DOI 10.5162/sensoren2019/P3.11

Recommendations for optimal sampling and filter rates in liquid chromatography

David Thomas Marehn[1,2,3], Detlef Wilhelm[1], Heike Pospisil[2], Roberto Pizzoerrato[3]
[1]AnaTox GmbH & Co. KG, Reinheimer Strasse 18b, 15517 Fuerstenwalde, Germany
[2]UAS Wildau, High Performance Computing in Life Sciences, Hochschulring 1, 15745 Wildau, Germany
[3]Università degli Studi Di Roma "Tor Vergata", Dipartimento Ingengeria Industriale, Via del Politecnico 1, 00133 Rome, Italy

Abstract

In the strong regulated field of pharmacy, all substance analysis are performed by very well defined experiments using specific techniques and parameters for each substance, which are written down in the pharmacopoeias, mainly the United State Pharmacopoeia (USP), Japanese Pharmacopoeia (JP) or European Pharmacopoeia (EP). Some of these experiments use techniques which belong to the field of chromatography especially the high performance liquid chromatography (HPLC) and gas chromatography (GC). Nearly all possible parameters are listed in the so called monographs of the active substances. Nevertheless two editable parameters of the detectors are not mentioned and not restricted here: the sampling rate and the signal filtration. Both parameters have a clear impact on the peak width, peak height and peak distortion. Therefore, it would be desirable to define recommendable setups for chromatography tests.

In this article we present how to find the optimal values for the sampling rate and signal filtration for a given substance of a monograph. Chromatograms with the highest possible sampling rate and lowest filtration have been altered by performing a sampling reduction and various signal filtration algorithms considering the detector firmware of the major HPLC manufacturers. For the evaluation, several chromatographic properties describing a chromatogram have been taken for an optimization process based on the multi objective (Pareto) optimization. The optimization has been tested on simulated and real HPLC data. The suitability has been proved by comparing the results against the usual approach of 20 points per peak.

The criterion allows to make further investigations on the substance monographs in order to develop the sampling rate and signal filtration recommendations depending on the HPLC system firmware.

Keywords: Chromatography, Pharmacopeia, Sampling Rate, Signal Filtration, Optimization

Introduction

Based on the increasing regulation in the field of pharmacy within the last century all substances that are used for medicinal products are very well defined and described in in the pharmacopoeias, whereby the United State Pharmacopoeia (USP), Japanese Pharmacopoeia (JP) and European Pharmacopoeia (EP) belong to the global player [1]. Each substance contains experiments which definitions resulted from method development processes in order to get a suitable setup for analyzing the active substances inside a medicinal product before selling it. All these setups are written down in the so called monographs and are describing specific techniques and parameters for each substance. Among other techniques some of the substances will be analyzed by the usage of chromatography like the high performance liquid chromatography (HPLC) and gas chromatography (GC). Nearly all parameters and conditions of the analytical system are defined and listed.

Nevertheless two editable parameters of the data acquiring module, the detector, are not mentioned and not restricted: the sampling rate and the signal filtration. Both parameters have a clear impact on the peak width, peak height and peak distortion [2, 3, 4]. In other words, two experiments under complete equal conditions may result in different chromatograms with respect to the mentioned detector parameters. This has a notable influence on the conclusion of a substance test and contradicts the strong legal requirement in pharmaceuticals. Therefore, it would be

desirable to define recommendable setups for chromatography tests.

Sampling Rate and Signal Filtration

Regardless of the detection type every detector used in the chromatography produces an analog, electrical signal like changes of voltage or current [5, p.19]. Usually an internally installed analog-to-digital converter (ADC) digitize the analog, continuous signal into a discrete signal and provide it to an external controller (e.g. computer) by a suitable interface. The sampling rate is a parameter of the ADC and specifies how many discrete data points per second will be sampled and provided. The unit of the sampling rate is given in Hz (1/s). Presently there are two approaches how an ADC operates. Either the analog signal will be exactly sampled by the given frequency or the highest available frequency will be used and a specific amount of data points will be averaged to one point. The second method is called data bunching [5, p.21].
The signal filtration is a routine during the data acquisition [2] or the data processing afterwards [4]. It is either based on physical hardware components, digital filters implemented in the firmware of the device or on algorithms within the chromatography data system (CDS) software installed on the computer which collects and interpret the resulting chromatogram. The filtration itself deals with noise that overlays the analog signal disturbing the chromatographic analysis [6].

