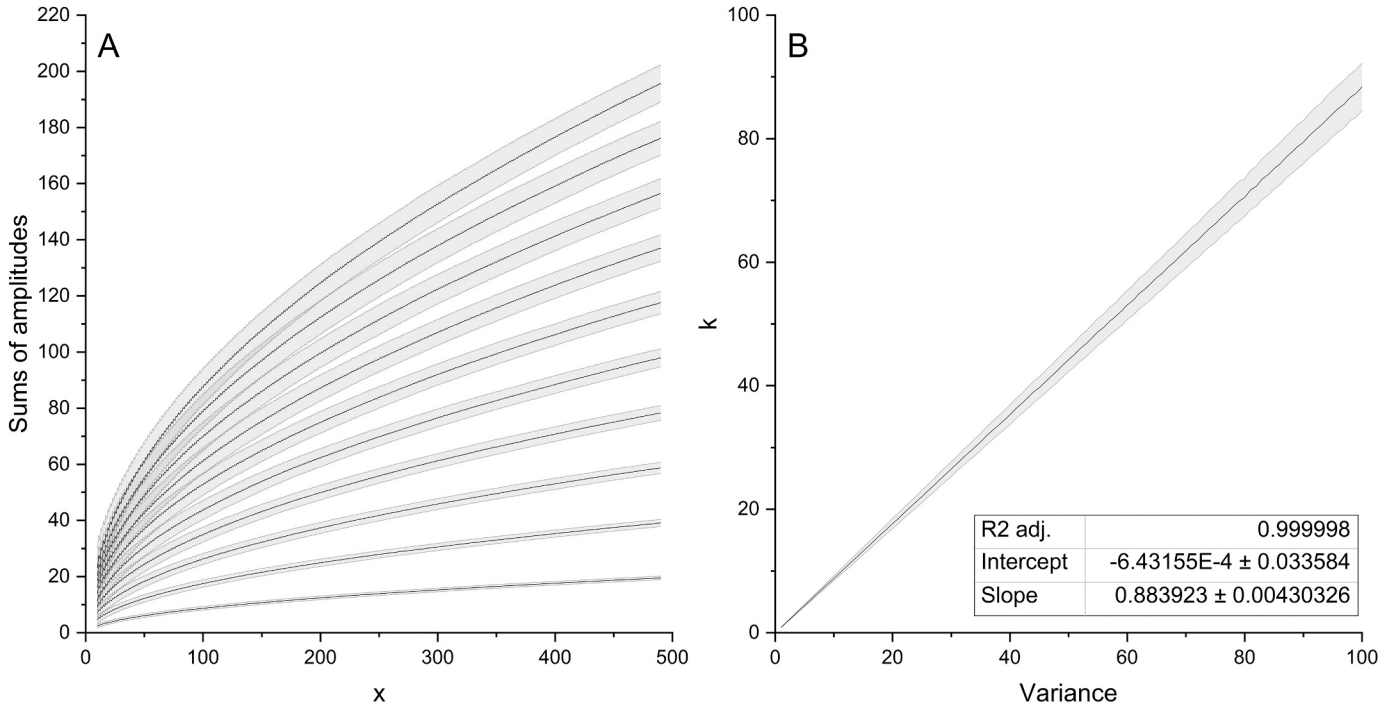


Fig. 2. A: SA of random normal distributed values of different variances, $\sigma^2 = \{1.0, 2.0, \ldots, 10\}$, $\bar{x} \pm s$, $n = 10^3$ B: k factor over the variance of the normal distributions, $\sigma^2 = \{1.0, 2.0, \ldots, 100\}$, $\bar{x} \pm s$, $n = 10^3$.

are equal to random generated values. Chaotic deviations appear as noise in data, but very complex processes like flowing powders on catch scales can also appear as chaotic. In real data, a mixture of these three deviations will be found while a clear demarcation is mostly not possible.

2.2. Sample rate dependency

The sample rate is of high importance regarding the outcome of the evaluation of time series data. Depending on the properties of the data, changing the sample rate can lead to changes in traditional statistical parameters like the standard deviation or the skewedness. This is particularly true for data that are derivatives from fundamental parameters. In these cases, the reduction of the sample rate reduces the magnitude in the time domain of the data. But even in fundamental time series data, a decrease in the sample rate reduces the Nyquist frequency, leading to higher aliasing due to an under-sampling effect [39]. The Nyquist frequency is the highest calculatable frequency of the respective spectrum, given as the half of the applied sample rate. Chaotic information and fluctuations of high frequency can be destructed by applying a low sample rate. In some cases, if the deviation's frequency and the sample rate match exactly, no information can be extracted. Taking the mass flow of a dosing process as an example of a derivative parameter, the change of the probability density function (PDF) of the measured values by the change in the sample rate is striking (Fig. 1A). Whereas the diameter of a filament produced via hot-melt extrusion illustrates the consistency of a fundamental parameter in its distribution (Fig. 1B). The PDF of fundamental parameters hardly changes, even if it is not normal distributed as demonstrates by the compression pressure over time data (Fig. 1C).

3. The sums of amplitude analysis (SAA)

3.1. Fast Fourier transformation

Every mathematical continuous signal or pattern of a certain function $g(x)$ can be described as a cumulation of an infinite number of sine curves of different amplitudes, phases and frequencies. These sine curves are describable by a function in a continuous spectrum named the continuous Fourier transform of this function $\hat{g}(x)$ (Eq. (2)). For real time series data, no exact function is available due to the finite sample rate or the time limitation of the signal. Nevertheless, the transformation of real data into a spectrum is possible by using discrete Fourier transformation (DFT) (Eq. (3)). The fast Fourier transformation algorithm of SciPy 1.7.2, a scientific open-source library for the programming language Python, was used to calculate the discrete frequency domain of time series data in this study. To simplify visualization, spectra will be shown in continuous line plots.

As the amplitude at a certain frequency denotes the amount of information of the time series in the regarding sinusoid, the cumulation of all amplitudes of the Fourier transform reflect the total information in the data.

$$\hat{g}(\xi) = \int_{-\infty}^{\infty} g(x) e^{-2\pi i x \xi} dx \tag{2}$$

$$\hat{g}_m = \frac{1}{N} \sum_{j=0}^{N-1} g_m e^{-2\pi i \frac{jm}{N}} \rightarrow D = \{m \in \mathbb{N} | m = \{0, \ldots, N-1\}\} \tag{3}$$

3.2. Sums of amplitudes analysis calculations

As described earlier, the properties of time series in frequency domain depends highly on the data type and the subgroup of deviations contained in the data. For sinusoidal waves, the frequency domain describes the full information of the data if the following conditions are met. Firstly, the sample rate must be at least greater than double the Nyquist frequency. Secondly, the chosen window size must be at least as large as the largest period in the data. If the window size is increased over this minimal window size, the sums of all amplitudes of one spectrum does not change. No information is added in the frequency domain in this case for larger data sets, as long the as the attributes of the sine functions do not change. For real data, containing not just sinusoidal information but also singular and chaotic events for larger window sizes, the information content obtained in the frequency domain

