AN EM APPROACH TO MINERAL ANALYSIS USING NATURAL GAMMA RAYs

Bill Moran\textsuperscript{a}, Du Q. Huynh\textsuperscript{b,∗}, Xuezhi Wang\textsuperscript{a}, Michael Edwards\textsuperscript{c}, Andrew Harris\textsuperscript{c}, Barbara F. La Scala\textsuperscript{a}

\textsuperscript{a}Department of Electrical and Electronic Engineering
The University of Melbourne, Australia

\textsuperscript{b}School of Computer Science and Software Engineering
The University of Western Australia, Australia

\textsuperscript{c}Scantech, Camden Park SA 5038, Australia

\textbf{Abstract}

We describe here a method for the analysis of materials on a conveyor belt using the natural gamma spectra collected with a BGO (Bismuth Germanate) gamma ray detector. This detector collects gamma ray emissions from the Potassium (K), Uranium (U), and Thorium (Th) atoms in the materials. Based on these data, and using a Poisson model for the data generation, a statistical model is proposed and an approximate maximum likelihood (ML) technique based on the expectation-maximization (EM) algorithm is then used to estimate the amount of each of the three elements in the material. The statistical model is further refined to incorporate parameters of drift in the detector and an estimation technique for this is developed and tested against real data. The Cramér-Rao Lower Bounds for the the estimators are calculated.

\textbf{Key words:} Poisson processes, Maximum likelihood estimation, Expectation-maximization, Statistical Signal Processing, Gamma-ray detectors.

\textsuperscript{∗}Corresponding author:
Du Q. Huynh (du@csse.uwa.edu.au)

Manuscript submitted to Digital Signal Processing March 10, 2009
1. Introduction

We describe, and compare with more conventional techniques, a method for analysis of minerals on conveyor belts using naturally occurring gamma ray emissions. The natural gamma analyzer is essentially a Bismuth-Germanate (BGO) detector housed in a temperature controlled lead box beneath a conveyor belt, with a lead shield above the belt (see Fig. 1). The lead serves to shield the detector from terrestrial and cosmic gamma radiation, which would otherwise add noise to the spectrum. The detector is connected to a digital multi-channel analyzer, which sorts the measured gamma rays into a 1024 channel spectrum by energy. For a coal application, the analyzer provides ash (non-combustible mineral matter) content typically every 2 to 15 minutes without contacting the coal or requiring the addition of sealed radioactive sources. This technique relies on strong linear correlations between concentrations of K, U and Th and ash in coal, and leads to a calibration equation for predicting ash content of coal. Thus it is useful to have a good estimation of the three elements.

In our application, gamma ray spectra are collected from a BGO detector with 1024 channels linearly spaced across energies up to 3 MeV. To maintain spectrum stability, the voltage applied to the detector is adjusted at the start of each sample collection. Analysis is based on the prior collection of elemental spectra for K (from potassium chloride), U (from torbenite rock) and Th (from a gas mantle), which we refer to as library spectra. These 3 chemical elements are the only ones to emit gamma radiation naturally [2]; that is without excitation by, for example, a neutron source. Fig. 2 shows the library spectra of these 3 elements. Counts using this technique are relatively low compared to active methods where emissions are stimulated using a neutron source. It is important therefore to have a method capable of producing good estimates from small counts. The method proposed here is applied to samples taken over a period of two minutes, and the results compare (favourably) with linear regression results for the same samples taken
over 15 minutes. While in our experiments the samples were static, in a real situation the time taken to collect results forces averaging over a considerable amount of minerals. The capability provided by our method of using shorter collection times while maintaining accuracy of estimation yields more localized information about the content of the minerals.

Figure 2: Library spectra for K, U, and Th. Each spectrum has a very distinctive signature: Spectrum K has a prominent peak in the channel interval [420,520] where spectra U and K are flat; spectra U and Th have unique peaks in the [530,630] and [750,950] intervals.

In addition to estimating these amounts, we demonstrate in this paper that the drift of the channels in the detectors can also be estimated. Conventional calibration methods simply sum the counts in wide windows around each major peak in the spectrum. Slight detector drifts will cause the peaks to move within these windows and so are assumed to have only a small impact on the analyzer accuracy. Of course, large detector drifts which cause these peaks to move partly or wholly outside the windows cause very inaccurate analysis. Because there is no ubiquitous reference peak in the spectrum, automatic adjustment of detector voltage to compensate for detector drift is not normally implemented. Over time, the gain of the photo-multiplier tube of the detector changes. This is minimized by controlling the detector temperature, but electronics drift is inevitable. The main deficiency of the window approach is that only parts of the spectrum are employed, which increases errors compared with a method which involves consideration of the entire spectrum. A calibration method for the detector that uses the entire spectrum will permit the use of a smaller, less expensive detector to achieve the same accuracy.

Here we describe a maximum likelihood (ML) approach using a Poisson model for channel counts (see [17] for justification of this model). One reason to do this is that in the natural gamma ray approach to mineral analysis, relatively few gamma rays are collected and so the usual statistical approximations and convergences to the mean associated with large numbers of samples are ineffective in these circumstances. We remark that a more realistic and
better defined model would take account of the known Gaussian shape of the peaks and the theoretically known effects of the Compton scattering on the spectra. We hope to discuss this refinement in a later paper.

A straightforward ML approach gives equations which appear to be difficult to solve directly, so instead we resort to an expectation-maximization (EM) approximation technique. The inclusion of a model for drift further complicates the calculations so that even the full EM approach becomes difficult. Instead we employ the formulae given by EM for estimation of the amounts of the elements, but a simpler method for drift estimation using a search for the drift parameters based on the Levenberg-Marquardt variant [11, 6, 15] of gradient descent. This is justified because drift parameters are usually within a very limited range and change slowly over time.

In the next section we describe the basic model for statistical estimation without drift and then in the following section how an EM approach can be used to provide a close to ML estimator. Then a modified model incorporating drift is described and an estimation algorithm is derived. In Section 5, the Fisher Information matrix is calculated leading to the Cramér-Rao lower bounds on the variance of the estimators, and simulations are given showing how closely the technique approaches these bounds. Finally, we describe results using both simulated and real data. A preliminary account of this work was published in ICASSP 2008 [14].

2. Problem Formulation

In this section we provide a model for data generation that ignores drift in the measured gamma ray spectrum over time. The three spectra for K, Th, and U, the only primary elements naturally emitting gamma radiation, as functions of the channels, are denoted by \( S^K \), \( S^U \), \( S^\text{Th} \) or more conveniently for writing formulae as \( S^i(c) \), for \( i = 1, \cdots, 3 \) and \( c = 1, \cdots, n_c \), where \( n_c \) denotes the number of channels (in our case, \( n_c = 1024 \)). These spectra are assumed known, since they can be measured in advance of the real-time estimation process. We write \( b \in \mathbb{R}^{n_c} \) for a measured datum from a material boat; that is, a spectrum from a mineral sample; and denote the counts of \( b \) at channel \( c \) by \( b(c) \). We write \( \alpha^K, \alpha^U, \alpha^\text{Th} \) for the amounts of potassium, uranium and thorium present in the sample. We use \( \alpha = (\alpha^1, \alpha^2, \alpha^3)^T \) to represent these three amounts when convenient.

