

Smoothing of X-ray diffraction data and $K\alpha_2$ elimination using penalized likelihood and the composite link model

Johan J. de Rooi,^{a,b*} Niek M. van der Pers,^b Ruud W. A. Hendrikx,^b Rob Delhez,^b Amarante J. Böttger^b and Paul H. C. Eilers^a

^aDepartment of Biostatistics, Erasmus Medical Centre, Rotterdam, The Netherlands, and ^bFaculty 3mE, Department of Materials Science and Engineering, Delft University of Technology, The Netherlands. Correspondence e-mail: j.derooi@erasmusmc.nl

X-ray diffraction scans consist of series of counts; these numbers obey Poisson distributions with varying expected values. These scans are often smoothed and the $K\alpha_2$ component is removed. This article proposes a framework in which both issues are treated. Penalized likelihood estimation is used to smooth the data. The penalty combines the Poisson log-likelihood and a measure for roughness based on ideas from generalized linear models. To remove the $K\alpha$ doublet the model is extended using the composite link model. As a result the data are decomposed into two smooth components: a $K\alpha_1$ and a $K\alpha_2$ part. For both smoothing and $K\alpha_2$ removal, the weight of the applied penalty is optimized automatically. The proposed methods are applied to experimental data and compared with the Savitzky–Golay algorithm for smoothing and the Rachinger method for $K\alpha_2$ stripping. The new method shows better results with less local distortion. Freely available software in MATLAB and R has been developed.

© 2014 International Union of Crystallography

1. Introduction

Smoothing of X-ray diffraction data and $K\alpha_2$ elimination are often applied before evaluation of the data in software packages as supplied with commercially available diffractometers. The scans consist of series of counts which appear to follow Poisson distributions. Smoothing, for instance using the method of Savitzky & Golay (1964), is used to reduce random noise, by combining the counts in adjacent channels to get better local estimates of the expected values. The resulting smooth curve is better suited for estimating, for example, the locations and heights of peaks than the observed data.

In addition to the random noise, analysis is hampered by the fact that the peaks show a doublet structure, caused by the $K\alpha_1$ and $K\alpha_2$ emission lines of the X-ray source. The $K\alpha_2$ reflection can be eliminated by utilizing information about the differences in wavelength, intensity and shape of the $K\alpha_1$ and $K\alpha_2$ lines. The $K\alpha_2$ component is often removed by applying the Rachinger (1948) correction.

In this article, we present a smoothing procedure using penalized likelihood estimation and extend it to include $K\alpha_2$ elimination. This results in a decomposition into two smooth components: a $K\alpha_1$ and a $K\alpha_2$ part. The amount of smoothness can be optimized automatically. Because the method is grounded in a well developed statistical theory, the generalized linear model (GLM) (Nelder & Wedderburn, 1972), standard errors can also be computed, if desired.

The next section briefly introduces the instrumental setup and software implementations. The theory is introduced in several steps. §3 starts with an introduction to smoothing,

using penalized least squares. Because we are dealing with counts, we switch to penalized likelihood smoothing, which utilizes the Poisson distribution. In §4 we estimate the $K\alpha_1$ and $K\alpha_2$ components. We use a P-spline basis (Eilers & Marx, 1996) in combination with the composite link model (Thompson & Baker, 1981), explained in §§4.1 and 4.2, respectively. Penalized estimation requires tuning of the penalty parameter; an automatic procedure is presented in §5. To illustrate the proposed methodology, a number of applications are presented. The article closes with a discussion of the main findings and possibilities for future research.

Although the proposed methodology is useful in a number of settings, we also note that for various programs to obtain a fit to the whole diffraction pattern (e.g. Rietveld analysis or MAUD; Lutterotti & Bortolotti, 2003) it is usually not appropriate to apply smoothing and/or $K\alpha_2$ elimination prior to data analysis.

2. XRD experiments

All measurements were carried out using a Bruker D8 Advance diffractometer in Bragg–Brentano geometry, equipped with a Vantec position-sensitive detector and a graphite monochromator in the diffracted beam. Data collection was carried out at room temperature using Co $K\alpha$ radiation ($\lambda_{\text{Co } K\alpha_1} = 0.179026$ nm). Diffractometer scans were made in θ – 2θ scanning mode with a fixed step size and a fixed counting time per step. The specimens were spun at 30 r min^{-1} .

Three specimens were used to test the smoothing and $K\alpha_2$ elimination:

(a) LaB_6 . For this specimen, the standard reference material for line position and line shape, SRM660a from NIST, has been used. The particle size of the LaB_6 powder was such that the effect of crystal statistics is virtually absent. The actual specimen consists of a thin layer of LaB_6 on an Si single-crystal substrate wafer with orientation (510). It was made by applying a sedimentation method (mass thickness $33\ \mu\text{m}$, diameter $24\ \text{mm}$). Scan parameters: step size $0.007^\circ\ 2\theta$, time per step $2\ \text{s}$. This specimen produces sharp isolated reflections on a low background and it was especially used to check the $K\alpha_2$ elimination procedure.

(b) FeO_x . This is a bulk specimen made from finely grained powder, pressed into a pellet with thickness $2\ \text{mm}$, diameter $31\ \text{mm}$. The powder consists mainly of haematite. Scan parameters: step size $0.041^\circ\ 2\theta$, time per step $1\ \text{s}$. This specimen produces reflections on a relatively high background and was especially used to check the residual background level after $K\alpha_2$ elimination.

(c) Ni-alumina. This is a thin layer of powder on an Si single-crystal substrate wafer with orientation (510). The powder contains boehmite, aluminum oxides and nickel oxides. Scan parameters: step size $0.035^\circ\ 2\theta$, time per step $4\ \text{s}$. This specimen produces a mixture of sharp and broad reflections.

Data evaluation was performed with the *DIFFRAC^{plus}* evaluation package program *EVA* (version 17.0.0.2; Bruker AXS Inc., Madison, Wisconsin, USA). The proposed smoothing algorithm and procedure for $K\alpha_2$ elimination are available in a MATLAB script (The Mathworks Inc., Natick, MA, USA) and the R programming language (R Development Core Team, 2011). In addition, functions to read and write .raw files from Bruker (version 3) are provided in the R language. R and MATLAB scripts are made available as supplementary material.¹

The penalized likelihood smoother is compared with the Savitzky–Golay filter as implemented in *EVA*. The latter is controlled by one parameter, the width of the sliding interval on which the filter is applied.

