J. Verbeeck, G. Bertoni, Ultramicroscopy 108 (2008) 74–83

Model-based quantification of EELS spectra: Treating the effect of correlated noise

J. Verbeeck[a,b,*], G. Bertoni[b]

[a]*Institut für Festkörperphysik, Technische Universität Wien, Wiedner Hauptstrasse 8-10, A-1040 Wien, Austria*
[b]*EMAT, University of Antwerp, Groenenborgerlaan 171, B-2020 Antwerp, Belgium*

Abstract

Correlated noise is generally present in experimentally recorded electron energy loss spectra due to a non-ideal electron detector. In this contribution we describe a method to experimentally measure the noise properties of the detector as well as the consequences it has for model-based quantification using maximum likelihood. The effect of the correlated noise on the maximum likelihood fitting results can be shown to be negligible for the estimated (co)variance of the parameters while an experimentally obtained scaling factor is required to correct the likelihood ratio test for the reduction of noise power with frequency. Both effects are derived theoretically under a set of approximations and tested for a range of signal-to-noise values using numerical experiments. Finally, an experimental example shows that the correction for correlated noise is essential and should always be included in the fitting procedure.

PACS: 79.20.Uv; 82.80.Pv; 06.20.Dk; 43.50.+y; 02.50.−z; 07.05.FB

Keywords: Correlated noise; Maximum likelihood; EELS quantification; Model-based quantification

1. Introduction

Model-based quantification of electron energy loss spectra makes use of a maximum likelihood estimator to estimate the parameters of a model by fitting this model to an experimental spectrum. If an appropriate model is chosen, the estimated parameters are measurements of physically interesting properties of the sample, like e.g. the concentration of certain chemical elements. This procedure was explained in detail in Refs. [1,2] and had the advantage of not only returning measurements of the parameters in the model but also returning estimates for the lowest expected variance on these parameters and returning an indication on whether the model was an appropriate description for the experimental observations. Using the maximum likelihood procedure also guaranteed that in practice the real (co)variance of the parameters for repeated experiments was approaching the lowest expected (co)variance asymptotically (for a large number of pixels in the spectrum) as calculated by the so-called Cramér Rao lower bound (CRLB). This achievement of the lower bound on the (co)variance of the parameters was shown to be valid for numerical calculations under situations similar to the experimental situation. In real experiments, however, the CRLB was only approached within a factor 2 or 3, indicating that some of the assumptions in the model were not met. The two basic assumptions were that we had:

1. A valid model for the expectation value λ_m of the experiment in each pixel m.
2. A valid noise model.

The first assumption is a typical problem in physics and we believe that we can come up with a sufficiently accurate model for the expectation values depending on a sufficient but limited amount of parameters. For the second assumption, however, we took the noise to be independent

*Corresponding author. EMAT, University of Antwerp, Groenenborgerlaan 171, B-2020 Antwerp, Belgium. Tel.: +32 32653249; fax: +32 32653257.
E-mail address: jo.verbeeck@ua.ac.be (J. Verbeeck).

Poisson-distributed noise with a factor allowing for gain in the detector system. In this paper we will describe how we introduce a more realistic noise model allowing for correlated noise due to a non-ideal detector system.

The original assumption of independent noise in each pixel would be perfectly valid for a serial-EELS instrument[1] but breaks down for parallel-EELS systems where the spectrum is recorded in parallel on a 1D or more commonly on a 2D detector [3,4]. Meyer et al. [5–7] have studied, both theoretically and experimentally, the signal and noise transfer through some common 2D detectors in great detail. In this paper we will make use of their findings and derive the consequences of the detector properties on the recording of EELS spectra and the effect this has on model-based quantification. We will show that the detector introduces correlation in the noise which can be taken approximately into account in the maximum likelihood process by experimentally obtained scaling factors. The approximations are shown to be very good for both realistic numerical experiments as well as for real experiments.

2. The effect of correlated noise on the Fisher information and on the likelihood

If we can treat the effect of the conversion between electrons and final observed counts in the detector as a gain G and a convolution with a point spread function h_m:

$$y_m = G x_m \otimes h_m, \tag{1}$$

with y_m the observed counts and x_m the original detected electrons and $\sum_m h_m = 1$. It is possible to come back to the original electron counts via deconvolution if the Fourier spectrum of h does not contain zeros. In practical cases, the point spread function h_m of the detector is a sharp function of only a few pixels wide with a smooth spectrum:

$$x_m = \frac{y_m}{G} \otimes g_m, \tag{2}$$

with g_m the point spread function of the deconvolution process given by

$$g_m = \mathscr{F}^{-1}\left[\frac{1}{\mathscr{F}(h_m)}\right]. \tag{3}$$

Note that noise amplification will occur, but in this case that is exactly what we want to make the noise power spectrum flat again as needed for independent Poisson noise. So, in principle we can deconvolve our experimental data to undo the effect of the detector and then apply the formulas for the independent pixel Poisson log likelihood from [1]:

$$\ln P = \sum x_m \ln(\lambda_m) - \lambda_m - \ln(x_m!), \tag{4}$$

$$\ln P = \sum x_m \ln\frac{\lambda_m}{x_m} + (x_m - \lambda_m), \tag{5}$$

[1] If the timing is chosen so that there are no memory effects.

with P the probability of obtaining an experimental spectrum with values x_m if the true model spectrum has values λ_m and pixel-independent Poisson noise is assumed. Stirling's formula was used for the logarithm of a factorial assuming a high number of electrons in each pixel ($x_m \gg 1$).