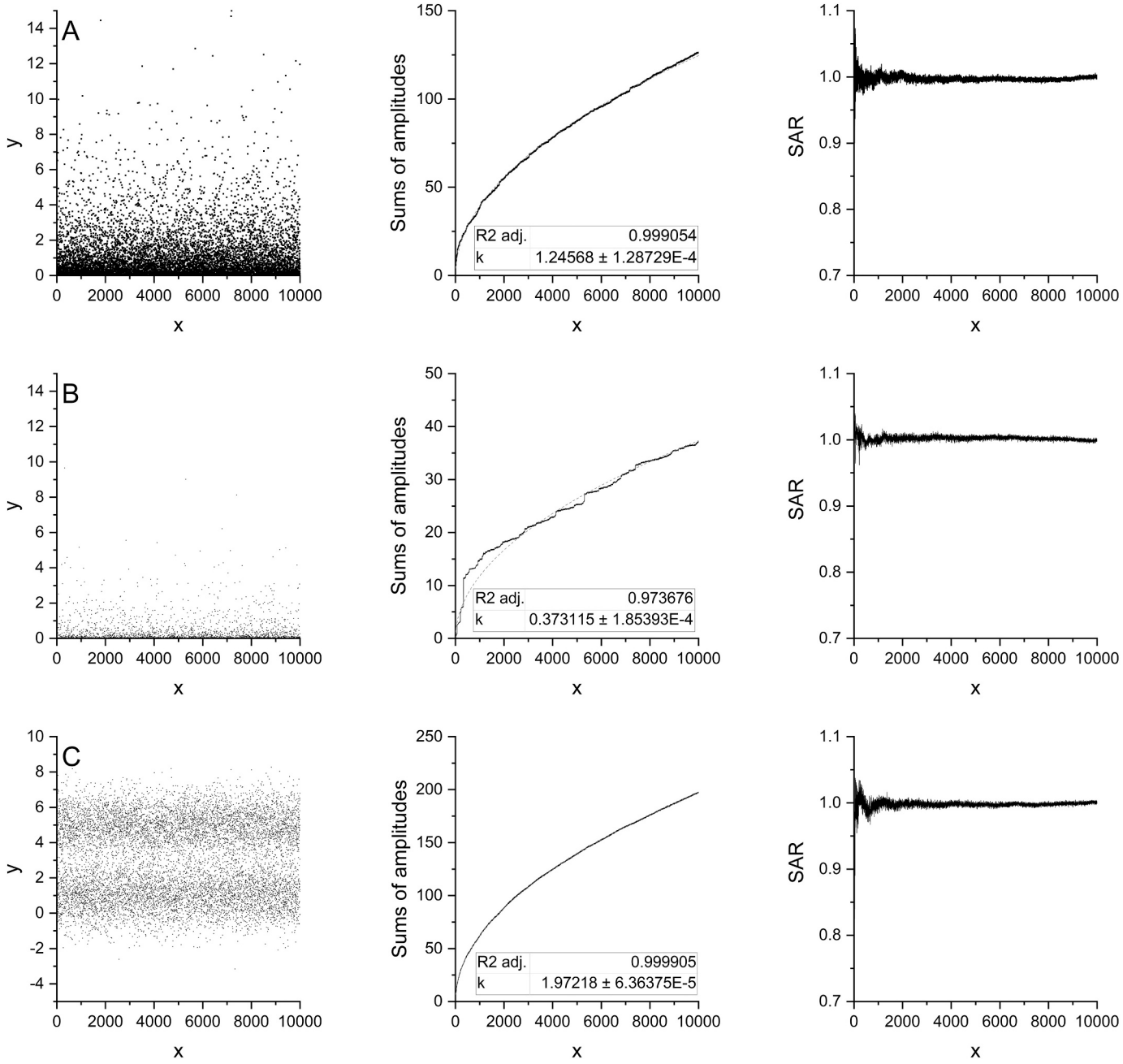


Fig. 3. Simulations of random data, left: time series over x, center: sums of amplitudes over x, right: sums of amplitudes ratio over x, A: $\chi^2(1)$, B: $\chi^2(0.1)$, C: bimodal data $\{\mathcal{N}_1(1, 1), \mathcal{N}_2(5, 1)\}$.

continues to increase with increasing window size. This means that the higher the amount of chaotic and singular events is in the data, the higher is the gain in the sums of amplitudes *SA* in the spectra.

We transferred the data to the frequency domain multiple times with increasing window size from a minimum size of 10 data points. The sums of amplitudes in all spectra were calculated (Eq. (4)). To generate reference data sets, the time series data inside the window size were permuted in time domain. In this permutation step, all sinusoidal information in the data was deconstructed. The sums of amplitudes in frequency domain of the permuted data were calculated as for the original data.

$$SA_N(g) = \sum_{m=1}^{N} \frac{1}{N} \left| \sum_{j=0}^{N-1} g_m e^{-2\pi i \frac{jm}{N}} \right| \rightarrow D = \{N \in \mathbb{N} | N = \{10, \ldots, N_{max}\}\} \quad (4)$$

$$SAR = \frac{SA(g)}{SA(\bar{\delta})} \rightarrow D = \{\delta | \delta \textit{ is the permutation of } g\} \quad (5)$$

Information can be drawn from the plot of the sums of amplitudes of the permuted and original data over the process time of the last included value. The sums of amplitudes ratio *SAR* of the original data and the permuted data over the process time (Eq. (5)) can be used to detect the nature of deviations in the process and indicates the percentage of chaotic and singular events. *1 - SAR* specifies the proportion of sinusoidal and repetitive singular information up to this specific point in time.

Because of the random occurrence of sinusoidal correlations in the permuted reference data, the mean of multiple different permutations of the time series data was used. To generate a “perfect” reference, the maximum SA of all permutations must be used, as this represents the