Optimize Sampling Rate Reduction and Signal Filtration of Raw Data

In this article we present how to find the optimal parameters for the sampling rate and signal filtration depending on up to four criterions that are usually used for the evaluating of a chromatogram during the method development. All the experiments are based on a self written tool implementing the two sampling rate reduction approaches, several signal filtration algorithms and the continuous update method for the multi objective optimization. The input for the tool are chromatograms with the highest available sampling rate and lowest filtration because these steps will be simulated by the tool itself.
For the optimization testing several sampling rates we focused on the reduction by data bunching because this seems to be the most realistic approach due to the integrator used in Delta-Sigma analog digital converters installed in most data acquisition systems [2]. So the reduction of the sampling rate from for example 100 Hz to 20 Hz leads to an averaging of five consecutive data points to a new single one.
The choice of the implemented and tested signal filtration algorithms consider the physical and digital filters installed in the detectors of the major HPLC manufacturers. As a physical filter so called RC filters using capacitive and ohmic resistances are installed and parameterized by a time constant (τ= R*C). An estimation equation for the RC filter exists [7] and shown in equation 1.

$$s_{out} = \frac{\Delta t \cdot s_{in}(t) + RC \cdot s_{out}(t - \Delta t)}{\Delta t \ + \ RC} \qquad (1)$$

With Δt as step width between two data points, RC as time constant τ, s_{in} as the raw signal and s_{out} as the resulting, filtered signal.
All other used filters belong to the digital filters and are implemented into the firmware of the detector devices. Additional all the following filters are windowed filters. That means a window with a specific size of N data points will pass the raw data and compute a filtered data point based on a weighting function applied on the data points in the window. The windows size (WS) is odd and the main parameter of all windowed filters. The most simple filter is the moving average. Here all data points in the window have the same weighting as shown in equation 2.

$$f_i = \frac{1}{WS} \sum_{j=-WS/2}^{WS/2} s_{i+j} \qquad (2)$$

With f as the filtered signal, s as the raw signal and WS as the window size.
The Hamming filter is based on the window function by Richard Hamming and weights the data points as visible in equation 3.

$$w_n = 0.54 - 0.46 \cdot \cos\left(\frac{2\pi n}{WS - 1}\right) \qquad (3)$$

With w_n as the filter weight of the n-th data index (beginning with 0) in the window and WS as the window size.
One of the most popular filter algorithms is Savitzky-Golay. Here the weighting is based on a polynomial regression of the data points in the window. So an additional parameter for this filter is necessary, the grade of polynomial. The weighting of each data point in the window can be computed [8] or prepared in a coefficient table as we do up to a maximal window size of 29 and grade of polynomial of 15.
The last implemented filter is the moving average filter with Gaussian kernel. Here the weighting uses the Gaussian distribution (equation 4).

$$f_i = \sum_{j=-WS/2}^{WS/2} s_{i+j} \cdot e^{\frac{-\pi \cdot j^2}{SD^2}} \quad (4)$$

With f as the filtered signal, WS as the window size and SD as the standard deviation of the Gaussian equation. So the filter here depends on the window size and the standard deviation.
For the optimization of the sampling rate and signal filtration parameters the multi objective optimization (Pareto optimization) has been chosen [9]. That way it is possible to optimize the result of the filtration on criterions that evaluate the chromatogram. These criterions are influenced by the sampling rate and signal filtration and are the peak width at half height, the symmetry factor, signal-to-noise ratio and resolution. The peak width and the distance of the symmetry factor to the ideal value of 1 will be minimized as where the signal-to-noise ratio and resolution will be maximized. This approach differs from other possible optimization methods using one specific evaluation value like the Durbin-Watson value [10] or largest negative chromatogram second derivative [4].

