

Signal Processing Methods for Capillary Electrophoresis

Robert Stewart¹, Iftah Gideoni² and Yonggang Zhu¹

¹CSIRO Materials Science and Engineering,

²CSIRO Information and Communication Technology Centre,
Australia

1. Introduction

Capillary electrophoresis (CE) is a separation technique that can be used as a sample pre-treatment step in the analysis of ionic analytes (Grossman and Colburn 1992; Stewart et al. 2008). Compared with other separation technologies it can offer advantages such as higher speed and sensitivity, smaller injection volumes and reduced consumption of solvent and samples, the possibility of miniaturisation, and reduced cost (Issaq 2001; Jarméus and Emmer 2008; Polesello and Valsecchi 1999; Wang 2005; Wee et al. 2008).

CE is based on the difference of the electrical mobilities of molecules within a capillary tube filled with electrolyte solution. When an electrical field is applied between the two ends of a capillary and a sample is introduced at one end, analytes are separated as they migrate towards the other end under the influence of the electrical field. These separated analytes are detected near the outlet by methods such as optical or electrochemical techniques (Polesello and Valsecchi 1999; Guijt et al. 2004; Kappes and Hauser 1999; Kubáň and Hauser 2004, 2009; Kuhn and Hoffstetter-Kuhn 1993; Marzilli et al. 1997; Tanyanyiwa et al. 2002; Zemann et al. 1998). The signal from a detector is digitised and typically presented in the form of voltage versus time, i.e. an *electropherogram*. Peaks evident in an electropherogram typically correspond to analytes in the sample, and with optimisation of the system parameters, the peaks can usually be resolved sufficiently. Fig. 1 shows an example electropherogram of data obtained from a practical trial reported earlier (Petkovic-Duran et al. 2008)

For analytical chemistry purposes, the operator's aim is to determine from the electropherogram what analytes are present and the corresponding concentrations. In this paper we assume that this is done by separating the task into two stages: Signal Processing, i.e., obtaining peak information from the electropherogram, and Pattern Matching, using this peaks' summary information to compare with established peak library of known chemicals. Whilst the process of identifying the peaks, removing the noise present and fitting curves for peak quantification is, to a large extent, done to-date manually by professionals, the operator would be greatly aided through fully automated techniques with little or no human input. This is particularly crucial for the development of field-deployable devices which could be operated by non-technical staff. Furthermore, automated signal processing techniques can allow results to be reproducible or consistent and can remove the subjectivity of a human evaluation. In addition, they can also detect features that may not

otherwise be obvious to the human eye, and enable the operation of detection devices by non-experts.

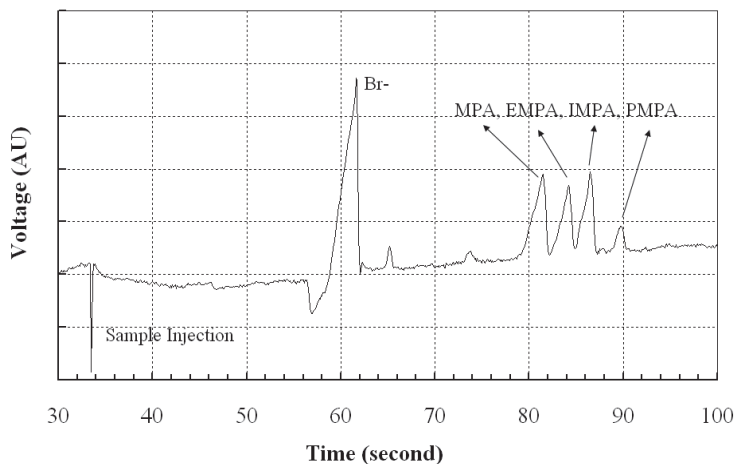


Fig. 1. An example electropherogram for separating a mixture of chemical warfare agent degradation products MPA, EMPA, IMPA and PMPA. Reproduced from Petkovic-Duran (2008). Conditions: 10mM MES/His (pH 6.1) buffer; separation field strength 340V/cm; injection field strength 625V/cm; Injection time 20s; frequency 360kHz; peak-to-peak-voltage 8V; sinus ac waveform. The first spike corresponds to the start of sample injection into the CE channel.

The task of automating the extraction of peaks, i.e. obtaining peak information such as peak shape, peak height, peak area and arrival time, from an electropherogram is, however, not an easy one. Not only must signal processing techniques be developed that can find the location of peaks, but they must do so in the presence of low and high frequency baseline noise that corrupts the signal. Furthermore, analytes may co-elute resulting in poorly resolved peaks that overlap. Even once the peak locations have been identified, the peaks need to be extracted and/or accurately quantified in the presence of interfering signal components.

The purpose of this paper is to review the current progress in signal processing relevant for capillary electrophoresis that are directed towards the quantification of peaks in electropherograms. We provide an overview of a signal processing strategy for a complete system, and then detail each signal processing step through examples cited from the literature. We are then able to draw some conclusions about the work needed to develop well defined and completely automated signal processing systems for CE. In the next section (Section 3) we detail a model for the electropherogram signal and the signal components to be analysed. Section 4 provides an overview of the steps in the signal processing strategy for CE signals. Pertinent examples from the literature are cited for each step (baseline noise removal, peak finding and peak extraction and quantification) to illustrate the different approaches adopted. Section 5 addresses how the performance of signal processing strategies/algorithms should be assessed, and on the need for benchmark testing and for the provision of system specifications. Section 6 discuss briefly some difficulties and requirements for the future Pattern Matching work needed for practically extracting

chemical identity information from the electropherogram's peak data. A concluding summary is provided in Section 7.

2. Modelling an electropherogram signal

An electropherogram is typically modelled as the superposition of a number of components under the assumption of system linearity (as are chromatograms (Dyson 1998) and mass spectrograms (Coombes et al. 2005)). These include the peaks corresponding to the analytes, system peaks and noise components. Mathematically, the model for an unfiltered/raw electropherogram signal can be expressed, e.g., in the form of voltage $v(t)$ (Kubáň and Hauser 2009), as follows,

$$v(t) = \sum_{i=1}^N p_i(t) + \sum_{j=1}^M s_j(t) + B + n(t) \quad (1)$$

where $p_i(t)$ is a peak that corresponds to the i^{th} analyte eluted, N is the number of analyte peaks, $s_j(t)$ is the j^{th} system peak, M is the number of system peaks, B is a constant baseline, and $n(t)$ is unwanted baseline noise. Here we consider a constant baseline, and any deviation from this is due to baseline noise, $n(t)$, which may itself contain a number of components. Eq. (1) is useful, not only for modelling electropherograms obtained from physical systems, but also for devising synthetic signals to test peak finding or peak extraction algorithms.

The aim of a peak extraction algorithm is to apply signal processing means to extract the separate peak components, $p_i(t)$, from an electropherogram. Each peak component can then be quantitatively analysed to provide information pertaining to the concentration of its corresponding analyte. When the sample being tested is unknown, the information obtained for the peaks can then be used in conjunction with prior knowledge or a database to identify the chemical compounds present in the sample (Reichenbach et al. 2009). In order to complete these tasks successfully, it is important to model or understand the characteristics of all the signal components in Eq. (1) and this will be discussed in detail in the following sections.

2.1 Peak models

Before discussing peak models, it is important to clarify the terminologies associated with a peak. On the basis of the peak definitions of The International Union of Pure and Applied Chemistry (IUPAC) and the Differential Thermal Analysis (DTA) (Inczédy et al. 1998), we propose a general definition for a peak with a view to it being widely applicable in different situations within a CE context, i.e. a peak is a portion of a signal or waveform with the characteristic of a rising and then a falling of the dependent variable with time. In particular, an analyte peak is a peak that is a signal component of an electropherogram (viz. $p_i(t)$ in Eq. (1)) resulting directly from the presence of an analyte and distinct from noise and other peaks [e.g. see (Vivó-Truyols et al. 2005a)] while a system peak is a peak directly resulting from the background electrolyte (viz. $s_j(t)$ in Eq. (1)) (Beckers and Everaerts 1997; Gaš et al. 2007; Gebauer and Boček 1997; Macka et al. 1997; Sellmeyer and Poppe 2002).

Fig. 2 provides an illustrative example of the qualitative definitions above. It can be seen how system and analyte peaks contribute to the electropherogram signal which also

contains low and high frequency baseline noise. It should be noted that we are only considering peaks that are at a coarser scale than the high frequency noise that is present. In this figure it is clear that if the low and high frequency baseline noise were removed, then the electropherogram peaks would approximate the analytes or system peaks provided they are fully resolved (don't overlap) and have the same constant baseline. It should also be mentioned that a peak can be either 'positive' or 'negative'. Negative peaks correspond to changes to below baseline and appear as valleys in an electropherogram. The profile of an analyte peak is dependent "on the physical-chemical processes inside the capillary, the heterogeneity of the capillary surface, capillary overload, solute mobility, and instrumental effects" (García-Alvarez-Coque et al. 2005).