This deconvolution process, however, has several disadvantages:

- Edge artefacts will arise from the fact that one only measures the spectrum in a limited range of energies. These artefacts are entirely unwanted because they alter the precious data.
- Although the deconvolution will flatten the noise spectrum to white noise as expected for Poisson noise, it does not completely undo the correlation since the scattering in the scintillator is a stochastic process which can in general not be undone unless one knows the path that each photon took, and not just the probabilities for each path.
- As a principle in parameter estimation, one should never apply transformations on the data, but always include their effect in the model.

Therefore we want to investigate whether it is possible to derive the effect of the induced correlation in the noise by the detector in a direct way. In the rest of this section, we assume the gain $G = 1$. This can always be achieved by simply scaling the data by G. We will first derive the effect of correlated noise on the elements of the Fisher information matrix and secondly we look at the effect on the log likelihood.

2.1. Correlated noise and Fisher information

We start by deriving the elements of the Fisher information matrix given by

$$F_{i,j} = -\frac{\partial^2 \ln P}{\partial\theta_i \partial\theta_j}, \tag{6}$$

with θ_i the parameters of the model. Assuming that we know the real electron counts x_m we can start from Eq. (5) and take a first derivative w.r.t. to a parameter:

$$\frac{\partial \ln P}{\partial\theta_i} = \sum_m \frac{\partial\lambda_m}{\partial\theta_i}\left[\frac{x_m}{\lambda_m} - 1\right]. \tag{7}$$

Taking the second derivative, we obtain the components of the Fisher information matrix:

$$F_{i,j} = -\frac{\partial^2 \ln P}{\partial\theta_i \partial\theta_j} = \sum_m \frac{\partial^2\lambda_m}{\partial\theta_i\partial\theta_j}\left[1 - \frac{x_m}{\lambda_m}\right] + \frac{\partial\lambda_m}{\partial\theta_i}\frac{\partial\lambda_m}{\partial\theta_j}\left[\frac{x_m}{\lambda_m^2}\right], \tag{8}$$

which is linear in x_m. In the optimum, the first derivative (Eq. (7)) is zero and therefore the factor multiplying the

double derivative vanishes. We get

$$F_{i,j,\text{opt.}} = \sum_m \frac{\partial\lambda_m}{\partial\theta_i}\frac{\partial\lambda_m}{\partial\theta_j}\left[\frac{x_m}{\lambda_m^2}\right]. \tag{9}$$

Rewriting this to make the presence of the noise clear gives

$$F_{i,j,\text{opt.}} = \sum_m \frac{\partial\lambda_m}{\partial\theta_i}\frac{\partial\lambda_m}{\partial\theta_j}\frac{x_m-\lambda_m+\lambda_m}{\lambda_m^2}, \tag{10}$$

$$F_{i,j,\text{opt.}} = \sum_m \frac{1}{\lambda_m}\frac{\partial\lambda_m}{\partial\theta_i}\frac{\partial\lambda_m}{\partial\theta_j}\left(\frac{n_m}{\lambda_m}+1\right), \tag{11}$$

with n_m the noise term.[2] We can split the information in two parts:

$$F_{i,j,\text{opt.}} = \sum_m \frac{1}{\lambda_m}\frac{\partial\lambda_m}{\partial\theta_i}\frac{\partial\lambda_m}{\partial\theta_j} + \sum_m \frac{1}{\lambda_m}\frac{\partial\lambda_m}{\partial\theta_i}\frac{\partial\lambda_m}{\partial\theta_j}\frac{n_m}{\lambda_m}. \tag{12}$$

There are two reasons why the last term can be neglected in practice:

- The noise over signal ratio n_m/λ_m should be much smaller than 1 for good experimental data.
- Usually the parameters are non-local in the sense that $\partial\lambda_m/\partial\theta_i \neq 0$ for a wide range of m and is smoothly varying with m. If we also assume that the model is smooth, the fluctuating noise term will cancel out after summing. This will happen regardless of whether the noise was correlated or not.

We get

$$F_{i,j,\text{opt.}} \approx \sum \frac{1}{\lambda_m}\frac{\partial\lambda_m}{\partial\theta_i}\frac{\partial\lambda_m}{\partial\theta_j}, \tag{13}$$

which is independent on the correlation of the noise. This means that for calculating the estimated covariance on the parameters via the CRLB theorem [1]:

$$VF \geqslant 1, \tag{14}$$

with V the covariance matrix. For a large number of pixels in the spectrum, the equal sign is approached as shown in Refs. [1,2]. This makes the estimated covariance matrix:

$$\hat{V} = F^{-1} \tag{15}$$

independent of the fact that the noise is correlated under the given assumptions. There are several intuitive reasons why these assumption should be valid:

- Reasonable models should have many fewer parameters than experimental pixels. The information in the whole spectrum is somehow divided over the parameters which cancels out the correlation between the pixels.
- If the signal-to-noise ratio is high, the requirement for non-local parameters and smoothness is not even necessary.
- The model will in general be smooth over the distances where correlation plays a role, since also the signal is at least smeared out by the modulation transfer function of the detector which in practice is wider than the noise transfer function (detector quantum efficiency (DQE) drops with frequency [7]). In practice the energy width of the source is another factor which smooths the model even more depending on the energy dispersion used.

A notable and obvious exception would be when the model would be just modelling the pixel values. In that case the parameters are highly local and then the Fisher information would depend strongly on the correlation. However, such a model would not be very useful since physics in general is of course looking for a reduction of the number of parameters by making use of appropriate models.

