Trends in Chromatography
Vol. 8, 2013

Original Communication

Physical characterization of column chromatography: Stringent control over equipment performance in biopharmaceutical production

C. Helling[1], T. Dams[2], B. Gerwat[2], A. Belousov[2] and J. Strube[1]

[1]Clausthal University of Technology, Institute for Separation and Process Technology, Leibnizstr. 15, D-53678 Clausthal-Zellerfeld, Germany, [2]Roche Diagnostics GmbH, Nonnenwald 2, D-82377 Penzberg, Germany

ABSTRACT

Column chromatography is of central importance for the manufacturing of biotechnologically derived active pharmaceutical ingredients. Stringent monitoring of the equipment used in those steps is crucial to successful and reliable manufacturing. Packed bed integrity in particular needs to be under continuous monitoring due to its criticality in the separation process. We propose a mathematical model that describes the inherent characteristics of the separation process, especially adapted to the circumstances of large-scale column chromatography. Parameters derived from that model allow a sensitive monitoring of the packed bed integrity and of the fluid dynamics involved. Changes in these parameters either seen as drifts, or sudden in nature can be interpreted mechanistically, because the circumstances, the respective parameter is influenced by, are known. Such knowledge leads to a better process comprehension for routine operations. The integrated form of the exponentially modified Gaussian distribution function is used to model the signal. Data from breakthrough curves from the column outlet can therefore be used to monitor the process continuously with high sensitivity towards perturbations within and around the packed matrix. Thus, orthogonal methods, in addition to the routine in-process controls, are installed in full agreement with health authorities' expectations towards improving process monitoring and process knowledge throughout the lifetime of the manufacturing processes.

KEYWORDS: biotechnology, chromatography, downstream processing, process analytical technology (PAT), statistical process control (SPC)

1. INTRODUCTION

In biotechnology manufacturing processes, the separation of active pharmaceutical ingredients (API), column chromatography is of central importance. Stringent monitoring of used equipment is crucial to successful and reliable manufacturing, associated with high process robustness. In the recent years this progress was promoted by the Food and Drug Administration (FDA) and the European Medicines Agency (EMA) with Process Analytical Technology (PAT) and Quality by Design (QbD) initiatives [1-4].

We propose a method for monitoring the physical conditions of packed bed integrity for chromatographic equipments by utilizing online trend data gathered in the routine process.

Currently, packing quality is often determined by a graphical evaluation of the HETP (height equivalent to a theoretical plate) values gathered from discrete HETP determinations using injections of tracer substances like solutions with high conductivity or UV absorbance [5]. Tracer experiments are time/cost consuming and results often depend on a manual/individual interpretation. Using breakthrough curves as an alternative has been proposed before [5], but the strict application of models to describe the shape of the breakthrough curve by using

parameters also used in HETP determinations has not been applied yet. The aim of this work is to make unambiguous decisions based on mathematically exact interpretations of detector signals from the column outlet.

With this method economic losses due to sudden bed decay and equipment errors can be identified early by applying statistical process control (SPC)-trend analysis or instituting scrutiny in analysis of breakthrough curves. Feed losses by loading a column with insufficient separation performance are avoided and time for individual tracer experiments is reduced, since replacement of separate HETP checks in the routine decreases efforts for extra steps. Long-term studies can give significant insight on process robustness and physical condition of the chromatography column. Detailed comprehension of arising trends or packed bed deficiencies can significantly speed up the reconstitution of regular processing condition.

Work avoidance and increase of production reliability lead to a positive business case for implementing such technologies.

In this work different methods are described and performed to analyse separation performance and column integrity. Mathematical functions that describe the breakthrough behaviour during buffer exchange steps are investigated. In cooperation with a routine manufacturer of 4 major biotechnology products on site and further products in development, several chromatography steps were investigated. These include several types of typical industrial chromatography resins. Long-term runs (up to 250 cycles) were examined as proof of concept and method validation. From those studies, predictive conclusions could be drawn with respect to the cycle number at which the resin lost the defined quality/performance and when the column had to be re-packed with regenerated or new resin.

2. MATERIALS AND METHODS

2.1. Mathematical fundamentals

Mathematical analysis was performed using two approaches. The first approach is based on an ideal packing where mass transport (pore diffusion and mass transport through the film) is not limited, which results in a bell-shaped Gaussian curve in HETP determinations. The second approach is a high accuracy description of the conditions where, besides the diffusion, other extra-column mixing effects occur as described in Guichon *et al.* [6].

2.1.1. Gaussian description of elution peaks

To describe the shape of symmetric elution peaks, the Gaussian function is implemented [7], see equation 1. When applying this equation, an ideal plug flow (non-mixing areas) is assumed outside the packed bed of the column. Based on the plate theory it can be concluded that after a certain length of time, the zone profile becomes Gaussian in shape [8].

$$f(x) = \frac{1}{s\sqrt{2\pi}} \cdot \exp\left[-\frac{(x-m)^2}{2s^2} \right] \tag{1}$$

The parameter s describes the standard deviation and parameter m is the centroid of the peak that is equal to the mean retention time. For symmetrical peaks this correlates to the abcissa value of the peak amplitude.

In Figure 1 the impact of both parameters is depicted. The black line shows the initial point regarding an m-value of 20 and an s-value of 5. Variations of the m-value lead to shifts in abscissa. The standard deviation has an impact on peak width and peak height. Decreasing s-values result in narrower peaks with higher amplitude and for increasing s-values the amplitude is reduced as well as the peak width is increased.

2.1.2. Exponentially modified Gaussian description of elution peaks

A model for describing the asymmetric shape of elution peaks that accounts for peak broadening induced by mixing effects outside the packed bed in addition to the diffusion controlled peak broadening was given by Guichon *et al.* [9]. In this study, a mathematical analysis of fronting was excluded. If fronting takes place during the chromatographic separation, the quality of the packed bed was defined as insufficient. For our purposes all columns are routinely packed and operated long term without fronting. This description of asymmetric elution peaks is modelled by the exponential modified Gaussian (EMG) distribution [6].

$$f(x) = \frac{A}{2t_0} \cdot \exp\left(\frac{1}{2} \cdot \left(\frac{s}{t_0} \right)^2 - \frac{x-m}{t_0} \right) \cdot erf\left(\frac{\left(\frac{x-m}{s} - \frac{s}{t_0} \right)}{\sqrt{2}} + 1 \right) \tag{2}$$

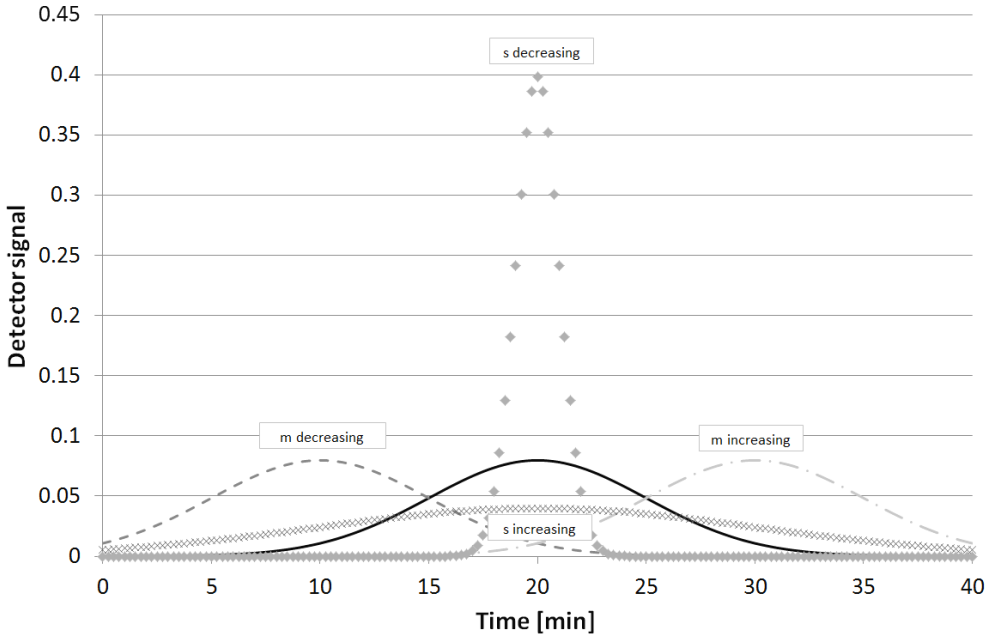

Figure 1. Influence of parameter m (mean retention time) and s (standard deviation) to Gaussian peaks.