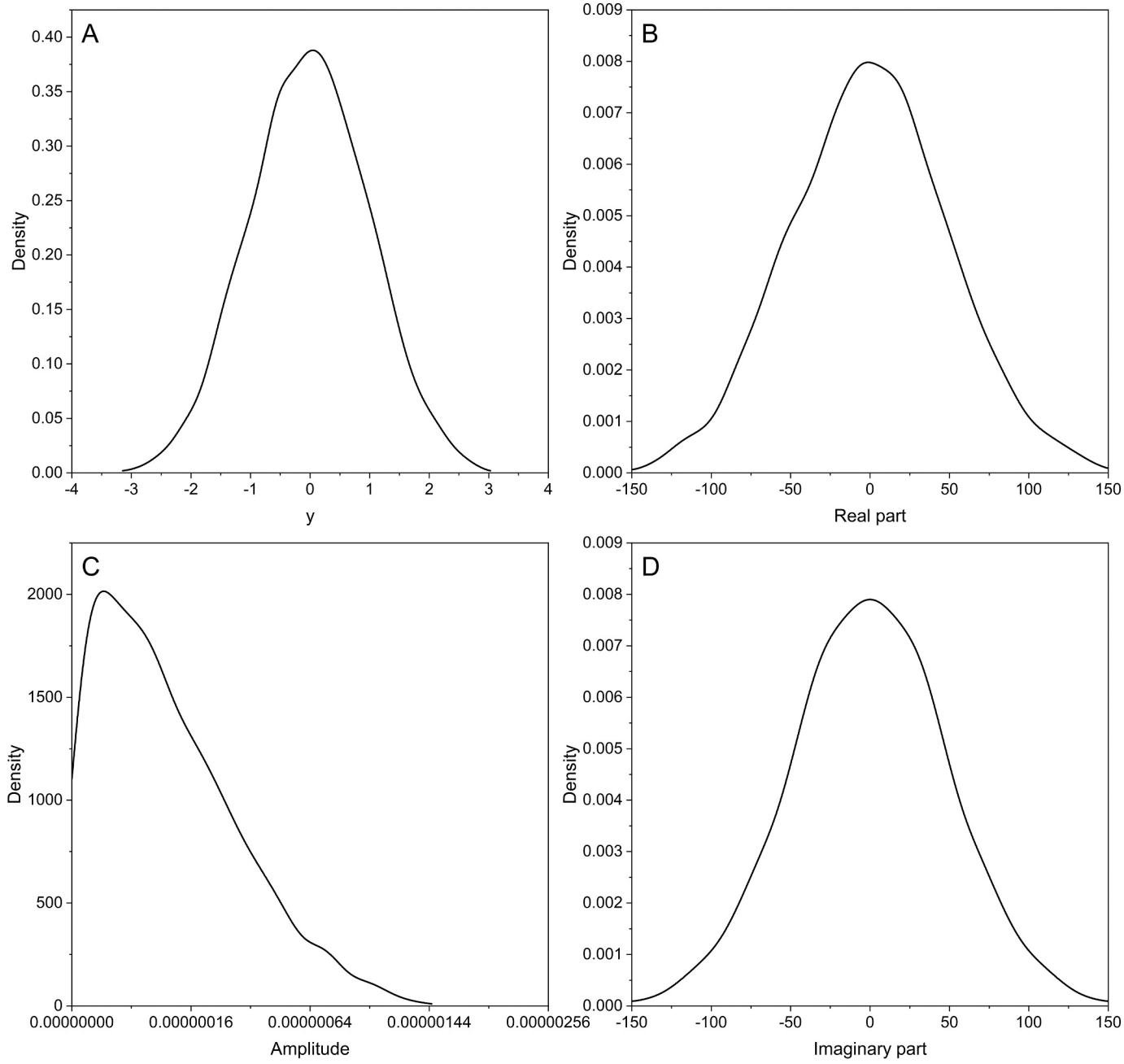


Fig. 4. kernel PDF of A: y ~ $\mathcal{N}$ (0,1), B: real part of ŷ, C: amplitudes of ŷ, D: imaginary part of ŷ.

most chaotic sequence of the values in the time domain. However, since not every possible permutation can be performed due to the required calculation time and the discrete character of the time domain data, the maximum of all calculated permutations is a poor estimator for the true maximum of all possible permutations and the mean is more reliable. Because of this, there is a chance, especially for smaller window sizes, that the sums of amplitudes are larger for the original data than for the mean of permutations. The latter is particularly important for time series of a high content of chaotic events.

4. Application of the sums of amplitudes analysis - material & methods

4.1. Simulation of SAA properties

To evaluate the behavior of data of different distributions, Monte Carlo simulations of the SAA were performed. Normal, equal and χ^2 distributed pseudo-random numbers were generated using NumPy 1.21.4, a mathematical library for Python. As seeds for the generators, the Fibonacci numbers $\{f_2, \ldots, f_{26}\}$ were used if the sample size was set up to 25. For larger sample sizes n, the seed were set to $\{1, \ldots, n\}$. The sums of amplitudes of the random data were calculated for increasing window sizes and evaluated for values included in the respective window size.

4.2. Powder dosing evaluation

The external validation of the dosing of a gravimetric 5 mm twin screw flat-tray feeder (ZD 5 FB, Three-Tec GmbH, Birren, Switzerland) was carried out by collecting the dosed material on a catch scale (CP 2245, Sartorius AG, Goettingen, Germany). Bulk solid material was collected in two different containers. They differ in base diameter d_b and

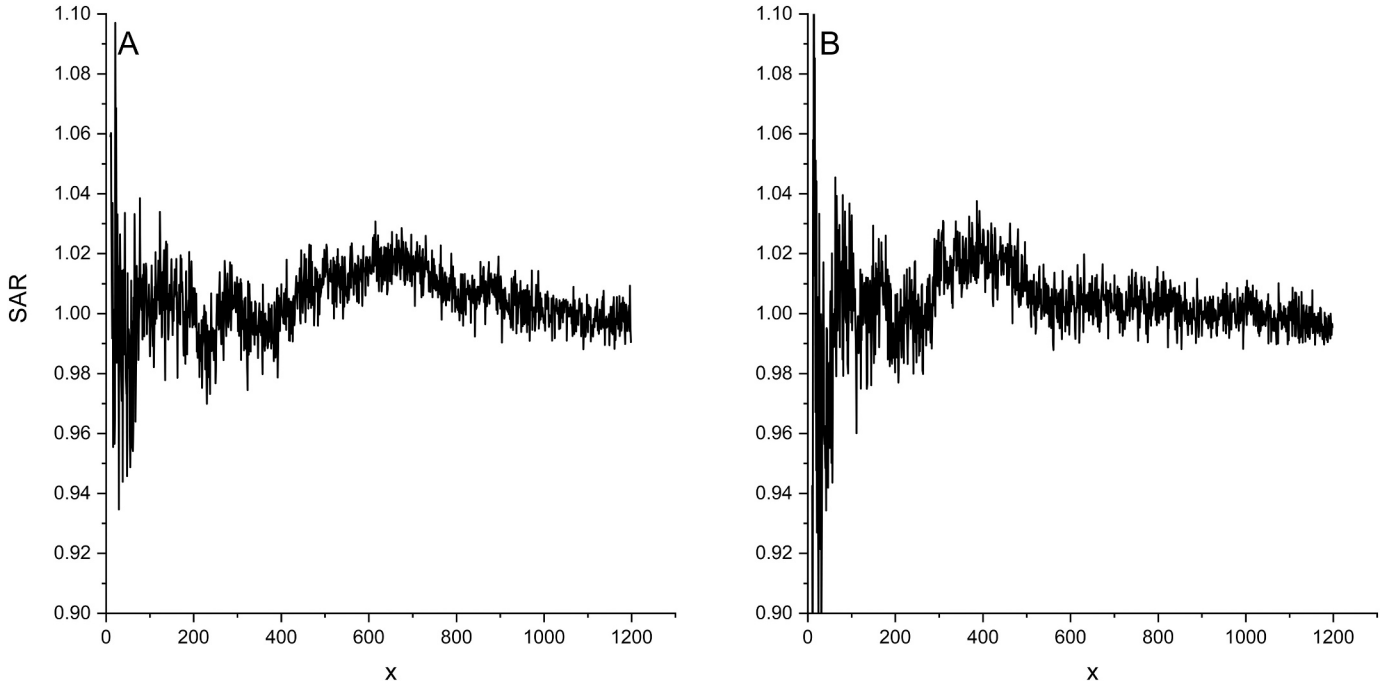


Fig. 5. SAR curves for random generated normal distributed data, A: $\mathcal{N}$ (0, 1), B: $\mathcal{N}$ (0, 10^4).

wall slope s_w: large cup, d_b(LC): 87.6 mm, s_w(LC): 73.7°, small cup, d_b(SC): 50 mm, s_w(SC): 87°. The experiments were named according to the following systematic: Cup size, material name, for example "large cup, Prosolv® ODT". Prosolv® ODT (co-processed microcrystalline cellulose, colloidal silicon dioxide, mannitol, fructose and crospovidone; JRS GmbH & Co. KG) and Flowlac® 100 (lactose; MEGGLE GmbH & Co. KG) were dosed in three different runs at a setpoint of 250 gh^{-1} for 20 min. No flowing aid was used. Data was sampled at a rate f_s of 5 Hz for the catch scale. The internal data of the feeding unit was sampled at 1 Hz and cut off at the Nyquist frequency (0.5 Hz) via a low pass filter. Prior to measurements, the PID controller of the feeding unit was autotuned. The coefficient of correlation (R^2) in the tuning was >0.998 for both materials.