2.2. Scaling the likelihood

Although we showed that the (co)variance of the parameters of a model is essentially unchanged when correlated noise is present, we will show now that the likelihood itself is strongly dependent on the presence of correlated noise. This can be seen by making a Taylor expansion around $x_m = \lambda_m$:

$$\ln P \approx -\sum_m \frac{(x_m-\lambda_m)^2}{2\lambda_m} - \frac{(x_m-\lambda_m)^3}{3!\lambda_m^2} + \frac{2(x_m-\lambda_m)^4}{4!\lambda_m^3} + \cdots. \tag{16}$$

We see that up to second order in the noise this is the same as for the log likelihood for Gaussian noise with a variance of λ_m. Neglecting the higher order terms we get

$$\ln P \approx -\sum_m \frac{(n_m)^2}{2\lambda_m} + \cdots, \tag{17}$$

with $n_m = x_m - \lambda_m$ the noise signal if the model is an adequate description of the experiment. We see that the log likelihood is given by the sum of the noise "power" (n_m^2) per pixel and per electron in that pixel. It is convenient to rewrite this in spectral form by noting that the average noise power spectrum of 1 pixel with 1 electron is constant and given by

$$n_f^2 = 1. \tag{18}$$

Noting that we can view the experiment as electrons arriving one after the other and that power (or variance) is additive, we can write the sum as

$$\ln P \approx -\frac{1}{2}\sum_f (n_f)^2 = -\frac{N}{2}. \tag{19}$$

For correlated noise, the average noise power spectrum is no longer flat and we have

$$\ln P' \approx -\frac{1}{2}\sum_f (n'_f)^2 = -\frac{N}{2C}, \tag{20}$$

[2] Note the qualitative resemblance with the so-called Shannon–Hartley theorem in communication theory which states that the channel capacity C (bits/s equivalent of information transferred) goes as $C = B\log_2(1 + (S/N))$ with B the bandwidth of the channel, S the total signal power and N the total noise power.

with n'_f the average noise power spectrum for 1 pixel with 1 electron. And

$$C = \frac{\sum n_f^2}{\sum n_f'^2}, \tag{21}$$

the ratio between the total average noise power in the spectrum without and with correlation. This shows that the log likelihood scales with $1/C$ when going from uncorrelated noise to correlated noise which is important when doing acceptance tests. It appears as if the total number of independent pixels is scaled down by C due to correlation. This affects the degree of freedom of the system from $N - I$ to $(N/C) - I$ with I the number of independent parameters in the model.

2.2.1. Effect on the variance for white illumination

Note also that C is the factor with which the variance for constant illumination is scaled. This can be seen from (Eq. (17)) since for white illumination $\lambda_m = \lambda$ is constant:

$$\ln P \approx -\frac{1}{2\lambda}\sum_m (n_m)^2 + \cdots, \tag{22}$$

$$\ln P \approx -\frac{C}{2\lambda}\sum_m (n'_m)^2 + \cdots. \tag{23}$$

The sample variance is given as

$$\text{var} = \frac{1}{N}\sum (n_m)^2, \tag{24}$$

$$\text{var}' = \frac{1}{N}\sum (n'_m)^2 = \frac{1}{C}\text{var}, \tag{25}$$

which therefore scales down by C if correlation is present. This is the reason why we have to correct the gain G_{var} measured as the experimental relationship between variance and mean for even illumination, with a correction factor C to obtain the true gain G of the detector.

2.2.2. Effect on model validation

The reduction of degrees of freedom of the system has a profound influence on model validation using the likelihood ratio [1,8,9]. Not only does it scale down the log likelihood but it also changes the χ^2 distribution with which we have to compare this output. We change the equations from

$$\text{LR} = -2\ln P \to \chi^2(N - I) \tag{26}$$

to

$$\text{LR}' = -2\ln P' \to \chi^2\left(\frac{N}{C} - I\right). \tag{27}$$

Note that it is not obvious that after correlation the distribution of the LR outcomes would still be a χ^2 distribution, and in principle a numerical simulation is needed to find out the new distribution function. We show, however, in Fig. 1 for a realistic numerical simulation (same model as in Section 3) that our assumption to just scale the χ^2 distribution for different degrees of freedom works quite accurately.

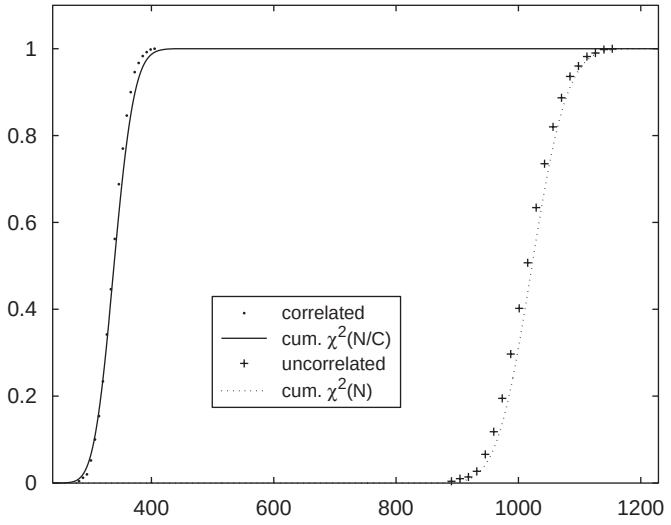

Fig. 1. Numerical calculation of the cumulative distribution of $\text{LR} = -2\ln P$ for $N = 1024$ and 1000 repeated experiments for both correlated and uncorrelated noise. Both results can be modeled quite well with a χ^2 distribution with N/C or N degrees of freedom. The fit is as least as good as that for the uncorrelated case. The main deviations are located in the right part of the curve which is unimportant since we typically want to use e.g. a 5% significance level.