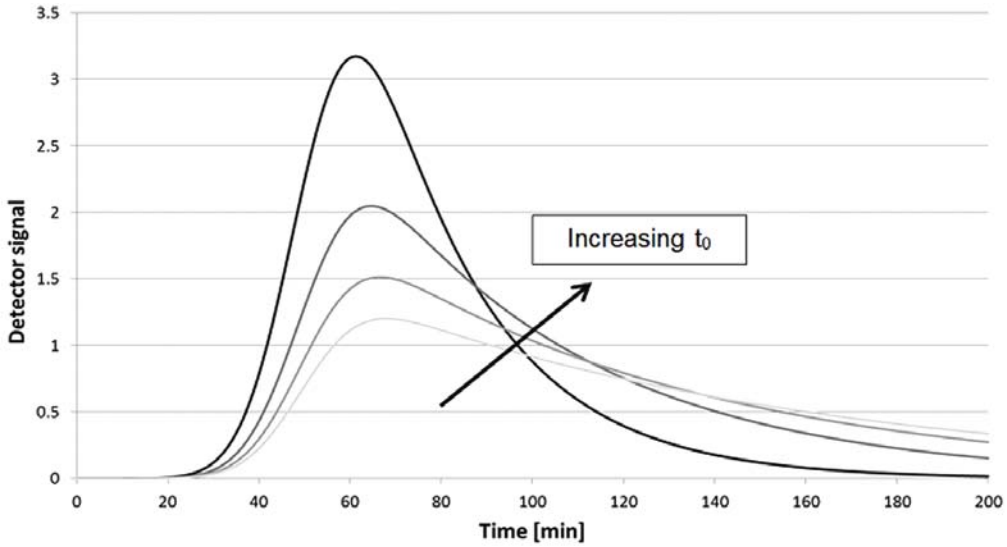

Figure 2. Influence of extra column mixing effects (t_0) to the peak shape.

In comparison to the Gaussian distribution, a time constant describing the decay time of extra-column mixing effects (t_0) is added. For a simplification, in the following text this parameter will be called tailing factor. Parameter A is proportional to the area under the peak. Other parameters are equivalent to the Gaussian distribution. The EMG function is able to describe tailing effects, in alignment with physical conditions of non-ideal steps outside and inside the column. Figure 2 depicts the influences of parameter t_0.

An increasing t_0-value leads to a higher tailing effect; thereby, the peak height decreases as well as the peak width increases. Note that only one more

mathematical constant is introduced to describe separate effects. Thus, the time constant of extra-column mixing (t_0), the retention time m and the standard deviation s of the Gaussian function (corresponding to the diffusion and dispersion terms, if mass transport limitation can be excluded) are sufficient to describe the shape of a breakthrough curve. The influence of the isotherm to the breakthrough curve can be neglected since for all runs the process conditions were kept constant and thus the isotherm has a constant impact to the peak shape. Furthermore, the usage of a breakthrough curve with salts in the washing step is preferable, since no significant adsorption occurs.

2.1.3. Integration of functions

A direct analysis of breakthrough curves is feasible with a multiplicity of equations proposed in literature [10-12]. Thereby, also a curve fitting prediction for multi component mixtures in the overlapping area [13] as well as for the non-overlapping area [14] is feasible. Mostly the skewness and the kurtosis are used to determine variations in the breakthrough curve [15].

As described previously, in this work we focused on the mathematical description of the breakthrough behaviour based on the Gaussian distribution and the EMG function.

The analysis of breakthrough curves of solvent changes at the column outlet can be performed by taking the first derivative of the signal at the column outlet and interpreting this bell-shaped curve similar to a HETP determination. Studies using this approach have shown a major drawback: Calculating the first derivative of conductivity signals leads to high noise in the bell-shaped curve due to the inflated noise of the physical signal itself after derivation. Although, this effect could be overcome by applying curve smoothing algorithms, this was not applied, because internal studies have shown that the smoothing algorithm leads to artificial modification of the raw data.

We took the reverse approach; instead of taking the first derivative of a breakthrough curve and fitting those data to a Gaussian (equation 1) or EMG (equation 2) function, we established the integrated forms of the Gaussian (equation 3) or EMG (equation 4) function, respectively and fitted the raw data to these equations.

The integration of the functions used above was performed using the online Version of Mathematica [16] http://integrals.wolfram.com/. Those integrations were used in the further development of the fitting functions.

The integrated form of the Gaussian distribution was calculated as:

$$F(x) = \frac{1}{2} \cdot A \cdot \left[1 + erf\left(\frac{x-m}{\sqrt{2} \cdot s} \right) \right] + y \qquad (3)$$

The introduced term y allows for an adjustment of the breakthrough curve in ordinate direction. The parameter A is correlated to the maximum height of the breakthrough curve. Additional parameters A and y are required solely to account for the absolute values of the physical signals (here: conductivity of starting and ending buffer). For cumulative functions the centre inflection point represents the centroid of the peak (parameter m).

The introduced error function is a non-elementary (special) function with a sigmoidal shape. The values of the error function are always between -1 and 1 [17].

The integrated form of the EMG function leads to the following equation:

$$F(x) = \frac{A}{2} \cdot \exp\left(-\frac{x}{t_0} \right) \left(\exp\left(\frac{2 \cdot m \cdot t_0 + s^2}{2 {t_0}^2} \right) \cdot \left(erf\left(\frac{\frac{x-m}{s} - \frac{s}{t_0}}{\sqrt{2}} \right) + 1 \right) \right) - \exp\left(\frac{x}{t_0} \right) \cdot erf\left(\frac{x-m}{\sqrt{2} s} \right) + y \qquad (4)$$

In similarity to the Gaussian function, the parameter y is added for a shift in ordinate direction accounting for the absolute values of solvent composition. Both parameters, s and t_0 interact with each other. Thus a prior analysis of this interaction has to be done for every application.

In Figure 3, on the right side a good match of the integrated form of the EMG function with the

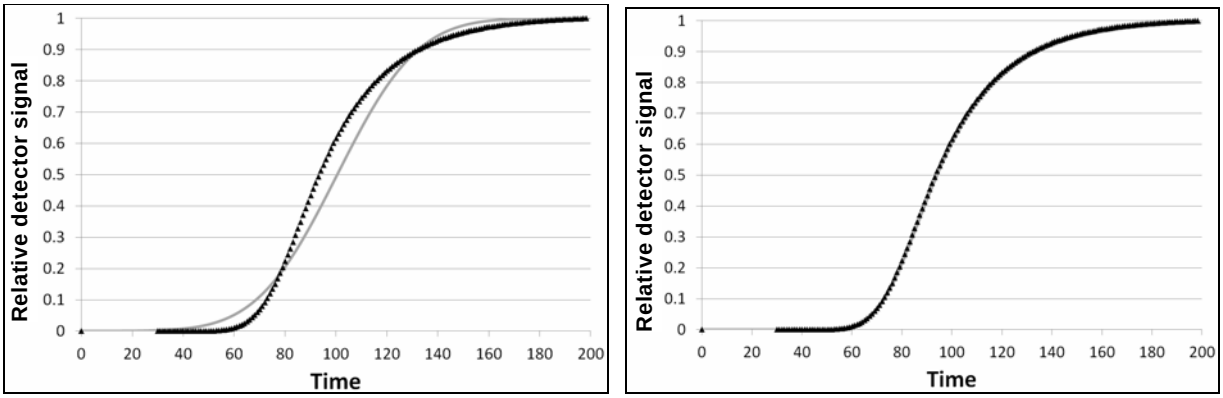

Figure 3. Ability to describe non-ideal breakthrough behaviour: a) Gaussian function (left side) and b) EMG function (right side).

experimental/process data is shown. On the left side of the figure the deviation from the cumulative Gaussian function of the experimental/process data is depicted.

