E11-2002-293

A. G. Soloviev*, E. I. Litvinenko, G. A. Ososkov,
A. K. Islamov, A. I. Kuklin

COMPARATIVE STUDY OF SMOOTHING TECHNIQUES WITH REFERENCE TO DATA TREATMENT FOR SMALL-ANGLE NEUTRON SCATTERING

*E-mail: solovjev@lynx.jinr.ru

Contents

1 Introduction

Small-angle neutron scattering (SANS) is a very popular method used by physicists, material scientists, chemists, and biologists. Determination of invariants for small-angle scattering curves allows one to analyze investigated object parameters. Raw data treatment includes preprocessing, i.e., calibration, normalization and conversion from the time-of-flight scale into the momentum transfer scale Q. Since data are registered by two different detectors, they must be combined and merged accounting for detector resolution. This merging is also the important part of preprocessing. The main problems of preprocessing arise due to a high data dispersion caused by relatively low statistics at high neutron wavelength, the presence of noise measurements, and due to the resolution variability of different detectors and even parts of them. This work is aimed to improve the quality of merged data and thereby to increase the range of used data, reduce the theoretical model uncertainty and to reduce measurement time consumption.

1.1 YuMO spectrometer

We analyzed the data measured on the YuMO spectrometer (Fig.1.1) operated on the 4-th channel of the IBR-2 fast pulsed reactor [1, 2]. The measurements were carried out in two-detector spectrometer configuration [3], which provides a wide range of momentum transfer Q, using the time-of-flight technique.

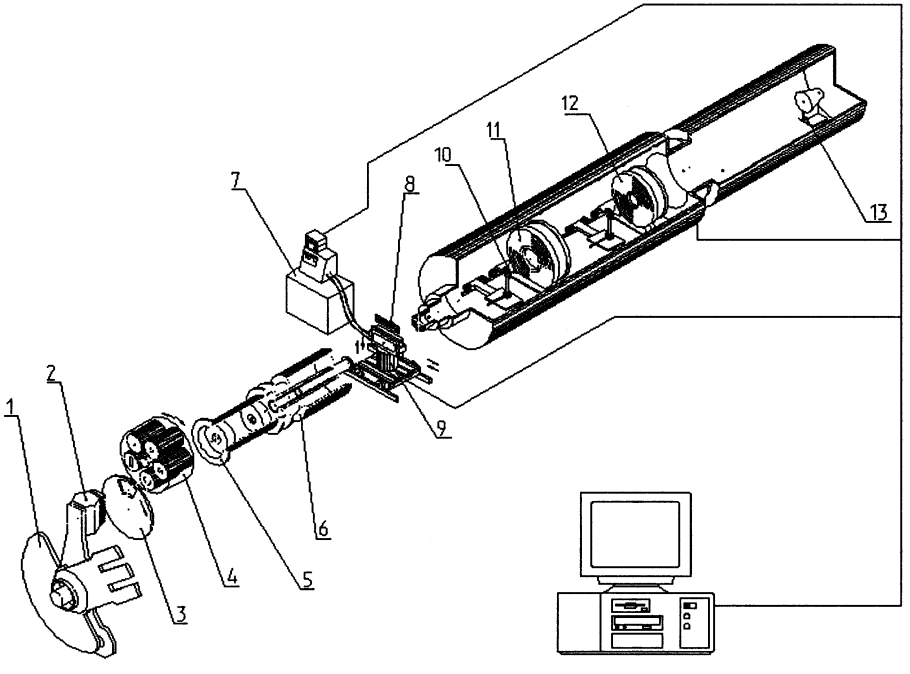

Figure 1.1: SANS YuMO spectrometer. 1 – reflectors, 2 – active zone with moderator, 3 – breaker, 4 – changeable collimator with different beam-holes, 5 – vacuum tube, 6 – adjustable collimator determining the size and position of the direct beam, 7 – thermostats, 8 – sample container, 9 – sample table, 10 – standard vanadium scatterer, 11, 12 – "OLD" and "NEW" detectors, 13 – direct beam detector

Data are registered by two ring-shaped detectors (11 and 12 in Fig.1.1) depending on the wavelength under several scattering angles. One of the detectors ("OLD") is positioned close to and the other one ("NEW") far from the sample.[1] Both detectors are used simultaneously during measurements. The software developed long ago to treat YuMO spectra cannot meet all the requirements, especially for its new set-up, cardinally upgraded recently [3].