A simple (and currently used) approach to the estimation of \( \alpha^K, \alpha^U, \alpha^\text{Th} \) involves a linear regression:

\[
\sum_{i=1}^{3} \alpha^i S^i = b.
\]  

This is, of course, a massively over-determined system and typically, in current practice, either a least-squares approach or a peak area windowing method is taken to solving it [16, 5, 7]. Our approach more closely models the physical processes generating the data and appears to provide significant improvement over the conventional approach.

We assume that the number of counts in the \( c^{th} \) channel is a Poisson random variable \( X(c) \), and that these random variables are independent both between channels and between elements. We calculate the likelihood in a form that will facilitate, at least approximately, ML estimation (see, e.g., [10]) of the parameters.
The likelihood of a given datum $x$ is then
\[
\Lambda_X(x) = P \left( X(c) = b(c) \text{ counts in the } c^{th} \text{ channel}, \forall c \right)
\]
\[
= \prod_{c=1}^{n_c} P \left( X(c) = b(c) \text{ counts in the } c^{th} \text{ channel} \right)
\]
\[
= \prod_{c=1}^{n_c} \sum_{b_K(c)=b_U(c)=b_{Th}(c)} \frac{1}{b_K(c)! b_U(c)! b_{Th}(c)!} \cdot \left( \alpha^K_S(c) \right)^{b_K(c)} e^{-\alpha^K_S(c)} \left( \alpha^U_S(c) \right)^{b_U(c)} e^{-\alpha^U_S(c)} \left( \alpha^{Th}_{S_{Th}}(c) \right)^{b_{Th}(c)} e^{-\alpha^{Th}_{S_{Th}}(c)}
\]
\[
= e^{-\sum_{c=1}^{n_c} \sum_{i=1}^{3} \alpha^i_S(c)} \prod_{c=1}^{n_c} \sum_{b(c)} \frac{1}{b(c)!} \left( \sum_{i=1}^{3} \alpha^i_S(c) \right)^{b(c)}
\]
which is a product of Poisson distributions with means $\alpha^K_S(c) + \alpha^U_S(c) + \alpha^{Th}_{S_{Th}}(c)$. Here, $b_K(c)$, $b_U(c)$, $b_{Th}(c)$, for $c = 1, \cdots, n_c$, are variables in the sum, i.e.,
\[
b^K(c) + b^{Th}(c) + b^{Th}(c) = b(c), \quad \forall c,
\]
and $\alpha^K$, $\alpha^U$, and $\alpha^{Th}$ are the only three unknown parameters. Throughout the paper, we use $\{b^K(c), b^U(c), b^{Th}(c)\}$ and $\{b^1(c), b^2(c), b^3(c)\}$ interchangeably.

Our justification for independence is as follows: each separate atom gives rise to its own independent decay chain. While there are dependencies within this chain because decay from a lower excited state is dependent on a previous decay from a higher excited state, it is unlikely in the context of large numbers of atoms of these materials being present in the sample that the analyzers will receive gamma rays from different parts of the same decay chain. Decays from different atoms, however, are statistically independent. As further evidence for independence, Fig. 3 shows the results of a correlation analysis for some coal data, where the inter-channel correlations are clearly negligible.
The log-likelihood is therefore
\[
\log \Lambda_{X}(a^K, a^U, a^\text{Th}) = -\sum_{c=1}^{n} \sum_{i=1}^{3} a^i S^i(c) + \sum_{c=1}^{n} \left[ b(c) \log \left( \sum_{i=1}^{3} a^i S^i(c) \right) - \log (b(c)!) \right].
\] (4)

Our aim is, given the counts in each channel, to choose values of these parameters that maximize the likelihood. The direct approach to this (differentiation with respect to the unknowns) gives the following equations for the ML estimator:
\[
\sum_{c=1}^{n} S^i(c) = \frac{\sum_{c=1}^{n} b(c) S^i(c)}{\sum_{j=1}^{3} a^j S^j(c)}, \quad \text{for } i = 1, 2, 3. \] (5)

Solution of these equations for the unknowns \( a^i \), for \( i = 1, 2, 3 \), directly appears to be difficult. We note first a straightforward calculation shows that the maximum likelihood estimator of a product of Poisson random variables is unbiased.

3. Expectation-Maximization Approach

The EM method provides an approximation to the ML solution of an estimation problem. We refer the reader to [13, 4] for a discussion of the EM technique and to [8, 9] for an example of EM for a multivariate Poisson distribution. The EM method works by assuming that there are measurements for certain “hidden variables”. The likelihood is computed on the basis of these hidden data and the ML solution obtained. By appropriate choice of the hidden variables, this technique can produce a more tractable approach than the direct ML method.

In our case the hidden data are the unknown counts \( b^K(c), b^U(c), b^\text{Th}(c) \) (this is the so-called \textit{complete data} in the EM literature). The distribution of these complete data is:

\[
\Lambda_{Y}(\alpha) = P \left( Y_c = (b^K(c), b^U(c), b^\text{Th}(c)) \text{ counts in } c^{th} \text{ channel, } \forall c \right) = \prod_{c=1}^{n} \left( e^{-\sum_{i=1}^{3} \alpha^i S^i(c)} \frac{\prod_{j=1}^{3} (a^j S^j(c))^{b(c)}}{\prod_{j=1}^{3} b(j)!} \right),
\] (6)

and the log-likelihood of which is

\[
\log \Lambda_{Y}(\alpha) = \sum_{c=1}^{n} \sum_{i=1}^{3} \left( -a^i S^i(c) - \log (b(c)!) \right) + b(c) \log (a^i S^i(c)). \] (7)

The EM method consists of two steps:

- **The Expectation step** (or the E-step). The conditional expectation of the log-likelihood with respect to the actual (incomplete) data \( X \) is computed.

- **The Maximization step** (or the M-step). Based on the imputed data obtained from the E-step, the unknown parameters, \( a^K, a^U, \) and \( a^\text{Th} \) are estimated to maximize the log-likelihood.