The grey line represents the fitted curves and the triangles represent the experimental/ process data. In comparison with the absolute value of the process data, the relative deviation of the EMG function for every individual data point is less than 3%. Thus, the EMG function is a good opportunity to describe non-ideal breakthrough behaviour in chromatographic processes.

2.1.4. Fitting procedure and quality of fit parameters

The previously described equations are added to the fitting algorithms of the commercial data analysis software Origin® [18]. This tool allows for a quick analysis of process data integrated in working routine. Furthermore a direct calculation of model parameters as well as statistical parameters is feasible.

From the fitting algorithm used in Origin®, statistical parameters are determined, that describe the quality of the fit, i.e. the deviation of the raw data from the mathematical model that is used in the fitting function. The quality of the used model can be tested and these parameters are needed as further parameters that detect aberrations of packing from physical descriptions (e.g. cracked bed with more than one breakthrough curve). For the analysis of the quality of the fit, Pearson product-moment correlation coefficient (R^2) is applied [19]:

$$R^2 = \frac{\sum_i^n (u_i - \bar{u}) \cdot (f(x_i) - \overline{f(x_i)})}{\sqrt{\sum_i^n (u_i - \bar{u})^2 \cdot \sum_i^n (f(x_i) - \overline{f(x_i)})^2}} \quad (5)$$

The correlation coefficient demonstrates linear deviations between two functions, the original breakthrough curve (u) and values of the fitted curve (f(x)). Overbars describe mean values. The closer the value is to one, the lower is the deviation between process data and fit function.

An additional statistical parameter given from the fitting algorithm is the Chi-squared distribution [20]. In our study the Chi-square distribution is independent of the degree of freedom (DoF). DoF is the difference between the number of independent values (n) and the number of values used as intermediate steps in the estimation of the parameter itself (p) [21].

$$\frac{Chi^2}{DoF} = \frac{\sum_i^n \left(\frac{u_i - f(x_i)}{s} \right)^2}{n - p} = \frac{\sum_i^n \left(\frac{u_i - f(x_i)}{s} \right)^2}{DoF} \quad (6)$$

In addition to those parameters, we introduced a third category of fit parameter: residuals difference (Res_{diff}) was added to the fitting algorithm. The highest (Res_{max}) and lowest (Res_{min}) deviation of raw data form the fitted function was determined. The difference is defined as maximum residual difference (Res_{diff}):

$$Res_{diff} = Res_{max} - Res_{min} \quad (7)$$

In relation to the maximum detector amplitude this value leads to a dimensionless and specific quantity, the percentage residual difference ($Res_{diff\%}$):

$$Res_{diff\%} = \frac{Res_{diff}}{A} \cdot 100\% \quad (8)$$

Further statistical parameters derived from the fitting algorithm are the standard error of each fitted parameter. Those can be used to estimate the reliability of the parameters and to estimate the reliability of indirectly calculated parameters.

2.1.5. Transformation of fitting parameters to HETP determination

Methods for calculation of HETP-values for chromatographic peaks, like full width at half maximum (FWHM), can be transformed to mathematical models, since the automatized curve fitting procedure is a more exact procedure compared to manual/graphical interpretation of HETP peaks (see results).

2.1.5.1. Calculation of HETP

For a constant bed height HETP-values correspond to the number of theoretical equilibrium plates (N) and can be calculated by the first and second moment of the peak [22]:

$$\mu = \frac{\int_0^\infty x \cdot c(x)dx}{\int_0^\infty x(t)dx} \qquad \sigma^2 = \frac{\int_0^\infty (x-\mu)^2 \cdot c(x)dx}{\int_0^\infty c(x)dx} \quad (9)$$

The first moment (μ) of a chromatographic peak represents the ordinate value of the area. In our study this is equivalent to the parameter m. The different peak behaviour on the left and right side of the peak can be expressed by the second moment ($\dot{\sigma}$). Thereby, c displays the abscissa value/used detector response at the consecutive time (x). Hence, the amount of theoretical equilibrium plates is calculated by [22]:

$$N = \frac{\mu^2}{\dot{\sigma}^2} \quad (10)$$

Since for symmetric peaks the second moment corresponds to the standard deviation, N can be calculated directly from the fitted parameters in Origin® using equation 11.

$$N = \left(\frac{m}{s}\right)^2 \quad (11)$$

Transformations to IUPAC definitions for symmetrical peaks show the interrelation of graphical methods to the direct calculation using equation 11 [23]:

$$N = 5.54 \cdot \left(\frac{m}{FWHM}\right)^2$$
$$FWHM = 2 \cdot s \cdot \sqrt{2 \cdot \ln(2)} = \sqrt{5.54} \cdot s \quad (12)$$

2.1.5.2. Calculation of asymmetry (tailing factor)

For a symmetrical (Gaussian) shape of the breakthrough curve, the asymmetry is one. But as described, asymmetric breakthrough behaviour curves are often observed, especially in large scale processes where extra column effects occur. Using the EMG as the basic function, the asymmetry can be also determined by parameters derived from equation 4.

IUPAC definitions are applied [23], i.e. the asymmetry of the peak is measured at 10% of total peak height. In order to calculate the asymmetry following the IUPAC definition from the data employing the fitting algorithm, equation 4 had to be simplified for analysis:

The following parameters in equation 2 have no impact on asymmetry:

- Multiplication with a constant factor A/t_0
- Parameter m that shifts the curve in horizontal direction
- Multiplication with quotient x/t_0

Thus, the asymmetry depends only on the factor K which is the ratio between the extra column mixing effects and the standard deviation (curve width):

$$K = \frac{t_0}{\sqrt{2} \cdot s} \quad (13)$$

A generic, simplified curve of the EMG function is described by equation 14:

$$f(x) = \exp(-x) \cdot \left(erf\left(x \cdot K - \frac{1}{2 \cdot K}\right) + 1\right) \quad (14)$$

As no analytical solution could be discovered for equation 14, a numerical solution for the asymmetry

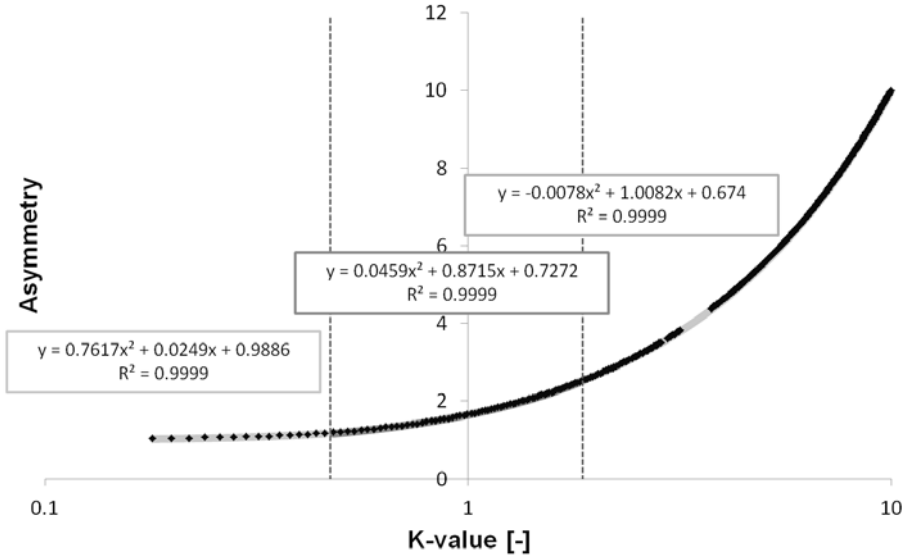

Figure 4. Mathematical calculation for the asymmetry based on EMG-function.

at 10% peak height of the *K-function* was done in a relevant range of K = 0.18 – 10. The *K-function* is non-linear, but empirically, the asymmetry can be deduced with sufficient precision by regression curves of second order by splitting the regression ranges in three parts.