To evaluate the background noise of the dosing process, data was collected while neither the screws nor the agitator at the bottom of the hopper of the device were running.

4.3. Tableting

The tableting data of a single compression cycle was captured on a compaction simulator (STYL'One Evo, Medelpharm, France). 200 mg of Fujicalin® (dibasic calcium phosphate anhydrous; Fuji Chemical Industry Co., LTD) were compressed at 3 mm s^{-1} at a targeted maximal compression pressure of 100 MPa, using 8 mm flat faced EU-D punches. The data was sampled with a rate of 10 kHz. An in-house written Python script was used to detect start and endpoint of the compression cycle.

4.4. Hot-melt extrusion

The hot melt extrusion data were collected in an extrusion process of Parteck® MXP PVA (polyvinyl alcohol; MERCK KGaA) on a 12 mm co-rotating twin-screw extruder (ZSE 12, Leistritz, Nuernberg) at 190 °C. The diameter of the produced filament was measured inline using a 3-axes gauge headed Laser system (LASER 2025 T, Sikora, Bremen) with a sample rate of 1 Hz.

5. Results and discussion

5.1. SAA and Monte Carlo simulation

The mean matrices of the sums of amplitudes of the normal distributed pseudo-random generated data sets follow a square root function (Fig. 2A).

The k-factor in the functional equation (Eq. (6)) correlates linearly with the variance σ^2 of the generated normal distributed data (Fig. 2B).

$$SA = k^*\sqrt{x} \tag{6}$$

The increase in the deviations of the factor k with the increase in σ^2 depends on the rather small sample size of $n = 10^3$ in relation to the respective population variance σ^2. While the intercept seems to be equal to 0, the slope of the function is around 0.884. The R^2_{adj} of the square root fit for the normal distributions was $1–10^{-6}$ and independent of the population mean μ. The sums of amplitudes of equal as well as χ^2 distributed data sets reveal a square root functionality likewise. Fig. 3A and B illustrate that the smaller the degrees of freedom in the χ^2 distribution the higher is the scattering of the sums of amplitudes around the square root function.

The basic square root correlation of the sums of amplitudes on the spectrum resolution seems to be independent of the width or skew of the distribution of the data. Even for bimodal distributed data of two merged normal distributions ($\mathcal{N}_1$(1,1), $\mathcal{N}_2$(5, 1)) this correlation is visible as Fig. 3C shows. Therefore, the permutation of data regardless of the kind of distribution seems to follow this function as long as the global shape of the curve is homogeneous.

A further understanding of this correlation function is gained by the investigation of the Fourier transformed time series in the complex number plain. The resulting point cloud is centered on the origin 0 + 0i. The two-dimensional density function approaches asymptotically a complex normal distribution [9]. As the amplitudes are calculated as the *N* scaled absolute values of the complex numbers by the Pythagorean theorem (Eqs. (7) and (8)), they follow a skewed distribution (Eqs. (9) and (10)).

$$|z| = |a + ib| = \sqrt{a^2 + b^2} \tag{7}$$

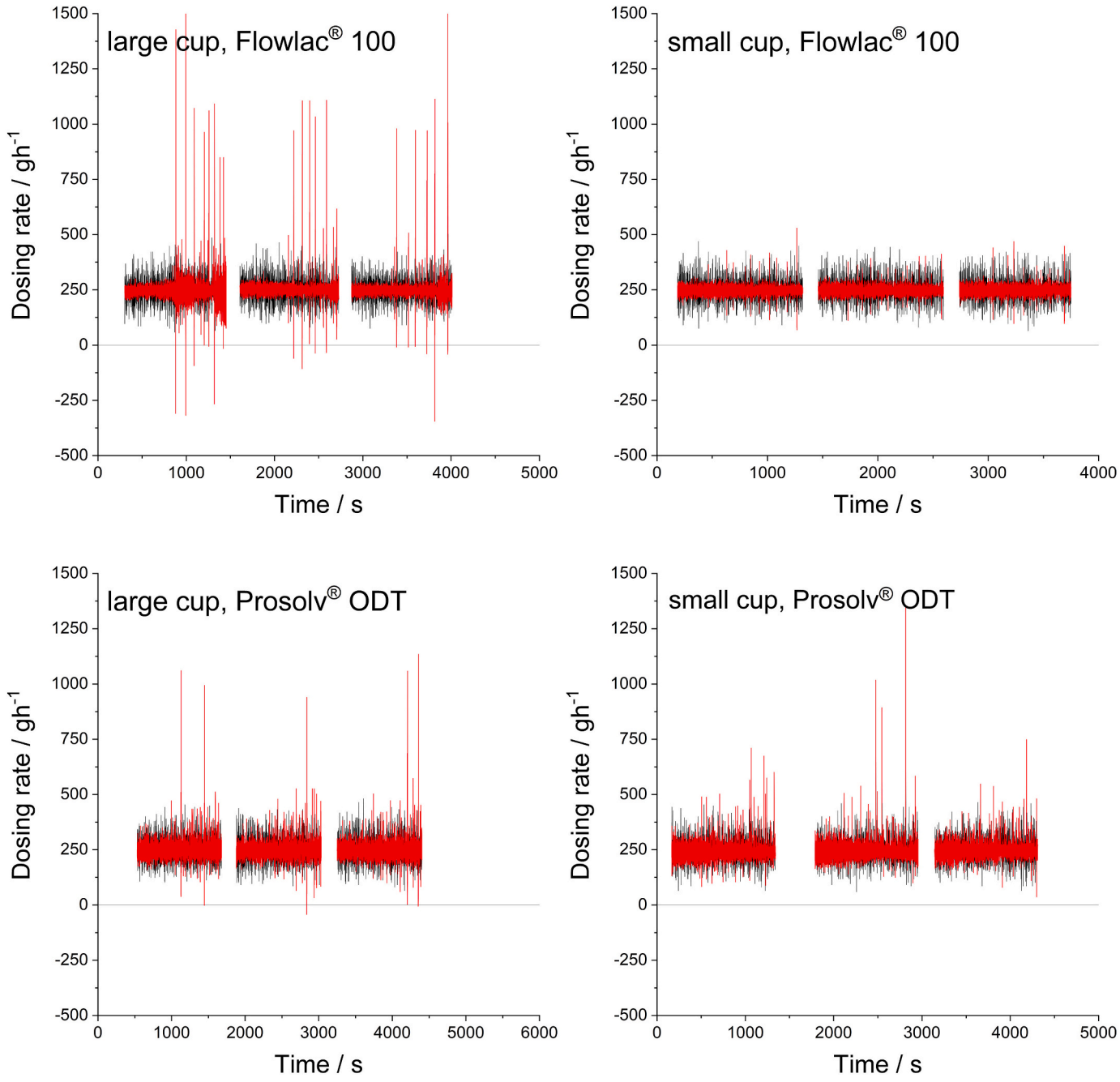


Fig. 6. Dosing curves of the dosing processes, black: intern data, red: extern data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$A = \frac{1}{N}|a + ib| \tag{8}$$

$$PDF = k_1 A e^{-k_2 * A^2} \tag{9}$$

$$CDF = -\frac{k_1 e^{-k_2 A^2}}{2k_2} \tag{10}$$

While the sums of the unscaled amplitudes form a function given by Eq. (11), the scaling results in the described square root dependency (Eq. (12)).

$$\sum_{i=0}^{N} |a + ib| = k^* N^* \sqrt{N} \tag{11}$$

$$\sum_{i=0}^{N} A = k^* \sqrt{N} \tag{12}$$

Fig. 4 illustrates the kernel density functions of the time domain, the real and imaginary part of the complex numbers of the Fourier transform as well as the distribution of the amplitudes of normal distributed random data.