Materials

Three different kinds of chromatograms were used for the optimization process. Two of them were simulated using the general creation equation of the HPLC simulator developed by Boswell et al. [11]. The function for one peak generation is:

$$C_i(t) = \frac{W_i}{2\sqrt{\pi}\sigma_{t,i}F} e^{-\left(\frac{t - t_{R,i}}{2\sigma_{t,i}}\right)^2} \quad (5)$$

With C_i as the molarity of compound i, W_i as the number of moles, σ as peak width, t_R as retention time and F as flow rate. Each compound i forms a peak function of t. For the final chromatogram all peak functions are summed up. That way a chromatogram containing one peak only and a sampling rate of 100 Hz was created. Another one containing 21 peaks corresponding to a separation of paracetamol, codeine, pitophenone and their impurities done by Vijta et al. [12] was created with a sampling rate of 40 Hz.
The real chromatogram of a HPLC isocratic test sample was acquired by an Agilent Technologies 1260 system that consists of a G4225A degasser, G1312B binary pump, G1367E wellplate autosampler, G1330B autosampler thermostat, G1316C column compartment and G4212B diode array detector. The isocratic test sample contains the four substances dimethyl phthalate, diethyl phthalate, biphenyl and o-terphenyl solved in methanol. The eluent consisted of a mix of 35 vol.% HPLC-grade water and 65 vol. % Acetonitrile. For the stationary phase Zorbax xDB-C8 column supplied by Agilent Technologies according to a reverse phase chromatography configuration was installed. The column had a length of 50mm, a diameter of 4.6 mm and a pore size of 1.8 µm. The analysis conditions where 1 ml/min flow, 1 µl injection volume, 40 °C column temperature, 254 nm detection wavelength and 80 Hz sampling rate.

Optimization Test Range

For the determination of the Pareto optimization front (POF) it is necessary to define a suitable test area for the parameters of the sampling rate and signal filters. For the sampling rates the common rates an Agilent Technologies detector provides were used: 100, 80, 40, 20, 10, 5, 2.5, 1.25 Hz. The given sampling rate of the raw chromatogram has been reduced only. The sampling rate has not ever been increased.
For the RC-Filter the time constant has been varied from 0 (no filtering) up to 1/2 * peak width at the half height ($W_{1/2}$) because it has been determined that the peak width exceeds an increasing of 25% at this moment when performing the filtration on an initial Gaussian peak [2]. The step of increasing the time constant depends on the current sampling period (1/sampling rate).
For all windowed filters the window size starts from 3 because a size of 1 will not filter. The size increases by 2 due to odd sizes up to a maximum that depends on the peak width at half height of the narrowest peak ($W_{1/2}$). The maximum will be calculated by equation 6.

$$WS_{max} = \frac{1}{2} \cdot sampling\ rate \cdot W_{1/2} \quad (6)$$

The Savitzky-Golay and Gaussian filter requires additional parameters with specific ranges. The grade of polynomial starts from 1 up to 1/2 * window size in order to avoid over fitting. The standard deviation of the Gaussian starts from 0.1 and increases by 0.1 as well as done by Lytle et al. [4] up to a 1/2 * window size because a higher standard deviation leads to a very close equal distribution.
The evaluation of the filtered chromatograms where done by a peak detection as a combination of local maxima detection by the persistence 1D algorithm [13] and the baseline expansion performed in the APEX Track algorithm of Waters Corporation [14]. The computations of the peak width, symmetry factor, signal-to-noise ratio and resolution are

based on the definitions in the general chapter <621> in the USP dealing with the chromatography [15].