With such a calculation in a selective K-value range, reasonable approximation can be achieved. For K approaching 0, the asymmetry is set to 1. The curve and the functions are plotted in Figure 4.

2.1.5.3. Calculation of standard error

Using the standard errors for each parameter as given from the fitting algorithm, the standard error of the derived parameter N can be calculated by employing Gauss law of error propagation, excluding covariance [24] according to the following equations:

$$\frac{\partial N}{\partial m}=\frac{2m}{s^2};\frac{\partial N}{\partial s}=\frac{-2m^2}{s^2} \tag{15}$$

$$\sigma_N' \approx \sqrt{\left(\frac{4m^2}{s^4}\right)\sigma_m^2+\left(\frac{4m^4}{s^6}\right)\sigma_s^2} \tag{16}$$

To calculate the standard deviation for the asymmetry (As), the empirical correlation of equation 14 was subject to Gauss law of error propagation, again excluding covariances, yielding the following equations.

$$As = a + bK + cK^2 = a + b\frac{t_0}{\sqrt{2}s} + ct^2\left(\frac{1}{2s^2}\right) \tag{17}$$

$$\frac{\partial As}{\partial t_0}=\frac{b}{\sqrt{2}s}+\frac{ct_0}{s^2} \tag{18}$$

$$\frac{\partial As}{\partial s}=-\frac{bt_0}{\sqrt{2}s}-\frac{ct_0^2}{s^2} \tag{19}$$

$$\sigma_{AS}' \approx \sqrt{\left(\frac{b}{\sqrt{2}s}+\frac{ct_0}{s^2}\right)^2\sigma_m^2+\left(\frac{bt_0}{\sqrt{2}s^2}+\frac{ct_0^2}{s^3}\right)^2\sigma_s^2} \tag{20}$$

The respective transformations were included in the post-processing steps of the fitting procedure in the Origin® script and are therefore reported automatically.

Consequently employing this analysis of HETP values, mathematical analysis of breakthrough behaviour can replace traditional graphical HETP determination. Further, mixing effects which play a major role in tailing can be deducted directly using the K-function. HETP values and asymmetry can be determined simultaneously.

2.2. Scale down experiments

In order to study the effects of flow rate and buffer components further, two characteristic resins were chosen from the high diversity of the 9 production processes used on the manufacturing site. Blue Sepharose is a representative of affinity resins and Superdex 75 is chosen as representative of size

exclusion resins. Experiments in lab-scale are essential to support the scale-independent applicability of the models. This is advantageous for new processes, since a validated method can be used to investigate the impact on packing integrity as an early step in process development.

HETP measurements were performed by pulse injection of highly concentrated (2 M) sodium chloride and acetone (2%), respectively. As an alternative method HETP values were calculated from breakthrough curves by changing the buffer from 0.5 M to 2 M sodium chloride. The flow rates were varied between 1 cm/h and 100 cm/h. To demonstrate the accuracy of the currently used manual method of HETP-value determination, a manual (graphical) analysis of HETP values by ten people was investigated with a self-packed column (ID 1 cm and length 12 cm). The column was packed with the cation exchanger Fractogel SO3. The flow rate was adjusted to 250 cm/h. Statistical reproducibility of all experiments was ensured by a three-time repetition of the experiments.

2.3. Large scale analysis

For both approaches (Gauss and EMG), data from historic production runs were analysed and a feasibility study was performed, employing 21 different types of columns and 9 different biotechnological downstream processes. Thereby, also processes currently under process development as well as different scales were analysed. The analysis included several types of resins:

- Affinity chromatography (AC)
- Cation exchange chromatography (CEX)
- Anion exchange chromatography (AEX)
- Hydrophobic interaction chromatography (HIC)
- Hydroxyapatite chromatography (HAC)
- Reversed Phase chromatography (RPC) as well as
- Size exclusion chromatography (SEC)

For a detailed oversight of examined processes (exact product names and chromatographic resins cannot be disclosed further) c.f. Figure 5.

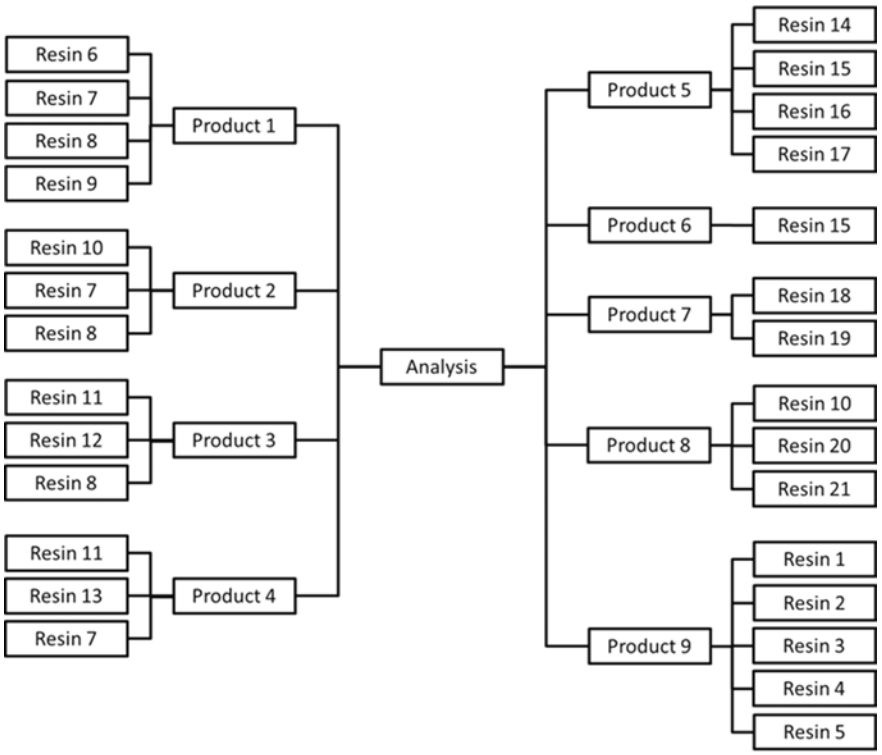

Figure 5. Several types of chromatography resins examined for different products in large scale manufacturing plants for market production.

The investigation included resins from several suppliers. An adequate applicability to a wide range of processes, scales, matrices, solvent conditions etc. could be shown for all routine purposes.

Method validation requires cases where potentially an insufficient column packing can be detected by the method. But that is not given for all processes because in most cases, the packed bed stability is given throughout the lifetime of the columns. Thus, this study focuses on an affinity resin with data of nearly 250 chromatographic cycles, where repacking due to losses in packed bed stability arose eight times. The reasons for repacking are based on the product quality. Manual analysis of peak shape as well as detected process irregularities led to a preventative decision for column repacking. In the case of the process investigated for 250 cycles, the elution, column washing and column regeneration steps are performed by isocratic buffer changes, which were analysed with the breakthrough curve fitting procedure. The detector signals of the column inlet were controlled for every chromatographic cycle to indicate irregularities before the mobile phase interacts with the resin that can lead to misinterpretations of the fitting results.