[1] "NEW" and "OLD" signatures are just names of detectors.

1.2 Spectra samples

In this work, we have used the results of neutron measurements, obtained on the apoferritin protein sample obtained from Aldrich corporation. This sample can be used as a standard for testing the SANS spectrometer, because the apoferritin solvent is mono-dispersive and its spectra are very distinctive.
The existing data treatment programs SAS [4] and OpenG2 [5] both provide the possibility to merge spectra, taking into account statistical errors. The resulting spectra with a decimated Q-grid still remain noisy (SAS results are shown in Fig. 1.1). As one can see, both spectra are deeply overlapped, but have rather different variability characters.

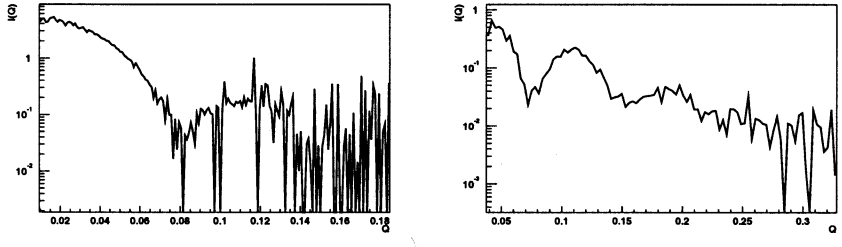

Figure 1.1: Apoferritin spectra samples from the "NEW" (left) and the "OLD" (right) detectors

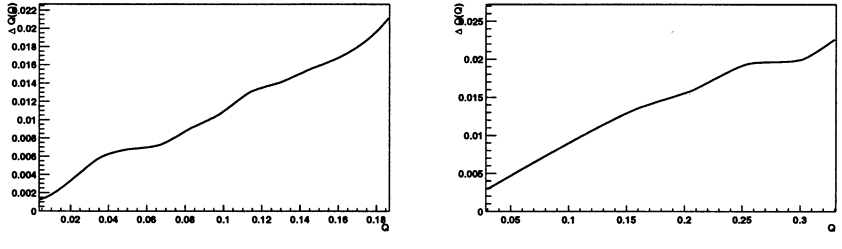

Figure 1.2: Q-resolution dependencies for the "NEW" (left) and the "OLD" (right) detectors

There is a lot of conventional software aimed to perform smoothing. But direct application of any of them to SANS spectra processing does not give a good result. It is caused by the experimental spectra character: their smoothness is substantially varied from the beginning of a spectrum to its end. Therefore, a new smoothing

approach is required. Our goal is to improve the quality of spectra produced by the SAS program by applying new non-traditional methods.
The main idea is to take into account the spectrometer resolution, ΔQ, while smoothing. The spectrometer resolution dependencies given by SAS program for the "NEW" and the "OLD" detectors at the first step of data treatment are shown in Fig.1.2.

2 Traditional smoothing techniques

One usually distinguishes white noise (having similar amplitudes over a wide frequency spectrum) and impulse noise which is a momentary perturbance, limited in the frequency band, and also limited at saturation (i.e., the maximum signal height permitted). Very often the spectrum of a noise contains higher frequencies than the spectrum of an original signal. So, in many cases a simple low-frequency filtering can be an effective technique for noise reduction. In principle, any filter with non-negative coefficients can be used for smoothing. In our work, we considered the most popular traditional filters: smoothing window and median ones [6].