To avoid errors of random collinearity in the random data sets, the mean value of different permutations should be used to deconstruct the sinusoidal information.

The SAR of the various random distributed time series in Fig. 5 show no clear differences in shape between each other. All SAR graphs scatter around SAR = 1. The residuals of this scattering depend on the respective x section. The higher the x values are the lower are the residuals. As

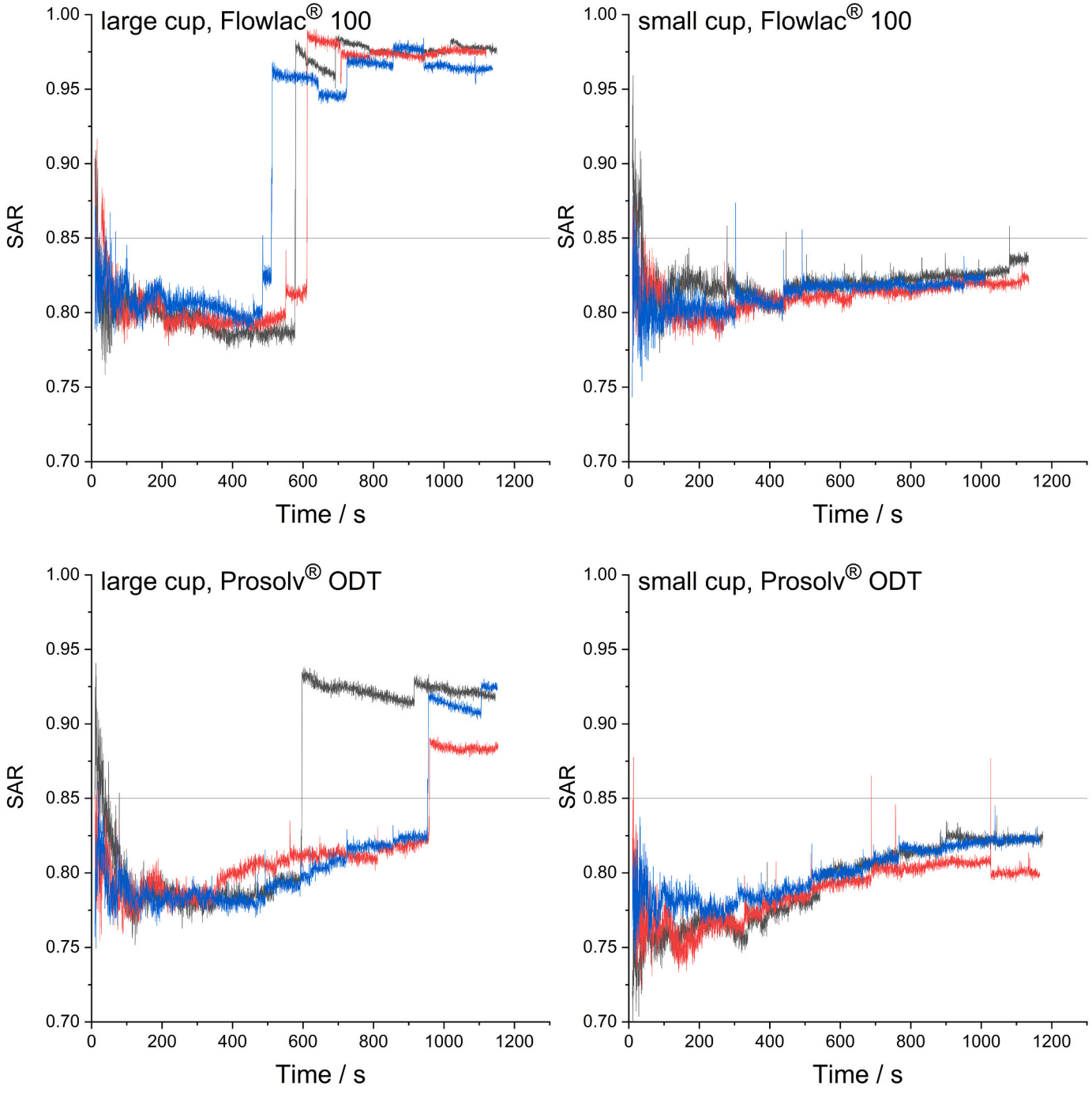


Fig. 7. SAR plots of the external data of the dosing processes.

expected, the conclusion based on these results is that the portion of the chaotic information in a given interval of random generated time series is ≈ 1.

5.2. SAA in dosing unit evaluation

The data of the dosing device evaluation in Fig. 6 reveal a different structure in time domain for the internal and external measurement. While the deviations in the internally measured data are uniform over the whole measurement time, the shape and pattern of deviations changes in the externally measured data. These changes are caused by the forming bulk solid cone inside the used collecting vessel on the external catch scale. Over time the base, lateral area, and the height of the cone enlarged. Therefore, material begins to slip and the quantity of material involved in these slipping increases, which magnifies the deviations measured by the scale. Due to differences in the growth of the bulk solid cones between diverse materials caused by their flow properties, the change in the detected deviations depends on the dosed material. Due to the different base areas of the used cups, the maximum base area of the bulk solid cone is fixed at two different values. This should result in differences in the data structure of the catch scale depending on the size of the vessel the bulk solid is dosed in.