Some prerequisites had to be checked before: To qualify for a routine monitoring step of such a breakthrough curve, buffer change has to be identified as an inert process, whereby no significant interactions other than diffusion (while passing the column) occur. This guarantees that there are no significant interactions of ions other than mass transport. The fitting procedure itself will detect fitting quality. If the model does not fit the real data, parameters for fitting quality will indicate a poor fit.

To reach higher parameter sensitivities, the influence of the chromatography equipment can be determined with a series of experiments that generate a breakthrough curve without the column. This correlates to a bypass of the equipment. Since the area and the adjustment to the ordinate direction are constant for the experiments with and without the column, only parameters with a direct influence to the slope of the breakthrough curve have to be considered. Consequently, the equipment parameters of standard deviation, inflection point and tailing factor have to be subtracted from the overall values. But this opportunity is limited by several parameter combinations, where a direct subtraction is not possible. A more accurate method is to use convolution integrals that allow consideration of non-idealities of the equipment. In our study, this is not requested, since process setup is kept constant and thus, the influence of the equipment is always the same for one system and the parameters show a sufficient sensitivity.

3. RESULTS AND DISCUSSION

3.1. Scale-down experiments

3.1.1. HETP determination (Van-Deemter studies)

For both resins the results indicated the same tendency. In Figure 6 the results for Blue Sepharose are given exemplarily. Equation 10 and 12 refer to both methods for HETP determination.

The minimum of the HETP-values is between 0.1 ml/min and 0.15 ml/min (7.5 cm/h-12 cm/h). These values are lower than the flow rates in production scale used (marked with a cross in Figure 6). The highest deviation between both equations occurs for the pulse injection of NaCl (approx. 20%). In general, following conclusions can be done:

- The HETP determination using the analysis of breakthrough curves leads to a similar behaviour as expected by the Van-Deemter equation
- the values derived by employing equation 10 are lower than those derived by application of equation 12 with regard to breakthrough curves and for the pulse injection and vice versa
- Breakthrough experiments lead to lower values than pulse injections.

It can be seen that the data from large scale (cross in Figure 6) are in a good agreement with the lab-scale experiments. Thus, general applicability of the new method is shown and future analysis of packing integrity can be done in lab-scale, early in process development. Of course, for a more detailed conclusion about the scalability, experiments with different velocities also should be done in the industrial scale. But this is not realizable for existing processes. The lab experiments demonstrate that the current measurements of the HETP values give comparable values (in the same order of magnitude) as the new method using direct analysis of

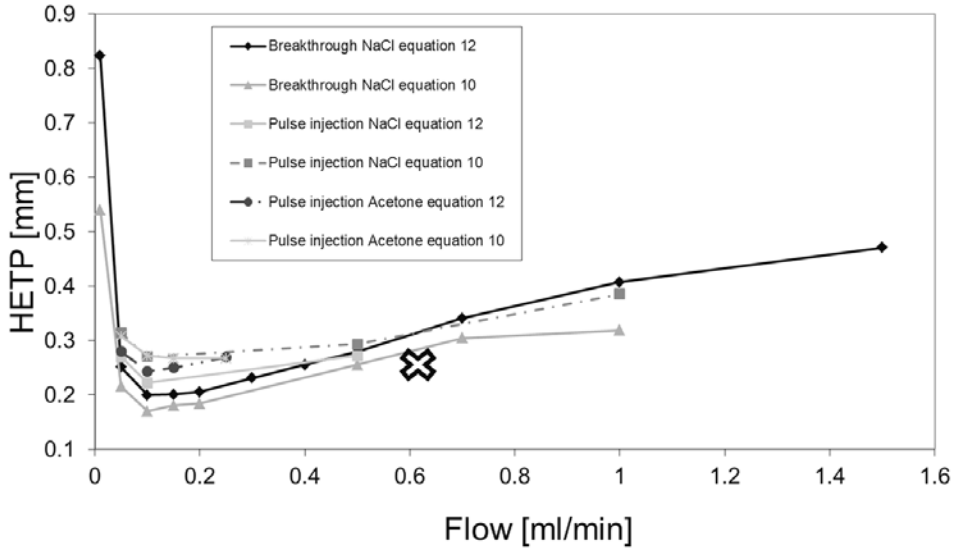

Figure 6. HETP determination for Blue Sepharose.

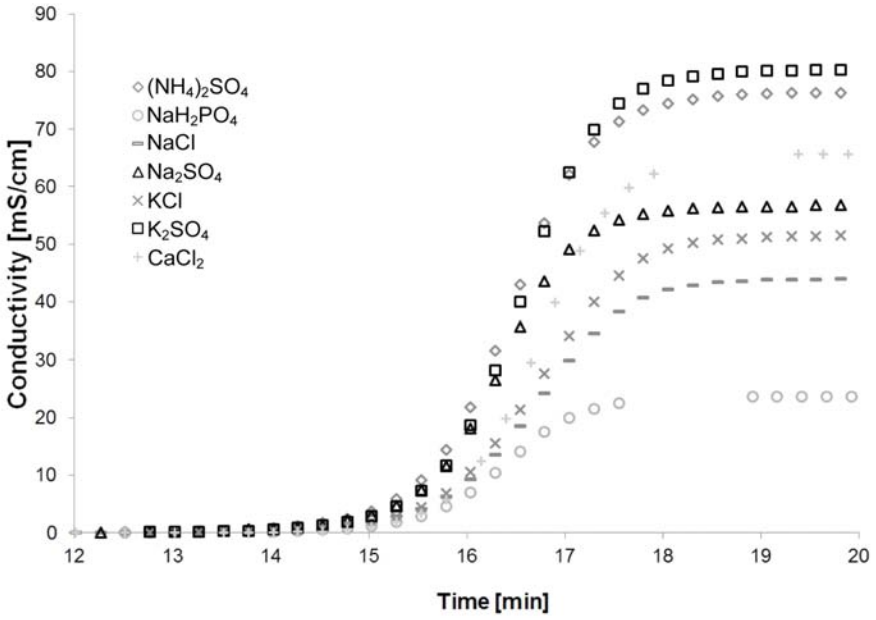

Figure 7. Salt breakthrough curves.

breakthrough curves. The new method also indicates a Van-Deemter behaviour. This fact supports the ability of the new methodology to replace the traditional graphical methods.

3.1.2. Dependence on ions involved

Breakthrough curves for a diversity of salt types with a concentration of 0.5 M are depicted in Figure 7.

Different salt types generate different breakthrough behaviour in respect to slope, skewness and kurtosis (data not shown). This depends on the ionic strength of the salts and on the free diffusion of the salt in the inner of the particle pore.

An agreement of breakthrough behaviour to Hofmeister series could not be measured. The analysis shows dependence between ion mobility and

breakthrough behaviour. A-priori prediction for other salts is not possible, especially if the solvent composition consists of more than 2 salts. In Table 1 the mean values and the relative deviations are given for each parameter.

Most significant deviation can be seen for the parameter A and y. Parameters A and y are representing the different solvents used, therefore the high variance is obvious.

The inflection point of the breakthrough curve is least dependent on the solvent composition; an impact of the ions involved can be excluded by that fact. A significant influence of the salt types can be seen for the standard deviation (width of the curve) and for the tailing. Differences of approx. 20% can occur. Consequently, this method has to be validated with original process media.

3.1.3. Comparison of manual HETP determination to fitting procedure

The demand for an automated analysis of monitored data is given, because an individualized graphical analysis leads to high deviations. In our test series, 10 participants (students) analysed a peak injection concerning retention time, asymmetry and HETP-values that are determined by equation 12.