2.1 Smoothing window filter

The smoothing window filter is an example of a linear filter. Linear filters are amenable to analysis in the Fourier domain. We used a simple uniform filter - the output signal (R) is based on a local averaging of the input values (A) having the same weight

$$R_j = (1/w) \sum_{j=0}^{w-1} A_{i+j-w/2} \ , \quad w/2 \leq i \leq n-1-w/2 \ ,$$

where n is the number of elements in A, and w is the width of the sliding window. The smoothing window filter provides an effective way of noise suppression. The known disadvantage of this filter is a blurred resulting signal due to suppression of high frequencies in both noise and original signal spectra.

2.2 Median filter

Median filter is a nonlinear filter applicable when the data distribution is not normal or unknown at all. The median filter is based also on the sliding window approach, but the data inside the window with an odd number of elements must be sorted, first, ascendingly in order to find their median, i.e. the middle element of the sorted set. Then this median replaces the central element of the window that slides to the next position, etc. In contrast to smoothing window filters, the median filter keeps better the contour of an image. The median filters are usually very effective for signals with sharp rejections or with step appearances that are non-normally distributed.

2.3 Modification of traditional smoothings for SANS spectra processing

We replace the constant width w for the smoothing window and median filters with the variable width according to the spectrometer resolution. According to Q-resolution dependencies for the "NEW" and the "OLD" detectors (Fig.1.2), w varied from 3 to 21 for the first one and from 3 to 35 — for the second one.

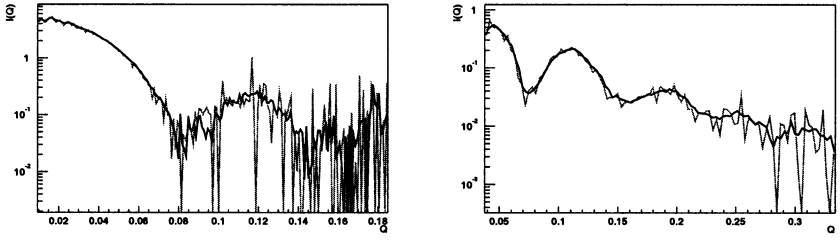

Figure 2.1: Spectra processing using an adapted smoothing window filter for the "NEW" (left) and the "OLD" (right) detectors

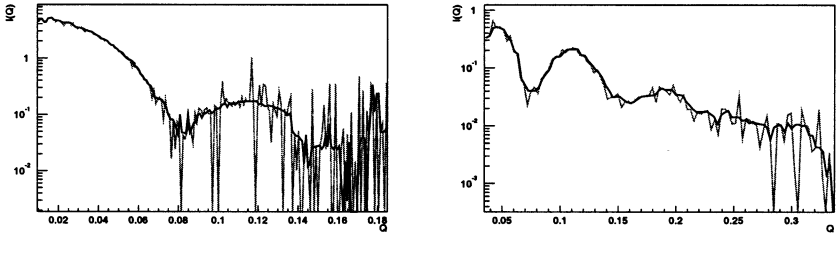

Figure 2.2: Spectra processing using an adapted median filter for the "NEW" (left) and the "OLD" (right) detectors

Such an adaptive procedure gives more suitable results, shown in Figs. 2.1-2.2.

3 Wavelet approach

Wavelets are versatile tools of harmonic analysis. Scientists always want more appropriate functions than sines and cosines which comprise the basis of Fourier analysis, to approximate signals with discontinuities. Sines and cosines are non-local and stretch out to infinity. Therefore, they do a very poor job in approximating signals

with sharp spikes. But with a wavelet analysis, we can use approximating functions that are contained in finite domains. Thus, the wavelet analysis gives a simultaneous knowledge of various signal frequencies and the time location of these frequencies as well.
Accomplishing wavelet analysis usually means the following:

- to make a proper choice of a wavelet type;
- to fulfill *a wavelet filtering* for denoising, removing pedestals, and extracting some features of analyzed data. It is carried out by

    1. transforming data to a wavelet domain;
    2. applying desirable cuts on wavelet 2D-spectrum;
    3. making an inverse transform.