Results and Discussion

By the usage of the self written optimization tool each kind of chromatogram visible in Fig. 1 has been processed by data bunching for reducing the sampling rate and by the filtration algorithms using the predefined parameter ranges. After the process of the raw data the peaks have been detected and evaluated by computing the values of the optimization criterions: peak width at half height, symmetry factor, signal-to-noise ratio and resolution (except for chromatogram A in Fig. 1 due to one single peak). The resulting values are used as input for the continuously updated method to find the non-dominated solutions forming the a Pareto optimization front (POF). At the end after processing all possible signal filtration parameters and sampling rates the optimal solution is determined by calculating the Euclidean distance to the utopia point. The utopia point itself corresponds to an ideal point which contains minimal or maximal values of each found solution depending on the optimization type. The Euclidean distance (ED) to the utopia point was calculated as follows:

$$ED = \sqrt{\sum_{object} \left(\frac{v_{object} - v^*_{object}}{v_{norm}} \right)^2} \quad (7)$$

Where object is the criterion like the peak width, v as the value of the object, v* as the utopia value of the object and v_{norm} as a normalization value that depends on the global minimal and maximal object values in all determined solutions. Additionally to the filter algorithms it has been determined which sampling rate is necessary to describe the narrowest peak in the chromatogram by 20 data points. The evaluation of this situation is shown in Tab. 1 beside the performed filtration algorithms. As visible in the table none of the solutions has the best value in all of the four criterions at the same time. That behavior applies for all three chromatograms. Each filter algorithm was optimized in an own run in order to get a solution one by one. A comparison of the six results by the Euclidean distance shows that for all three chromatograms the Savitzky-Golay algorithm with the given parameter lead to the best solution at all. It is presented as the rows in bold format in Tab. 1. But it has to be considered that for all three chromatograms the best grade of polynomial seems to be 1 here. That corresponds to a moving average filtration. The reason why the moving average

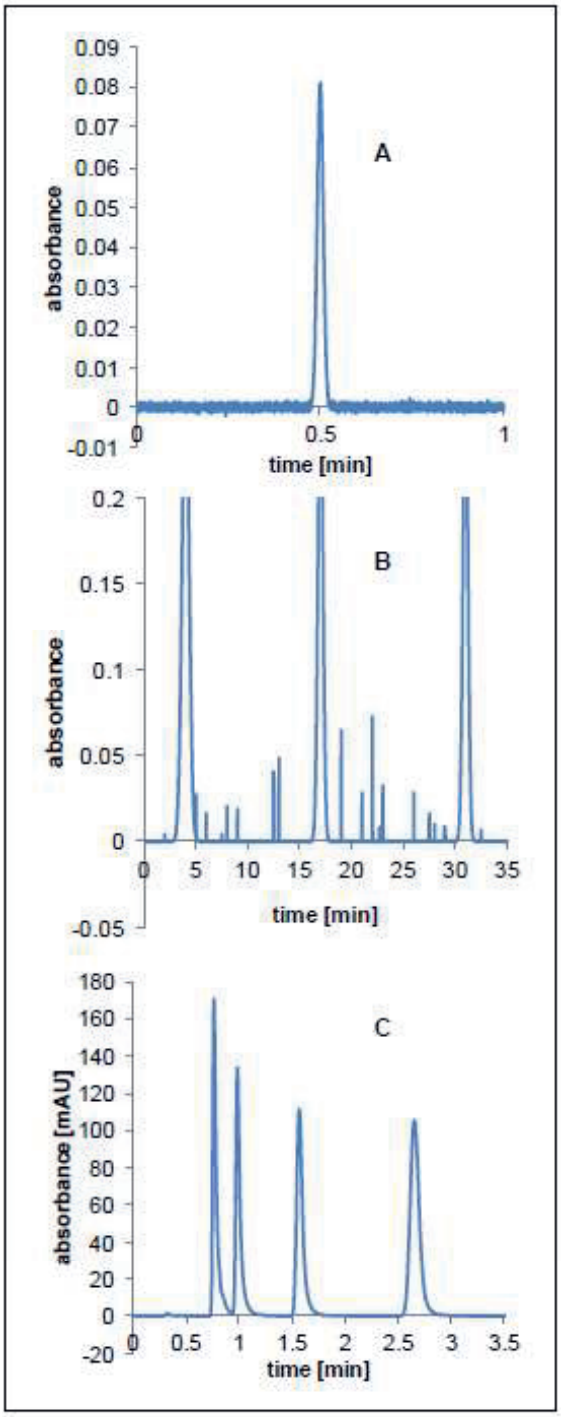

Fig. 1: A - Simulated chromatogram of one peak width sampling rate of 100 Hz, B - Simulated chromatogram based on separation of paracetamol, codeine, pitophenone and impurities in [12] with 40 Hz sampling rate, C - Real chromatogram of isocratic test sample acquired by diode array detector at 254 nm and with 80 Hz sampling rate.