2.3. Conclusions

To summarise the findings of the previous sections we can state that under certain circumstances the effect of correlated noise on the parameter estimation problem is twofold:

- The Fisher information for a reasonable model does not change with correlated noise. This means that for the calculation of the estimated (co)variance via the CRLB we can use the formulas for uncorrelated noise. As long as we correct properly for the gain of the detector.
- The log likelihood on the other hand scales with a factor C relating the total noise power with and without correlation which is important to take into account when doing acceptance tests. Intuitively this can be understood by the fact that correlation causes a reduction in the high frequency components of the noise which will make the experiment look too good when calculating the likelihood assuming uncorrelated noise.

At first sight there might seem to be a contradiction in both conclusions since the Fisher information matrix is the double derivative of the log likelihood, so if the log likelihood scales, so must the $F_{i,j}$. The reason for this apparent contradiction can be seen as follows. The $\ln P$ is a function of n_m^2 while the $F_{i,i}$ contain a sum over n_m. For $\ln P$ the summing will depend on the correlation because it

affects the total noise power by a factor C, while the sum over n_m in the $F_{i,i}$ formula cancels out approximately because of the fact that the noise has a zero mean regardless of whether it is correlated or not.

3. Numerical simulations

A series of numerical experiments is carried out to test the validity of the approximations above on a simple model consisting of a constant background and a Lorentz peak having four parameters:

$$\lambda_m = \theta_0 + \theta_1 \frac{(\theta_2/2)^2}{(m-\theta_3)^2 + (\theta_2/2)^2}, \tag{28}$$

with θ_0 the height of the constant background, θ_1 the height of the Lorentz peak, θ_2 the FWHM and θ_3 the position. The constant background is chosen in order to avoid edge artefacts when calculating convolutions. Several realisations of the experiment can now be obtained adding Poisson noise n_m from an appropriate random generator:

$$x_m = \lambda_m + n_m. \tag{29}$$

The effect of the detector is then modeled as convolution:

$$y_m = \lambda_m \otimes h_S + n_m \otimes h_N, \tag{30}$$

with h_S and h_N the point spread functions for signal and noise, respectively. We take for these point spread functions Lorentzians with FWHM f_S and f_N. In general $f_S > f_N$ because the DQE drops with higher frequencies (see Section 4). To obtain DQE values comparable to an existing detector [6] we use $f_S = 2.2$ pixels and $f_N = 2$ pixels. We can deconvolve the y_m to regain a white noise spectrum and get an approximation for the original electron counts $\hat{x}_m$:

$$\hat{x}_m = \lambda_m \otimes h_S \otimes g_N + n_m \otimes h_N \otimes g_N, \tag{31}$$

with

$$g_N = \mathscr{F}^{-1}\left[\frac{1}{\mathscr{F} h_N}\right], \tag{32}$$

which immediately shows a problem when $h_N \neq h_S$:

$$\hat{x}_m = \lambda_m \otimes h_S \otimes g_N + n_m \tag{33}$$

in the sense that the deconvolution of the noise cannot cancel the point spread function of the signal properly. Fortunately, since $f_S > f_N$ this will only cause a blurring of the signal. The gain of the detector is taken to be the same for signal and noise as

$$G = \sum h_N = \sum h_S. \tag{34}$$

We can now test the different ways of treating the effect of correlated noise:

1. Do model fitting on the x_m assuming Poisson-distributed independent pixels.
2. Do model fitting on the y_m assuming Poisson-distributed independent pixels but correcting for the gain.
3. Do model fitting on the $\hat{x}_m$ assuming Poisson-distributed independent pixels.
4. Do model fitting on the y_m assuming Poisson-distributed independent pixels but correcting for the likelihood scaling.

With 2 being the way which was used in EELSMODEL up to now and 4 the way which is implemented in a new version of EELSMODEL.

The results for 200 repeated experiments for different signal-to-noise ratios are given in Tables 1–3. Note that all models find a reasonable estimate for the CRLB. Model 2 is always too optimistic in terms of acceptance tests which is corrected in model 4. Model 4 finds an average $\ln P/(N/C - I)$ very close to 1, but the amount of accepted

Table 1
Results for an average of 10 electrons per pixel

	θ_0	Stdev	CRLB	θ_1	Stdev	CRLB	θ_2	Stdev	CRLB	θ_3	Stdev	CRLB	Chisq	Accepted (%)
Model 1	8.04	0.204	0.210	31.6	3.349	2.945	10.2	1.147	1.123	127.02	0.463	0.412	1.015	95.50
Model 2	7.96	0.202	0.209	31.8	3.235	2.941	10.2	1.094	1.117	127.02	0.448	0.410	0.324	100.00
Model 3	8.06	0.203	0.210	32.6	3.407	3.019	9.92	1.105	1.083	127.02	0.445	0.399	1.012	95.50
Model 4	7.96	0.202	0.209	31.8	3.235	2.941	10.2	1.094	1.117	127.02	0.448	0.410	1.016	96.50

Table 2
Results for an average of 100 electrons per pixel

	θ_0	Stdev	CRLB	θ_1	Stdev	CRLB	θ_2	Stdev	CRLB	θ_3	Stdev	CRLB	Chisq	Accepted (%)
Model 1	80.6	0.657	0.664	321	10.481	9.391	10.1	0.390	0.346	127.01	0.128	0.127	1.003	94.00
Model 2	80.4	0.655	0.666	316	10.008	9.210	10.3	0.383	0.355	127.01	0.128	0.130	0.321	100.00
Model 3	80.7	0.657	0.664	323	10.514	9.429	10.0	0.388	0.343	127.01	0.127	0.127	1.002	94.00
Model 4	80.4	0.655	0.666	316	10.008	9.210	10.3	0.383	0.355	127.01	0.128	0.130	1.009	96.50