The procedure for $K\alpha_2$ stripping is compared with the algorithm available in the *EVA* software, which is an implementation of the Rachinger method. This method assumes that the $K\alpha_1$ and $K\alpha_2$ line profiles are identical in shape and that both profiles are tied by a fixed intensity ratio τ : $I_{K\alpha_2}/I_{K\alpha_1} = 0.5$.

3. Smoothing

Smoothing is a frequently applied tool and many different algorithms are available. Very popular in analytical chemistry is the Savitzky–Golay filter (Savitzky & Golay, 1964), which tunes smoothness by varying the order of the local polynomial and the number of basis functions. Other methods use wave-

lets or Fourier transforms, which are both reviewed by Artursson *et al.* (2000). Chen *et al.* (2005) used smooth principal components to exploit the shared variation in a set of different patterns. This can be valuable in some cases but also limits the applicability of the method, because often only a single scan is obtained. A more recent article (Hogg *et al.*, 2012) proposes a Bayesian method for de-noising signals. The Poissonian nature of the data is accounted for by the use of the Anscombe transform. Long computation times make this procedure less practical.

In contrast, we present an automatic, fast, likelihood-based method. The method is presented in a stepwise manner, alongside applications to experimental XRD patterns. For didactic reasons, we start with the case where the sum of squares of differences is used to measure the fit of a model to data (the ‘least-squares’ principle) and subsequently turn to the likelihood approach.

3.1. Penalized least-squares smoothing

We indicate the counts by y_i , for $i = 1 : m$, with m the number of data points of the X-ray pattern. We try to find a smooth series μ of the same length that does not deviate too much from \mathbf{y} . We have to balance two goals: (1) the fidelity of μ to the data \mathbf{y} , and (2) the roughness (the opposite of smoothness) of μ . Intuitively it is clear that the smoother we make μ , the larger will be the difference between \mathbf{y} and μ .

To measure the fidelity of μ to \mathbf{y} , we use the familiar sum of squares:

$$S = \sum_i^m (y_i - \mu_i)^2. \quad (1)$$

A convenient way to measure the roughness of μ is to take the sum of squared differences of neighbouring values:

$$R = \sum_{i=2}^m (\mu_i - \mu_{i-1})^2 = \sum_{i=2}^m (\Delta\mu_i)^2. \quad (2)$$

In the last term of (2), Δ is an operator, a shorthand notation for taking differences. Actually a better measure of roughness uses second-order differences, differences of differences:

$$\begin{aligned} R &= \sum_{i=3}^m [(\mu_i - \mu_{i-1}) - (\mu_{i-1} - \mu_{i-2})]^2 \\ &= \sum_{i=3}^m (\mu_i - 2\mu_{i-1} + \mu_{i-2})^2 = \sum_{i=3}^m (\Delta^2\mu_i)^2. \end{aligned} \quad (3)$$

The elegance of the Δ operator becomes clear this way: Δ^2 means that it is applied two times.

The balancing act between fidelity and roughness can now be written as an objective function that has to be minimized:

$$Q = \sum_i^m (y_i - \mu_i)^2 + \kappa \sum_{i=3}^m (\Delta^2\mu_i)^2. \quad (4)$$

This is known as the Whittaker (1923) smoother, which was more recently discussed by Eilers (2003). Because the second term works as a penalty on μ , driving it away from the minimal sum of squares solution (but in the desirable direction of more smoothness), this procedure is also called penalized least

¹ Supporting information for this article is available from the IUCr electronic archives (Reference: FS5067).

squares. The parameter κ tunes the balance. If κ is small the second term has little influence and $\boldsymbol{\mu}$ will be near to \mathbf{y} , but it will be relatively rough. On the other hand, when κ is large, $\boldsymbol{\mu}$ will be (too) smooth, sacrificing fidelity to \mathbf{y} .

The loss function Q can be rewritten using matrices and vectors as

$$Q = \|\mathbf{y} - \boldsymbol{\mu}\|^2 + \kappa \|D_d \boldsymbol{\mu}\|^2. \quad (5)$$

Here $\|\mathbf{x}\|^2 = \sum_i x_i^2$ is shorthand notation for the quadratic norm, the sum of squares of elements, of any arbitrary vector \mathbf{x} . D_d is the difference matrix: it computes the d th-order differences over $\boldsymbol{\mu}$. For a signal with length $m = 5$, D_2 is defined as

$$D_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \end{pmatrix}. \quad (6)$$

In the remainder of this article we drop the subscript d and only use second-order differences. The minimizing vector $\hat{\boldsymbol{\mu}}$ follows as the solution the linear system of equations

$$(I + P)\hat{\boldsymbol{\mu}} = \mathbf{y}, \quad (7)$$

with $P = \kappa D'D$ and I the identity matrix. This is a large system: m linear equations with m unknowns. Yet it can be solved extremely easily and quickly, using sparse matrices, in both MATLAB and R.

3.2. Penalized likelihood smoothing

Penalized least squares, as explained above, is very successful in various applications (Eilers, 2003), but it is not a good choice when we are dealing with counts. The least-squares procedure does not guarantee that $\boldsymbol{\mu}$ will be positive everywhere. Also we do not account for the fact that higher counts have a larger variance. In addition, this smoother rounds off sharp peaks too much (strong smoothing) or gives a wavy baseline (lighter smoothing). We can solve these problems by switching to penalized likelihood.

The X-ray photon counts in a diffraction pattern are assumed to follow a Poisson distribution, with an expected value that varies with the diffraction angle. The photon counting sets a limit to the precision that can be obtained in a given amount of time. According to the Poisson distribution, the variance of the counts is equal to the expected value, $\boldsymbol{\mu}$, so the standard deviation, σ , is equal to the square root of $\boldsymbol{\mu}$. The relative error is thus equal to $\boldsymbol{\mu}^{-1/2}$.