filter and Savitsky-Golay did not have the same optimal solutions lies in the range of test area. The moving average has a wider range of settable window size because of the limited performance of the Savitsky-Golay filter with a maximal window size of 29. That way further solutions have been added to the POF for the moving average filter and the utopia point differs. That means comparing the final results between the algorithms is not as practicable as performing an optimization with all algorithms and parameters at once. Doing so is very time consuming and do not fulfill the demands in this article for the optimal solutions separated by each filter which corresponds with one implemented digital or installed physical filter in a data acquiring system. Nevertheless it is visible that representing the peaks by 20 data points is a good approach but if all three or four criterions are considered it is possible to find a better solution using one of the filter algorithms.

Conclusion

Using the written tool and apply the sampling reduction and all the known filtration algorithms on three test chromatograms has shown that it is possible to define a suitable test area for the filter parameter and optimize them by evaluation criterions that are usually determined during the method development and validation of a chromatographic experiment. That way it is possible to select a filter algorithm that belongs to a specific HPLC detector distributed by the different manufacturers and optimize the parameters.

For a general optimization comparing the different filter algorithms an all in one approach has to be performed by finding all solutions of the Pareto optimal front within one run.

During the current experiments the separated peaks in the chromatograms have been handled equivalent. In a following step the recommendations for the filter parameter can be improved by the distinction of the chromatograms in determination of active substance quantity or purity. This will influence the peaks that will be used for the evaluation. In Fig. 1 (b) for example a substance quantity determination would consider the three huge peaks of paracetamol, codeine and pitophenone only whereby the impurities become to the field of interest during a purity determination.

Tab. 1: Solutions of the optimization process divided by the processed chromatograms in Fig. 1 and the sampling rate and signal filtration parameters. The optimal solutions were determined by Euclidean distance to the utopia point. Each algorithm has its own utopia point. $W_{1/2}$=Peak width at half height, S/N-Ratio=Signal-to-Noise ratio, τ=time constant, WS=window size, GP=grade of polynomial, SD=standard deviation.

Sampling Rate and Filter Parameter	*∑ ($W_{1/2}$) [msec]*	*∑ |1 - Symmetry factor|*	*∑ (S/N-Ratio)*	*∑ (Resolution)*
Chromatogram A				
10.58 Hz - 20 data points	1170.233	0.007124312	156.0675	/
20 Hz - RC (τ = 150 ms)	1224.812	0.03504925	315.3293	/
40 Hz - Moving Average (WS = 11)	1173.835	0.06105919	257.8538	/
20 Hz - Hamming (WS = 5)	1170.628	0.04447394	226.5203	/
20 Hz - Savitzky-Golay (WS = 3, GP = 1)	**1168.979**	**0.006469480**	**206.7768**	**/**
20 Hz - Gaussian (WS = 5, SD = 2.3)	1170.858	0.04453622	230.6013	/
Chromatogram B				
11.55 Hz - 20 data points	134435.0	0.4056107	379.5420	1082.554
40 Hz - RC (τ = 50 ms)	125097.0	0.2609492	372.0598	1067.215
40 Hz - Moving Average (WS = 5)	126390.2	0.6281030	373.8028	1071.499
40 Hz - Hamming (WS = 5)	122863.3	0.5511815	359.7738	1079.187
40 Hz - Savitzky-Golay	**123327.1**	**0.2881788**	**362.7819**	**1078.525**