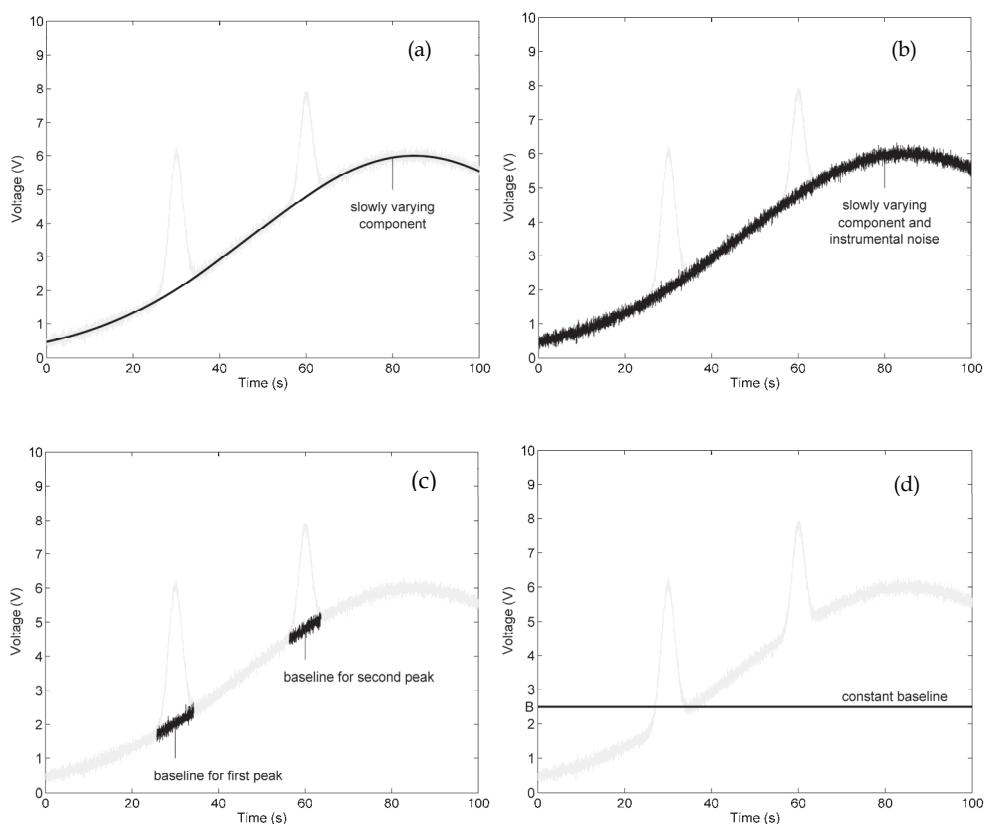


Fig. 2. A synthetic electropherogram composed from the superposition of peaks and baseline noise. Each interval in the electropherogram shown in bold corresponds to a successful or unsuccessful peak candidate.

A number of different peak models have been proposed in the literature. The triangle is likely to be the earliest peak model used (Dyson 1998) and is perhaps the simplest. It has been used as a peak model in a number of studies (Barclay and Bonner 1997; Stewart et al. 2008). One definition for a triangle function is (Couch 1990),

$$\Lambda\left(\frac{t}{T}\right) = \begin{cases} 1 - \frac{|t|}{T}, & |t| \leq T \\ 0, & |t| > T \end{cases} \quad (2)$$

where t is time, and T is half the width of the triangle which has unity height and is centred about the y -axis with an apex at $(0,1)$. This function can be scaled, translated and sampled to give a suitable digital signal representation of a peak.

A more common approach, however, has been to model a peak as a Gaussian curve or variant thereof (Bocaz-Beneventi et al. 2002; Oda and Landers 19970. Such curves are likely to be more realistic given the underlying physical-chemical process. For example, Solis et al. (Solis et al. 2007) define a peak to be of the form (Graves-Morris et al. 2006):

$$p_i[t] = A_i \exp\left[-4\left(\frac{T_s t - T_{o_i}}{W_i}\right)^2\right] \quad (3)$$

where A_i is the peak's amplitude, T_{o_i} is the migration time, W_i the width of the peak and T_s the sampling interval. This model could be applied to both $p_i(t)$ and $s_j(t)$ in Eq. (1). More complex peak models have also been introduced for CE. For example, to account for skewed peaks, a "Combination of Square Roots (CSR)" model has been proposed (García-Alvarez-Coque et al. 2005) and compared to other models. Another example is the resonance model for peaks described in (Graves-Morris et al. 2006).

As with CE, there is a need for peak models in other analytical chemistry techniques including chromatography, spectroscopy and gel and zone electrophoresis. The signals obtained in these techniques are generally alike, often owing to the similarity between the underlying physical-chemical processes (the mechanisms in chromatography and electrophoresis are particularly similar (Johansson et al. 2003; Poppe 1999). Gaussian based peak models are a popular choice for various techniques in analytical chemistry. For example, in chromatography the peaks in chromatograms are expected in theory to be close to Gaussian as they result from the dispersion of sample bands (Dyson 1998; Parris 1984). However, in practice, as with CE, Gaussian models must be modified or replaced to capture other effects that impact on peak shape (including asymmetry). Examples of modified Gaussian peak models include Exponentially Modified Gaussian (EMG) (Grushka 1972; Naish and Hartwell 1988; Poole 2003) and Exponential-Gaussian Hybrid (EGH) functions (Lan and Jorgenson 2001; Poole 2003). Numerous other peak models for chromatography have also been proposed (see the references cited in (García-Alvarez-Coque et al. 2005) for some further examples).

2.2 Baseline noise

Noise is usually the result of "spontaneous fluctuations which occur within matter at the microscopic level" (e.g. thermal noise and shot noise). In a CE system, noise components in a baseline signal may be due to electrical noise, chemical noise originating in the underlying physical-chemical processes of separation and so on. In general, baseline noise could contain both low frequency and high frequency components. The high frequency noise is from the instrument/detector which results from "incomplete grounding or from the signal amplification system" (Kuhn and Hoffstetter-Kuhn 1993) or from other sources such as electronic components including the Analogue to Digital Converter (ADC) (Jacobsen 2007;

Solis et al. 2007; Xu et al. 1997). The low frequency noise is mainly generated from temperature variations, impurities of the background electrolyte ('chemical noise') and air bubbles (Kuhn and Hoffstetter-Kuhn 1993; Xu et al. 1997). The unsatisfactory sample injection could also contribute to the low frequency noise due to the background buffer solution variations. The term *baseline drift* can refer to very low frequencies (Kuhn and Hoffstetter-Kuhn 1993) or to low frequencies in general (Solis et al. 2007). It should be mentioned that the detector type could also affect the baseline signal characteristics. For example, there are a range of detectors which can be used for CE or other techniques such as UV/Visible, fluorescence, electrochemical, conductivity, light scattering, mass spectral techniques. Some of the detectors measure bulk property of samples (e.g. conductivity and refractive index techniques) while some are for measuring solute properties (e.g. UV/Vis, fluorescence, electrochemical techniques). Bulk property detectors tend to have higher background signal which could vary due to background condition change. Solute detectors, on the other hand, usually have less background signal.

Since understanding the nature of the noise present in a system is important for an appropriate denoising of a signal (Perrin et al. 2001; Szymańska et al. 2007), numerous quantitative noise characterisation studies have been carried (Katsumine et al. 1999; Smith 2000; Smith 2007; Vaseghi 2008) to understand the spectral characteristics of the noise present in practical systems. The outcome of such studies may indicate the need for a unified noise model that does not partition the noise into low and high frequencies. For example, a brown noise model may be appropriate (Vaseghi 2008) or other models, e.g. a $1/f$ noise model (Katsumine et al. 1999; Smit and Walg 1975), or a correlated noise model (Perrin et al. 2001). The simplest method for noise modelling is to estimate the noise statistics from the signal-inactive periods (Vaseghi 2008). For instance, the noise in a system may be characterised by its power spectral density which can be estimated using a variety of methods (Cruz-Marcelo et al. 2008; Smith 2007). Of course, a noise process should really also be checked for stationarity by confirming its statistical parameters are constant over a sufficient interval of time. If a noise process is non-stationary, for example, the heteroscedastic noise (Li et al. 2006; Mittermayr et al. 1999; Mittermayr et al. 1997), then modelling techniques for time-varying stochastic processes will need to be employed (Vaseghi 2008). Where noise characterisation is impractical, adaptive filters or filters whose parameters can be adjusted or tuned manually in-situ may need to be employed for effective removal of noise.

It is worth mentioning that, while various models have been proposed for the low and high frequency noise, a unified approach is also required to handle the noise as a whole to simplify the signal processing process. The baseline drift and noise contributions could be modelled as interferences from different ends of frequency spectrum domain. Approaches may include Fourier and wavelet transforms. Some of these techniques will be reviewed in the next section

3. Signal processing techniques for peak extraction

3.1 Processing approaches

The aim of a peak extraction algorithm is to extract the separate peak components, $p_i(t)$, from an electropherogram. In general, when trying to identify the peaks in an unknown sample we aren't privy to information about when and where the different signal components will occur and in what measure. Only a measured signal is available and we have to solve an inverse problem (Mammone et al. 2007; Tarantola 2005) to identify the parameters in our

model. Approaches to solving single-channel source separation problems must rely on additional information, with filtering, decomposition and grouping, and source modelling approaches having been used (Schmidt 2008). A typical signal processing strategy might consist of a number of steps including: (i) the removal/suppressing of the baseline noise, (ii) finding the peaks, and (iii) extracting and/or quantifying the peaks. Some approaches combine some of these steps into a single step. In the following sections we detail each of these steps in further detail.