The result of one EM step will give an estimate whose likelihood is at least as large as its predecessor [13]. Typically the EM process is iterated several times.
Now we compute the conditional expectation of the log-likelihood with respect to the actual (incomplete) data \( X \), for an initial guess \( \alpha_0 = (\alpha^K_0, \alpha^U_0, \alpha^{Th}_0)^\top \) of the parameters. In other words, we calculate

\[
E\left[ \log \Lambda_Y(\alpha^K, \alpha^U, \alpha^{Th}) | X, \alpha_0 \right] = \sum_{c=1}^{n} \left\{ -\sum_{j=1}^{3} \alpha'S'(c) - E \left[ \log(b^K(c) \cdot b^U(c) \cdot b^{Th}(c)) | X, \alpha_0 \right] + \sum_{j=1}^{3} E \left[ b'(c) | X, \alpha_0 \right] \log(\alpha'S'(c)) \right\}.
\]

(8)

Now we note that the term \( E \left[ \log(b^K(c) \cdot b^U(c) \cdot b^{Th}(c)) | X, \alpha_0 \right] \) does not involve the variables, \( \alpha^K, \alpha^U, \) and \( \alpha^{Th}, \) that we are maximizing over, and so we can discard them. Thus we only calculate

\[
\sum_{c=1}^{n} \sum_{i=1}^{3} (-\alpha'S'(c) + E \left[ b'(c) | X, \alpha_0 \right] \log(\alpha'S'(c)) ),
\]

and to do this it is enough to calculate \( E \left[ b^K(c) | X, \alpha_0 \right] \) and the other terms of this kind. This is a simple calculation with the trinomial distribution, which is the conditional distribution of a triple of Poisson random variables given their sum. So we have

\[
b'_j(c) = E \left[ b'(c) | X, \alpha_0 \right] = \frac{\alpha'_0 S'(c)}{\sum_{j=1}^{3} \alpha'_j S'(c)} X_j = \frac{\sum_{j=1}^{3} \alpha'_j S'(c)}{\sum_{j=1}^{3} \alpha'_j S'(c)} b(c), \quad \text{for } i = 1, 2, 3,
\]

(9)

where the subscript denotes the iteration number.

At the M-step, we attempt to estimate \( \alpha^K, \alpha^U, \) and \( \alpha^{Th} \) such that the log-likelihood function is maximized. Simple differentiation yields that the point at which the log-likelihood function in (7) achieves its maximum is

\[
\alpha' = \frac{b'_j(c)}{\sum_{c=1}^{n} S'(c)} \frac{\sum_{c=1}^{n} E \left[ b'(c) | X, \alpha_0 \right]}{\sum_{c=1}^{n} S'(c)}, \quad \text{for } i = 1, 2, 3.
\]

(10)

Combining this with equation (9) we obtain an estimate of \( \alpha'_n \) at iteration \( n \) based on the values from the previous iteration:

\[
\alpha'_n = \frac{n}{\sum_{c=1}^{n} \sum_{j=1}^{3} \alpha'_{n-1} S'(c) b(c)} \frac{\sum_{c=1}^{n} E \left[ b'(c) | X, \alpha_0 \right]}{\sum_{c=1}^{n} S'(c)}, \quad \text{for } i = 1, 2, 3.
\]

(11)

In fact, then, we can merge the E-step and M-step. The EM method yields the following iterative scheme for calculation of the parameters \( \alpha^K, \alpha^U, \alpha^{Th} \):

EM-1) Take the counts \( \{b(c) | c = 1, \cdots, n_c\} \);

EM-2) Guess a value for \( \alpha'_i \), for \( i = 1, 2, 3 \) — this might be based on a simple use of estimates of the mean for the Poisson distributions;

EM-3) Use the formulae (11) to iteratively find \( \alpha'_n \), for \( i = 1, 2, 3 \), at the \( n \)th iteration.

EM-4) In accordance with the EM theory this iteration never decreases the likelihood, and typically converges to the maximum likelihood solution.
We note that Eqs. (11) correspond to an iterative approximation algorithm for solving Eqs. (5).

4. Incorporating drift

At this point we incorporate drift of the detectors into the model. Empirical evidence suggests that, to a reasonable approximation, drift is linear in the channels. That is, the data originally going into channel \( c \) now enters channel \( \beta c + \gamma \), where \( \beta \) and \( \gamma \) are two new unknown parameters that are independent of the channel. Occurrences of \( b'(c) \) are now substituted by \( b'(\beta c + \gamma) \). This problem of drift is due to the instrumental instability of the detectors, resulting in a gain and threshold shift on the actual channels that collect the counts (see [16, Secs. 6.4 and 10.1]). It is convenient to use a spline interpolation of the channel counts so that the channel number becomes a continuous variable. This facilitates rescaling of channel numbers by amounts that do not adhere to the rigid integer count constraint. In any case, sampling of the energy levels by the sensors leads naturally to approximations of this kind. Let \( \beta_0 \) and \( \gamma_0 \) be the initial estimates of drift. The log-likelihood function given in (7) becomes:

\[
\log \Lambda_Y(\alpha, \beta, \gamma) = \sum_{i=1}^{3} \sum_{c=1}^{n} \left( -\alpha' S'(c) - \log(b'(\beta c + \gamma)) + b'(\beta c + \gamma) \log(\alpha' S'(c)) \right).
\]

(12)

EM calculations for the estimation of \( \alpha \) yield

\[
\alpha' = \frac{\sum_{c=1}^{n} \mathbb{E} \left[ b'(\beta c + \gamma) | X, \alpha_0, \beta_0, \gamma_0 \right]}{\sum_{c=1}^{n} S'(c)},
\]

(13)

where

\[
\mathbb{E} \left[ b'(\beta c + \gamma) | X, \alpha_0, \beta_0, \gamma_0 \right] = \frac{\alpha_0 S'(\tau(c))}{\sum_{j=1}^{3} \alpha_0 S'(\tau(j))} \beta(\tau(c)),
\]

(14)

\[
\tau(c) = \frac{\beta c + \gamma - \gamma_0}{\beta_0}.
\]

(15)

The conditional expectation of the log-likelihood function, with respect to these initial estimates and the data, becomes:

\[
\mathbb{E} \left[ \log \Lambda_Y(\alpha, \beta, \gamma | X, \alpha_0, \beta_0, \gamma_0) \right] = -\sum_{i=1}^{3} \alpha' \int S'(c) dc - \sum_{i=1}^{3} \mathbb{E} \left[ \log \left( b'(\beta c + \gamma) | X, \alpha_0, \beta_0, \gamma_0 \right) \right] \int S'(c) dc + \sum_{i=1}^{3} \mathbb{E} \left[ b'(\beta c + \gamma) | X, \alpha_0, \beta_0, \gamma_0 \right] \log(\alpha' S'(c)) \int S'(c) dc.
\]

(16)

We seek \( \alpha, \beta, \) and \( \gamma \) such that the above expectation is maximized. This gives an estimator for \( \alpha' \) as follows:

\[
\alpha' = \frac{\int \mathbb{E} \left[ b'(\beta c + \gamma) | X, \alpha_0, \beta_0, \gamma_0 \right] \int S'(c) dc}{\int S'(c) dc},
\]

(17)

where

\[
\mathbb{E} \left[ b'(\beta c + \gamma) | X, \alpha_0, \beta_0, \gamma_0 \right] = \frac{\alpha_0 S'(\frac{\beta c + \gamma - \gamma_0}{\beta_0})}{\sum_{j=1}^{3} \alpha_0 S'(\frac{\beta c + \gamma - \gamma_0}{\beta_0})}.
\]

(18)
It remains to find estimators of the drift parameters. These parameters must satisfy
\[
\frac{\partial}{\partial \beta} \mathbb{E} [\log \Lambda_Y(\alpha, \beta, \gamma | X, \alpha_0, \beta_0, \gamma_0)] = 0 \quad \text{and}
\frac{\partial}{\partial \gamma} \mathbb{E} [\log \Lambda_Y(\alpha, \beta, \gamma | X, \alpha_0, \beta_0, \gamma_0)] = 0.
\]
The full derivation of them is given in Appendix A.