Table 3
Results for an average of 1000 electrons per pixel

	θ_0	Stdev	CRLB	θ_1	Stdev	CRLB	θ_2	Stdev	CRLB	θ_3	Stdev	CRLB	Chisq	Accepted (%)
Model 1	807	2.315	2.101	3215	29.502	29.719	10.0	0.111	0.109	127.00	0.040	0.040	1.004	95.50
Model 2	806	2.322	2.101	3157	28.100	29.154	10.3	0.110	0.112	127.00	0.040	0.041	0.331	100.00
Model 3	807	2.314	2.099	3229	29.623	29.839	9.99	0.111	0.108	127.00	0.040	0.040	1.004	95.50
Model 4	806	2.322	2.101	3157	28.100	29.154	10.3	0.110	0.112	127.00	0.040	0.041	1.038	96.00

spectra is slightly too high (95% expected). This could be due to the fact that we approximate the distribution of the likelihood ratio with a chi-square distribution. We see that also model 3 performs well although in practice we want to avoid deconvoluting the spectra to avoid introducing edge artefacts. On the other hand model 4 gives almost identical results without deconvolution even for situations with a very poor signal-to-noise ratio.

4. Noise properties of the detector

The detector can be characterised by measuring separately the transfer function for the noise and for the signal as was shown in Refs. [5–7]. The fact that noise and signal transfer are different is quite confusing but leads to the physically observed effect of a spatial frequency-dependent DQE. The DQE is connected to the transfer of signal and noise as given in Ref. [5]:

$$\mathrm{DQE}(u,v) = \mathrm{DQE}(0,0)\left(\frac{\mathrm{MTF}(u,v)}{\mathrm{NTF}(u,v)}\right)^2, \tag{35}$$

with NTF the noise transfer function and MTF the modulation or signal transfer function written as a function of spatial frequency (u,v). Experimentally this is a function which decreases with spatial frequency which means that the NTF has more high-frequency content than the MTF. Or that the point spread function for noise is more localised than the point spread function for the signal. This was explained in Ref. [5] as the fact that for a single electron hitting the detector the distribution of the few photons created by this event is likely to be more localised than the overall distribution of photons when a large number of electrons hit the detector. This affects the noise power spectrum since each electron creates independently its own noise contribution while for the signal the expectation value is used. In our case we are, however, mainly interested in the noise transfer which is easier to measure than the signal transfer as described in Ref. [7].

4.1. Experimental determination of the noise power spectrum

To obtain an estimate of the noise transfer function the following procedure is used:

- The CCD is illuminated with even illumination with a mean number of counts $\lambda_{m,n} = \bar{x}_{m,n}$.
- Two exposures taken under the same conditions are subtracted to obtain the noise signal $\sqrt{2}n_{m,n}$. This avoids artefacts in the gain correction or dark correction to be regarded as noise.
- The noise signal is summed in vertical direction over the same area as is used for the conversion of the CCD image into an EELS spectrum: $n_m = \sum_n n_{m,n}$.
- The power spectrum of this integrated noise signal is calculated as $|\mathscr{F} n_m|^2$, with $\mathscr{F}$ the Fourier transform.
- The previous steps are repeated and an average power spectrum is calculated. Iteration is stopped when the power spectrum has an acceptable smoothness, convergence is slow especially for the low frequencies. Averaging the power spectra and not first averaging the noise is necessary because of the fact that only the noise variances (related to power) are additive (if subsequent exposures are independent).

An improvement in speed can be obtained when using the oversampling technique described in Ref. [7]. In this technique, power spectra taken at different binning settings are combined. This has the advantage that readout times are lower for binned readout while the spectra still contain the low-frequency information of the noise power spectrum.

The obtained averaged power spectrum can be numerically smoothed to obtain a useful estimate of the power spectrum. The power spectrum can be seen as the square of the noise transfer function $\mathrm{NTF}^2(u)$ since the incoming independent Poisson noise has a white power spectrum which is altered by the noise transfer function to create the observed power spectrum. Fourier transforming the noise power spectrum gives the autocorrelation function of the noise according to the Wiener–Khinchine theorem.

4.2. Experimental determination of the gain

To determine the gain of the detector system (counts/electron) as far as the noise is concerned one can make a plot of the variance as a function of the mean for even illumination by collecting a set of data with increasing exposure times.[3] The procedure is similar to the noise power spectrum determination:

[3]The dark current increases with exposure time, but in practice this has a negligible effect on the result. The effect can be quantified by repeating the procedure with the illumination fully blanked so that only dark counts are measured.

- Record the noise signal n_m as above, under conditions of white illumination.
- Repeat this for different exposure times.
- Determine the factor G_{var} relating the experimental variance of the noise $\text{var}(n_m) = G_{\text{var}}\text{mean}(x_m)$ with linear interpolation.

The gain factor G_{var} obtained in this way is not the true gain of the system since the correlation lowers the variance. It was shown in Eq. (25) that the reduction of the variance is given by a correlation factor C which can be determined from the noise power spectrum. The true gain of the system is therefore given by

$$G = G_{\text{var}}C. \tag{36}$$

Both procedures for the noise power spectrum and the determination of gain are combined in a Digital Micrograph™ script, freely available together with an updated version of EELSMODEL [10]. Note that in previous papers [1,2] we used $G_{\text{var}} = G$ since we neglected correlation ($C = 1$). In the following sections we will discuss the case when correlation is not negligible as is the case for currently available detectors.