The Poisson distribution states that the probability of observing y_i , given the expected value μ_i , is

$$\Pr(y_i | \mu_i) = \mu_i^{y_i} \exp(-\mu_i) / y_i!. \quad (8)$$

The logarithm of this probability, as a function of μ_i , given y_i , is the log-likelihood

$$L(\mu_i | y_i) = y_i \log \mu_i - \mu_i - \log(y_i!). \quad (9)$$

The last term is a fixed number once y_i is given, so we will drop it from here. If the elements of \mathbf{y} , conditional on their expected values, $\boldsymbol{\mu}$, are independent, we can sum the log-likelihood over all observations and use it as a measure of fit, replacing the

sum of squares in (1). Because a better fit gives a larger log-likelihood, we flip its sign, so that minimization becomes the goal again. Also, for technical reasons, to get simpler equations later on, we multiply it by 2. We then have

$$G = 2 \sum_i^m (\mu_i - y_i \log \mu_i), \quad (10)$$

which is known as the deviance. In principle it is possible to combine equation (10) with the old measure of roughness, R , as in (4). Instead we introduce a transformation of $\boldsymbol{\mu}$,

$$\boldsymbol{\eta} = \log \boldsymbol{\mu}, \quad (11)$$

and insert that into the roughness expression. This step is inspired by the theory of generalized linear models (GLM) (Nelder & Wedderburn, 1972), where it is called the link function. It guarantees positive values of $\boldsymbol{\mu}$. The objective function to minimize now becomes

$$Q = 2 \sum_i^m [\exp(\eta_i) - y_i \eta_i] + \kappa \sum_{i=3}^m (\Delta^2 \eta_i)^2, \quad (12)$$

with the second part being the roughness penalty. Minimization of (12) leads to the system of equations

$$P\boldsymbol{\eta} = \mathbf{y} - \boldsymbol{\mu}. \quad (13)$$

Estimation is usually performed using iteratively reweighted least squares (IRWLS). This boils down to iteratively solving the linear system

$$(\tilde{W} + P)\tilde{\boldsymbol{\eta}} = \mathbf{y} - \tilde{\boldsymbol{\mu}} + \tilde{W}\tilde{\boldsymbol{\eta}}, \quad (14)$$

where a tilde as in $\tilde{\boldsymbol{\eta}}$ indicates the current approximation to the solution. $W = \text{diag}(\boldsymbol{\mu})$ is a diagonal matrix with weights, for Poisson data, defined as

$$w_{ii} = \frac{1}{\mu_i} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 = \mu_i. \quad (15)$$

Convenient starting values are obtained by taking $\tilde{\eta}_0 = \log(y_i + 1)$. Usually about five iterations are needed to obtain estimates for $\boldsymbol{\eta}$ that are near enough to a solution for practical purposes (four significant figures or better in the solver residuals). More details on the theoretical properties of the GLM and its estimation routines have been given by McCullagh & Nelder (2000) and Dobson & Barnett (2008).

3.3. Comparison: penalized likelihood smoothing versus the Savitzky–Golay smoother from EVA

Fig. 1 shows the results of smoothing the XRD pattern of a sample of Ni-alumina. The upper panel shows the complete scan, with the observed data (black) and the smoothed pattern (green) obtained using the penalized likelihood method (PL smoother). The optimal smoothing parameter according the Akaike information criterion (AIC) is 33932. The AIC is further explained in §5.

For a more detailed inspection of the Ni-alumina scan, we zoom in on two specific regions, which are presented in the lower two panels of Fig. 1. In both panels the penalized likelihood approach gives satisfactory results. The Savitzky–Golay filter from EVA (SG filter), which is tuned manually, seems to

be too flexible in the right panel. For the likelihood smoother we included 95% confidence bands, which are depicted as the grey areas surrounding the fit of the smoother.

Given the smooth fit $\hat{\mu}$, we can check whether the Poisson assumption is reasonable. The variance of y_i should be equal to μ_i , so the normalized (or Pearson) residuals $(y_i - \hat{\mu}_i)/\hat{\mu}_i^{1/2}$

should have a standard deviation close to 1. In the case of the penalized likelihood smoother, it is 0.92, close to the expected value, showing that the Poisson assumption holds well for the present data. Checking the residuals of the pattern derived using the SG filter gives a standard deviation of 0.80 and might indicate too little smoothing is applied.

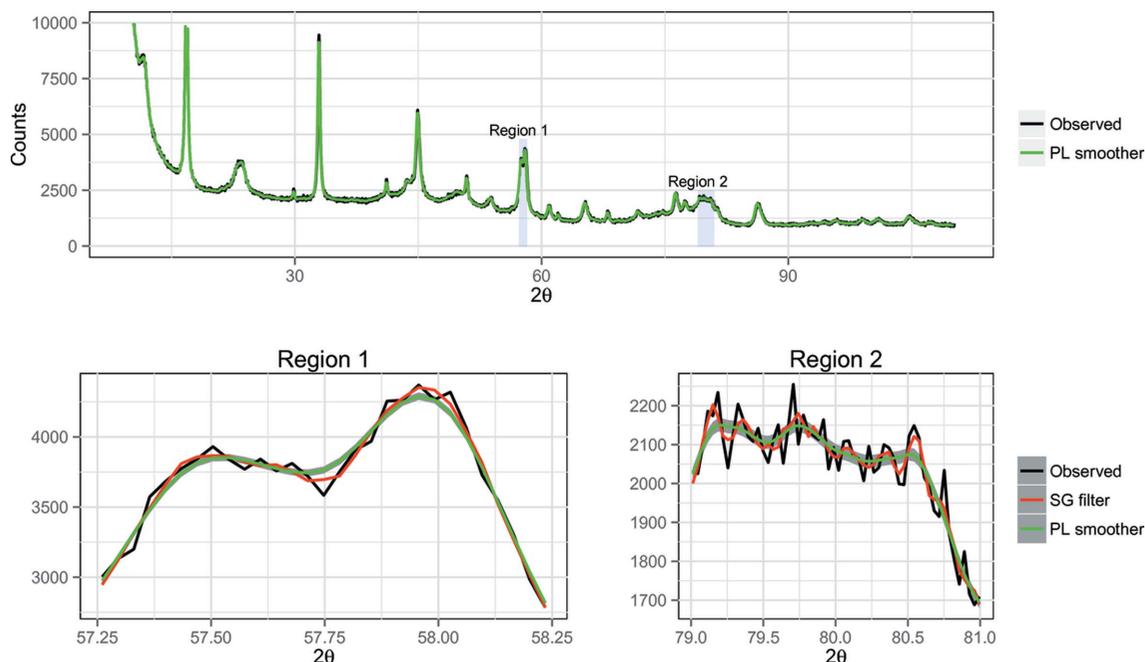


Figure 1

Results of smoothing the Ni-alumina scan. The upper panel shows the complete pattern, with the observed data (black) and the smoothed data (green) obtained using the penalized likelihood method. The two lower panels provide more detail on two regions of the scan. They include the penalized likelihood smoother with 95% confidence intervals (grey shaded regions) and the results of applying the Savitzky–Golay filter.