In principle, these estimators can be solved to find new estimates of the unknown parameters \(\alpha, \beta, \gamma\). Unfortunately, there appears to be no simple way to do that, and numerical algorithms would require numerical calculation of the integrals involved. This is not a computationally effective process. In consequence, we resort to a simpler approach which in the circumstances appears to yield good results. We retain the formulae for the EM estimators for the amounts of materials given in the preceding section but simplify the estimation of the drift parameters \(\beta\) and \(\gamma\). This is done next.

4.1. Simplified Drift Parameter Estimation

As we have seen, calculation of drift parameters via the EM algorithm is difficult, and we adopt a simplified approach. With the detectors regularly calibrated, the drift involved is quite small, but not insignificant enough to be neglected, as our experiments with real data will show. We propose a standard optimization approach, the Levenberg-Marquardt (LM) method [6, 15], to iteratively refine the drift parameters \(\beta\) and \(\gamma\), using the initial estimates \(\beta_0 = 1\) and \(\gamma_0 = 0\).

Let \(d = (\beta, \gamma)^\top\) and define
\[
\begin{align*}
    f(c; d) &= b'(\beta c + \gamma) - \sum_{i=1}^{3} \alpha^i S^i(c) \\
    f(d) &= (f(1; d), \cdots, f(n_c; d))^\top.
\end{align*}
\]

We design the following objective function to be minimized by LM:
\[
\min_{\beta, \gamma} f^\top(\beta, \gamma) f(\beta, \gamma).
\]

The LM optimization process is carried out in an alternating fashion with the EM approach for the estimation of \(\alpha\). The entire procedure of our algorithm is thus:

1. Set \(d = (\beta_0, \gamma_0)^\top = (1, 0)^\top\).
2. Use the current estimate \(d\) for the drift parameters, apply the EM algorithm to estimate \(\alpha\), i.e., the amounts for K, U, and Th.
3. By using the current estimate of \(\alpha\), apply the LM steps above to refine \(d\).
4. Go back to step 2.
We found that the above procedure needs only to be iterated 4 times to obtain good alignments of the peaks of the spectra. We call this algorithm of estimating \( \alpha \) and \( d \) in an alternate fashion the EM-LM algorithm.

5. Cramér-Rao Lower Bound

The Fisher Information Matrix provides information about the limits on the accuracy of any unbiased estimator of a parameter [10]. Let \( \mathbf{I}(\alpha) \) be the Fisher Information Matrix:

\[
\mathbf{I}(\alpha)_{ij} = -\mathbb{E}_X \left[ \frac{\partial^2 \log \Lambda_X(\alpha)}{\partial \alpha_i \partial \alpha_j} \right].
\]

(22)

Let \( \mathbf{C} \) be the covariance matrix of an unbiased estimator \( \hat{\alpha} \) for \( \alpha \). Then

\[
\mathbf{C} - \mathbf{I}(\alpha)^{-1} \geq 0,
\]

(23)

where this is interpreted as meaning that the left hand side is positive semi-definite. Since the diagonal terms of a positive semi-definite matrix are always non-negative, the inequality (23) yields

\[
\text{var}(\alpha_i) \geq \left( \mathbf{I}(\alpha)^{-1} \right)_{ii},
\]

(24)

though in general the cross terms contain information also. The term \( \mathbf{I}(\alpha)^{-1} \) shown in (24) is the Cramér-Rao lower bound for \( \alpha \). For more information about this bound, see, e.g., [12, 10].

In our case, the log-likelihood is given in equation (4). Hence,

\[
\frac{\partial \log \Lambda_X(\alpha)}{\partial \alpha_j} = -\sum_{c=1}^n \left( S_j(c) - \frac{b(c)S_j(c)}{\sum_{i=1}^3 \alpha_i S_i(c)} \right)
\]

(25)

and

\[
\frac{\partial^2 \log \Lambda_X(\alpha)}{\partial \alpha_i \partial \alpha_k} = -\sum_{c=1}^n \frac{b(c)S_j(c)S_k(c)}{\left( \sum_{i=1}^3 \alpha_i S_i(c) \right)^2}.
\]

(26)

It follows that the elements of the Fisher information matrix with respect to \( \alpha \) are

\[
\mathbf{I}_{jk} = -\mathbb{E}_X \left[ \frac{\partial^2 \log \Lambda_X(\alpha)}{\partial \alpha_j \partial \alpha_k} \right] = \sum_{c=1}^n \mathbb{E}[b(c)] \frac{S_j(c)S_k(c)}{\left( \sum_{i=1}^3 \alpha_i S_i(c) \right)^2} = \sum_{c=1}^n \frac{\left( \sum_{i=1}^3 \alpha_i S_i(c) \right)S_j(c)S_k(c)}{\left( \sum_{i=1}^3 \alpha_i S_i(c) \right)^2}
\]

\[
= \sum_{c=1}^n \frac{S_j(c)S_k(c)}{\sum_{i=1}^3 \alpha_i S_i(c)}, \quad \text{for } j, k = 1, 2, 3.
\]

(27)

The above derivation does not take into account detector drift. Using the same linear model of detector drift as described in Section 4, the channel number \( c \) is replaced by \( \beta c + \gamma \) for the \( b \) spectrum in the log-likelihood function (4). This leads to a Fisher information matrix \( \mathbf{I}(\alpha, \beta, \gamma) \) whose elements with respect to \( \alpha \) are the same as Eq. (27) given
above. By differentiating the log-likelihood function w.r.t. \( \beta \) twice, we obtain

\[
\frac{\partial^2}{\partial \beta^2} \log(\Lambda_X(\alpha, \beta, \gamma)) = \sum_{c=1}^n c^2 \left\{ \log \left( \sum_{i=1}^3 \alpha_i S^i(c) \right) - \Psi(b(\beta c + \gamma) + 1) \right\} b''(\beta c + \gamma) - \Psi_1(b(\beta c + \gamma) + 1)(b'(\beta c + \gamma))^2 \right\},
\]

(28)

where \( b' \) and \( b'' \) denote the first and second derivatives of \( b \), and

\[
\Psi(z) \triangleq \frac{\partial}{\partial z} \log(\Gamma(z)) = \frac{\partial}{\partial z} \log((z - 1)!) \\
\Psi_1(z) \triangleq \frac{\partial^2}{\partial z^2} \log(\Gamma(z)) = \frac{\partial^2}{\partial z^2} \log((z - 1)!) \\

\]

are, respectively, the digamma function and the trigamma function [1].