5. Experiments

5.1. Experimental noise properties

The noise properties for two different detectors were measured: A 1K CCD with phosphor scintillator on a GIF200 mounted on a Phillips CM30 and a 1K CCD with phosphor scintillator on a GIF2000 mounted on a JEOL 3000F. A power spectrum for the GIF2000 detector is given in Fig. 2. The results for gain and correlation factor obtained by the script described in Section 4 are given in the following table:

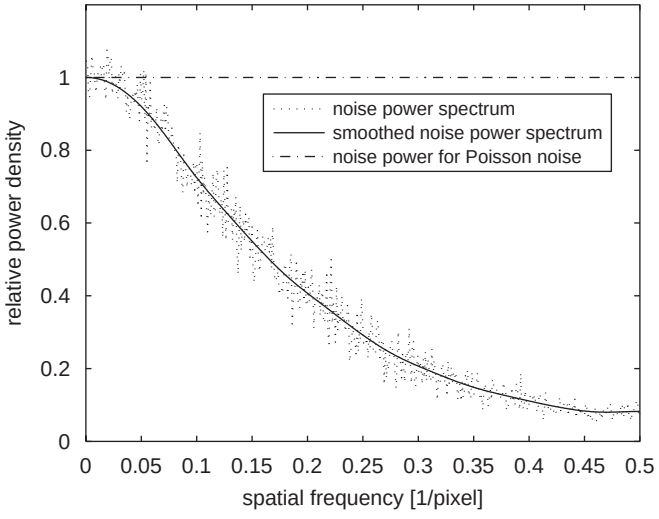

Fig. 2. Experimental noise power spectrum for a 1K phosphor GIF2000 CCD on a JEOL 3000F together with a smoothed version and the flat theoretical noise power spectrum for independent Poisson noise. Use is made of the oversampling technique to improve the low-frequency noise as described in Ref. [7] with 50 recordings at binning 1, 200 at binning 2 and 600 at binning 4.

	Type	G_{var}	G	C	G_{T}
1K Phosphor GIF200	CM30	1.80	5.45	3.03	2.4
1K Phosphor GIF2000	3000F	2.83	7.12	2.52	3.1

Note that the newer detector on the GIF2000 system apparently has a higher gain and less noise correlation. This makes it a more desirable detector for core loss EELS measurements where you want to get most out of noisy spectra.

The gain G used in this paper is the value with which the noise power spectrum has to be divided to have the same total noise power as an ideal uncorrelated electron detector. The signal gain on the other hand is denoted by G_{T} and is measured by using the GIF drift tube as a Faraday cup coupled to a load resistor and a voltmeter. In the terminology of Meyer et al. [7] our $G_{\text{T}} = \text{MTF}(0)$ and $G_{\text{var}} = \text{NTF}^2(0)$. Note that for a detector where the noise is only formed by a single correlated Poisson detection step[4] $G_{\text{T}} = G$. Apparently, in a real detector $G > G_{\text{T}}$ meaning that more noise power is present as would be expected from the signal gain G_{T}. In our case both detectors have about double the total noise power as an ideal detector with the same correlation properties. It is interesting to note that the gain ratio between both detectors is approximately 1.3 in favour of the detector on the 3000F regardless of whether one takes G or G_{T}.

5.2. Repeated experiments on $SrTiO_3$

A set of 100 spectra from a $SrTiO_3$ crystal was taken keeping exactly the same conditions with a JEOL 3000F microscope and GIF 2000 spectrometer. The detector chip in the GIF is a 1K CCD with phosphor scintillator. A tilt of $\approx 2^\circ$ from the 100 zone axis is used to avoid channelling. One spectrum for this set is shown in Fig. 3 together with a model consisting of a second-order log polynomial, two Hartree Slater edges for the $TiL_{2,3}$ and O K-edges and a fine structure component [2] of 65 points for Ti and 45 points for O. Multiple scattering is included by a convolution with a recorded low loss spectrum. The obtained results for the Ti/O ratio can be tabulated with their 95% confidence intervals between brackets calculated from Eqs. (26) and (27) in Ref. [1]:

Ti/O ratio	0.359	(0.353, 0.365)
Stdev	0.0316	(0.0278, 0.0367)
CRLB	0.0306	
Acc. at 5%	83%	

[4] Imagine a direct conversion of electrons into electron hole pairs in a semiconductor and the scattering inside the detector gives rise to the point spread function and noise correlation.

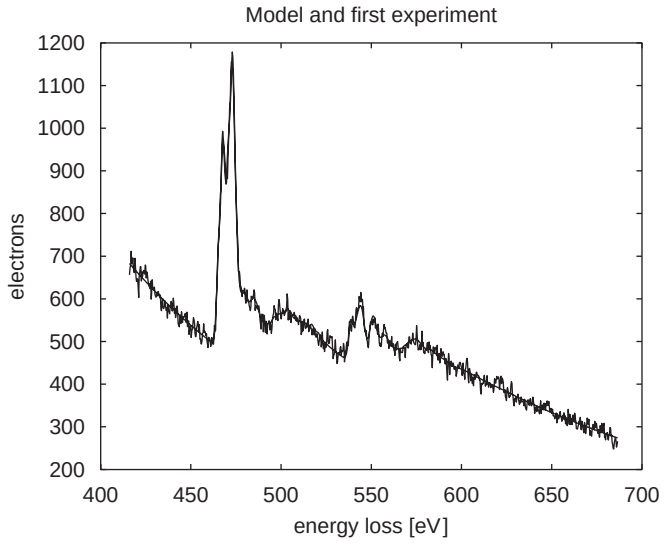

Fig. 3. Experimental spectrum and the average model obtained from fitting 100 spectra.