3.2 Baseline noise removal

To remove the baseline noise from an electropherogram, it is often convenient to assume the noise components are confined within certain spectral ranges. Linear filters can then be used to suppress the content in those ranges (Horlick 1972; Rabiner et al. 1975). There are many standard filters that can be used to do this, some of which have been listed (Yang et al. 2009). In their review on peak detection algorithms for mass spectrometry, various open source software packages were compared and the filters employed were identified.

3.2.1 High frequency noise removal

For removing the high frequency baseline noise in analytical chemistry data, the moving average filter is a popular choice. Perhaps the most intuitive filter to understand, this Finite Impulse Response (FIR) filter (non-recursive filter) performs local averaging to attenuate the rapidly fluctuating components of a signal (Lyons 2004; Oppenheim et al. 2007). It is used for smoothing in the software PROcess (Li et al. 2005; Yang et al. 2009) for smoothing. However, with this filter, peaks tend to be flattened out. A widely cited seminal paper by Savitzky and Golay (Savitzky and Golay 1964) provides the details for a popular alternative for filtering out high frequency baseline noise. The filter, known as the Savitzky-Golay (SG) filter, has filter coefficients that implement least squares polynomial smoothing and can denoise signals and calculate derivatives with reduced peak degradation (Leptos et al. 2006; Peters et al. 2007; Vivó-Truyols and Schoenmakers 2006; Vivó-Truyols et al. 2005a). The filter has been extended and improved in many different ways (Browne et al. 2007). Other digital filters have also been used including the Kaiser filter (Mantini et al. 2007) and Gaussian related filters (Leptos et al. 2006; Yang et al. 2009). In general, digital filters have different time and frequency domain characteristics that are appropriate for different situations and often have parameters that must be optimized so that peak distortion is minimised (Hamming 1983) and the maximal amount of noise removed. It is possible to custom design a digital filter based on an ideal transfer function which, for low-pass (high-frequency attenuating) filter designs, can entail choosing an appropriate cut-off frequency (Lam and Isenhour 1981).

3.2.2 Low frequency noise removal

Digital linear filters can also be used for removing low frequency baseline noise (drift). However, the standard linear filtering approaches can have problems when the baseline noise is not confined to specific frequency bands or overlaps with the frequency bands of peaks. Similar problems arise in the processing of ECG signals, and are of principal concern since filtering can cause significant distortion to important key features (Mozaffary and Tinati 2005). Whilst non-linear filters may be one alternative strategy (Kiryu et al. 1994), the most common approach for filtering out the low frequency baseline noise in CE signals (and

signals in related areas) is to estimate the low-frequency baseline and then subtract the estimate from the signal. Windowing and interpolation based methods are commonly used. For example, a moving window method, retaining minimums, was used to estimate the low frequency baseline noise component in chromatograms (Quémener et al. 2007). The estimated component was then subtracted from the original signal to give a corrected chromatogram. The method appeared successful but required the width of the moving window to be set experimentally. A similar strategy was employed for baseline correction of CE data (Coombes et al. 2005; Szymańska et al. 2007). In another published work (Gras et al. 1999), selected values from windows were interpolated using cubic splines to estimate the low frequency baseline noise in MALDI-TOF mass spectra. A different strategy was employed with the signal trend being estimated by removing the peaks from a mass spectrum signal (Mantini et al. 2007). Other researchers have also applied curve fitting techniques (Coombes et al. 2005; Bernabé-Zafón et al. 2005; Gillies et al. 2006; Mazet et al. 2005).

3.2.3 Wavelet transformation for noise removal

Despite some successful demonstrations of (low and high frequency) baseline noise removal techniques, no single approach has proved so undeniably successful across various conditions that it has been universally adopted. Recently, however, there has been some interest in the use of relatively new signal processing techniques such as wavelets.

Wavelets (Burrus et al. 1998; Daubechies 1988; Grossmann and Morlet 1984; Mallat 1989, 1999) allow the simultaneous analysis of a signal's time and frequency (or scale) properties, which is particularly useful for the processing of transient signals. The continuous wavelet transform (CWT) of a signal, $f(t) \in L^2\mathbf{R}$, at a scale, s , and time u , is given by (Mallat 1999),

$$Wf(u, s) = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{s}} \psi^* \left(\frac{t-u}{s} \right) dt \quad (4)$$

The “mother wavelet”, $\psi(t) \in L^2\mathbf{R}$, is a function with zero average and “is well localised in both time and frequency” (Cohen and Kovačević 1996). Under certain conditions an inverse continuous wavelet transform also exists that allows the reconstruction of $f(t)$ (e.g. Burrus et al. 1998; Cohen and Kovačević 1996). The coefficients from the wavelet transformation provide an indication of the signal energy contained at various scales over time. It is therefore possible to calculate the energy density (scalogram) of a signal which can be used in signal analysis and/or graphically depicted in a time-frequency heat map.

For a discrete wavelet transform (DWT), the wavelet coefficients can be computed for a discrete grid of points $(s, u)_{n \in \mathbf{Z}}$ (Cohen and Kovačević 1996). Practically though, a multiresolution analysis (MRA) (Mallat 1989) is most often used to give a series expansion for $f(t) \in L^2\mathbf{R}$, in terms of a scaling function, $\varphi(t)$, and wavelets, $\psi_{j,k}$ (Burrus et al. 1998),

$$f(t) = \sum_{k=-\infty}^{\infty} c(k) \varphi_k(t) + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} d(j, k) \psi_{j,k}(t) \quad (5)$$

where orthogonality constraints are placed on the expansion functions which form a basis. The first summation in Eq. (5) provides a low resolution approximation to $f(t)$, and each

successive index, j , in the second summation adds additional detail to the signal. A fast DWT algorithm that employs the use of filter banks has been developed by Mallat for the calculation of the approximation coefficients $c(k)$ and detail coefficients $d(j,k)$ (Mallat 1989), and is frequently used in practice. Once the coefficients have been obtained for a signal, they are thresholded or processed in various ways, before the reconstruction part of Mallat's algorithm can be used to reconstruct the signal.

One of the most widely used applications of wavelets is in denoising. Donoho and Johnstone developed a wavelet denoising technique based on the thresholding of coefficients (Donoho 1995; Donoho and Johnstone 1994). The denoising works on the premise that the underlying signal can have a sparse representation where it is approximated by a small number of relatively large-amplitude coefficients, whereas noise will have its energy spread across a large number of small-amplitude coefficients (Burrus et al. 1998; Mallat 1999). Hence, thresholding can remove noise whilst keeping the underlying signal largely intact. Different thresholding strategies such as hard thresholding and soft thresholding can be applied (Burrus et al. 1998; Donoho 1995).

A comprehensive study on wavelet denoising in CE using thresholding based methods was recently performed (Perrin et al. 2001). In the study, a number of wavelets from different wavelet families were tested, with the Haar wavelet found to perform the best. High frequency noise was removed by filtering the detail coefficients using hard thresholding and the low frequency baseline noise was removed by filtering the approximation coefficients using soft thresholding. Soft thresholding was used so as to reduce peak distortion. After thresholding, the inverse wavelet transform was calculated to reconstruct the signal with impressive results. The strategy was developed to accommodate larger baseline drifts (Liu et al. 2003). Fig. 3 shows an example of signal denoising using wavelet transform. This line of work was further developed by implementing spatially adaptive thresholding for the denoising of DNA capillary electrophoresis signals (Wang and Gao 2008). Wavelet denoising strategies were also used (Ceballos et al. 2008) on pattern recognition in CE. Similar wavelet based denoising strategies to those cited above have also been applied to liquid chromatography data (Barclay and Bonner 1997; Shao et al. 2004), Raman spectroscopy (Hu et al. 2007), mass spectrometry data (Barclay and Bonner 1997; Coombes et al. 2005), as well as numerous other areas of research (Jagtiani et al. 2008; Komsta 2009).

The DWT is sufficient in many scenarios when removing high and low frequency noise from signals. However, in some situations it may be appropriate to use the continuous wavelet transform so that finer grained control over the scales used can be gained. In particular, Jakubowska has done some interesting work using the CWT for removing high and low frequency noise in voltammetry signals (Jakubowska 2008). After finding the CWT, certain scale bands were identified as containing noise, and these were excluded in the reconstruction of a signal using the inverse CWT. The method was also suitable for resolving overlapping peaks which will be discussed in the next sections.