Similarly, we can derive that

\[
\frac{\partial^2}{\partial \gamma^2} \log(\Lambda_X(\alpha, \beta, \gamma)) = \sum_{c=1}^n \left\{ \log \left( \sum_{i=1}^3 \alpha_i S^i(c) \right) - \Psi(b(\beta c + \gamma) + 1) \right\} b''(\beta c + \gamma) - \Psi_1(b(\beta c + \gamma) + 1)(b'(\beta c + \gamma))^2 \right\},
\]

(29)

By definition, the diagonal elements of the Fisher information matrix for the parameters \( \beta \) and \( \gamma \) are

\[
I_{4,4} \triangleq -E \left[ \frac{\partial^2}{\partial \beta^2} \log \Lambda_X(\alpha, \beta, \gamma) \right] \\
I_{5,5} \triangleq -E \left[ \frac{\partial^2}{\partial \gamma^2} \log \Lambda_X(\alpha, \beta, \gamma) \right].
\]

(30)

(31)

Substitution of (28) and (29) into the above equations yields

\[
I_{4,4} = \sum_{c=1}^n c^2 \left\{ \Xi_1(c, \beta, \gamma) + \Xi_2(c, \beta, \gamma) - \log \left( \sum_{i=1}^3 \alpha_i S^i(c) \right) E \left[ b''(\beta c + \gamma) \right] \right\}
\]

(32)

\[
I_{5,5} = \sum_{c=1}^n c^2 \left\{ \Xi_1(c, \beta, \gamma) + \Xi_2(c, \beta, \gamma) - \log \left( \sum_{i=1}^3 \alpha_i S^i(c) \right) E \left[ b''(\beta c + \gamma) \right] \right\}
\]

(33)

where

\[
\Xi_1(c, \beta, \gamma) = E \left[ \Psi(b(\beta c + \gamma) + 1)b''(\beta c + \gamma) \right] \\
\Xi_2(c, \beta, \gamma) = E \left[ \Psi_1(b(\beta c + \gamma) + 1)b'^2(\beta c + \gamma) \right].
\]

(34)

(35)

We have used some approximations to calculate these terms in the next section. In these approximations the

\[\text{Non-diagonal elements can be similarly derived. We omit this here.}\]
derivatives of \( b \) are approximated using finite differences. So

\[
b'(bc + \gamma) \approx b(bc + \gamma + 1) - b(bc + \gamma)
\]  
(36)

\[
E[b'(bc + \gamma)] \approx \sum_{i=1}^{3} a_i \left( S'(c+1) - S'(c) \right)
\]  
(37)

\[
E[b''(bc + \gamma)] \approx \sum_{i=1}^{3} a_i \left( S'(c+1) - 2S'(c) + S'(c-1) \right)
\]  
(38)

Also, the expressions for \( \Xi_1 \) and \( \Xi_2 \) defined in (34)-(35) cannot be evaluated in closed form. In the simulation study, we approximated them using Monte Carlo simulation. That is, the expectations were approximated using a sample mean via \( N \) samples:

\[
\Xi_1(c, \beta, \gamma) \approx \frac{1}{N} \sum_{j=1}^{N} \Psi \left( b_j'(bc + \gamma + 1) \right) b_j''(bc + \gamma)
\]  
(39)

\[
\Xi_2(c, \beta, \gamma) \approx \frac{1}{N} \sum_{j=1}^{N} \Psi_1 \left( b_j'(bc + \gamma + 1) \right) b_j''(bc + \gamma)
\]  
(40)

and each sample \( b_j \) was drawn from the Poisson distribution described by the parameters \( \alpha, S, \beta \) and \( \gamma \):

\[
b_j(c) \sim \text{Poisson}(\cdot | \alpha, S, \beta, \gamma),
\]  
(41)

for \( j = 1, \ldots, N \) and \( c = 1, \ldots, n_c \).

In practice, the detector drift parameter \( \beta \) is very close to 1 (e.g., 1.001) and parameter \( \gamma \) is very small (e.g., in the range \([0, 5]\)). For such drift parameter values, the computed Cramér-Rao lower bounds are close to zero. This is consistent with our intuition as small changes in the values of both \( \beta \) and \( \gamma \) do not significantly affect the spectra, though as our results show incorporation of these parameters in the estimation process does significantly improve accuracy of estimation of \( \alpha \) (see Table 1).

6. Results

An independent chemical analysis of the samples used is not readily available to produce “true” values for the \( \alpha \) parameters or for the drift parameters for our real data. As a result, we have tested the algorithm in the following two ways:

1. Application to synthetically generated data based on the library spectra and prescribed values for the \( \alpha, \beta, \) and \( \gamma \) parameters.

2. Application to collections of real data, where performance is measured in two ways:
   - by fidelity of the reconstructed spectra using the estimated parameters.
   - by comparison with estimates using measurements taken over much longer time intervals, where counts are sufficiently high that a linear regression technique can reasonably be used to make the estimation.
We compared the accuracy of the EM approach for estimating \( \alpha \), with the technique which involves replacement of the EM step (i.e., step 2) of the EM-LM algorithm by the linear least squares solution given by (1). We term this algorithm the LLS-LM algorithm.

6.1. Simulated Data

In the simulations, a number of spectra \( b \) were drawn from a Poisson distribution with the mean parameter being \( \sum_{i=1}^{3} \alpha_i S^i(c) \) for each channel \( c = 1, \cdots, 1024 \). A drift of channel counts from \( c \) to \( \beta c + \gamma \) was simulated using the known ground truth amounts \( \hat{\alpha} \) and the ground truth drift parameters \( \hat{\beta} \) and \( \hat{\gamma} \). Both the EM-LM and LLS-LM algorithms were applied to estimate \( \alpha, \beta \) and \( \gamma \). For the LM part of both algorithms, 4 iterations were imposed.

The relative root-mean-squared (RRMS) errors of these estimated parameters were computed and compared with the derived Cramér-Rao lower bounds (CRLBs).