3.1 Continuous wavelet transform

The *wavelet transform* of function $f(x) \in L^2(\mathbb{R})$ is its projection to the *basic wavelet* dilated by factor a and shifted by b

$$W_\psi(a,b)[f] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{|a|}} \psi\left(\frac{x-b}{a}\right) f(x)dx. \tag{1}$$

The reconstruction, or *inverse wavelet transform*, is

$$f(x) = \frac{1}{C_\psi} \int\int \psi\left(\frac{x-b}{a}\right) W_\psi(a,b) \frac{da\,db}{a^2}, \tag{2}$$

where C_ψ is the normalization constant for a given basic wavelet ψ.
As a first example of how the wavelets work, let us take a harmonic signal constructed by superposing the low-frequency one with a small fraction of the high-frequency one and then contaminating it by uniformly distributed random noise, see Fig.3.1 (left).

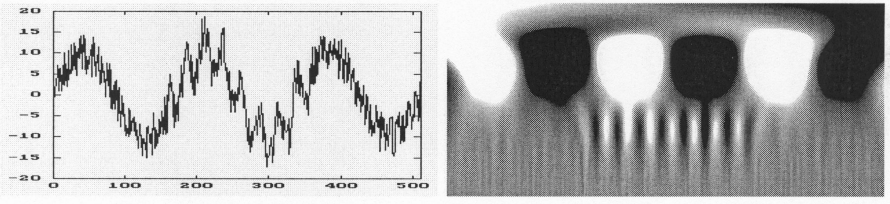

Figure 3.1: A sample of a signal comprising two harmonic components contaminated by noise (left) and its wavelet spectrum (right, horizontal axis corresponds to the time parameter b, vertical one — to the scale parameter a)

Fig.3.1 (right) presents the wavelet spectrum. The shade-plot provides a powerful tool, which helps to display the structure of the signal. The set of wavelet coefficients can be presented as a projection of the 3-dimensional surface $W = W(a, b)$ onto the a-b plane. Coefficients with higher values are shown in light colors, and the lower ones in dark colors.
The next figure shows the results of the final filtering of the same signal.

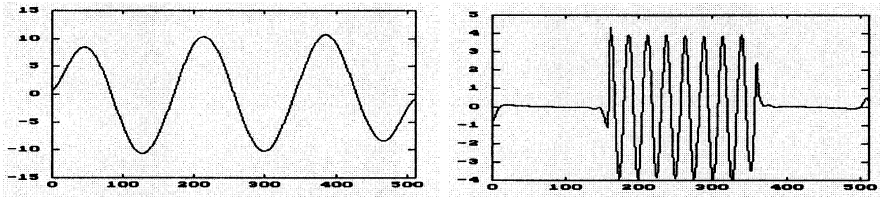

Figure 3.2: Extracting data from the signal of Fig.3.1: low frequency component (left) and high frequency component (right)

It can be seen that the filter allows one to extract both components of the original signal. Thus, selecting properly the scales of the wavelet transformation, it is possible to highlight the components of the desired scales.

3.2 Adoptive modification of CWT for SANS spectra processing

Any SANS spectrum is just a histogram

$$f(x) = \frac{1}{N_{bins}} \sum_{k=1}^{N_{bins}} h_k \delta(x - x_k) \ . \tag{3}$$