3.3 Peak detection

Once the noise has been filtered, the next task is to find the peaks from the filtered electropherogram $v_f(t) = v(t) - n(t)$. Many different criteria can be used to determine whether a point corresponds to the apex of a peak (Yang et al. 2009). One simple strategy for finding the location of peaks in $v_f(t)$ might be to find those points that are above a threshold and/or correspond to a local maximum found by looking for positive to negative zero crossings in

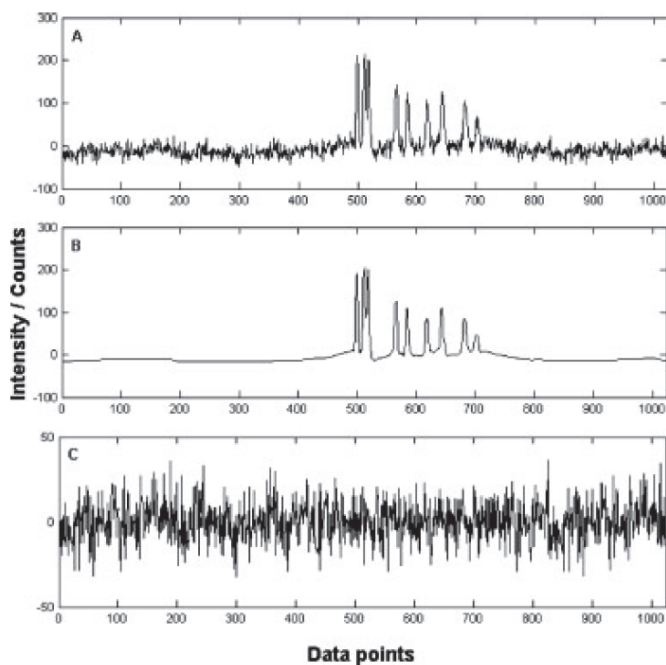


Fig. 3. An example of signal denoising by a discrete wavelet transform technique. (A) Original signal; (B) Denoised signal by db5 at decomposition level; (C) Removed noise. Reproduced from Liu et al. (2003) with permission from Wiley-VCH Verlag GmbH & Co. KGaA.

the signal derivative (or looking at just the second derivative). Thresholding and derivative based peak detection strategies are popular in various areas outside of CE. For example, an auto-threshold based peak detection algorithm for analyzing electrocardiograms was developed (Jacobson 2001); peaks in mass spectrometry data were found by selecting local maximum points with a signal-to-noise ratio above a certain value (Coombes et al. 2005; Mantini et al. 2007), and points of local maximum with an intensity sufficiently greater than neighbouring points were classified as peaks (Yasui et al. 2003). Zero-crossings of the derivative of signal were used to find the peaks in the analysis of partial discharge amplitude distributions (Carrato and Contin 1994). Derivatives of a smoothed signal are often used for peak detection in chromatography (Poole 2003). Signal peaks were detected in chromatograms after the Savitzky-Golay (SG) method was used to calculate derivatives (Peters et al. 2007). For CE, it was detailed how peaks can be identified through the location of inflection points (calculated using the second forward difference) (Graves-Morris et al. 2006).

Whilst derivative and threshold based techniques may be appropriate for some signals, low frequency baseline noise remaining in an electropherogram signal can disrupt threshold based peak detection (Wee et al. 2008), and high frequency noise can also cause problems especially for derivative based techniques, as discussed in (Lu et al. 2006). Recently, other approaches have been developed that are less susceptible to the inevitable unsuppressed noise that remains in the filtered electropherogram signal.

It is well established that wavelets can be used to detect singularities (points not differentiable) and edges (Mallat 1999; Struzik 2000). Singularities are detected by following across scales the local maxima of the wavelet transform. Using related multiscale zooming procedures, it is possible to detect discrete-time peaks even though the corresponding peaks may be continuous and differentiable. This idea has been applied to mass spectrometry data (Du et al. 2006). The CWT was firstly evaluated at different scales using the Mexican Hat wavelet to give a matrix of coefficients. Patterns evident in the matrix of coefficients were then analysed, with local maxima at each scale being linked across scales to give “ridges”. Ridges meeting certain conditions were then used to indicate the location of peaks. The algorithm was later extended to peak detection in CE (Petkovic-Duran et al. 2008; Stewart et al. 2008). An example of peak detection using the technique is shown in Fig. 4, with comparison with three other techniques (Du et al. 2006; Mantini et al. 2007; Morris et al. 2005). Similar peak detection approaches have also been adopted in other areas, such as detecting evoked potentials (EPs) of the brain in electroencephalograms (EEGs) using the DWT (McCooley et al. 2005; McCooley and Kumar 2007).

A potential benefit of these wavelet approaches is that the pre-processing step of removing low and high frequency baseline noise is unnecessary, as the process of selecting ridge lines can account for the presence of noise. Reconstruction of signals based on the local maxima of the wavelet-transform modulus is also possible (Mallat and Hwang 1992), so it would be interesting to investigate whether reconstruction of peaks in CE could be performed in this way. Such a technique has already been applied for extracting EPs of the brain from EEGs (Zhang and Zhen 1997) in addition to detection of EPs (mentioned earlier).

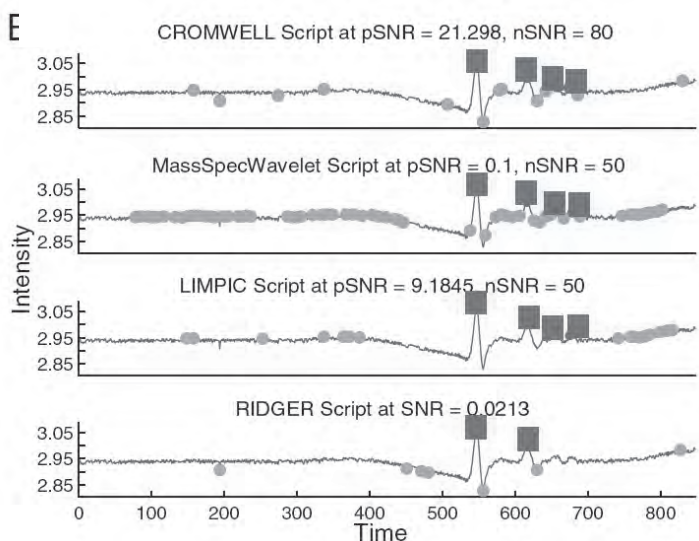


Fig. 4. An example of peak detection using a continuous wavelet transform technique and its comparison with other techniques. The techniques shown are, from top to bottom, CROMWELL (Morris et al. 2005), MassSpecWavelet Script (Du et al. 2006), LIMPIC (Mantini et al. 2007) and Ridger (Wee et al. 2008), respectively. Reproduced from (Wee et al. 2008) with permission from Wiley-VCH Verlag GmbH & Co. KGaA.

Whilst it is beyond the scope of this paper, it should be noted that when multiple data sets are available (i.e. data from multiple trials) statistical peak alignment or finding techniques can be applied (Ceballos et al. 2008; Coombes et al. 2005; Cruz-Marcelo et al. 2008; Dixon et al. 2006; Liu et al. 2008; Morris et al. 2005; Yu et al. 2008).

3.4 Peak resolution, peak extraction and quantification

After the baseline noise in the signal is suppressed and the peak locations are identified, peak components can be readily extracted by taking the portion of the signal above the constant baseline (B in Eq. 1), or $B + \delta$ where δ can account for a threshold that is above residual noise or tailing peaks) around the identified peak locations. A peak model can be fitted to the extracted peaks and measure important peak parameters such as area, height, skewness and so on. However, a single model may be insufficient and the curve fitting process is complicated when there are overlapping peaks that are not baseline resolved (Dyson 1998).

Usually in developing a CE method, the experimental conditions are optimised so as to ensure that all sample components are separated (Hanrahan et al. 2008; Vera-Candioti et al. 2008). However, such optimisation may not always be practicable and/or complete separation may be difficult, if not impossible, to achieve (Mammone et al. 2007; Sentellas et al. 2001; Zhang et al. 2007). As a result, unresolved peaks may be present in an electropherogram and this means quantitative measurements made directly on the peaks may be inaccurate (Dyson 1998). There is thus great need for signal processing techniques that are able to separate overlapping peaks especially in situations where conducting additional trials would be undesirable, infeasible or ineffectual.

Were the data acquired multi-dimensional (e.g. from multiple identical or dissimilar detectors or resulting from multiple trials under similar or dissimilar conditions), numerous statistical techniques would be at the researchers disposal to try and resolve overlapping peaks (Bocaz-Beneventi et al. 2002; Li et al. 2006; Sentellas et al. 2001; Zhang et al. 2007; Zhang and Li 2006). However, different approaches are necessary when the data from only a one-dimensional data vector (that results in a single electropherogram trace) is available. There are two main signal processing approaches that can be followed to analyse overlapping peaks.

The first approach is to try and extract the peak components from the signal using curve fitting. An accurate peak model is required and, if the number of peaks is known, the model can then be fitted to the peaks that are overlapping using non-linear least squares curve fitting techniques. Such techniques have been applied to CE (Jarméus and Emmer 2008; Vera-Candioti et al. 2008), as well as chromatography (Dasgupta 2008; Jin et al. 2008; Vivó-Truyols et al. 2005a, b), voltammetry (Huang et al. 1995) and gel electrophoresis (Kitazoe et al. 1983; Shadle et al. 1997) to name but a few. Fig. 5 shows an example of deconvolution technique (Vivó-Truyols et al. 2005b) for extracting overlapping peaks from a chromatographic signal. Such technique is also applicable for CE and other similar signals. However, the above-mentioned methods require a good peak model and initial parameter estimates must be sufficiently accurate for convergence of the curve fitting routines (Jarméus and Emmer 2008; Olazábal et al. 2004). In addition, in some cases the methods may be sensitive to noise and it may be difficult to automate the whole process (Jarméus and Emmer 2008; Vivó-Truyols et al. 2005a; Zhang et al. 2000). If there is poor separation, quantitative measurements made on the peaks may be inaccurate (Du et al. 2006).