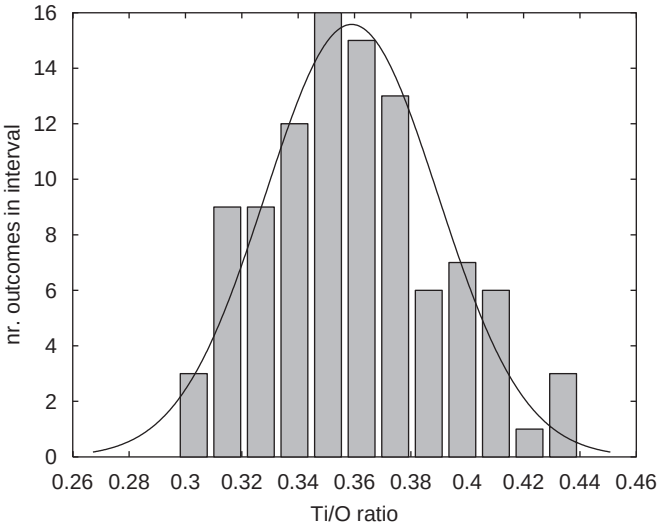

Fig. 4. Histogram of the distribution of the estimated Ti/O ratio for 100 spectra together with the theoretical Gaussian prediction. The experimentally obtained mean = 0.359 and standard deviation = 0.0316, while the CRLB = 0.0306. An excellent agreement with theory is obtained, indicating a good description of the noise.

We see from the table how remarkably well the standard deviation is predicted by the CRLB. The obtained ratio is biased with respect to the expected 0.333, but this is a matter of accuracy which will be discussed in a future paper. In principle the theory only describes the expected standard deviation but contains no indication on the accuracy of the obtained results. Note also that the standard deviation in this case is quite high ($\approx 10\%$) which is due to the rather noisy spectra. Clearly if one would take only a single spectra to estimate the stoichiometry of this sample with such a low signal-to-noise ratio, large errors are to be expected. The acceptance of 83% is a little bit lower as the theoretical expected acceptance of 95%. This can have two origins: either the model is not good enough, or the noise model is wrong. To discriminate between these two options we fitted a power law background $A(\frac{E}{E_{\text{start}}})^{-r}$ in a region of the spectrum behind the O K-edge (588–680 eV). This fit resulted in the next table with 95% confidence levels between brackets:

A	1891	(1878,1904)
Stdev	63	(55,73)
CRLB	66	
r	3.32	(3.31,3.33)
Stdev	0.066	(0.058, 0.077)
CRLB	0.071	
Acc. at 5%	94%	

We see from this table that the noise model is quite accurate in the sense that a model with 94% acceptance is possible with standard deviation on the two parameters being very close to the predicted CRLB. The fact that both CRLB are higher than the real obtained standard deviations can be seen as a statistical effect: the 95% confidence intervals for 100 repeated experiments are given between brackets.

This leads to the conclusion that the model must be the origin of the slightly lower acceptance for the complete model including the excitation edges. One of the weak points of the fine structure component is the fact that it does not take into account life-time broadening. Therefore we need an extremely high number of parameters (65 and 45) to come close to an acceptable model while this high density of points is only needed near the onset of the edges because of life time broadening. Future work is needed to design a fine structure component which takes this life time effect into account by e.g. sampling on a quadratic mesh. If we look at the distribution of the Ti/O ratio in Fig. 4 we see how well the theoretical Gaussian curve describes the observed results. In comparison with previous publications [1,2], we see that including the correlation gives a much better estimate of the standard deviation. The problem of creating acceptable models still remains, although with the use of the fine structure component [2] we come very close.

Another useful tool to study the noise in these spectra is by creating a residual map. The residual map is calculated as the difference between the average model (model evaluated at the mean of the parameters $\bar{\theta}_i$ for 100 spectra) and the spectra. If the model is statistically correct then the residual should be a pure noise signal which can then be used to compare this noise with the expected behaviour of the noise for a given detector. This is a consistency check of the noise model from the data itself, but it can also reveal problems with the model. The residual map is shown in Fig. 5 and shows a smooth noisy behaviour as expected. Only near the white lines of Ti, some outliers seem to exist which can be attributed to energy drift. If the energy drifts in comparison to the fixed average model, an error is made

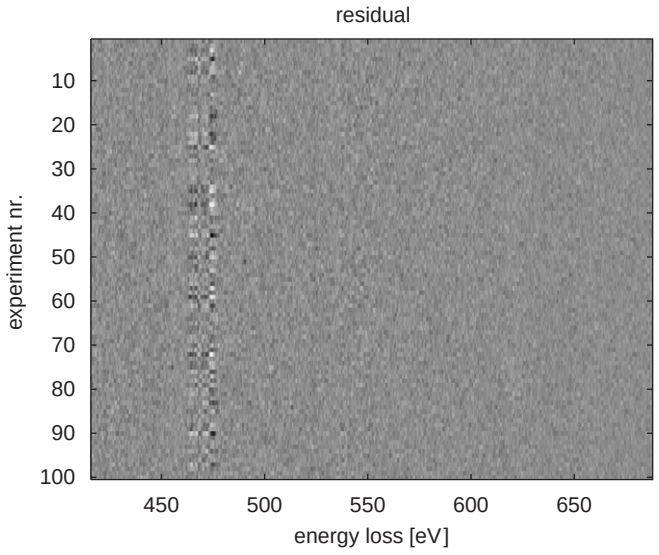

Fig. 5. Residual map for the 100 spectra calculated as the difference between the average model and each experiment. This map should show the noise if the model was a good description of the experiment and is therefore a good tool to study whether this noise behaves as expected for a given detector. If deviations exist they must either point to a bad model or to a bad noise model.