For a discretized signal like (3) CWT (1) is written as

$$W_\psi(a, b)[f] = \frac{1}{N_{bins}} \sum_{k=1}^{N_{bins}} \frac{1}{\sqrt{|a|}} h_k \psi \left(\frac{x_k - b}{a} \right) \ , \tag{4}$$

and it looks like the Parsen–Rozenblatt estimates of the unknown probability density over a sample (see [7]). Equation (4) is nothing but the "averaged sum" of the wavelets $\psi\left[(x_k - b)/a\right]$ compressed to the size a and "placed" at points x_k.
Let us take a Gaussian

$$g(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

instead of a true basic wavelet ψ in (4). Then we have

$$W_g(a,b)[f] = \frac{1}{N_{bins}} \sum_{k=1}^{N_{bins}} \frac{h_k}{\sqrt{2\pi}a} \exp\left(-\frac{(x_k-b)^2}{2a^2}\right)$$

and no inverse transform is needed.
The direct application of the last formula to SANS spectra processing does not give a good result yet, because their smoothness is substantially varied from the beginning of the spectrum to its end. We choose a dilation factor (RMS of Gaussian) depending on a point, according to a given Q-resolution of YuMO spectrometer at this point (see Fig.1.2). Thus, we have constructed a *new* kernel transform with the adaptable gaussian halfwidth

$$\tilde{f}(b) = \frac{1}{N_{bins}} \sum_{k=1}^{N_{bins}} \frac{h_k}{\sqrt{2\pi}a(x_k)} \exp\left(-\frac{(x_k-b)^2}{2a^2(x_k)}\right) . \tag{5}$$

Finally, let us write (5) in a notation suitable for SANS:

$$\tilde{I}_n = \frac{1}{N_{bins}} \sum_{k=1}^{N_{bins}} \frac{I_k}{\sqrt{2\pi}\Delta Q_k} \exp\left(-\frac{(Q_k-Q_n)^2}{2(\Delta Q_k)^2}\right) .$$

This transform appears to be good for the SANS spectra smoothing. Results of it application are showh in Fig.3.3.

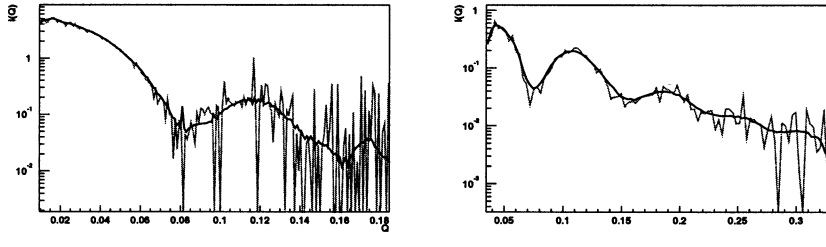

Figure 3.3: Spectra processing for the "NEW" (left) and the "OLD" (right) detectors

3.3 Discrete wavelet transform: lifting scheme

The idea of a discrete wavelet transform is to represent given data as a decomposition using basis functions $\psi_{j,k} = 2^{-j/2}\psi(2^{-j}t - k)$ and $\phi_{j,k} = 2^{-j/2}\phi(2^{-j}t - k)$. They are constructed as scaled and shifted two basic functions:

- a scaling function ϕ;

- a wavelet function (or "mother wavelet") ψ.

Both of these functions should have a locality in time/space and frequency domains. The purpose is to get the following decomposition of a source data

$$f(x_i) = \sum_{k=1}^{N} s_{Lk}\phi_{Lk}(x_i) + \sum_{j=1}^{L}\sum_{k=1}^{N} d_{jk}\psi_{jk}(x_i) \ ,$$

where N is a number of samples in the data sequence, and L is a decomposition level. As the result, we can describe our signal in terms of coefficients s_{Lk} and $d_{j,k}$, where s_k are called approximation coefficients and $d_{j,k}$ are details at the j-th level of the decomposition. The way how to get this representation of the signal is shown in Fig.3.4.

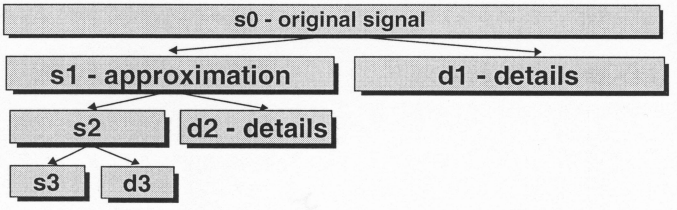

Figure 3.4: Wavelet decomposition scheme

Lifting scheme is a fundamentally other approach to building wavelet decompositions, proposed by W.Sweldens [8, 9]. Constructing wavelets using lifting consists of three phases.