Flowlac® 100 shows a clear dependency on the used cup size on the catch scale. While the dosing in the small cup leads to fewer and smaller deviations and a more uniform pattern, the dosing in the large cup displays a changing point. After a certain time, the deviations increase in

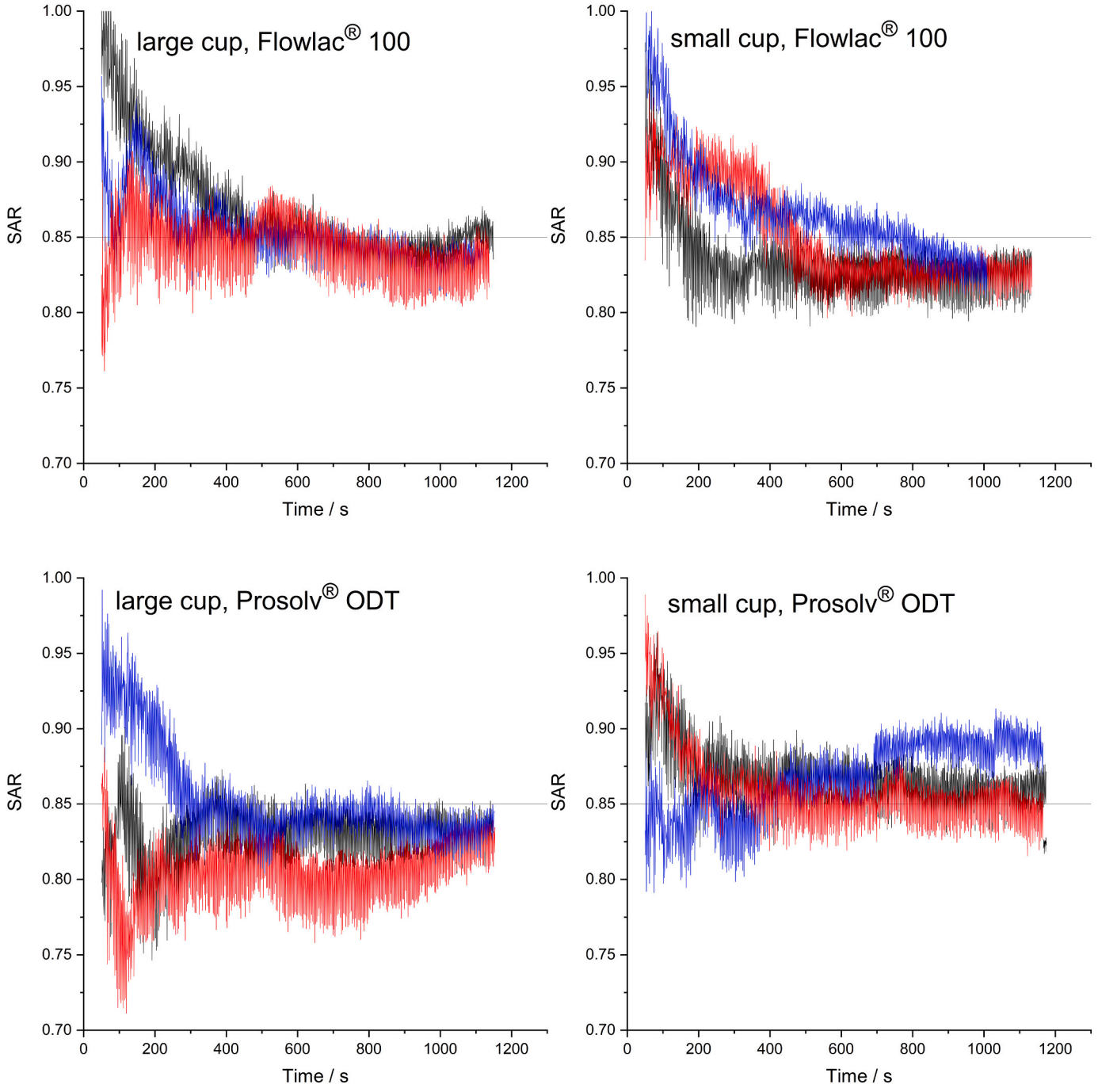


Fig. 8. SAR plots of the internal data of the dosing processes.

magnitude and the pattern changes. The new oscillations are much higher in positive direction than in negative relative to the mean of the dosing curve. This results in a skewed distribution of the data. The external data of Prosolv® ODT show no clear differences between both cup sizes. There seem to be changing points after which the shape and pattern of the deviations change. Similar to Flowlac® 100, a clear difference between the positive and negative amplitudes of the dosing curve is found after the changing point in Prosolv® ODT data.

The SAR curves in Fig. 7 indicate distinct differences in the material dosing processes on the external scale. Most striking is the clear correlation of the SAR graph if the materials are dosed in the large cup. For Flowlac® 100, after 500 to 600 s an increase of the SAR from 0.80 up to 0.95–0.975 indicates the same changing points as in time domain. The same is found for Prosolv® ODT dosed in the large cup. The spread for the onset of the changing point is much wider here, however, the nature of SAR increase is similar to the one for Flowlac® 100. The SAR graphs for both materials are of a different shape if dosing is performed with the small cup on the external catch scale. No clear changing point can be detected even for the times with high deviations in the Prosolv® ODT time domain. If both of the latter are compared with each other, the difference in the slope between the two materials can be found. While the graphs for Flowlac® 100 show some peaks and a small slope up to an SAR of approximately 0.825 after 1000 s, the SAR of Prosolv ODT increases from 0.75 to 0.80 within 1100 s with a higher slope. This indicates a more ordered forming of the bulk solid cone in the beginning of the process for Prosolv® ODT, while in the later process the differences between both materials fade. Therefore the differences seem to be most important in the artification of the data in the beginning of the

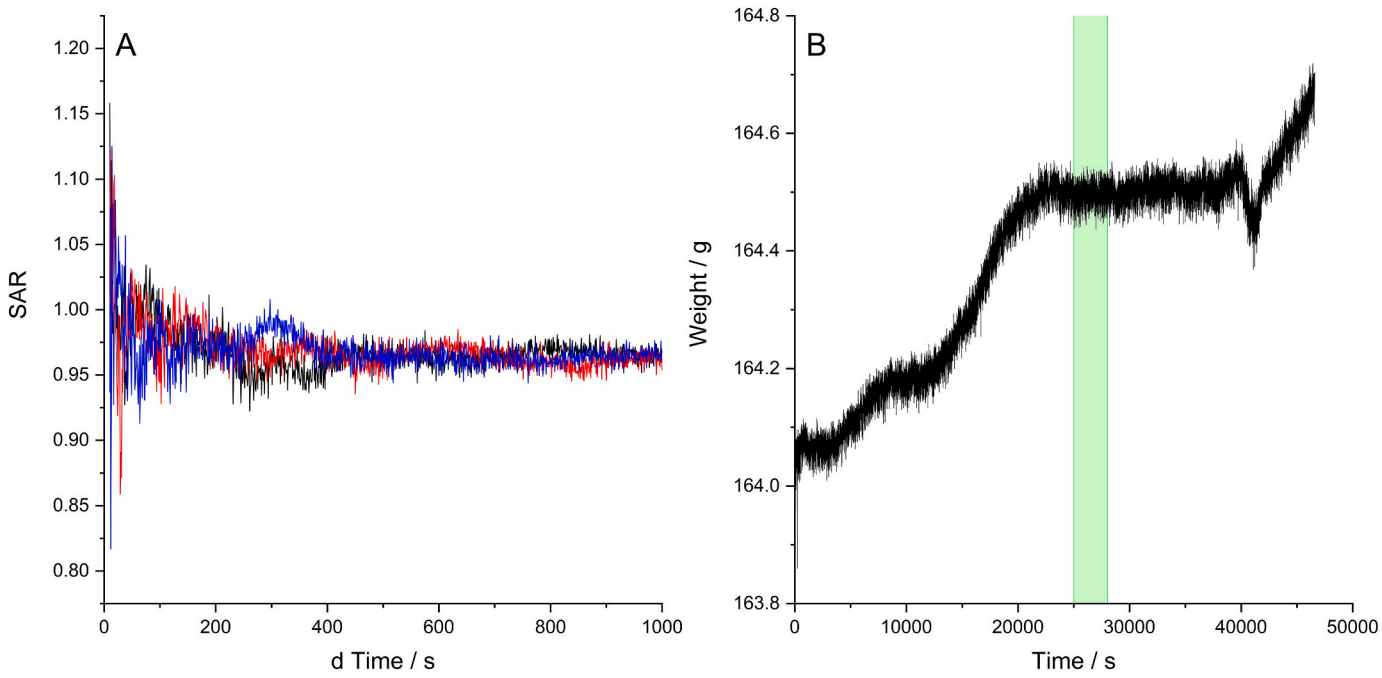


Fig. 9. A: SAR graph of the background noise of the ZD 5 FB device, B: weight over time plot, green area: data used for the SAA. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

evaluation process.