Table 1. Parameter analysis of the salt breakthrough curves.

Parameter	Mean value	Relative deviation [%]
y [cm/s]	28.49	41.17
A [cm/s]	56.52	40.00
m [s]	16.43	1.29
w [s]	0.716	11.02
t_0 [s]	0.11	19.20

The confidence interval was determined with the student distribution [25]:

$$\beta_{1,2} = \left[\bar{z} \pm t_{st}\left(1-\frac{\alpha_{Ex}}{2}, n-1\right)\frac{\sigma}{\sqrt{n}} \right] \quad (21)$$

where n is the number of participants, σ the variance, α_{Ex} the values from student-distribution for each probability and $\bar{z}$ the mean value. Results for an analysis of a typical elution peak are summarized in Table 2.

For all values high variation between the individual analysers could be detected. Thus, reproducibility is limited. For the HETP values this discrepancy results in a standard deviation of approx. 50%. This requires highly educated operators to minimize these effects. An automated analysis can support operators and can lead to a standardized process control with higher reproducibility. But still offline analysis is a useful tool in production processes.

The error of an automatizied control system is less, since the variance between the different analysis is not significant (<0.1%) and the error of the fit is mostly in an accurancy of more than 99%.

3.2. Analysis under manufacturing conditions/ large-scale analysis

In most cases studied from historical large scale data, the resulting breakthrough curve of the conductivity at the column outlet could be used for a mathematical analysis of packed bed integrity. This allows an establishment of this additional process control method to large scale manufacturing in biotechnological manufacturing. In all cases studied here, conductivity signals were used for analysis of the breakthrough curves, but other signals can also be employed.

Table 2. Statistical determination of N-values and asymmetry by an individualized graphical analysis.

	Number of theoretical separation plates [-]	Asymmetry [-]	Retention time [h]
Mean value	401.93	1.81	0.0131
Standard deviation	200.47	0.14	0.0026
Percent deviation	49.88	7.86	19.96
Confidence interval 95%	122.48	0.087	0.0016
Confidence interval 97.5%	151.15	0.107	0.0020
Confidence interval 99.9%	287.14	0.204	0.0038

Table 3. Statistical analyse of the Gaussian and EMG parameters.

Parameter	Average Gaussian distribution	Average EMG function
R^2	0.996	0.998
Chi^2/DoF	0.50	0.38
$Res_{\%}$	7.36	3.21
s [h]	0.00092	0.00047
y [ms/cm]	2.184	23.68
A [ms/cm]	38.13	36.80
M [h]	0.01168	0.01105
t_0 [h]	-	0.00096

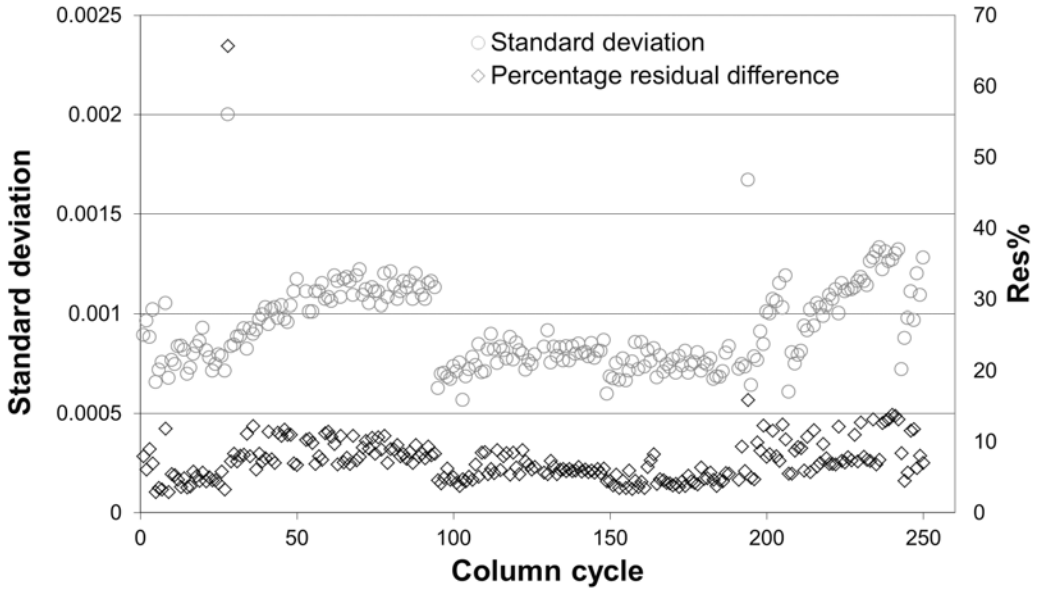

Figure 8. Process analysis of standard deviation and residuals regarding Gaussian distribution.

3.2.1. Fitting of Gaussian distribution

In the first approach, the potential to describe the breakthrough curves by fitting an ideal Gaussian distribution were analysed. The resulting values of the statistic parameters for the robustness analysis by the Gaussian distribution are given in Table 3.

Figure 8 indicates a high sensitivity of all statistic parameters. But, high discrepancy between the average value and the maximum/minimum value results from a defect of the column packing, where air bubbles passed the column and thus, the resin packing lost integrity.

Significant differences for the inflection point exist only for new resin packings. Deviations of parameters A and y are equivalent to deviations of the solvent system. These values and the variations of the HETP-values do not lead to sufficient predictions of column integrity. As mentioned, the number of theoretical plates in the Gaussian distribution is based on the assumption of an ideal peak shape. This can lead to inaccurate results by cross interactions to the standard deviation.

The trend of standard deviation and percentage residuals is depicted in Figure 8. These parameters show the highest sensitivity to the column packing.

The percentage residuals have no predictable effect. However, the standard deviation indicates a trend to characterize process robustness. After every re-packing a certain level is reached that is a benchmark for every run with this column packing. Afterwards, standard deviation increases. But, a prediction of the chromatographic cycle with insufficient column packing is not possible, since in most cases the column was repacked before the limit of packing quality was reached. Furthermore, spontaneous errors as air bubbles in the column occur and these errors are not linked to a predictive behaviour of trends in the standard deviation. A statistical analysis of the cross interactions regarding all parameters did not lead to a predictive model.

3.2.2. Fitting of the EMG distribution

While using the integrated form of the Gaussian distribution function with data from real chromatography columns in large scale (up to 140 cm diameter), we found, that in some cases, the raw data will not be perfectly described by such a model (i.e. poor fitting results), mostly in cases, where columns show asymmetric peaks (i.e. tailing). Especially in large scale production plants, extra-column mixing effects can have a strong contribution. Consequently, the simple model of perfectly diffusion driven peak broadening (if mass transfer is not limiting) was extended to a mathematical model, where extra-column mixing effects are reflected. A clear advantage of this approach is the possibility to discriminate between the influence of the equipment to the breakthrough curve and effects inside the packed bed of the column. The resulting EMG and statistical parameters of the robustness analysis of 250 chromatographic cycles are depicted in Table 3.

In contrast to the Gaussian distribution, the y value of the EMG function expresses not the adjustment of the ordinate, but rather the ordinate value of the inflection point. The average values of the statistical parameters in comparison to the Gaussian fit demonstrate a higher accuracy and hence indicate a better fitting routine. The better fitting of the integrated form of the EMG function was already illustrated in Figure 3. Thus, minimum and maximum values decrease or increase, respectively in comparison with the Gaussian parameters. The lower R^2 values of the EMG function results from the better ability to describe non-ideal breakthrough curves. Analysis of variance is done for all parameters to identify statistical relevance of the parameters. For the parameter R^2 the comparison between the Gaussian distribution (left side) and the EMG function (right side) is depicted in Figure 9.