1. Lazy wavelet *splits* the data $\{\lambda_{0,k}\}$ into two subsets, even and odd

$$\lambda_{-1,k} = \lambda_{0,2k} \ , \quad k \in \mathbf{Z} \ , \tag{6}$$

and *wavelet coefficients*

$$\gamma_{-1,k} = \lambda_{0,2k+1} \ , \quad k \in \mathbf{Z} \ .$$

2. The second step calculates the wavelet coefficients (high pass) as the failure to *predict* the odd set based on the even

$$\gamma_{-1,k} = \gamma_{-1,k} - P(\lambda_{-1,k}) \ . \tag{7}$$

A simple example of (7) is

$$\gamma_{-1,k} = \gamma_{-1,k} - (\lambda_{-1,k} + \lambda_{-1,k-1})/2 \ ,$$

i.e., P is a function piecewise linear over intervals of length 2, and wavelet coefficients measure to which extent the original signal *fails to be linear*. In terms of the frequency content, the wavelet coefficients capture high frequencies in the original signals.

3. The third step *updates* the even set using the wavelet coefficients (low pass). The idea is to find a better $\lambda_{-1,k}$ so that a certain scalar quantity $Q()$, i.e. mean, is preserved, or $Q(\lambda_{-1,k}) = Q(\lambda_{0,k})$. We construct operator U and update $\{\lambda_{-1,k}\}$ as

$$\lambda_{-1,k} = \lambda_{-1,k} + U(\gamma_{-1,k}) \ . \tag{8}$$

In a simple case, for a long signal, we can update the λ coefficients with the following equation:

$$\lambda_{-1,k} = \lambda_{-1,k} + (\gamma_{-1,k} + \gamma_{-1,k-1})/4 \ .$$

The three stages of lifting described by (6), (7), and (8) are combined and iterated to generate a fast lifted wavelet transform algorithm

$$\begin{cases} \{\lambda_{j,k}, \gamma_{j,k}\} & = & Split(\lambda_{j+1,k}) \ , \\ \gamma_{j,k} & -= & P(\lambda_{j,k}) \ , \\ \lambda_{j,k} & += & U(\gamma_{j,k}) \ . \end{cases}$$

Inverse transform

$$\begin{cases} \lambda_{j,k} & -= & U(\gamma_{j,k}) \ , \\ \gamma_{j,k} & += & P(\lambda_{j,k}) \ , \\ \lambda_{j+1,k} & = & Join(\lambda_{j,k}, \gamma_{j,k}) \ . \end{cases}$$

3.4 Application of DWT to SANS spectra processing

The filtering using lifting is carried out, as usual, in three steps:

1. calculate the set of wavelet coefficients;
2. cut off the spectrum (e.g., set to zero all coefficients which values are less then the prescribed threshold);
3. perform the inverse wavelet transform of the obtained spectrum.

The processing was produced by using a LIFTPACK [10] based program [11].
The absence of a structure in the removed component confirms the validity of this processing.

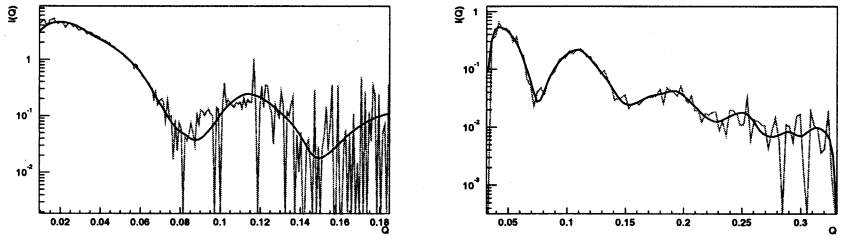

Figure 3.5: Spectra processing for the "NEW" (left) and the "OLD" (right) detectors