In contrast, the internal data show no changing point, neither in the time domain nor in the SAR. As Fig. 8 shows, the scattering of the SAR data is higher than the external data SAR plot while the terminal SAR values are a little higher compared to the external data after approx. 1100 s. Sinusoidals of high frequency, resulting from the rotating screws of the device, in the internal data might be one explanation for the broader scattering. Both materials hardly differ in the SAR curves indicating comparable dosing results. This is in direct contrast to the SAR data of the external catch scale. It can be concluded that this commonly used method for bulk solid dosing evaluation generates different artificial data structures depending on the dosed material and, even more so, the size of the collecting vessel. The magnitude of these data artefacts depends on the duration on the evaluation process as the change in the shape of the bulk solid cone and therefore the character of the material slipping changes over time. All SAR show definite differences to the simulated graphs of random generated time series, so that the portion of sinusoidal information in all evaluation is significantly higher than zero. Therefore, all time series of internally collected data of the dosing process consists of a certain portion of sinusoidal deviations. This may result from the uniform revolution of the screws or the wave shaped response of the used PID controller.

The comparison of these results to the SAR graphs of the background noise indicates a higher amount of sinusoidal deviations in the actual dosing process. However, the noise SARs in Fig. 9 fluctuate around ca. 0.95 showing that the portion of chaotic noise is not 100%. This is explainable by the respective weight over time plot. The noise measurement was performed overnight. Changes in temperature or in the ventilation of the location of the device might result in this SAR. As shown before in the simulation with random data, the SAR of real chaotic processes will show the total absence of sinusoidal information by a value of 1. This indicates the sensitivity of the sums of amplitudes analysis to detect systematic changes in time series data, even if they have a high proportion of chaotic noise.

6. Conclusion

A new systematic and terminological approach to classify different process deviations and describe their properties in time and frequency domain is presented and substantiated by the introduction of a novel, non-parametric classification method, the SAA. Due to the independence from the distribution of the data the SAA can be used for all kind of time series. The cause for artification in most classical methods, the collinearity of the data, is investigated to classify the time series regarding their portion of sinusoidals. Possible future use cases are the implementation of SAA in controlling loops of continuous processes or the evaluation of causes for the process deviations. The upcoming awareness for the special properties of time series data is a step forward in the rectification of misconceptions in the assessment of process data in pharmaceutical context. This does not only concern the evaluation of bulk solid dosing processes but various standard procedures like twin-screw extrusion and granulation, tableting, mixing, or milling. The authors would like to highlight that the impact of the information obtained from the SAA on the consistency of the evaluation results may differ considerably between various studies. This depends on the properties of the evaluated time series in the individual study.

Based on the information drawn from the development of the SAA the authors want to supplement the workflow for considerations and parameter selection for bulk solid dosing evaluations provided by Kruisz, J. et al. [35] with the following points:

- The sample rate of all devices should be kept the same for different measurements to ensure comparability of different processes and devices. This is particularly important for derivative parameters such as rates.
- The measurement system must be able to detect the expected change of the measured system in the given sample rate. A scale for example must have an accuracy which is sufficient to measure the change in mass over the reciprocal of the sample rate.
- The duration of dosing evaluations can change the properties of the bulk solid cone on the external scale. This must be considered, to avoid misinterpretations. As long as the SAR of the actual process is not in a steady state, not all characteristic deviations of the process are captured.

CRediT authorship contribution statement

Stefan Klinken: Conceptualization, Methodology, Software, Validation, Writing – original draft, Visualization. **Julian Quodbach:** Funding acquisition, Supervision, Project administration, Conceptualization.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Julian Quodbach reports financial support was provided by Bundesministerium für Wirschaft und Energie.

Data availability

Data will be made available on request.

Acknowledgement

The authors would like to thank the Bundesministerium für Wirtschaft und Energie, the Arbeitsgemeinschaft industrieller Forschungsvereinigungen “Otto von Guericke” e.V. (AiF), and the Forschungsvereinigung Arzneimittel-Hersteller e.V. for funding and organizing the project (IGF project nr. 20574 N / 1). The authors would also like to acknowledge Jörg Breitkreutz, Peter Kleinebudde, Valentinë Elezaj and Marcel Kokott for their fruitful comments and words of encouragement and Sabrina Berkenkemper for providing tableting data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.powtec.2022.118003.