The second approach to analysing overlapping peaks is to apply signal processing to increase the resolution of the peaks by decreasing their width but preserving their height or

area. For example, voltammetric or polarographic peaks were resolved using a deconvolution method (Engblom 1990). In the technique, signals were transformed to the Fourier domain where the width of the peak (in the time domain) could be modified by dividing/multiplying the Fourier transform by a suitable function (note: multiplication in the Fourier domain corresponds to convolution in the time domain). By an appropriate function choice, peaks can be resolved in the time domain by having their widths reduced but their height or area preserved. A deconvolution approach was also detailed in Kauppinen et al. 1981 and more recently deconvolution/convolution approaches have been used in combination with wavelets (Wang et al. 2004; Zhang et al. 2000; Zhang et al. 2001; Zheng et al. 2000).

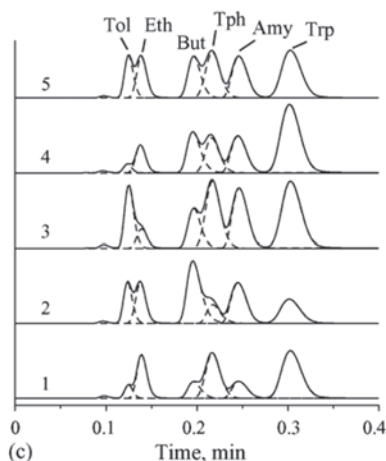


Fig. 5. An example of deconvolution of multi-overlapped chromatographic signal peaks. The five runs correspond to five mixtures eluted with 80% (m/m) methanol. The compounds are toluene (Tol), ethylbenzene (Eth), butylbenzene (But), o-terphenyl (Tph), amylbenzene (Amy), and triphenylene (Trp), respectively. Reprinted from Vivó-Truyols et al. (2005b) with permission from Elsevier.

Perhaps the most popular approaches for dealing with overlapping peaks involve the use of wavelet transforms. These can be used to transform a signal with overlapping peaks into a new waveform with resolved peaks, where, in so doing, some peak parameters (e.g. area and location) can be preserved. Such approaches are suitable where quantification of certain peak parameters is of interest but the complete peak waveform is not required. For example, the discrete wavelet transform (DWT) was used (Shao et al. 1997) to process chromatograms by decomposing them into detail coefficients at different decomposition levels. The overlapping peaks for the chromatograms analysed had peaks in the detail coefficients that were well resolved at a specific level of decomposition. After baseline correction (which was needed as a secondary step) the areas under the peaks were used to provide an accurate quantitative measure of the concentration.

The continuous wavelet transform (CWT) has also been used to resolve overlapping peaks in CE. Unlike the dyadic levels of decomposition typically used with the DWT's orthogonal wavelets, in non-orthogonal wavelet analysis the choice of scales is arbitrary (Torrence and

Compo 1998). For example, the maximum wavelet scale was chosen to resolve coefficient peaks Jiao et al. 2008. At the selected scale, the resolved coefficient peaks could then be quantitatively analysed. Other researchers have applied similar approaches using the CWT (Jakubowska and Kubiak 2008; Shao et al. 1997; Shao and Sun 2001; Wang et al. 2004). In choosing the scale, the aim is to minimise the noise but ensure peaks are baseline resolved, and the mother wavelet chosen as well as the signals being analysed impact upon the results. Where resolution at specific scales in the wavelet coefficients is not evident, information across scales can also be utilised (Xiao-Quan et al. 1999) as it is for peak finding (see Section 4.3).

In difficult problems, peaks may still not be resolved by the CWT coefficients. In such cases it may be useful or necessary to select a range of scales and reconstruct a time-domain signal using the inverse CWT (Jakubowska 2008). An appropriate choice for the scale band used can then lead to resolution of peaks in the reconstructed signal (Torrence and Compo 1998). Whilst it is clear that a range of techniques exist for extracting quantitative information about overlapping peaks, there is still a need for techniques that are able to extract the peak components in their entirety. The process of resolving peaks with wavelet transforms does not generally permit the peak components to be extracted directly, as usually some of the peak parameters are modified during transformation. Even though this isn't a problem for obtaining some quantitative information, it may be useful to have a complete representation for the peaks. In addition, most of the techniques demonstrated are for particular problems or setups. Generalised approaches to resolving overlapping peaks are needed that make minimal assumptions about the peak components and baseline noise.

4. Assessing algorithm and system performance

As this review has revealed, there are numerous approaches available for processing signals in capillary electrophoresis, each with a different signal processing strategy and demonstration on synthetic and/or real test data. However, little effort has been devoted to the comparison of performance from these different methods in spite of the few published work (Barclay and Bonner 1997; Cruz-Marcelo et al. 2008; Wee et al. 2008; Yang et al. 2009). In order to select or develop most efficient method, comparative studies are needed to test the methods on the same data (Leptos et al. 2006). One way to enable this happening is to make the software algorithms free and publically available so that other researchers can use and compare. One good example is that of Mantini and co-workers who provide both the algorithm source code and test data as additional material accompanying their paper (Mantini et al. 2007). Perhaps the more practical way is to standardise the test data sets so that different algorithms can be compared using the same data for their performance in key parameters such as noise removal, peak detection, peak resolution, peak extraction and quantification, and computing efficient, speed and power requirements. This can be viewed as a step towards reproducible research (Arora and Barak 2009; Deb 2001; Zheng et al. 1998). This type of approach is common in other areas of research. For example, in the area of multi-objective evolutionary optimisation (MOEA), there are standard test problems that algorithms are compared against (Deb 2001). Since that algorithm performance is likely to be dataset dependent, there should be a range of standardised benchmark datasets which could be obtained from real experimental setups or generated synthetically, and made publically available.

Standard performance indicators are needed in addition to standardised test data, so that algorithms can be quantitatively assessed and compared. For example, peak detection algorithms could be assessed by their false discovery rate (FDR) and sensitivity (Cruz-Marcelo et al. 2008; Wee et al. 2008; Yang et al. 2009) or receiver operating characteristic (ROC) curve (Mantini et al. 2007). For assessing noise removal efficiency or for evaluating the preservation of peak properties after peak resolution, different measures that might be used include: root square error (RSE) or integrated square error (ISE) (Jagtiani et al. 2008), root mean square (RMS) (Barclay and Bonner 1997), relative error (RE) (Zhang et al. 2001; Zheng et al. 1998), individual sum of squared residuals (Vivó-Truyols et al. 2005b), signal-to-noise ratio (SNR) and correlation coefficient (Jakubowska and Kubiak 2008). In addition to these performance indicators, an analysis of an algorithm's computational complexity (Arora and Barak 2009) would also be worth reporting. If this approach were widely adopted, then the performance of algorithms could be readily compared.

Assessing algorithm performance is one step towards complete system characterisation and performance assurances. With appropriate quantification of the noise processes present in a system and bounded variation in the peak model, in conjunction with knowledge regarding the performance of the signal processing algorithms used, a researcher should be able to specify the quantitative properties of peaks along with a determined uncertainty measure. This would also allow systems to be compared with other (including commercial) systems and would ensure objective assessment of results could be made.

5. Pattern matching - inferring chemical identity

Given the successful extraction of peaks from an electropherogram, the next step is to use this summary information to identify the present chemicals. Substantial effort has been devoted to the algorithm development for such a purpose (Fruetel et al. 2006; García-Pérez et al. 2008; Liu et al. 2008; Stein and Scott 1994). As this process requires reference to a library of peak information, extracted from known chemicals measured in similar environment, we will refer to it as *Pattern Matching*.

Due to uncertain inputs, the pattern matching is an inference process, resulting in probabilities. Inferring about the general composition of a sample, without reference to particular list of chemicals, is beyond what we consider possible from electrophoresis, as the solution for this question, based on electrophoresis results, is materially under-constraint. In general, we attempt to answer the following question:

What are the odds that a particular chemical (i) is present in the sample, given the peaks expected from chemical i and other chemicals in the same environment?

$$\frac{P(\text{Chemical}_i / DI)}{P(\text{Chemical}_i / DI)} = ? , \quad (6)$$

where D is the extracted peaks data, and I is the background library information.

The pattern matching is encumbered by several real life constraints: the accuracy of peaks extraction is not assured; It is not certain that the peaks extracted from two measurements, performed on same sample and in the same environment, will be the same. The list of known chemicals may not be exhaustive; chemicals for which we do not have peaks information (library) may be present in the sample. It is possible that the measurement environment is not identical to the environment in which the library was created, leading to

different peaks extracted for same sample. It is also possible that several other chemicals will be present in the sample, influencing peaks extracted and attributed to the chemical of interest.