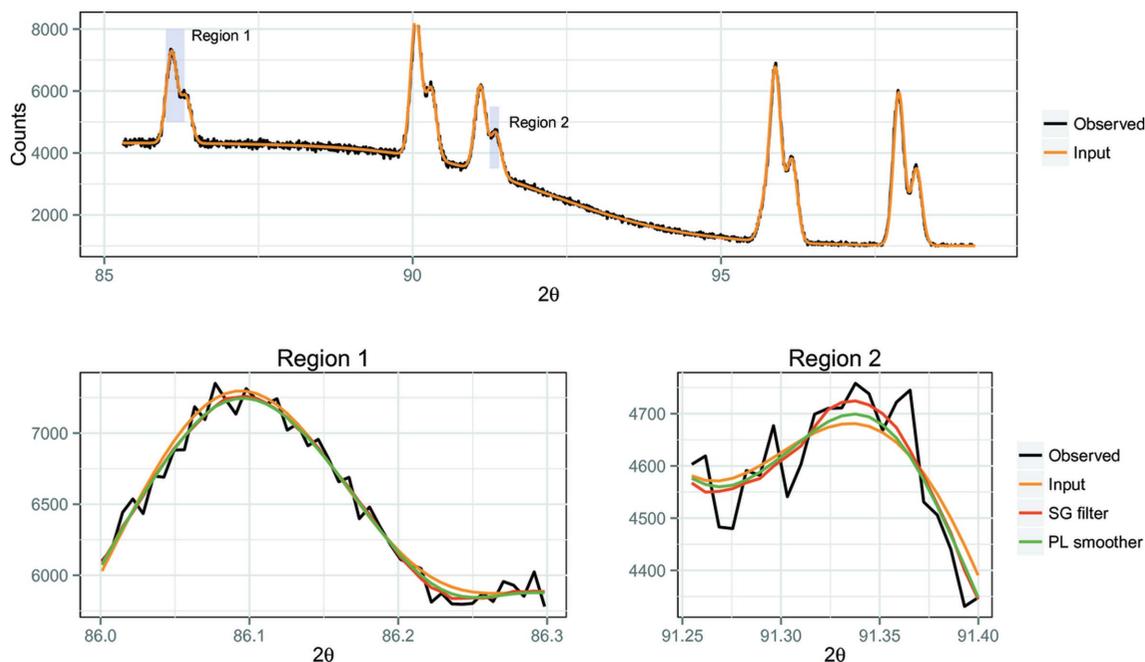


Figure 2

Comparison of the SG filter and the PL smoother using a simulated data set. The upper panel shows the complete scan and the observed data (black) and the true, noise-free, input data (orange). The two smaller panels show more details about two regions of the scan. In both cases the input, the output and the smooth curves of both procedures are provided.

For a more solid comparison of both smoothers, a simulation was conducted. The simulated XRD pattern is presented in the upper panel of Fig. 2. The noise-free (input) pattern is depicted in orange. The observed data, including Poisson noise, are shown in black. The pattern consists of a number of peaks on a drifting baseline.

Parts of the smoothed data are shown in the lower panels of Fig. 2. The smooth pattern from the penalized likelihood smoother (green) is shown together with the pattern coming from the SG filter using $\omega = 17$ (red), the true profile (orange) and the observed data (black). The optimal amount of smoothness when using the penalized likelihood smoother is obtained using the AIC.

The SG filter is generally tuned manually. Fig. 3 shows the performance of this filter for different values of the window width. The upper panel shows the distance between the simulated data without noise and the obtained pattern applying the SG filter. The distance is expressed as the root mean squared error (RMSE). The window width is given on the x axis. The red horizontal line is the optimal value obtained using the PL smoother and the AIC. The lower panel is similar to the upper one, the only difference being that it shows the distance towards the data with noise (the ‘observed data’). The performance of the optimal model of the SG filter is similar to that of the model obtained using the penalized likelihood smoother. However, in the absence of a proper criterion it is difficult to identify the optimal model. In contrast, the PL smoother is tuned automatically and in addition can be extended to $K\alpha_2$ elimination (see §4).

4. $K\alpha_2$ elimination

X-ray diffraction patterns usually contain several different wavelength components: $K\alpha_1$, $K\alpha_2$, $K\beta$ and bremsstrahlung. For analysis of X-ray diffraction patterns it is attractive to work with monochromatic patterns. Usually, physical filtering of the X-ray radiation is applied. Often a $K\beta$ filter or a monochromator is placed in the X-ray beam path in order to absorb the unwanted radiation. Also detectors with high energy resolution can be applied to distinguish between $K\beta$ and $K\alpha$, but at present these are not able to separate $K\alpha_1$ lines from $K\alpha_2$ lines.

In (our) daily practice $K\beta$ is removed by physical filtering. The classical procedure for $K\alpha_2$ elimination is the Rachinger (1948) correction. This method relies on three assumptions.