(WS = 3, GP = 1)				
40 Hz - Gaussian (ws = 5, sd = 1.6)	118718.9	0.6011666	342.7808	1084.018
Chromatogram C				
9.17 Hz - 20 data points	16984.34	7.781151	18247.30	62.51906
20 Hz - RC (τ = 100 ms)	16957.61	6.961686	22655.59	63.36911
80 Hz - Moving Average (ws = 39)	16367.06	6.170660	26634.29	76.55633
40 Hz - Hamming (ws = 19)	16175.93	6.374446	24793.98	80.99163
40 Hz - Savitzky-Golay (ws = 21, gp = 1)	**16397.73**	**6.288150**	**27415.59**	**76.67735**
80 Hz - Gaussian (ws = 51, sd = 20.9)	16206.84	6.237541	25445.40	80.37572

References

[1] Dedecke Handelsgesellschaft mbH. EP, USP und JP; *URL: http://www.dedecke-gmbh.de/epuspjp.html* (visited on 2017-10-12).

[2] M. F. Wahab, P. K. Dasgupta, A. F. Kadjo, D. W. Armstrong. Sampling frequency, response times and embedded signal filtration in fast, high efficiency liquid chromatography: A tutorial, *Analytica Chimica Acta* 907, 31-44 (2016); doi: 10.1016/j.aca.2015.11.043

[3] A. Felinger, A. Kilár, B. Boros. The myth of data acquisition rate, *Analytica Chimica Acta* 854 178-182 (2015), doi: 10.1016/j.aca.2014.11.014

[4] F. E. Lytle, R. K. Julian. Automatic Processing of Chromatograms in a High-Throughput Environment, *Clinical Chemistry* 62.1, 144-153 (2016); doi: 10.1373/clinchem.2015.238816

[5] D. Stauffer. Das Chromatogramm, *Chromatogramme richtig integrieren und bewerten*, Wiley-VCH Verlag GmbH & Co. KGaA, (2008); isbn: 978-3-527-62222-1 978-3-527-31774-5

[6] J. W. Dolan. The Role of the Signal-to-Noise Ratio in Precision and Accuracy, *LCGC Europe* 19.1, 12-16 (2006)

[7] G. Lungu. Introduction to Numerical Methods – Basic Filters #1; *URL: http://www.excelunusual.com/basic-filters-1/* (visited on 2018-04-12).

[8] P. A. Gorry. General Least-Squares Smoothing and Differentiation by the Convolution (Savitzky-Golay) Method, *Analytical Chemistry* 62(6), 570–573 (1990); doi: 10.1021/ac00205a007

[9] N. Gunantara. A review of multi-objective optimization: Methods and its applications, *Cogent Engineering* 5, 1-16 (2018); doi: 10.1080/23311916.2018.1502242

[10] G. Vivó-Truyols, J. R. Torres-Lapasió, A. M. van Nederkassel, Y. Vander Heyden, D. L. Massart. Automatic program for peak detection and deconvolution of multi-overlapped chromatographic signals: Part I: Peak detection, *Journal of Chromatography A* 1096, 133-145 (2005); doi: 10.1016/j.chroma.2005.03.092

[11] P. G. Boswell, D. R. Stoll, P. W. Carr, M. L. Nagel, M. F. Vitha, G. A. Mabbott. An Advanced, Interactive, High-Performance Liquid Chromatography Simulator and Instructor Resources, *Journal of chemical education* 90, 198-202 (2013); doi: 10.1021/ed300117b

[12] J. Vojta, P. Hanzlík, A. Jedlička, P. Coufal. Separation and determination of impurities in paracetamol, codeine and pitophenone in the presence of fenpiverinium in combined suppository dosage form, *Journal of Pharmaceutical and Biomedical Analysis* 102, 85-92 (2015); doi: 10.1016/j.jpba.2014.09.002

[13] T. Weinkauf, Y. Kozlov. Extracting and Filtering Minima and Maxima of 1D Functions, 2013, *URL: https://www.csc.kth.se/~weinkauf/notes/persistence1d.html*

[14] Waters Corporation. ApexTrack Integration: Theory and Application, *Document*, (2016)

[15] The United States Pharmacopeial Convention. <621> Chromatography, *The United States Pharmacopeia and National Formulary*, (2017); isbn: 978-3-7692-6821-8