An effective pattern matching will need to reliably account for these uncertainties, and consistently indicate the probabilities of chemicals presence, using our understanding that several other known and unknown chemicals may be present in the sample. This, in general, cannot be solely done using naive matching of the extracted peaks to the library peaks of the chemical of interest, as it ignores the possible presence of other chemicals. And, comparison of separate calculation of likelihoods of the peaks measured for each of the known chemicals ignores other information in hand, such as the possibility of presence of unknown chemicals and concurrent presence of chemicals. The pattern matching process will also benefit from ability to mount the evidence, or in other words, to learn: given peaks measured in one environment, the adequate process should combine this information with information extracted from measurement of the same sample in a different environment. Such a pattern matching process for electrophoresis is, to the best of our knowledge, not defined yet, and is the subject of future work.

6. Concluding remarks

In this paper we have provided an overview of the signal processing methods that have been used in capillary electrophoresis and other related areas. We firstly discuss the various models proposed in the literature for modelling peak shapes and baseline noise. Signal processing techniques for extracting peaks from the signal are then reviewed. This covers noise removal such as digital filters and wavelet transforms, peak detection, peak extraction and quantification. We also discuss possible approaches for assessing algorithm development and system performance.

The problem of identifying peaks could be regarded as feature extraction, and problems of such a nature are not confined to analytical chemistry. There exists significant opportunity to apply ideas and adapt techniques from other disciplines (such as pattern recognition, adaptive control, particle swarm optimisation, evolutionary multi-objective optimisation, machine learning, artificial neural networks (ANN) and artificial intelligence generally), to the processing of signals in CE. Indeed, for real systems that may have drifting parameters or characteristics, techniques from other disciplines may be necessary to produce adaptable signal processing algorithms.

With benchmark testing and performance assessment, we should be able to develop algorithms to realise effective and efficient automatic signal analysis and quantification. Peaks quantified could then be used as the input to an inference pattern matching stage to determine the analyte components present in a sample. These steps will help facilitate the development of CE systems that reliably perform to specification and allow users to focus on the experimental results of separation.

7. Acknowledgments

Financial support from the Australian Government Department of the Prime Minister and Cabinet through the National Security Science and Technology Branch, the Australian Federal Police (AFP) and the Australian Customs Service (ACS) is greatly acknowledged. The authors would also like to thank Ms. Karolina Petkovic-Duran for providing the data used in Fig. 1.

8. References

- Arora S, Barak B (2009) Computational Complexity: A Modern Approach. Cambridge University Press, Cambridge
- Barclay VJ, Bonner RF (1997) Application of Wavelet Transforms to Experimental Spectra: Smoothing, Denoising, and Data Set Compression. *Anal Chem* 69:78-90
- Beckers JL, Everaerts FM (1997) System peaks in capillary zone electrophoresis What are they and where are they coming from? *J Chromatogr A* 787:235-242
- Bernabé-Zafón V, Torres-Lapasíó JR, Ortega-Gadea S, Simó-Alfonso EF, Ramis-Ramos G (2005) Capillary electrophoresis enhanced by automatic two-way background correction using cubic smoothing splines and multivariate data analysis applied to the characterisation of mixtures of surfactants. *J Chromatogr A* 1065:301-313
- Bocaz-Beneventi G, Latorre R, Farková M, Havel J (2002) Artificial neural networks for quantification in unresolved capillary electrophoresis peaks. *Anal Chim Acta* 452:47-63
- Browne M, Mayer N, Cutmore TRH (2007) A multiscale polynomial filter for adaptive smoothing. *Digital Signal Process* 17:69-75
- Burrus CS, Gopinath RA, Guo H (1998) Introduction to Wavelets and Wavelet Transforms: A Primer. Prentice-Hall, Upper Saddle River
- Carrato S, Contin A (1994) Application of a peak detection algorithm for the shape analysis of partial discharges amplitude distributions. Conference Record of the 1994 IEEE International Symposium on Electrical Insulation:288-291
- Ceballos GA, Paredes JL, Hernández LF (2008) Pattern recognition in capillary electrophoresis data using dynamic programming in the wavelet domain. *Electrophoresis* 29:2828-2840
- Cohen A, Kovačević J (1996) Wavelets: the mathematical background. *Proc IEEE* 84:514-522
- Coombes KR, Tsavachidis S, Morris JS, Baggerly KA, Hung M-C, Kuerer HM (2005) Improved peak detection and quantification of mass spectrometry data acquired from surface-enhanced laser desorption and ionization by denoising spectra with the undecimated discrete wavelet transform. *Proteomics* 5:4107-4117
- Couch LW, II (1990) Digital and Analog Communication Systems. 4th edn. Macmillan, New York
- Cruz-Marcelo A, Guerra R, Vannucci M, Li Y, Lau CC, Man T-K (2008) Comparison of algorithms for pre-processing of SELDI-TOF mass spectrometry data. *Bioinformatics* 24:2129-2136
- Dasgupta PK (2008) Chromatographic peak resolution using Microsoft Excel Solver: The merit of time shifting input arrays. *J Chromatogr A* 1213:50-55
- Daubechies I (1988) Orthonormal bases of compactly supported wavelets. *Commun Pure Appl Math* 41:909-996
- Deb K (2001) Multi-Objective Optimization using Evolutionary Algorithms. John Wiley and Sons, Chichester
- Dixon SJ, Brereton RG, Soini HA, Novotny MV, Penn DJ (2006) An automated method for peak detection and matching in large gas chromatography-mass spectrometry data sets. *J Chemom* 20:325-340
- Donoho DL (1995) De-noising by soft-thresholding. *IEEE Trans Inf Theory* 41:613-627
- Donoho DL, Johnstone IM (1994) Ideal spatial adaptation by wavelet shrinkage. *Biometrika* 81:425-455

- Du P, Kibbe WA, Lin SM (2006) Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics* 22:2059-2065
- Dyson N (1998) *Chromatographic Integration Methods*. Royal Society of Chemistry Chromatography Monographs, 2nd edn. Royal Society of Chemistry, Cambridge
- Engblom SO (1990) The Fourier transform of a voltammetric peak and its use in resolution enhancement. *J Electroanal Chem* 296:371-394
- Fruetel JA, West JAA, Debusschere BJ, Hukari K, Lane TW, Najm HN, Jose Ortega J, Renzi RF, Shokair I, VanderNoot VA (2006) Identification of Viruses Using Microfluidic Protein Profiling and Bayesian Classification. *Anal Chem* 80:9006-9012
- García-Alvarez-Coque MC, Simó-Alfonso EF, Sanchis-Mallols JM, Baeza-Baeza JJ (2005) A new mathematical function for describing electrophoretic peaks. *Electrophoresis* 26:2076-2085
- García-Pérez I, Vallejo M, García A, Legido-Quigley C, Barbas C (2008) Metabolic fingerprinting with capillary electrophoresis. *J Chromatogr A* 1204:130-139
- Gaš B, Hruška V, Dittmann M, Bek F, Witt K (2007) Prediction and understanding system peaks in capillary zone electrophoresis. *J Sep Sci* 30:1435-1445
- Gebauer P, Boček P (1997) System peaks in capillary zone electrophoresis I. Simple model of vacancy electrophoresis. *J Chromatogr A* 772:73-79
- Gillies P, Marshall I, Asplund M, Winkler P, Higinbotham J (2006) Quantification of MRS data in the frequency domain using a wavelet filter, an approximated Voigt lineshape model and prior knowledge. *NMR Biomed* 19:617-626
- Gras R, Müller M, Gasteiger E, Gay S, Binz P-A, Bienvenut W, Hoogland C, Sanchez J-C, Bairoch A, Hochstrasser DF, Appel RD (1999) Improving protein identification from peptide mass fingerprinting through a parameterized multi-level scoring algorithm and an optimized peak detection. *Electrophoresis* 20:3535-3550
- Graves-Morris PR, Fell AF, Bensalem M (2006) Parameterisation of symmetrical peaks in capillary electrophoresis using $[3/2]$ -type rational approximates. *J Comput Appl Math* 189:220-227
- Grossman PD, Colburn JC (eds) (1992) *Capillary Electrophoresis: Theory and Practice*. Academic Press, San Diego
- Grossmann A, Morlet J (1984) Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM J Math Anal* 15:723-736
- Grushka E (1972) Characterization of Exponentially Modified Gaussian Peaks in Chromatography. *Anal Chem* 44:1733-1738
- Guijt RM, Evenhuis CJ, Macka M, Haddad PR (2004) Conductivity detection for conventional and miniaturised capillary electrophoresis systems. *Electrophoresis* 25:4032-4057
- Hamming RW (1983) *Digital Filters*. Prentice-Hall, Englewood Cliffs
- Hanrahan G, Montes R, Gomez FA (2008) Chemometric experimental design based optimization techniques in capillary electrophoresis: a critical review of modern applications. *Anal Bioanal Chem* 390:169-179
- Horlick G (1972) Digital data handling of spectra utilizing Fourier transformations. *Anal Chem* 44:943-947
- Hu Y, Jiang T, Shen A, Li W, Wang X, Hu J (2007) A background elimination method based on wavelet transform for Raman spectra. *Chemom Intell Lab Syst* 85:94-101