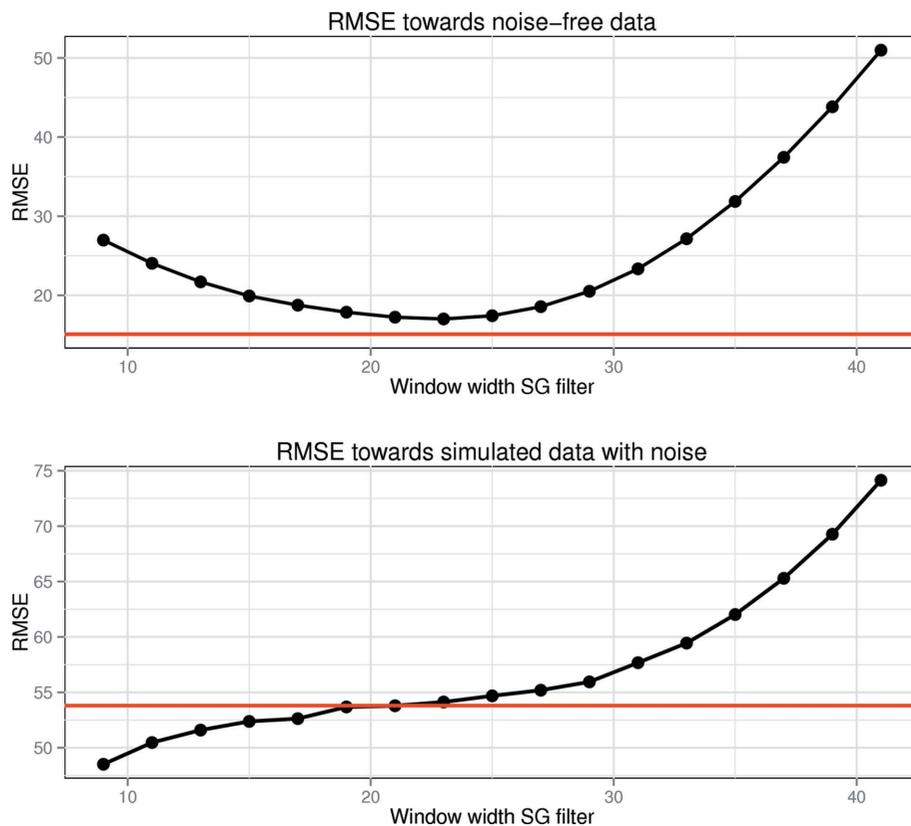


Figure 3 Simulation results, comparing the Savitzky–Golay filter with the penalized likelihood smoother. The upper panel shows the distance (expressed as the RMSE) between the noise-free simulated profile and the obtained pattern applying the SG filter. The window width of the SG filter is given on the x axis. The red horizontal line is the optimal value obtained using the PL smoother and the AIC. The lower panel is similar to the upper one but shows the differences towards the data with noise (the ‘observed data’).

First, the peak shape is assumed to be the same for the $K\alpha_1$ and $K\alpha_2$ components. Second, the α_2 line is expected to be half the intensity of the α_1 line. Third, the doublet distance is assumed to be a known function of the diffraction angle 2θ and the characteristic wavelengths of the anode material.

Here we take a similar approach: the X-ray diffraction pattern is considered to be the sum of two smooth latent components. Both components can be estimated using the penalized composite link model (Eilers, 2007). Smoothness is imposed using P-splines, which are explained below.

4.1. Penalized likelihood using P-splines

In the previous section, smoothness was obtained by using simple differences in the roughness penalty of the smoother. This is possible because the scans we use have data that are equally spaced on the diffraction angle (2θ) scale. Because the doublet distance δ does not coincide with the data points, estimating a smooth component for both $K\alpha_1$ and $K\alpha_2$ is impossible using this approach. Instead, a model is required that enables interpolation. This can be realized using a basis representation, with the additional advantage that the

dimensions of the estimating equations are reduced, which leads to faster calculations.

Here we use P-splines (Eilers & Marx, 1996), which are B-splines on a regular grid, combined with a penalty. The penalty is a natural extension of the difference penalty as discussed in the previous section, with the only difference that, instead of penalizing differences of adjacent data points, the penalty is imposed on adjacent spline coefficients. As advocated by Eilers & Marx (1996) we use a richly defined basis and smoothness is tuned with the penalty.

A B-spline of degree q is a smooth bell-shaped curve, composed of $q + 1$ polynomial segments each of degree q which join smoothly at the knots. The left panel of Fig. 4 shows the individual polynomial pieces, the resulting smooth spline function and the positions of the knots for a single B-spline. A simple example where a series of P-splines are fitted to the data is presented in the right panel of Fig. 4. Notice that it shows the general principle of P-splines using least squares and does not resemble the likelihood case, which involves a transformation of the data. We refer to Dierckx (1995) and de Boor (2001) for a thorough discussion of B-splines.

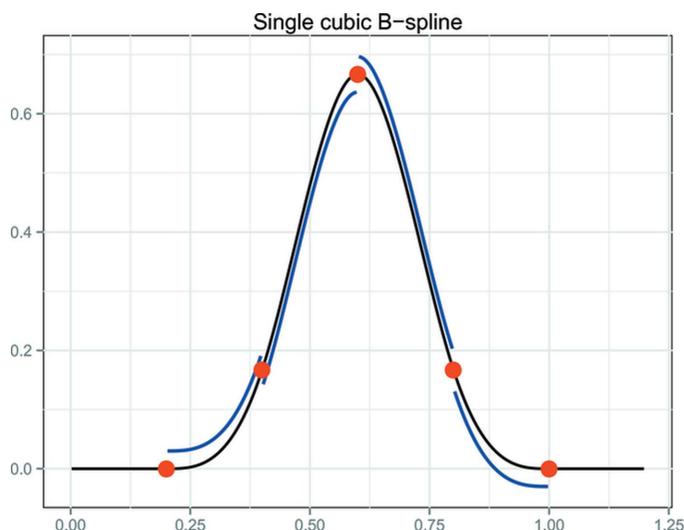
Using splines for smoothing the data (meaning no $K\alpha_2$ elimination) the model can be written as

$$\eta = B\beta, \quad (16)$$

with B the B-spline basis and β the vector of coefficients. Here we again use the transformation of μ : $\eta = \log \mu$. Parameters are estimated by using the penalized deviance; adding a penalty to equation (10) results in

$$G + \kappa \|D\beta\|^2 = G + \kappa \beta' P \beta. \quad (17)$$

Equation (17) can be solved using the iteratively reweighted least-squares algorithm (IRWLS) for estimating GLMs (Nelder & Wedderburn, 1972), with the penalty included:



$$(B'\tilde{W}B + P)\beta = B'\tilde{W}B\tilde{\beta} + B'(y - \tilde{\mu}), \quad (18)$$

where $\tilde{\beta}$ and $\tilde{\mu}$ denote current approximations to the solution.

4.2. The penalized composite link model

Here we are interested in estimating not a single smooth component but a separate smooth pattern for $K\alpha_1$ and $K\alpha_2$. To do this we combine the P-spline framework with the composite link model (CLM) (Thompson & Baker, 1981). The CLM is an extension of the GLM, modelling so-called 'indirect observations'. The combination of P-splines and the CLM is described by Eilers (2007), where it is termed the penalized composite link model (PCLM).