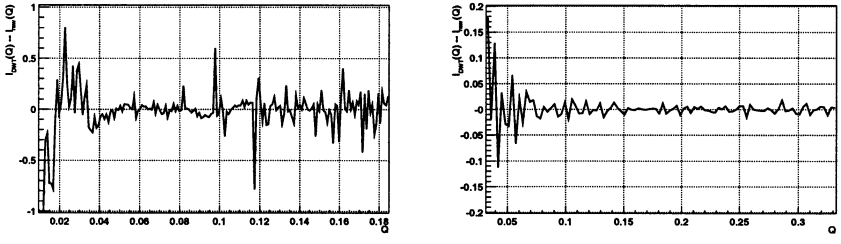

Figure 3.6: Components, removed from the spectra for the "NEW" (left) and the "OLD" (right) detectors

4 Influence of smootings on the next processing stage

Determination of invariants for small-angle scattering curves allows one to analyze the structure of an object under study. Upon the first step of this analysis the shape of the object (which is the apoferritin complex) is approximated by simple geometrical bodies — spheres, ellipsoids, cylinders or prisms [12].
After filtering the resulting curves were fitted by a spherical shell model, which is one of the most adequate for apoferritin

$$I(q) = A\left[\frac{4}{3}\pi R_1^3\Phi(qR_1) - \frac{4}{3}\pi R_2^3\Phi(qR_2)\right]^2 + B\ ,$$
$$\Phi(t) = 3\frac{\sin t - t\cos t}{t^3}\ ,$$

where R_1 and R_2 are inner and outer radiuses and A and B are the amplitude and background parameters. As a quality criterion to compare our results, the χ^2

criterion could be used

$$\chi^2 = \frac{1}{N_{bins} - N_{parms}} \sum_{k=1}^{N_{bins}} \left(\frac{I_{theory}(Q_k) - I_k}{\Delta I_k} \right)^2 ,$$

where $N_{parms} = 4$ is a number of model parameters.

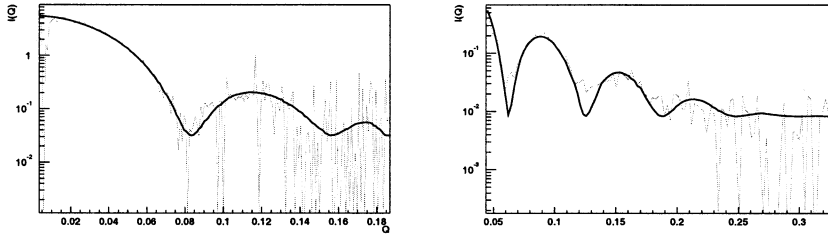

Figure 4.1: Theoretical spectra for the "NEW" (left) and the "OLD" (right) detectors

An example of the theoretical spectra is shown in Fig.4.1, the χ^2 values are shown in the following table.

Detector:	"NEW"	"OLD"
Raw spectra	**1.52**	**2.13**
Smoothing window	**0.76**	**1.36**
Median filter	**0.77**	**1.45**
CWT	**1.12**	**1.59**
DWT	**0.97**	**1.33**

The wavelet filtering looks preferable due to obtained χ^2 values. Moreover, the wavelet filtering gives smoother resulting curves, including their edges, with no loss of experimental information.

5 Conclusion

- A *new* kernel transform with an adaptable gaussian halfwidth is constructed. It is a sutable tool for SANS spectra smoothing. This transform is ready-to-use in SANS data processing:
  - it is applied automatically (no manulal work is needed);
  - it takes spectrometer resolution correctly into account;
  - it improves the resulting scattering spectra quality.

- The lifting scheme has a number of advantages in comparison with the classical discrete wavelets. This transform works for signals of an arbitrary size with correct treatment of the boundaries. In addition, all computations can be carried out in-place. Moreover, the lifting scheme makes them optimum, sometimes increasing the speed of calculations by a factor of two.