Two cases are presented here. In the first case, we consider the situation of zero drift (i.e., \( \hat{\beta} = 1 \) and \( \hat{\gamma} = 0 \)) and \( \hat{\alpha}_K = \hat{\alpha}_U = \hat{\alpha}_Th = \hat{\alpha} \), where the values of \( \hat{\alpha} \) varied from 0.01 to 25. At each sampling point, \( N = 500 \) simulations were performed and the RRMS errors of \( \alpha \) defined as

\[
\frac{\sqrt{\sum_{i=1}^{N} (\hat{\alpha}_i - \bar{\alpha}_i)^2}}{\bar{\alpha}_i}, \quad \text{for } i = 1, 2, 3,
\]

were computed. As the Cramér-Rao lower bound is the lowest bound of variance that can be achieved, to compare it with the RRMS errors defined above, we need to apply the square-root function and division in a similar fashion. In the simulations, the quantities \( \sqrt{I(\alpha, \beta, \gamma)^{-1}_{ij}} \sqrt{\bar{\alpha}_i} \), for \( i = 1, 2, 3 \), were used and the simulation results are shown in Fig. 4.

In the second case, drifts from the detector were simulated. A constant drift (\( \hat{\beta} = 1.002 \) and \( \hat{\gamma} = 4 \)) was assumed throughout the entire simulation. The same range of \( \hat{\alpha} \) values as in the previous case was used here. The RRMS errors of \( \alpha \) were computed as given in (42) and the simulation results are shown in Fig. 5. We also illustrate in Fig. 6 the RRMS of \( \beta \) and \( \gamma \) against, respectively, the quantities \( \sqrt{I(\alpha, \beta, \gamma)^{-1}_{ij}} \sqrt{\bar{\beta}} \) and \( \sqrt{I(\alpha, \beta, \gamma)^{-1}_{55}} \sqrt{\bar{\gamma}} \).

Comments on Simulation Results

1. The computed CRLBs for \( \hat{\alpha} \) with and without drifts from the detector are almost identical, suggesting that the drift is adequately compensated for by the estimation process.

2. It is observed that the EM-LM algorithm consistently gave smaller RRMS errors of \( \alpha_K, \alpha_U \), and \( \alpha_Th \) than did the LLS-LM algorithm. In particular, the RRMS errors of \( \alpha_K \) almost coincide with the blue curve for the Cramér-Rao lower bound (see Fig. 5). This is as expected as an inspection of the library spectra given in Fig. 2 reveals that the count rate of K at its peak, i.e., channels [420, 520], is almost three times as high as the peaks at the respective channels [530, 630] and [750, 950] of U and Th. If we were to decrease \( \hat{\alpha}_K \) proportionally in our simulations then the curves for the three elements would be identical. For the necessary and sufficient conditions for the achievability of CRLB for the Poisson distribution, we refer the interested reader to [3].
We chose to present Fig. 5 (similarly for Figs. 4 and 6) with the vertical axes of all subplots in (base 10) logarithmic scale so that the three curves for LLS-LM, EM-LM, and CRLB are more distinguishable. Because of the logarithmic scale, although it appears that, as $\tilde{\alpha} \to \infty$, the red curves in the subplots for $\alpha^U$ and $\alpha^{Th}$ in Fig. 5 maintain a fixed distance from the blue solid curves, the RRMS errors of $\alpha$ from the EM-LM algorithm actually approach zero on the linear scale in the same way as the blue curves.

3. The EM-LM algorithm has much smaller overall RRMS errors compared to those from the LLS-LM algorithm, especially when detector drifts are present. We note that the data generated in these simulations were based on the Poisson distribution, which corresponds to the model captured in the EM-LM algorithm.

4. As in comment 2, the vertical axis of each subplot of Fig. 6 is shown in logarithmic scale, so that while the RRMS errors from the EM-LM algorithm for both drift parameters, $\beta$ and $\gamma$, over the interval $[10, 25]$ of $\tilde{\alpha}$ seem to be flat, these errors actually become smaller and closer to the CRLBs as $\tilde{\alpha} \to \infty$. The poor performance of the LLS-LM algorithm in estimating these parameters is evident from both figures. This reflects the robustness of the EM-LM approach in the estimation of drift. Large relative errors are observed in the interval where $\tilde{\alpha}$ was smaller. This may be explained by the fact that, when $\tilde{\alpha}$ is smaller, the randomness of relative errors dominates the underlying true spectrum counts, making it more difficult to distinguish the true counts from channel to channel.

5. In simulations, we also looked at the cases where the values of $\alpha^K$, $\alpha^U$, and $\alpha^{Th}$ are in different ratios. Although the values of RRMS errors and CRLBs may vary, the results consistently agree with what is summarized above and have the same performance trends as those in Figs. 4-6.

To demonstrate that the incorporation of drift estimation significantly improves the accuracy of estimates of $\alpha$, results are presented from 3 sets of 100 tests with synthetic data, simulating the cases of high, medium, and low channel counts. In each test, the input boat spectrum was firstly passed to the LLS and EM algorithms, which estimated the amounts $\alpha$ using, respectively, linear least-squares and EM while ignoring the drift of the detector. The EM-LM and LLS-LM algorithms (including the LM optimization step for estimating the drift parameters) were then applied. To compare the performances of these algorithms, the following error measures were computed: the mean (denoted by $\epsilon$) and standard deviation (denoted by $\sigma$) percentage errors of $\alpha$ and $d$ from the known ground truth values in the simulation.

Table 1 shows the means and standard deviations of the percentage errors of $\alpha$ and $d$ from the algorithms mentioned above. It is evident from the table that EM outperformed LLS for all the 3 datasets. When LM was incorporated to estimate the drift parameters, EM-LM outperformed all the other three algorithms for all cases. Our simulation results demonstrate that the EM algorithm alone (i.e. ignoring detector drift) yields only sub-optimal results. By incorporating the estimation of drift parameters, the EM-LM algorithm yields more accurate estimates of the amounts $\alpha$. In addition, we found that using LLS to estimate the amounts $\alpha$ when these entities are small often leads to negative
Figure 4: Test results of simulations with zero drift (i.e., $\bar{\beta} = 1$ and $\bar{\gamma} = 0$): Comparison of the relative root-mean-squared (RRMS) errors of $\hat{\alpha}_K$, $\hat{\alpha}_U$, and $\hat{\alpha}_{Th}$ from the EM-LM and LLS-LM algorithms versus the computed Cramér-Rao lower bounds. We set $\bar{\alpha}_K = \bar{\alpha}_U = \bar{\alpha}_{Th} = \bar{\alpha} \in (0, 25]$ in the simulations. All the plots have the vertical axes in logarithmic scale.