Using the generalized linear model the counts would be modelled as $\mathbf{E}(y) = \mu$. Assuming that the expected counts $\mathbf{E}(y)$ are the sum of two components, $K\alpha_1$ and $K\alpha_2$, we add a composition matrix C to the GLM. The elements of C describe how the latent expectations γ_1 and γ_2 , pertaining to the $K\alpha_1$ and $K\alpha_2$ components, respectively, are combined. Here we assume the same shape for both components, which leads to the composition matrix $C = [I \mid \tau I]$. Applying a PCLM, the expected counts are modelled as

$$\mathbf{E}(y) = \mu = C\gamma. \quad (19)$$

This results in $\mu = \gamma_1 + \gamma_2\tau$, where γ_1 pertains to the $K\alpha_1$ reflection and γ_2 to the $K\alpha_2$ reflection. Parameter τ defines the relative intensity of the two components and is set to 0.5 (Rachinger, 1948). The expectations are modelled using P-splines:

$$\gamma_1 = \exp(B_1\beta) = \exp\left[\sum_k \beta_k B_k(\rho_j)\right] \quad (20)$$

and

$$\gamma_2 = \exp(B_2\beta) = \exp\left[\sum_k \beta_k B_k(\rho_j + \delta_j)\right], \quad (21)$$

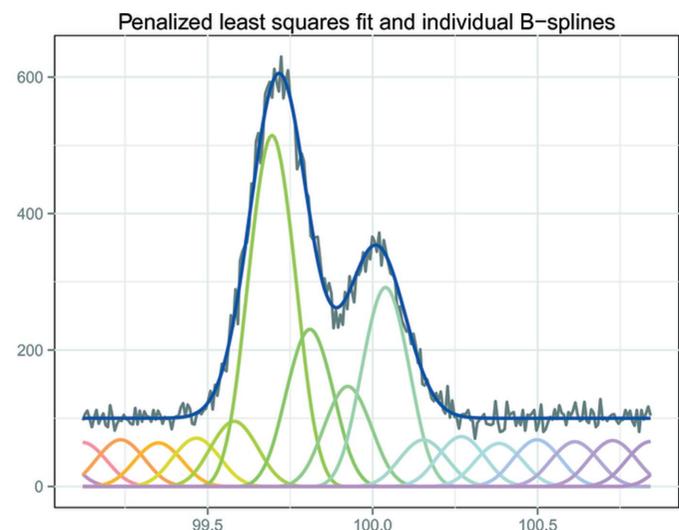


Figure 4

(Left) A single cubic B-spline function, including the knots and the (shifted) segments. (Right) A series of P-splines fitted to some data, using penalized least squares.

with $\rho = 2\theta$ and δ being the doublet distance. It is calculated as

$$\delta = 2 \arcsin\left(\frac{\lambda_2}{\lambda_1 \sin \theta}\right) - 2\theta, \quad (22)$$

with λ_2 and λ_1 the wavelength of the $K\alpha_2$ and $K\alpha_1$ emission, respectively.

Similarly to a GLM, the PCLM is estimated in an iterative manner solving

$$(\tilde{X}'\tilde{W}\tilde{X} + P)\tilde{\beta} = \tilde{X}'(\mathbf{y} - \tilde{\mu} + \tilde{W}\tilde{X}\tilde{\beta}), \quad (23)$$

with $\tilde{\beta}$ indicating the current approximation to the solution. The latent expectations γ are combined in $\Gamma = \text{diag}(\gamma)$ and the matrix X is calculated as

$$X = \tilde{W}^{-1}C\tilde{\Gamma}B_c, \quad (24)$$

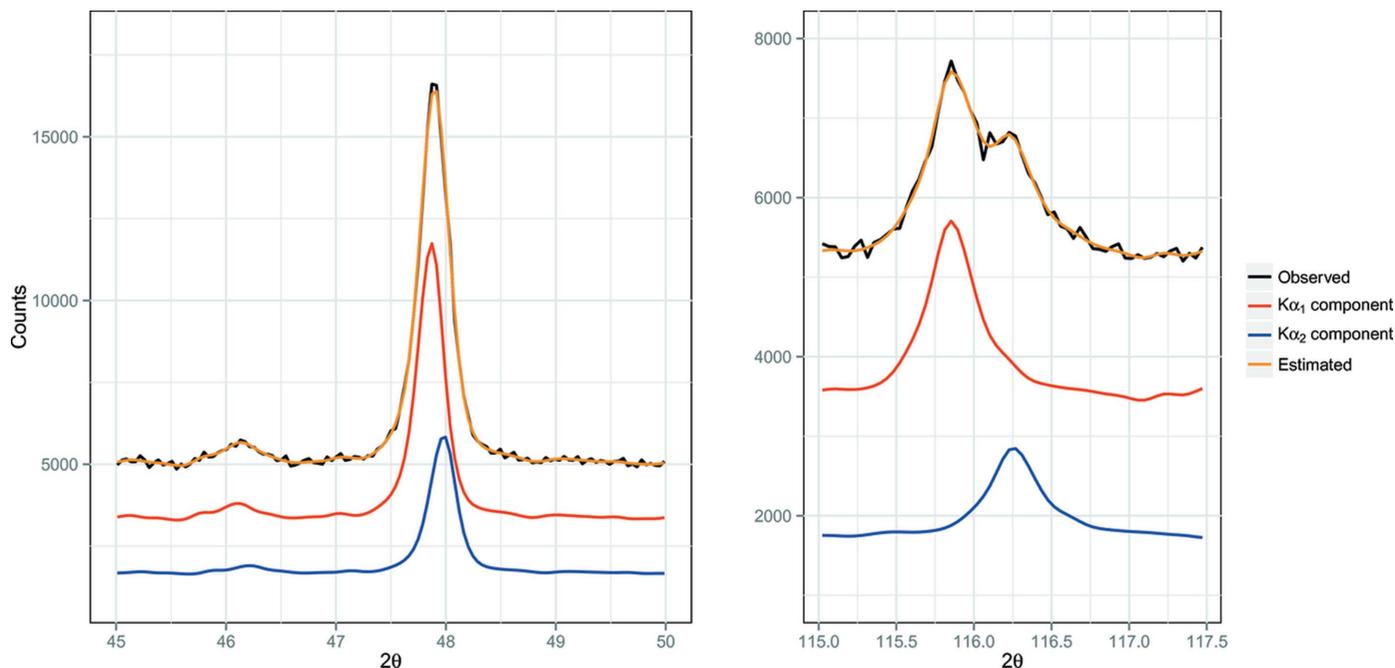


Figure 5 $K\alpha_2$ elimination using the PCLM, applied to the FeO_x data. The two panels show the estimated $K\alpha$ components and the reconstructed data at two different diffraction angles.