  However, to be applicable for SANS data processing, the lifting scheme filtering needs further developments, since it requires some manual preadjusment in its present implementation.

References

[1] *Yu.M.Ostanevich.* Time-of-flight small-angle scattering on pulsed neutron sources. Makromol. Chem., Macromol. Symp. **15**, 91-103 (1988)

[2] http://nfdfn.jinr.ru/ibr2.html

[3] *Kuklin A.I., Islamov A.Kh. and V.I. Gordeliy.* Two Detectors System for Small Angle Neutron Scattering Instrument. Submitted to J. Appl. Cryst.

[4] http://www.jinr.ru/~tsap/Koi/jinrlib/Xw012.htm

[5] http://nfdfn.jinr.ru/~litvin/openg2

[6] *Young I.T., Gerbrands J.J., Van Vliet L.J.* Fundamentals of Image Processing. Delft University of Technology. **ISBN 90-5691-01-7** (1998) 1–112

[7] *Soloviev A.G., Sosnin A.N., Zhidkov E.P.* Improving the accuracy of estimation of unknown random variable probability density over empirical data. Computer Physics Communications **137** (2001) 366–379;

[8] *W.Sweldens, P.Schroder.* "Building your own wavelets at home" Wavelets in Computer Graphics, pp. 15–87, ACM SIGGRAPH Course Notes, 1996.

[9] *W.Sweldens.* "The lifting scheme: A new philosophy in biorthogonal wavelet construction". Proc. SPIE, 1995.

[10] http://www.cs.sc.edu/~fernande/liftpack
http://www.cs.sc.edu/~fernande/liftpack/liftbibl.html#spie96
http://cm.bell-labs.com/who/wim/papers/papers.html#spie96

[11] http://www.jinr.ru/~tsap/Koi/jinrlib/xu008/Xu008e.htm
http://www.jinr.ru/~tsap/Koi/jinrlib/xu008/Xu008.htm

[12] *Feigin, L.A. & Svergun, D.I.* (1987) Structure analysis by small-angle X-ray and neutron scattering. New York: Plenum Press, 335 pp.

Received on December 27, 2002.

Соловьев А. Г. и др. E11-2002-293
Сравнительный анализ методов сглаживания применительно к обработке данных малоуглового рассеяния нейтронов

Представлен метод улучшения качества спектров рассеяния, которое достигается благодаря использованию разрешения спектрометра как при вейвлет-фильтрации, так и при традиционных методах сглаживания данных малоуглового рассеяния нейтронов. На следующем этапе обработки данных этот результат приводит к лучшим подгоночным кривым, задаваемым формфактором теоретической модели. Кроме того, вейвлет-анализ позволяет выделять и анализировать фоновую составляющую (шум) и вводить инструментальные поправки.

Работа выполнена в Лаборатории нейтронной физики им. И. М. Франка ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна, 2002

Soloviev A. G. et al. E11-2002-293
Comparative Study of Smoothing Techniques with Reference to Data Treatment for Small-Angle Neutron Scattering

This work presents an improvement of the resulting scattering spectra quality due to the use of the spectrometer resolution during both wavelet filtering and traditional smoothing of the small-angle neutron scattering data. This result leads to a better fitting of the form factor curves at the next step of data analysis. In addition, the wavelet analysis permits one to extract and analyze a background (noisy) component and to carry out instrumental hardware corrections.

The investigation has been performed at the Frank Laboratory of Neutron Physics, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna, 2002

Макет *Т. Е. Попеко*

Подписано в печать 24.01.2003.
Формат 60 × 90/16. Бумага офсетная. Печать офсетная.
Усл. печ. л. 1,06. Уч.-изд. л. 1,54. Тираж 310 экз. Заказ № 53724.

Издательский отдел Объединенного института ядерных исследований
141980, г. Дубна, Московская обл., ул. Жолио-Кюри, 6.
E-mail: publish@pds.jinr.ru
www.jinr.ru/publish/