Table 1: Comparison of the LLS, EM, LLS-LM and EM-LM algorithms. Sets 1, 2, and 3, respectively, correspond to the datasets of high, medium, and low channel counts of $b$. The variables $\epsilon_\alpha$ and $\epsilon_d$ denote the percentage errors of the $\alpha$ and drift values for K, U, and Th.
Figure 5: Test results of simulations with $\bar{\beta} = 1.002$ and $\bar{\gamma} = 4$: Comparison of the relative root-mean-squared (RRMS) errors of $\hat{\alpha}^K$, $\hat{\alpha}^U$, and $\hat{\alpha}^{Th}$ from the EM-LM and LLS-LM algorithms versus the computed Cramèr-Rao lower bounds. We set $\bar{\alpha}^K = \bar{\alpha}^U = \bar{\alpha}^{Th} = \bar{\alpha} \in (0, 25]$ in the simulations. All the plots have the vertical axes in logarithmic scale.
Figure 6: Test results of simulation with $\bar{\beta} = 1.002$ and $\bar{\gamma} = 4$: Comparison of the relative root-mean-squared (RRMS) errors of $\hat{\beta}$ and $\hat{\gamma}$ from the EM-LM and LLS-LM algorithms versus the computed Cramèr-Rao lower bounds, with the true amounts $\bar{\alpha}^{K} = \bar{\alpha}^{L} = \bar{\alpha}^{Th} = \bar{\alpha} \in (0, 25]$. All the plots have the vertical axes in logarithmic scale.
amounts, which is non-physical.

6.2. Estimation for Real Data

We tested our EM-LM algorithm on many real spectra varying from coal to iron ore. Two of the real experiments are presented here. In the first experiment, the measured spectrum from the material boat was from an iron ore dataset; in the second experiment, the datum was from a coal dataset.

In Fig. 7(a), the red curve shows the input boat spectrum from an iron ore dataset and the blue curve shows the reconstructed boat spectrum using the library spectra \( S_i | i = 1, \ldots, 3 \) and the amounts \( \alpha \) estimated from EM without taking drift into account. The two curves clearly misalign. Fig. 7(b) shows a much better reconstructed boat spectrum after drift of the detector is corrected using EM-LM. Note that, because of the different \( \alpha \) values, the green curves in Fig. 7(a) and (b) are different even though visually they look almost identical. To distinguish these two cases in Fig. 7, we use \( \alpha \) and \( \alpha_\# \) to denote the estimated amounts with and without drift correction incorporated. Similarly, \( b \) and \( b_\# \) denote, respectively, the input and the drift corrected boat spectra. In Fig. 7(c), we can see that the drift corrected boat spectrum (the blue curve) is slightly to the left of the original boat spectrum (the red curve). The two error curves, \( b - \sum_{i=1}^{3} \alpha S_i \) (red curve, with no drift correction) and \( b_\# - \sum_{i=1}^{3} \alpha_\# S_i \) (blue curve, with drift correction), are shown in Fig. 7(d). The quantity \((e - e_\#)/e\), where \( e \) and \( e_\# \) are, respectively, the total errors from all channels for these red and blue curves, can be used as a measure of error reduction when drift estimation is incorporated into the algorithm. In this experiment, the reduction of error in channel count differences after drift correction was 34%.

Fig. 8 shows the testing of EM-LM on a sample spectrum from a coal dataset. The results are similar as the previous figure. The reduction of error in channel count difference after drift correction is 15% in this experiment. Note that in both experiments \( \sum_{i=1}^{3} \alpha \ll 1 \) as the boat spectra were collected in less than 15 minutes duration.

As mentioned before, it is impossible to obtain the ground truth of \( \alpha \). Neither is it possible to obtain the ground truth of the drift parameters, \( d \). However, our experiments confirm that the reconstructed boat spectra coincide much better with the library spectra when drift correction is incorporated into the algorithm. From our tests on synthetic data, the EM-LM outperforms the the linear least squares method. In an attempt to achieve something close to ground truth, we have compared results obtained using our method for data collected over 2 minutes with results for the same sample using 15 minute data and applying linear regression. The longer collects of data result in counts that are acceptable for the application of linear regression methods since, by the law of large numbers, the ratios of counts approach the ratios of the means of the count rates.

7. Conclusion

We have described an estimation method for on-belt analysis of materials using their natural gamma spectra. The technique involves the use of a Poisson model for the generation of the spectra, based on the experimentally determined library spectra for each of these elements. We have incorporated parameters in the model to describe the drift of the detector performance over time. We have described the ML Estimation problem for all of the parameters,
With no drift correction: $\alpha = (0.1314, 0.0863, 0.0928)$

With drift correction: $\alpha_# = (0.1336, 0.1092, 0.0699)$

Figure 7: Results of the EM and EM-LM algorithms on the EM020911 boat spectrum from an iron ore dataset. (a) Original input boat spectrum, $b$, and the reconstruction, $\sum_{i=1}^{3} \alpha_i S_i$ (using EM only); (b) Drift corrected boat spectrum, $b_#$, and the reconstruction, $\sum_{i=1}^{3} \alpha_i # S_i$ (using EM-LM); (c) The input and drift corrected boat spectra; (d) $b - \sum_{i=1}^{3} \alpha_i S_i$ and $b_# - \sum_{i=1}^{3} \alpha_i # S_i$. 
With no drift correction: \( \alpha = (0.1418, 0.0225, 0.0643) \)

\[ \sum_{i=1}^{3} \alpha_i S_i \]

With drift correction: \( \alpha_# = (0.1531, 0.0324, 0.0496) \)

\[ \sum_{i=1}^{3} \alpha'_i S'_i \]

Figure 8: Results of the EM and EM-LM algorithms on the STCKCP boat spectrum from a coal dataset. (a) Original input boat spectrum, \( b \), and the reconstruction, \( \sum_{i=1}^{3} \alpha_i S_i \) (using EM only); (b) Drift corrected boat spectrum, \( b_# \), and the reconstruction, \( \sum_{i=1}^{3} \alpha'_i S'_i \) (using EM-LM); (c) The input and drift corrected boat spectra; (d) \( b - \sum_{i=1}^{3} \alpha_i S_i \) and \( b_# - \sum_{i=1}^{3} \alpha'_i S'_i \).
its approximation using an EM approach, and finally a technique in which EM is used to estimate the amount of each element present, but employs a more direct approach to the estimation of the detector drift parameters.

The last approach has been implemented on both synthetically generated and real data and been shown, insofar as tests available to us will permit, to be significantly superior to the more conventional approach using a linear regression to estimate the amount of each material present.

While we believe that the results obtained here represent a significant advance on previous techniques used in this form of analysis, this paper is a report on ongoing work. We are continuing to investigate alternative and better methods for the overall estimation process associated with on-belt gamma ray spectrometric methods of elemental analysis and this work will be reported elsewhere.

Appendix A. Derivation of the EM estimators for $\beta$ and $\gamma$

In this appendix, we show the derivation of the EM estimators for $\beta$ and $\gamma$. As will be seen later, while the analytical form can be derived, it is difficult to implement in practice.