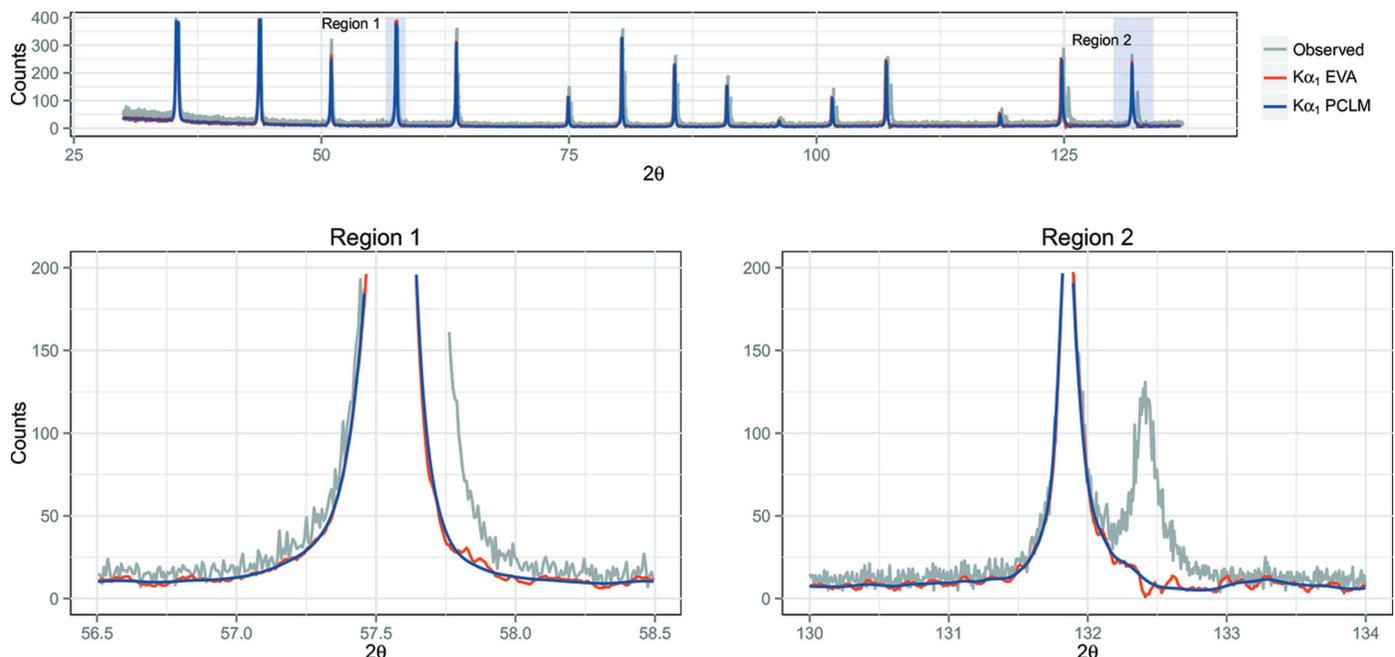


Figure 6 $K\alpha_2$ elimination applied to the LaB_6 data. All panels present the observed data (grey), the estimated $K\alpha_1$ part using EVA (red) and the same component estimated using the PCLM (blue). The lower-left panel shows a low-angle part of the data, while the right panel is taken at the high-angle side.

where $B'_c = [B'_1 | B'_2]$: the composition of the two B-spline bases.

The result is illustrated in Fig. 5, which shows two parts of an FeO_x diffraction pattern, a low-angle part and the high-angle part. In each panel, four curves are displayed: the observed data (black), the estimated smooth curve (orange), and the α_1 (red) and α_2 (blue) components. Irrespective of the angle, the peaks are well separated, while the resulting components and the fitted total curve are smooth.

4.2.1. Comparison: PCLM versus the EVA–Rachinger correction. The top panel in Fig. 6 shows the complete scan of the LaB₆ sample. In grey, the observed data are presented; the red line shows the estimated $K\alpha_1$ using the EVA software, while the blue line is the $K\alpha_1$ component obtained using the PCLM. At this scale it is difficult to observe differences between the two procedures. Zooming in on one of the peaks (as done in the lower panel of Fig. 6) does unveil some differences. Especially on the high-angle side, EVA generates unwanted artefacts.

5. Automated selection of the penalty parameter

In the previous sections we have presented a method for smoothing and subsequently we have extended the method to enable $K\alpha_2$ elimination. Both models can be tuned by visual inspection, yet it is attractive to have a method for automatic tuning of the amount of smoothing. We applied two fitting criteria to determine the optimal penalty parameter, the AIC proposed by Akaike (1974) and the L-curve, as introduced by Hansen (1992).

In the case of the AIC one searches for the value of κ that minimizes the criterion. This minimum is supposed to be the best trade-off between fidelity to the data and model complexity. The AIC is defined as

$$\text{AIC} = G + 2\text{ED}. \tag{25}$$

The first part of the equation is the deviance, which is described in equation (10). The ED is the effective model

dimension, a measure of the effective number of parameters in a model. The ED is computed as

$$\text{ED} = \text{tr}[(X\hat{W}X + P)^{-1}(X\hat{W}X)], \tag{26}$$

where $\hat{W} = \text{diag}(\hat{\mu})$ and tr denotes the trace of a matrix, which is the sum of the elements on its diagonal. A simple grid search can be used to find the value of κ that minimizes the criterion.

The left panel of Fig. 7 shows, for $K\alpha_2$ removal from the LaB₆ data, how the AIC changes with κ . Note that a linear grid has been used for the logarithm of κ , because a range of several orders of magnitude has to be investigated.

Although the AIC performs satisfactorily, the computation of the effective dimension is costly, because a complete matrix of m rows by m columns is to be computed. For large m this takes much memory and computation time. A simple solution is to use only a segment of a scan and use that to search for the optimal value for κ . Once κ has been determined in this way, it can be used for processing the whole scan.

A more elegant solution is to use a fast alternative fitting criterion: the L-curve, presented for ridge regression by Hansen (1992) and recently used in the setting of P-splines by Frasso & Eilers (2012). Here the algorithm is successfully applied to Poisson data. The basic idea is simple: one plots the logarithm of the size of the penalty term, φ , against the logarithm of the residuals, ψ :

$$\varphi = \log(\|D\beta\|^2), \quad \psi = \log(G). \tag{27}$$

Notice that both quantities are a function of κ . The figure will show a more or less L-shaped curve. The optimal penalty parameter is obtained in the corner of the L shape, where the curvature is largest. A simple alternative, presented by Frasso & Eilers (2012), is to search for the minimum of the squared step size $(\Delta\psi)^2 + (\Delta\varphi)^2$ and take the average of the two κ values related to that step. The right panel of Fig. 7 shows these steps as a function of the penalty parameter, in the case of the LaB₆ data.