- Huang W, Henderson TLE, Bond AM, Oldham KB (1995) Curve fitting to resolve overlapping voltammetric peaks: model and examples. *Anal Chim Acta* 304:1-15
- Inczédy J, Lengyel T, Ure AM (eds) (1998) *Compendium of Analytical Nomenclature: Definitive Rules 1997*. 3rd edn. Blackwell Science, Oxford
- Issaq HJ (2001) The role of separation science in proteomics research. *Electrophoresis* 22:3629-3638
- Jacobsen NE (2007) *NMR Spectroscopy Explained: Simplified Theory, Applications and Examples for Organic Chemistry and Structural Biology*. John Wiley and Sons, Hoboken
- Jacobson ML (2001) Auto-threshold peak detection in physiological signals. *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*.
- Jagtiani AV, Sawant R, Carletta J, Zhe J (2008) Wavelet transform-based methods for denoising of Coulter counter signals. *Meas Sci Technol* 19:065102
- Jakubowska M (2008) Inverse continuous wavelet transform in voltammetry. *Chemom Intell Lab Syst* 94:131-139
- Jakubowska M, Kubiak WW (2008) Signal processing in normal pulse voltammetry by means of dedicated mother wavelet. *Electroanalysis* 20:185-193
- Jarméus A, Emmer Å (2008) CE Determination of monosaccharides in pulp using indirect detection and curve-fitting. *Chromatographia* 67:151-155
- Jiao L, Gao S, Zhang F, Li H (2008) Quantification of components in overlapping peaks from capillary electrophoresis by using continuous [sic] wavelet transform method. *Talanta* 75:1061-1067
- Jin G, Xue X, Zhang F, Zhang X, Xu Q, Jin Y, Liang X (2008) Prediction of retention times and peak shape parameters of unknown compounds in traditional Chinese medicine under gradient conditions by ultra performance liquid chromatography. *Anal Chim Acta* 628:95-103
- Johansson G, Isaksson R, Harang V (2003) Migration time and peak area artifacts caused by systemic effects in voltage controlled capillary electrophoresis. *J Chromatogr A* 1004:91-98
- Kappes T, Hauser PC (1999) Electrochemical detection methods in capillary electrophoresis and applications to inorganic species. *J Chromatogr A* 834:89-101
- Katsumine M, Iwaki K, Matsuda R, Hayashi Y (1999) Routine check of baseline noise in ion chromatography. *J Chromatogr A* 833:97-104
- Kauppinen JK, Moffatt DJ, Mantsch HH, Cameron DG (1981) Fourier transforms in the computation of self-deconvoluted and first-order derivative spectra of overlapped band contours. *Anal Chem* 53:1454-1457
- Kiryu T, Kaneko H, Saitoh Y Artifact elimination using fuzzy rule based adaptive nonlinear filter. In: *Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference on, 19-22 Apr 1994*. pp III/613-III/616
- Kitazoe Y, Miyahara M, Hiraoka N, Ueta H, Utsumi K (1983) Quantitative determination of overlapped proteins in sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *Anal Biochem* 134:295-302
- Komsta Ł (2009) Suppressing the charged coupled device noise in univariate thin-layer videoscans: A comparison of several algorithms. *J Chromatogr A* 1216:2548-2553

- Kubáň P, Hauser PC (2004) Contactless conductivity detection in capillary electrophoresis: A review. *Electroanalysis* 16:2009-2001
- Kubáň P, Hauser PC (2009) Ten years of axial capacitively coupled contactless conductivity detection for CZE - a review. *Electrophoresis* 30:176-188
- Kuhn R, Hoffstetter-Kuhn S (1993) *Capillary Electrophoresis: Principles and Practice*. Springer-Verlag, Berlin
- Lam RB, Isenhour TL (1981) Equivalent width criterion for determining frequency domain cutoffs in fourier transform smoothing. *Anal Chem* 53:1179-1182
- Lan K, Jorgenson JW (2001) A hybrid of exponential and gaussian functions as a simple model of asymmetric chromatographic peaks. *J Chromatogr A* 915:1-13
- Leptos KC, Sarracino DA, Jaffe JD, Krastins B, Church GM (2006) MapQuant: Open-source software for large-scale protein quantification. *Proteomics* 6:1770-1782
- Li H, Hou J, Wang K, Zhang F (2006) Resolution of multicomponent overlapped peaks: A comparison of several curve resolution methods. *Talanta* 70:336-343
- Li X, Gentleman R, Lu X, Shi Q, Iglehart JD, Harris L, Miron A (2005) SELDI-TOF mass spectrometry protein data. In: Gentleman R, Irizarry RA, Carey VJ, Dudoit S, Huber W (eds) *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer, New York, pp 91-109
- Liu B-F, Sera Y, Matsubara N, Otsuka K, Terabe S (2003) Signal denoising and baseline correction by discrete wavelet transform for microchip capillary electrophoresis. *Electrophoresis* 24:3260-3265
- Liu J, Yu W, Wu B, Zhao H (2008) Bayesian mass spectra peak alignment from mass charge ratios. *Cancer Inform* 6:217-241
- Lu W, Nystrom MM, Parikh PJ, Fooshee DR, Hubenschmidt JP, Bradley JD, Low DA (2006) A semi-automatic method for peak and valley detection in free-breathing respiratory waveforms. *Med Phys* 33:3634-3636
- Lyons RG (2004) *Understanding Digital Signal Processing*. Prentice Hall, Upper Saddle River
- Macka M, Haddad PR, Gebauer P, Boček P (1997) System peaks in capillary zone electrophoresis 3. Practical rules for predicting the existence of system peaks in capillary zone electrophoresis of anions using indirect spectrophotometric detection. *Electrophoresis* 18:1998-2007
- Mallat S, Hwang WL (1992) Singularity detection and processing with wavelets. *IEEE Trans Inf Theory* 38:617-643
- Mallat SG (1989) A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans Pattern Anal Mach Intell* 11:674-693
- Mallat SG (1999) *A Wavelet Tour of Signal Processing*. 2nd edn. Academic Press, San Diego
- Mammone N, Fiasché M, Inuso G, Foresta FL, Morabito FC, Versaci M (2007) Information theoretic learning for inverse problem resolution in bio-electromagnetism. *LNAI* 4694 (4694):414-421
- Mantini D, Petrucci F, Pieragostino D, Boccio PD, Nicola MD, Ilio CD, Federici G, Sacchetta P, Comani S, Urbani A (2007) LIMPIC: a computational method for the separation of protein MALDI-TOF-MS signals from noise. *BMC Bioinf* 8:101
- Marzilli LA, Bedard P, Mabrouk PA (1997) Learning to Learn: An Introduction to Capillary Electrophoresis. *Chem Educ* 1 (6):1-12

- Mazet V, Carteret C, Brie D, Idier J, Humbert B (2005) Background removal from spectra by designing and minimising a non-quadratic cost function. *Chemom Intell Lab Syst* 76:121-133
- McCooley C, Kumar DK, Cosic I Decomposition of evoked potentials using peak detection and the discrete wavelet transform. In: *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, 2005. pp 2071-2074
- McCooley CG, Kumar D Automated peak decomposition of evoked potential signals using wavelet transform singularity detection. In: *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007. pp 23-26
- Mittermayr CR, Lendl B, Rosenberg E, Grasserbauer M (1999) The application of the wavelet power spectrum to detect and estimate 1/f noise in the presence of analytical signals. *Anal Chim Acta* 388:303-313
- Mittermayr CR, Rosenberg E, Grasserbauer M (1997) Detection and estimation of heteroscedastic noise by means of the wavelet transform. *Anal Commun* 34:73-75
- Morris JS, Coombes KR, Koomen J, Baggerly KA, Kobayashi R (2005) Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum. *Bioinformatics* 21 (9):1764-1775
- Mozaffary B, Tinati MA (2005) ECG Baseline Wander Elimination using Wavelet Packets. *World Acad Sci Eng Tech* 3:14-16
- Naish PJ, Hartwell S (1988) Exponentially Modified Gaussian Functions - a Good Model for Chromatographic Peaks in Isocratic HPLC? *Chromatographia* 26:285-296
- Oda RP, Landers JP (1997) Introduction to Capillary Electrophoresis. In: Landers JP (ed) *Handbook of Capillary Electrophoresis*. CRC Press, Boca Raton, pp 1-47
- Olazábal V, Prasad L, Stark P, Olivares JA (2004) Application of wavelet transforms and an approximate deconvolution method for the resolution of noise overlapped peaks in DNA capillary electrophoresis. *Analyst* 129:73-81
- Oppenheim AV, Willsky AS, Young IT (2007) *Signals and Systems*. Prentice-Hall, Englewood Cliffs
- Parris NA (1984) *Instrumental Liquid Chromatography: A Practical Manual on High-Performance Liquid Chromatographic Methods*. 2nd edn. Elsevier, Amsterdam
- Perrin C, Walczak B, Massart DL (2001) The Use of Wavelets for Signal Denoising in Capillary Electrophoresis. *Anal Chem* 73:4903-4917
- Peters S, Vivó-Truyols G, Marriott PJ, Schoenmakers PJ (2007) Development of an algorithm for peak detection in comprehensive two-dimensional chromatography. *J Chromatogr A* 1156:14-24
- Petkovic-Duran K, Zhu Y, Chen C, Swallow A, Stewart R, Hoobin P, Leech P, Ovenden S Hand-Held Analyzer Based on Microchip Electrophoresis with Contactless Conductivity Detection for Measurement of Chemical Warfare Agent Degradation Products. In: Nicolau DV, Metcalfe G (eds) *Biomedical Applications of Micro- and Nanoengineering IV and Complex Systems*, 2008. *Proceedings of SPIE Vol. 7270* (SPIE, Bellingham, WA), p 72700Q
- Polesello S, Valsecchi SM (1999) Electrochemical detection in the capillary electrophoresis analysis of inorganic compounds. *J Chromatogr A* 834:103-116
- Poole CF (2003) *The Essence of Chromatography*. Elsevier, Amsterdam