Due to the closed range of both confidential intervals (maximum 0.0006), a visualisation on a single y-axis is not feasible. Thus, in Figure 9 the ranges on the y-axis differ. The non-overlapping of both diamonds will be pointed out clearly by the difference between average values (0.00277) and the maximum confidence interval (0.0006). The difference between the average values is about 4 times higher than the biggest confidence interval. It is obvious, that in this case no overlapping of the confidence interval can occur.

Since the sample size of both analyses is the same, the marks that appear not to overlap indicate that two group means could be significant at the 95% confidence interval [26]. A similar, significant difference could be seen for all parameters and better fitting could be achieved with the EMG basic function. Thus, the EMG function is a good possibility for correct process description with both the ability to detect sudden decay and trends in the process.

For the Gaussian distribution a sensitivity of the standard deviation is shown (Figure 8); the standard deviation does not show this sensitive behaviour for the EMG function. The tailing factor is more sensitive for that case. In Figure 10 the standard deviation (circles) and the tailing factor (diamonds) are depicted.

As shown, the tailing factor indicates the cycle of repacking sufficiently and it is sensitive to the packing quality. The characteristic behaviour of the tailing factor equals almost the characteristic trend of the standard deviation for the Gaussian distribution, but the impact of the tailing factor is more sensitive. A retrospective analysis of those packing instabilities revealed that, during the process, air accumulated in the upper part of the column, destabilizing the system over time. This observation supports the mathematical model, that t_0 represents effects that are in correlation with

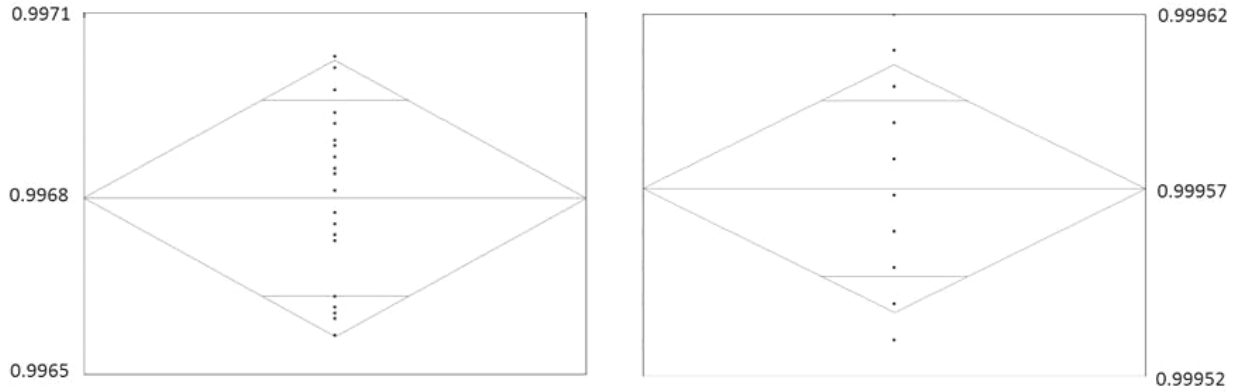

Figure 9. Analyse of Variance for R^2: a) Gaussian distribution (left side) b) EMG function (right side).

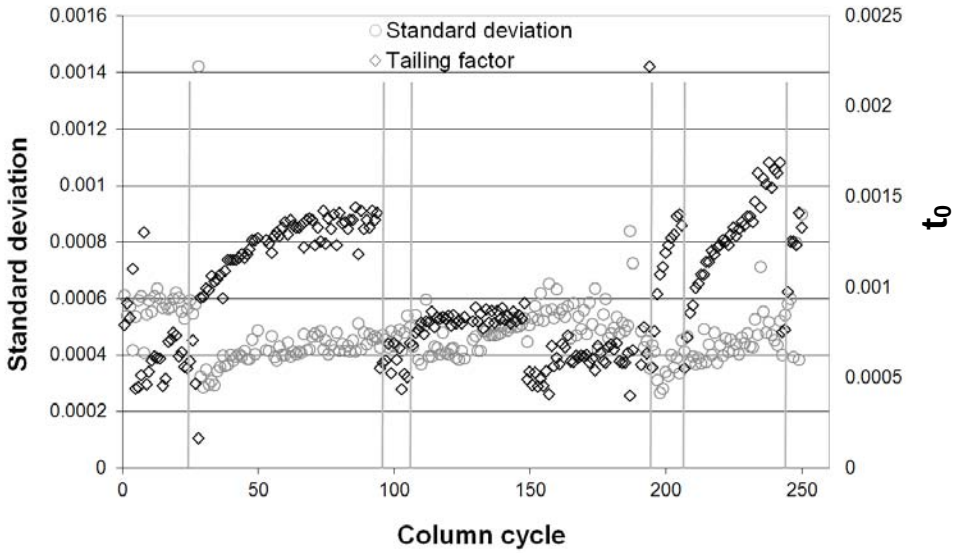

Figure 10. Process analysis of standard deviation and tailing factor regarding EMG function.

extra-column mixing effects, in this case void volumes introduced by accumulated air. Since the inflection point varies only in a little range, no conclusion about a correlation between the void volumes and the packing quality can be done. The whole data set demonstrates that the tailing factor is a sensitive parameter to ensure and predict process robustness, if a certain data amount is available to validate the process. Furthermore, the trend of the significant parameters is more sensitive for the EMG function than for the Gaussian distribution. Detection of outliers, assignment of parameters to physical conditions and thus analysis of errors in the chromatographic system is simpler in that case. Asymmetry as well as the HETP-values alone does not show such a strong predictive behaviour. For both parameters, the first cycle with a repacked column is a benchmark. The asymmetry does not vary significantly for one column packing but can vary significantly between separate column packings.

Based on those results it can be concluded that the EMG function has a higher sensitivity and thus a better applicability for large scale production plants than the Gaussian distribution.

3.2.3. Data rate analysis

Besides prediction of column integrity the amount of required data is also essential for an effective process robustness analysis by such a fitting procedure.

This has to be done for all parameters. The data rate analysis started with 3600 data points. The amount of data points was halved and for every data rate both fitting routines were performed. For the fit with the Gaussian distribution a significant deviation occurs, if the data rate is reduced by a factor of 32. The same phenomena occur also for the parameters of the EMG function. Since this function describes the breakthrough curve with a higher accuracy, the required data amount is lower than for the Gaussian distribution concerning the function parameters and not the statistical parameters. A more sensitive parameter to validate the necessary amount of data points is the percentage residual. The advantage is that both fitting routines are comparable to each other. It is shown, that the EMG function requires a higher amount of data points than the Gaussian distribution for constant percentage residuals. Of course, the magnitude of the percentage residuals is lower for the EMG function than for the Gaussian distribution, reasonable due to the better accuracy of the fit. This also describes the higher required data rate. Consequently, as a general number, for the fit of the Gaussian distribution at least 300 data points are required and for the fit of the EMG function a data rate at least 500 data points are required.

4. CONCLUSION

The scope of this work was to figure out a method that can lead to predictive conclusion about the robustness of biopharmaceutical processes. This is essential to stabilize process performance and speed of trouble shooting (SPC, PAT). The aim is to achieve orthogonal data for process monitoring. Thereby, the work is focused on the column integrity of chromatographic separations as an example for the whole variability of the unit operations in bio-manufacturing processes. The robustness analysis requires an analysis of processes, where the packed bed integrity was not stable throughout the column lifetime. Using such process excursions, the sensitivity of this extended process monitoring technology can be evaluated. The analysis was performed using an example for an affinity resin over 250 chromatographic cycles. Acceptance criteria for process robustness can be deducted from historical data; no prolonged validation time is required, because the data are already available and can be leveraged retrospectively.