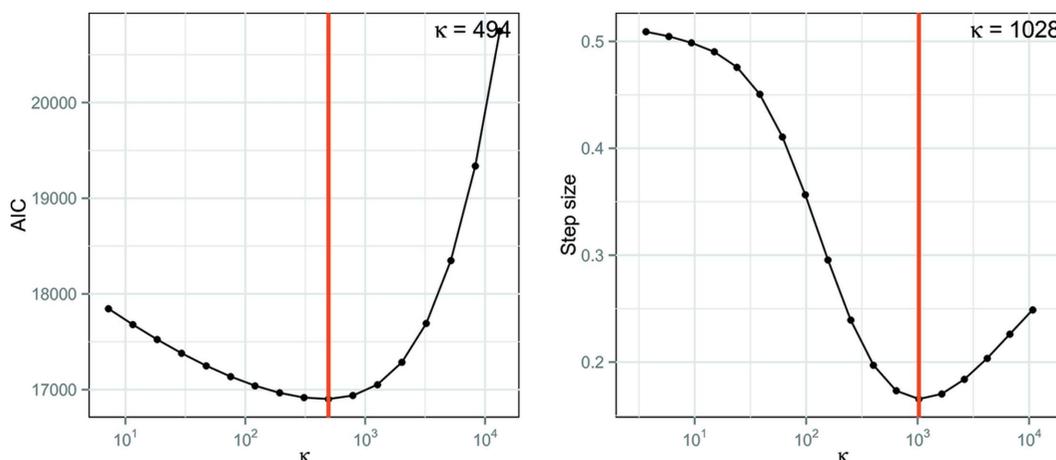


Figure 7 The AIC curve (left) and the sequence of step sizes connected to the L-curve (right), used to obtain the optimal penalty parameter for estimating the $K\alpha_1$ component of the LaB₆ data.

6. Discussion

Although our results are gratifying, we see areas that may need improvement. We assumed that the $K\alpha_1$ and $K\alpha_2$ components have the same shape. In the applications presented this assumption appears to hold well, but in the literature examples exist in which this is not the case. Currently we are investigating ways to incorporate these broadening effects into the PCLM.

The PCLM method is insensitive to the point spacing, meaning that one can always estimate the two components. However, when the observed data are very coarse, the results will not be useful, because too many details are lost. Consequently it is preferable to use a relatively small step size and at the same time keep the scanning times small as well. The increased noise level is compensated by the smoother.

Using the penalized likelihood procedure, smooth derivatives can be determined for peak finding purposes. In the case of very coarse data the P-splines can be used to obtain smoother results.

In the current setup, baseline removal is not required. The present baseline component is distributed over both parts. This is theoretically incorrect but does not cause any practical problems. After $K\alpha_2$ removal, baseline estimation can still be performed on the individual components for further analysis. One of the aims for future investigations is to estimate the baseline as part of the model.

7. Conclusion

We have presented a new method for smoothing and for smoothing combined with $K\alpha_2$ elimination. The general procedure is grounded in the generalized linear model, which enables Poisson statistics to be used for modelling the counts. The optimal smoothing parameter is determined using the Akaike information criterion or the L-curve. If desired, manual tuning of the model is possible as well.

An attractive property of the procedure for $K\alpha_2$ removal is that the pattern is modelled non-parametrically. No functional

form for the peaks has to be assumed. In essence only the characteristic wavelengths of the anode material are required.

Results obtained for XRD scans from our laboratory show satisfactory results, both for smoothing and for smoothing combined with $K\alpha$ doublet removal. This is confirmed by inspection of the estimated components.

References

- Akaike, H. (1974). *IEEE Trans. Autom. Control*, **19**, 716–723.
- Artursson, T., Hagman, A., Björk, S., Trygg, J., Wold, S. & Jacobsson, S. (2000). *Appl. Spectrosc.* **54**, 1222–1230.
- Boor, C. de (2001). *A Practical Guide to Splines*. Applied Mathematical Sciences, Vol. 21. New York: Springer.
- Chen, Z. P., Morris, J., Martin, E., Hammond, R. B., Lai, X., Ma, C., Purba, E., Roberts, K. J. & Bytheway, R. (2005). *Anal. Chem.* **77**, 6563–6570.
- Dierckx, P. (1995). *Curve and Surface Fitting with Splines*. Monographs on Numerical Analysis. Oxford University Press.
- Dobson, A. J. & Barnett, A. (2008). *An Introduction to Generalized Linear Models*. Boca Raton: CRC Press.
- Eilers, P. H. (2003). *Anal. Chem.* **75**, 3631–3636.
- Eilers, P. H. C. (2007). *Stat. Modell.* **7**, 239–254.
- Eilers, P. H. C. & Marx, B. D. (1996). *Stat. Sci.* **11**, 89–102.
- Frasso, G. & Eilers, P. H. C. (2012). *Proceedings of the 27th International Workshop on Statistical Modelling*, pp. 402–405. Utrecht University.
- Hansen, P. C. (1992). *SIAM Rev.* **34**, 561–580.
- Hogg, C. II, Mullen, K. & Levin, I. (2012). *J. Appl. Cryst.* **45**, 471–481.
- Lutterotti, L. & Bortolotti, M. (2003). *IUCr Commission on Crystallographic Computing Newsletter*, No. 1, pp. 43–50.
- McCullagh, P. & Nelder, J. (2000). *Generalized Linear Models*. London: Chapman and Hall/CRC.
- Nelder, J. A. & Wedderburn, R. W. M. (1972). *J. R. Stat. Soc. Ser. A*, **135**, 370–384.
- Rachinger, W. A. (1948). *J. Sci. Instrum.* **25**, 254–255.
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Savitzky, A. & Golay, M. J. E. (1964). *Anal. Chem.* **36**, 1627–1639.
- Thompson, R. & Baker, R. J. (1981). *Appl. Stat.* **30**, 125–131.
- Whittaker, E. T. (1923). *Proc. Edinb. Math. Soc.* **41**, 63–75.