- Poppe H (1999) System peaks and non-linearity in capillary electrophoresis and high-performance liquid chromatography. *J Chromatogr A* 831:105-121
- Quémener B, Bertrand D, Marty I, Causse M, Lahaye M (2007) Fast data preprocessing for chromatographic fingerprints of tomato cell wall polysaccharides using chemometric methods. *J Chromatogr A* 1141:41-49
- Rabiner LR, Sambur MR, Schmidt CE (1975) Applications of a nonlinear smoothing algorithm to speech processing. *IEEE Trans Acoust Speech Signal Process ASSP* 23:552-557
- Reichenbach SE, Carr PW, Stoll DR, Tao Q (2009) Smart Templates for peak pattern matching with comprehensive two-dimensional liquid chromatography. *J Chromatogr A* 1216:3458-3466
- Savitzky A, Golay MJE (1964) Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal Chem* 36:1627-1639
- Schmidt MN (2008) Single-channel source separation using non-negative matrix factorization (PhD thesis), Technical University of Denmark,
- Sellmeyer H, Poppe H (2002) Position and intensity of system (eigen) peaks in capillary zone electrophoresis. *J Chromatogr A* 960:175-185
- Sentellas S, Saurina J, Hernández-Cassou S, Galceran MT, Puignou L (2001) Resolution and quantification in poorly separated peaks from capillary zone electrophoresis using three-way data analysis methods. *Anal Chim Acta* 431:49-58
- Shadle SE, Allen DF, Guo H, Pogożelski WK, Bashkin JS, Tullius TD (1997) Quantitative analysis of electrophoresis data: novel curve fitting methodology and its application to the determination of a protein-DNA binding constant. *Nucleic Acids Res* 25:850-860
- Shao X, Cai W, Sun P, Zhang M, Zhao G (1997) Quantitative determination of the components in overlapping chromatographic peaks using wavelet transform. *Anal Chem* 69:1722-1725
- Shao X, Sun L (2001) An application of the continuous wavelet transform to resolution of multicomponent overlapping analytical signals. *Anal Lett* 34 (2):267-280
- Shao X, Wang G, Wang S, Su Q (2004) Extraction of mass spectra and chromatographic profiles from overlapping GC/MS signal with background. *Anal Chem* 76:5143-5148
- Smit HC, Walg HL (1975) Base-Line noise and detection limits in signal-integrating analytical methods. Application to Chromatography. *Chromatographia* 8:311-323
- Smith AD (ed) (2000) *Oxford Dictionary of Biochemistry and Molecular Biology*. Rev. ed. edn. Oxford University Press, Oxford
- Smith JO, III (2007) *Spectral Audio Signal Processing* (March 2007 Draft)
- Solis A, Rex M, Campiglia AD, Sojo P (2007) Accelerated multiple-pass moving average: A novel algorithm for baseline estimation in CE and its application to baseline correction on real-time bases. *Electrophoresis* 28:1181-1188
- Stein SE, Scott DR (1994) Optimization and testing of mass spectral library search algorithms for compound identification. *J Am Soc Mass Spectrum* 5:859-866
- Stewart R, Wee A, Grayden DB, Zhu Y Capillary Electrophoresis (CE) Peak Detection using a Wavelet Transform Technique. In: Nicolau DV, Metcalfe G (eds) *Biomedical Applications of Micro- and Nanoengineering IV and Complex Systems*, 2008. *Proceedings of SPIE Vol. 7270* (SPIE, Bellingham, WA), p 727012

- Struzik ZR (2000) Determining local singularity strengths and their spectra with the wavelet transform. *Fractals* 8:163-179
- Szymańska E, Markuszewski MJ, Capron X, van Nederkassel A-M, Heyden YV, Markuszewski M, Krajka K, Kaliszan R (2007) Increasing conclusiveness of metabonomic studies of cheminformatic preprocessing of capillary electrophoretic data on urinary nucleoside profiles. *J Pharm Biomed Anal* 43:413-420
- Tanyanyiwa J, Galliker B, Schwarz MA, Hauser PC (2002) Improved capacitively coupled conductivity detector for capillary electrophoresis. *Analyst* 127:214-218
- Tarantola A (2005) *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, Philadelphia
- Torrence C, Compo GP (1998) A Practical Guide to Wavelet Analysis. *Bull Am Meteorol Soc* 79:61-78
- Vaseghi SV (2008) *Advanced Digital Signal Processing and Noise Reduction*. 4th edn. John Wiley and Sons, Chichester
- Vera-Candioti L, Culzoni MJ, Olivieri AC, Goicoechea HC (2008) Chemometric resolution of fully overlapped CE peaks: Quantification of carbamazepine in human serum in the presence of several interferences. *Electrophoresis* 29:4527-4537
- Vivó-Truyols G, Schoenmakers PJ (2006) Automatic selection of optimal Savitzky-Golay smoothing. *Anal Chem* 78:4598-4608
- Vivó-Truyols G, Torres-Lapasió JR, van Nederkassel AM, Heyden YV, Massart DL (2005a) Automatic program for peak detection and deconvolution of multi-overlapped chromatographic signals: Part I: Peak detection. *J Chromatogr A* 1096:133-145
- Vivó-Truyols G, Torres-Lapasió JR, van Nederkassel AM, Heyden YV, Massart DL (2005b) Automatic program for peak detection and deconvolution of multi-overlapped chromatographic signals: Part II: Peak model and deconvolution algorithms. *J Chromatogr A* 1096:146-155
- Wang J (2005) Electrochemical detection for capillary electrophoresis microchips: A review. *Electroanalysis* 17:1133-1140
- Wang Y, Gao Q (2008) Spatially adaptive stationary wavelet thresholding for the denoising of DNA capillary electrophoresis signal. *J Anal Chem* 63:768-774
- Wang Y, Mo J, Chen X (2004) 2nd-order spline wavelet convolution method in resolving chemical overlapped peaks. *Sci China, Ser B Chem* 47:50-58
- Wee A, Grayden DB, Zhu Y, Petkovic-Duran K, Smith D (2008) A Continuous Wavelet Transform Algorithm for Peak Detection. *Electrophoresis* 29:4215-4225
- Xiao-Quan L, Xi-Wen W, Jin-Yuan M, Jing-Wan K, Jin-Zhang G (1999) Electroanalytical signal processing method based on B-spline wavelets analysis. *Analyst* 124:739-744
- Xu X, Kok WT, Poppe H (1997) Noise and baseline disturbances in indirect UV detection in capillary electrophoresis. *J Chromatogr A* 786:333-345
- Yang C, He Z, Yu W (2009) Comparison of public peak detection algorithms for {MALDI} mass spectrometry data analysis. *BMC Bioinf* 10:4
- Yasui Y, Pepe M, Thompson ML, Adam B-L, Wright GL, Jr., Qu Y, Potter JD, Winget M, Thornquist M, Feng Z (2003) A data-analytic strategy for protein biomarker discovery: profiling of high-dimensional proteomic data for cancer detection. *Biostatistics* 4:449-463
- Yu W, He Z, Liu J, Zhao H (2008) Improving mass spectrometry peak detection using multiple peak alignment results. *J Proteome Res* 7:123-129

- Zemann AJ, Schnell E, Volgger D, Bonn GK (1998) Contactless conductivity detection for capillary electrophoresis. *Anal Chem* 70:563-567
- Zhang F, Chen Y, Li H (2007) Application of Multivariate curve resolution-alternating least square methods on the resolution of overlapping CE peaks from different separation conditions. *Electrophoresis* 28:3674-3683
- Zhang F, Li H (2006) Resolution of overlapping capillary electrophoresis peaks by using chemometric analysis: improved quantification by using internal standard. *Chemom Intell Lab Syst* 82:184-192
- Zhang J, Zhen C (1997) Extracting evoked potentials with the singularity detection technique. *IEEE Eng Med Biol Mag* 16:155-161
- Zhang XQ, Zheng JB, Gao H (2000) Comparison of wavelet transform and Fourier self-deconvolution (FSD) and wavelet FSD for curve fitting. *Analyst* 125:915-919
- Zhang Y, Mo J, Xie T, Cai P, Zou X (2001) Application of spline wavelet self-convolution in processing capillary electrophoresis overlapped peaks with noise. *Anal Chim Acta* 437:151-156
- Zheng J, Zhang H, Gao H (2000) Wavelet-Fourier self-deconvolution. *Sci China, Ser B Chem* 43:1-9
- Zheng X-P, Mo J-Y, Cai P-X (1998) Simultaneous application of spline wavelet and Riemann-Liouville transform filtration in electroanalytical chemistry. *Anal Commun* 35:57-59