References

[1] M. Fonteyne, J. Vercruysse, F. de Leersnyder, B. van Snick, C. Vervaet, J.P. Remon, T. de Beer, Process analytical technology for continuous manufacturing of solid-dosage forms, TrAC Trends Anal. Chem. 67 (2015) 159–166.
[2] J. Gough, D. Nettleton (Eds.), Managing the Documentation Maze, John Wiley & Sons, Inc, Hoboken, NJ, USA, 2010.
[3] V. Elezaj, A. Lura, L. Canha, J. Breitkreutz, Pharmaceutical development of film-coated mini-tablets with losartan potassium for epidermolysis bullosa, Pharmaceutics 14 (2022).
[4] J. Krieghoff, J. Rost, C. Kohn-Polster, B.M. Müller, A. Koenig, T. Flath, M. Schulz-Siegmund, F.-P. Schulze, M.C. Hacker, Extrusion-printing of multi-channeled two-component hydrogel constructs from gelatinous peptides and anhydride-containing oligomers, Biomedicines 9 (2021).
[5] M. Verstraeten, D. van Hauwermeiren, K. Lee, N. Turnbull, D. Wilsdon, M. am Ende, P. Doshi, C. Vervaet, D. Brouckaert, S.T.F.C. Mortier, I. Nopens, T. de Beer, In-depth experimental analysis of pharmaceutical twin-screw wet granulation in view of detailed process understanding, Int. J. Pharm. 529 (2017) 678–693.
[6] R. Lehtonen, K. Djerf, Eurostat Sampling Reference Guidelines: Introduction to Sample Design and Estimation Techniques, 2008.
[7] R. Latpate, J. Kshirsagar, V. Kumar Gupta, G. Chandra, Advanced Sampling Methods, Springer Singapore, Singapore, 2021.
[8] World Health Organization, WHO Guidelines for Sampling of Pharmaceutical Products and Related Materials. https://www.who.int/publications/m/item/who-guidelines-for-sampling-of-pharmaceutical-products-and-related-materials, 2005. Accessed 26 September 2022.
[9] D.R. Brillinger, Time Series: Data Analysis and Theory, Society for Industrial & Applied Mathematics, 2001.
[10] H. Ponsar, R. Wiedey, J. Quodbach, Hot-melt extrusion process fluctuations and their impact on critical quality attributes of filaments and 3D-printed dosage forms, Pharmaceutics 12 (2020).
[11] R. Meier, M. Thommes, N. Rasenack, K.-P. Moll, M. Krumme, P. Kleinebudde, Granule size distributions after twin-screw granulation - do not forget the feeding systems, Eur. J. Pharm. Biopharm. 106 (2016) 59–69.
[12] W.E. Engisch, F.J. Muzzio, Method for characterization of loss-in-weight feeder equipment, Powder Technol. 228 (2012) 395–403.
[13] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, P.-A. Muller, Deep learning for time series classification: a review, Data Min. Knowl. Disc. 33 (2019) 917–963.
[14] M.M. Crowley, F. Zhang, M.A. Repka, S. Thumma, S.B. Upadhye, S.K. Battu, J. W. McGinity, C. Martin, Pharmaceutical applications of hot-melt extrusion: part I, Drug Dev. Ind. Pharm. 33 (2007) 909–926.
[15] C.A. Blackshields, A.M. Crean, Continuous powder feeding for pharmaceutical solid dosage form manufacture: a short review, Pharm. Dev. Technol. 23 (2018) 554–560.
[16] R. Meier, J. Harting, J. Happel, P. Kleinebudde, Implementation of microwave sensors in continuous powder feeding – a novel tool to bridge refill phases, Pharmazeutische. Industrie. (2017) 576–582.
[17] W.E. Engisch, F.J. Muzzio, Feedrate deviations caused by hopper refill of loss-in-weight feeders, Powder Technol. 283 (2015) 389–400.
[18] Göb Pöppel, Kopp Hellberg, Nuding Mundil, Vetter Rabe, NA 40: Dosing Accuracy of Continuous Scales, NAMUR-Arbeitskreis 3.3 “Wägetechnik”, 2006.
[19] S.-P. Simonaho, J. Ketolainen, T. Ervasti, M. Toiviainen, O. Korhonen, Continuous manufacturing of tablets with PROMIS-line - introduction and case studies from continuous feeding, blending and tableting, Eur. J. Pharm. Sci. 90 (2016) 38–46.
[20] B. van Snick, J. Holman, V. Vanhoorne, A. Kumar, T. de Beer, J.P. Remon, C. Vervaet, Development of a continuous direct compression platform for low-dose drug products, Int. J. Pharm. 529 (2017) 329–346.
[21] N. Bostijn, J. Dhondt, A. Ryckaert, E. Szabó, W. Dhondt, B. van Snick, V. Vanhoorne, C. Vervaet, T. de Beer, A multivariate approach to predict the volumetric and gravimetric feeding behavior of a low feed rate feeder based on raw material properties, Int. J. Pharm. 557 (2019) 342–353.
[22] C. Muehlenfeld, M. Thommes, Miniaturization in pharmaceutical extrusion technology: feeding as a challenge of downscaling, AAPS PharmSciTech 13 (2012) 94–100.
[23] M. Gyürkés, L. Madarász, Á. Köte, A. Domokos, D. Mészáros, Á.K. Beke, B. Nagy, G. Marosi, H. Pataki, Z.K. Nagy, A. Farkas, Process design of continuous powder blending using residence time distribution and feeding models, Pharmaceutics 12 (2020).
[24] L. Madarász, Á. Köte, M. Gyürkés, A. Farkas, B. Hambalkó, H. Pataki, G. Fülöp, G. Marosi, L. Lengyel, T. Casian, K. Csorba, Z.K. Nagy, Videometric mass flow control: a new method for real-time measurement and feedback control of powder micro-feeding based on image analysis, Int. J. Pharm. 580 (2020), 119223.
[25] B. Santos, F. Carmo, W. Schlindwein, G. Muirhead, C. Rodrigues, L. Cabral, J. Westrup, K. Pitt, Pharmaceutical excipients properties and screw feeder performance in continuous processing lines: a quality by design (QbD) approach, Drug Dev. Ind. Pharm. 44 (2018) 2089–2097.
[26] B.J. Johnson, M. Sen, J. Hanson, S. García-Muñoz, N.V. Sahinidis, Stochastic analysis and modeling of pharmaceutical screw feeder mass flow rates, Int. J. Pharm. 621 (2022), 121776.
[27] B. Bekaert, L. Penne, W. Grymonpré, B. van Snick, J. Dhondt, J. Boeckx, J. Vogeleer, T. de Beer, C. Vervaet, V. Vanhoorne, Determination of a quantitative relationship between material properties, process settings and screw feeding behavior via multivariate data-analysis, Int. J. Pharm. 602 (2021), 120603.
[28] M. Beretta, T.R. Hörmann, P. Hainz, W.-K. Hsiao, A. Paudel, Investigation into powder tribo-charging of pharmaceuticals, Part I: Process.-Induc. Charge Via Twin-Screw Feed., Int J. Pharm. 591 (2020), 120014.
[29] M.S. Escotet-Espinoza, J.V. Scicolone, S. Moghtadernejad, E. Sanchez, P. Cappuyns, I. van Assche, G. Di Pretoro, M. Ierapetritou, F.J. Muzzio, Improving Feedability of highly adhesive active pharmaceutical ingredients by Silication, J. Pharm. Innov. 16 (2021) 279–292.
[30] S. Fathollahi, S. Sacher, M.S. Escotet-Espinoza, J. DiNunzio, J.G. Khinast, Performance evaluation of a high-precision low-dose powder feeder, AAPS PharmSciTech 21 (2020) 301.
[31] B. El Kassem, T. Brinz, V. Jenkouk, Y. Heider, B. Markert, Design of a vertical loss-in-weight feeder prototype with experimental proof of concept validation, Pharm. Dev. Technol. 26 (2021) 559–575.
[32] S. Sacher, S. Fathollahi, J.G. Khinast, Comparative study of a novel Micro-feeder and loss-in-weight feeders, J. Pharm. Innov. (2021), https://doi.org/10.1007/s12247-021-09599-6.
[33] B. Bekaert, W. Grymonpré, A. Novikova, C. Vervaet, V. Vanhoorne, Impact of blend properties and process variables on the blending performance, Int. J. Pharm. 613 (2022), 121421.
[34] E.-J. Wagenmakers, A. Sarafoglou, B. Aczel, One statistical analysis must not rule them all: any single analysis hides an iceberg of uncertainty. Multi-team analysis can reveal it, Nature (2022) 423–425.
[35] J. Kruisz, J. Rehrl, T.R. Hörmann-Kincses, J.G. Khinast, Effects of signal processing on the relative standard deviation in powder feeding characterization for continuous manufacturing, Powder Technol. 389 (2021) 536–548.
[36] J.J. Cartwright, J. Robertson, D. D'Haene, M.D. Burke, J.R. Hennenkamp, Twin screw wet granulation: loss in weight feeding of a poorly flowing active pharmaceutical ingredient, Powder Technol. 238 (2013) 116–121.
[37] R. Chamberlain, H. Windolf, S. Geissler, J. Quodbach, J. Breitkreutz, Precise dosing of Pramipexole for low-dosed filament production by hot melt extrusion applying various feeding methods, Pharmaceutics 216 (2022) 1–17.
[38] Cambridge Dictionary. https://dictionary.cambridge.org/de/worterbuch/englisch/fluctuation, 2014.
[39] H. Nyquist, Certain topics in telegraph transmission theory, Proc. IEEE 90 (1928) 280–305.