Since in most regeneration, equilibration or wash steps during a chromatographic cycle, the buffers are changed in composition and concentration, the breakthrough curve type of conductivity can be chosen for the robustness analysis.

Two mathematical approaches are compared for the robustness analysis.

On the one hand the breakthrough curve is characterized by the integrated form of the Gaussian distribution function. In that case, the deviations between the ideal symmetric function and the asymmetric process data are determined by the standard deviation and poor quality of fit parameters.

On the other hand, the EMG function is used to describe the breakthrough curve in a more precise way. The fit of the function matches sufficiently to the process data. In the example studied, the tailing factor is the most sensitive parameter to indicate the quality of column packing, in perfect agreement with the underlying theory of void volumes in the column yielding mixing effects outside the packed bed.

Both the Gaussian and the EMG model cannot be used to describe breakthrough curves with a fronting behaviour. If fronting appears, it reflects cracks in the packed matrix and will lead to poor fitting results (poor quality of fit parameters).

It can be concluded that specific parameter sensitivity is given for both fit functions with the potential to predict column integrity. For the EMG approach also a sensitivity analysis of the parameters was performed, since cross interactions could lead to a prediction of the column packing. Currently prediction potential is limited by two facts:

Spontaneous errors of the equipment lead to a drastic increase or decrease of the fit parameters that could not be predicted.

Most industrial separation processes run stable throughout the column lifetime, therefore, only one example could be studied where packing integrity was lost, still proving the applicability of the concept. In order to establish monitoring of other processes where columns are used long-term, the acceptance criteria have to be deducted from the successful range of parameters from normal operating conditions. Additionally, a trend analysis of the parameters can be established to detect process drifts early on a continuous basis.

Moreover, the required amount of data points was analysed to give a sufficient prediction of column integrity. Thereby, it was pointed out that the fit of the EMG function requires only a little amount of data points to estimate the function parameters with a high accuracy. But, individual outlier of the process data influences the statistical parameters as percentage residuals. If these parameters have to be constant, the necessary data amount is about 500 data points. The data rate for the Gaussian distribution is less, but also the function parameters and not only the statistic parameters are drastically affected by a decreasing data rate.

This new method will be directly implemented into routine process monitoring as an add-on to automation. The positive business case results mainly in better process surveillance and reliability as well as cost savings by avoiding separate HETP determinations. An evaluation of the influence of column integrity for e.g. other parameters like purity or yield (QbD approach) has additional potential.

SYMBOLS

A	Amplitude of the breakthrough (equivalent to peak area)	[kg/m³]
As	Asymmetry	[-]
c	concentration (detector signal)	[kg/m³]
Chi^2	Chi-squared value from the distribution	
f	function	[s]
FWHM	Full width at half maximum	[s]
K	K-factor	[-]
m	Inflection point (equivalent to centroid of a peak)	[s]
n	Number of participants	[-]
N	Amount of theoretical plates	[-]
R^2	Pearson product-moment correlation coefficient	[-]
Res	Residual	[-]
s	Standard deviation	[s]
t	time	[s]
t_0	time constant for extra-column mixing effects (tailing factor)	[s]
t_{St}	Probability parameter student distribution	[-]
u	y-value (concentration) of original breakthrough signal	[kg/m³]
x	Process time	[s]
y	Adjustment in ordinate direction	[kg/m³]
$\bar{z}$	Mean value	
DoF	degrees of freedom	
Greek		
α_{Ex}	Probability from student distribution	[-]
$\beta_{1/2}$	Upper and Lower boundaries of confidence interval	[kg/m³]
μ	First moment	[s]
$\dot{\sigma}$	Second moment	[s]
σ	Variance	
σ'	Standard error	
Indices		
Diff	Difference	
i	specific time	
max	Maximum	
min	Minimum	
%	percentage/relative	

ACKNOWLEDGEMENTS

The authors do like to thank their colleagues at the Institute of Separation and Process Technology at Clausthal University of Technology. Special thanks go to Christian Borrmann, Andreas Köppen and Annemarie Einenkel for their previous work and data analysis. Furthermore the authors want to thank Michael Pohlscheidt, Bernhard Liessem, Ulrich Opitz, Katharina Schiffl (Roche) and Christopher Bork (Genentech) for great support of the project and fruitful discussions.

REFERENCES

1. US Food and Drugs Administration, Guidance for Industry PAT – A framework for innovative pharmaceutical manufacturing and quality assurance, 2004, US Department of Health and Human Services, Rockville.
2. International conference on harmonization of technical requirements for registration of pharmaceuticals for human use, 2005, ICH harmonized tripartite guideline: Pharmaceutical Development Q8 (R2).
3. International conference on harmonization of technical requirements for registration of pharmaceuticals for human use, 2009, ICH harmonized tripartite guideline: Quality Risk Management Q9.

4. International conference on harmonization of technical requirements for registration of pharmaceuticals for human use, 2008, ICH harmonized tripartite guideline: Pharmaceutical Quality System Q10.
5. Rathore, A. S., Kennedy, R. M., O'Donnell J. K., Bemberis, I. and Kaltenbrunner, O. 2003, Biopharm int., 16(3), 30-41.
6. Felinger, A. and Guichon, G. 2001, J. Chromatogr. A, 913(1-2), 221.
7. Casella, G. and Berger, R. L. 2001, Statistical inference (2nd ed.), Duxbury, Pacific Grove.
8. Moscariello, J., Lightfoot, E. and Rathore, A. S. 2005, Biopharm int., 18(8), 58-64.
9. Guichon, G., Shirazi, D. G. and Felinger, A. 2006, Fundamentals of Preparative and Nonlinear Chromatography (2nd ed.), Elsevier Academic Press, San Diego.
10. Di Marco, V. B. and Bombi, G. G. 2001, J. Chromatogr. A, 931, 1.
11. Howerton, S. B., Lee, C. and McGuffin, V. L. 2003, Anal. Chim. Acta, 478, 99.
12. Caballero, R. D., García-Alvarez-Coque, M. C. and Baeza-Baeza, J. J. 2002, J. Chromatogr. A, 127, 494.
13. Pápai, Z. and Pápai, T. L. 2002, Analyst, 127, 494.
14. Baeza-Baeza, J. J. and García-Alvarez-Coque, M. C. 2004, J. Chromatogr. A, 1022(1-2), 17.
15. Cox, J., Cunnien, P., Ganguly, J., Ghosh, B., Ladiwala, A., Song, R. and Thommes, J. 2009, Patent WO2009094203 A2, Biogen Idec Inc.
16. Wolfram, S. 1999, The MATHEMATICA Book (4th Edition), Cambridge University Press, Cambridge.
17. Andrews, L. C. 1992, Special functions of mathematics for engineers, SPIE Press, Bellingham.
18. http://www.originlab.com/
19. Rodgers, J. L. and Nicewander, W. A. 1988, Am. Stat., 42(1), 59.
20. Lancaster, H. O. 1969, Chi-squared Distribution (Probability & Mathematical Statistics), John Wiley & Sons Inc, New York.
21. Morden, S. 2011, Degrees of Freedom, Orbit, Bonn.
22. Seidel-Morgenstern, A. 1995, Mathematische Modellierung der präparativen Flüssigchromatographie, Deutscher Universitätsverlag, Wiesbaden.
23. Ettre, L. S. 1993, Pure Appl. Chem., 65(4), 819.
24. Hartung, J., Elpelt, B. and Klösener, K.-H. 2005, Statistik – Lehr und Handbuch der angewandten Statistik, Oldenbourg Verlag, Munich.
25. Ekstrom, C. T. and Sorensen, H. 2010, Introduction to Statistical Data Analysis for the Life Sciences, CRC Press, Boca Raton.
26. Proust, M. 2009, JMP™ Introductory Guide (2nd Edition), Cary.