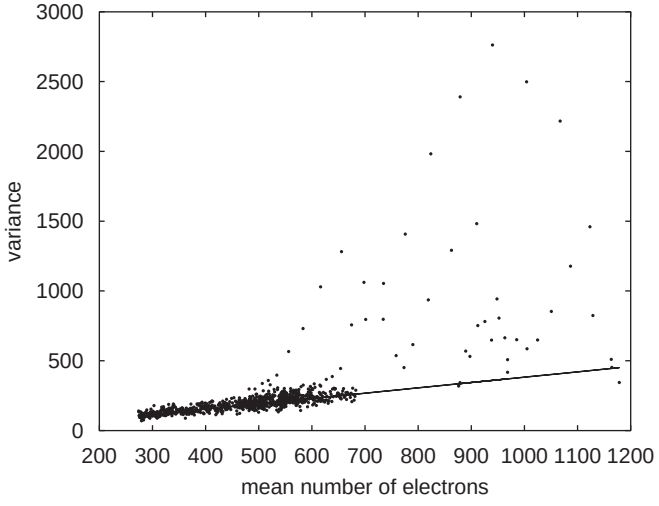

Fig. 6. Variance vs. mean calculated from the residual. We see that the variance follows closely the expected linear behaviour with a slope give by a previous measurement for this detector indicating that the model is quite a good description of the experiments. Some outlier points exist which can be attributed to energy drift. If the energy drifts, the model will not fit very well especially near the sharp white line features of the $TiL_{2,3}$ edge.

which cannot be attributed to noise. If we plot the variance of the residual w.r.t. the mean we get Fig. 6. This plot shows that most points lie close to the line $\mathrm{var}_m = G_{\mathrm{var}}\bar{x}_m$, obtained from previous measurements of the noise properties of the detector, with the exception of some outlier points which were attributed to energy drift. This shows that indeed the noise behaves as described by the noise characterisation of the detector.

6. Conclusion

We have shown that correlated noise due to the non-ideal detector can be characterised using automated scripts which measure the noise power spectrum and the gain factor, relating the variance of the noise for white illumination with the mean number of counts in such a spectrum. The scripts are available on the EELSMODEL website within the new distribution [10].

The obtained noise power spectrum is also essential as an input for a Wiener type deconvolution process where deconvolution only is performed for those frequency components where the signal-to-noise ratio is greater than 1. A detailed study describing this is currently being carried out.

Knowing the noise power spectrum and the gain factor we have established approximate ways of describing the effect on a maximum likelihood fitting process. We showed that under reasonable assumptions, the effect of the correlated noise on the (co)variance estimates of the parameters is negligible, while the gain factor correction is essential. The model validation, using the likelihood ratio test, is very sensitive to the correlated noise but its effect can be approximated quite well by a scaling of the number of independent pixels by a correlation factor C. This correlation factor can be obtained from the experimental noise power spectrum as the ratio of the total noise power, if independent Poisson noise is assumed, over the experimental total noise power. This corrects for the fact that correlation reduces the amount of high-frequency noise which would lead to an unrealistically high probability that a certain experiment could have occurred for a given model.

The approximate corrections were tested on a large set of numerical experiments showing that the corrections work extremely well even for spectra with very low signal-to-noise ratios. Finally, the approach is tested on a set of experimental $SrTiO_3$ spectra showing that the corrections are essential and lead to a realistic estimation of the variance of the parameters and a useful indication for the model acceptance tests.

The corrections have been included in the model-based quantification program EELSMODEL and a new version is available for download in Ref. [10]. Users can create entries for the measured noise properties of their specific detector making use of an automated script.

It is clear that not only for imaging purposes, but also for quantitative EELS, the TEM community would profit from the development of better detectors with a reduced point spread function for signal as well as noise. But until then, it will be necessary to include the effects of correlated noise when applying model-based fitting techniques on spectral as well as on image data.

Acknowledgements

G.B. wants to thank FWO-Vlaanderen for financial support under contract nr. G.0147.06. J.V. wants to thank Sandra Van Aert for stimulating discussions. The authors acknowledge financial support from the European Union under the Framework 6 program under a contract for an Integrated Infrastructure Initiative. Reference 026019 ESTEEM. Both authors appreciate the help of R. Van Ginderen in measuring the beam current.

References

[1] J. Verbeeck, S. Van Aert, Ultramicroscopy 101 (2004) 2.
[2] J. Verbeeck, S. Van Aert, G. Bertoni, Ultramicroscopy 106 (2006) 976.
[3] A.J. Gubbens, O.L. Krivanek, Ultramicroscopy 51 (1–4) (1993) 146.
[4] O.L. Krivanek, Ultramicroscopy 28 (1–4) (1989) 118.
[5] R.R. Meyer, A.I. Kirkland, Ultramicroscopy 75 (1) (1998) 23.
[6] R.R. Meyer, A.I. Kirkland, R.E. Dunin-Borkowski, J.L. Hutchison, Ultramicroscopy 85 (2000) 9.
[7] R.R. Meyer, A.I. Kirkland, Microsc. Res. Tech. 49 (2000) 269.
[8] A.J. den Dekker, S. Van Aert, A. van den Bos, D. Van Dyck, Ultramicroscopy 104 (2) (2005) 83.
[9] T. Radetic, U. Dahmen, Mat. Res. Soc. Symp. 695 (2002) 47.
[10] ⟨http://webhost.ua.ac.be/eelsmod/eelsmodel.htm⟩.