Let

$$
\psi_i'(\beta, \gamma | X, \alpha_0, \beta_0, \gamma_0) \triangleq \mathbb{E} \left[ \log(b(i(\beta c + \gamma)!) | X, \alpha_0, \beta_0, \gamma_0) \right],
$$

and observe that $\psi_i'(\beta, \gamma | X, \alpha_0, \beta_0, \gamma_0)$ can be written in terms of $\beta$ and $\gamma$, as given below:

$$
\psi_i'(\beta, \gamma | X, \alpha_0, \beta_0, \gamma_0) \triangleq \mathbb{E} \left[ \log(b(i(\beta c + \gamma)!) | X, \alpha_0, \beta_0, \gamma_0) \right] = \sum_{r=0}^{n(c)} \log(r!) \frac{n(c)!}{r!(n(c) - r)!} \left( \frac{\bar{S}^i(c)}{\beta_0} \right)^r \left( \sum_{j \neq i} \frac{\bar{S}^j(c)}{\beta_0} \right)^{n(c) - r}, \tag{43}
$$

where

$$
n(c) = b(\beta_0 c + \gamma_0) \tag{44}
$$

and

$$
\bar{S}^i(c) \triangleq \frac{\alpha_i^0 S^i(c)}{\sum_{j=1}^{3} \alpha_j^0 S^j(c)}. \tag{45}
$$

We apply the following change of variables. Let $\tilde{c} = \left( \frac{\beta c + \gamma - \gamma_0}{\beta_0} \right)$. Then $d\tilde{c} = \left( \frac{\beta}{\beta_0} \right) dc$ and

$$
\int \psi_i'(\beta, \gamma | X, \alpha_0, \beta_0, \gamma_0) dc = \frac{\beta_0}{\beta} \int \sum_{r=0}^{n(c)} \log(r!) \frac{n(c)!}{r!(n(c) - r)!} \left( \frac{\bar{S}^i(c)}{\beta_0} \right)^r \left( \sum_{j \neq i} \frac{\bar{S}^j(c)}{\beta_0} \right)^{n(c) - r} d\tilde{c} \triangleq \frac{\beta_0}{\beta} \Delta', \tag{46}
$$

where $\Delta'$ is defined to be the integral term on the right hand side.
Similarly, let
\[ \sigma_{c}^{*}(\beta, \gamma | X, \alpha_{0}, \beta_{0}, \gamma_{0}) \triangleq \mathbb{E} \left[ \nu(\beta c + \gamma) | X, \alpha_{0}, \beta_{0}, \gamma_{0} \right] = \tilde{S} \left( \frac{\beta c + \gamma - \gamma_{0}}{\beta_{0}} \right) \]

Then
\[ \int \sigma_{c}^{*}(\beta, \gamma | X, \alpha_{0}, \beta_{0}, \gamma_{0}) \log(\alpha' S'(c))dc = \int \tilde{S} \left( \frac{\beta c + \gamma - \gamma_{0}}{\beta_{0}} \right) \log(\alpha' S'(c))dc 
= \frac{\beta_{0}}{\beta} \int \tilde{S}(\tilde{c})b(\tilde{c}) \log(\alpha' S'((\beta_{0}\tilde{c} - \gamma + \gamma_{0})/\beta)d\tilde{c}, \]
where a change of variable from \( c \) to \( \tilde{c} \) has been applied.

The conditional expectation of the log-likelihood function now becomes:
\[ \mathbb{E} \left[ \log \Lambda_{Y}(\alpha, \beta, \gamma | X, \alpha_{0}, \beta_{0}, \gamma_{0}) \right] = \sum_{i=1}^{3} \alpha' \int S'(c)dc - \sum_{i=1}^{3} \frac{\beta_{0}}{\beta} \Lambda' 
+ \frac{\beta_{0}}{\beta} \sum_{i=1}^{3} \int \tilde{S}(\tilde{c})b(\tilde{c}) \log(\alpha' S'((\beta_{0}\tilde{c} - \gamma + \gamma_{0})/\beta)d\tilde{c}. \] (49)

Differentiation of (49) with respect to \( \beta \) and setting to 0 yields
\[ \frac{\partial}{\partial \beta} \mathbb{E} \left[ \log \Lambda_{Y}(\alpha, \beta, \gamma | X, \alpha_{0}, \beta_{0}, \gamma_{0}) \right] = 0 \]
\[ \Rightarrow \sum_{i=1}^{3} \left[ \beta \Lambda' - \beta \int \tilde{S}(\tilde{c})b(\tilde{c}) \log(\alpha' S'((\beta_{0}\tilde{c} - \gamma + \gamma_{0})/\beta)d\tilde{c} - \int (\beta_{0}\tilde{c} - \gamma + \gamma_{0})S'(\tilde{c})b(\tilde{c}) \frac{S'((\beta_{0}\tilde{c} - \gamma + \gamma_{0})/\beta)}{S'((\beta_{0}\tilde{c} - \gamma + \gamma_{0})/\beta)}d\tilde{c} \right] = 0, \] (50)

Differentiation of (49) with respect to \( \gamma \) and setting to 0 yields
\[ \frac{\partial}{\partial \gamma} \mathbb{E} \left[ \log \Lambda_{Y}(\alpha, \beta, \gamma | X, \alpha_{0}, \beta_{0}, \gamma_{0}) \right] = 0 \]
\[ \Rightarrow \sum_{i=1}^{3} \int \tilde{S}(\tilde{c})b(\tilde{c}) \frac{S'((\beta_{0}\tilde{c} - \gamma + \gamma_{0})/\beta)}{S'((\beta_{0}\tilde{c} - \gamma + \gamma_{0})/\beta)}d\tilde{c} = 0. \] (51)

Eqs. (50) and (51) show the conditions that the parameters \( \beta \) and \( \gamma \) must satisfy for the EM updates, given the initial estimates \( \alpha_{0}, \beta_{0}, \) and \( \gamma_{0} \). These equations for \( \beta \) and \( \gamma \) are difficult to solve, even numerically, and so we have resorted to the Levenberg-Marquardt approach described in the main text.

Acknowledgements

This research work was supported by a Linkage Project grant (LP0562337) funded by the Australian Research Council. The authors would like to thank the anonymous reviewers for their suggestions for improving the manuscript.

References

Bill Moran is the Research Director of Melbourne Systems Laboratory (MSL), a research institute in the Department of Electrical and Electronic Engineering at The University of Melbourne, Australia. He was elected to the Fellowship of the Australian Academy of Science in 1984. He holds a Ph.D. in Mathematics from the University of Sheffield, UK ('68).

He has been a Principal Investigator on numerous research grants and contracts, in areas spanning pure mathematics to radar development, from both Australian and US Research Funding Agencies, including DARPA, AFOSR, AFRL, Australian Research Council (ARC), Australian Department of Education, Science and Training, DSTO. He is a currently a member of the Australian Research Council College of Experts.

His main areas of research are in signal processing, both theoretically and in applications to radar, waveform design, and radar theory, sensor networks, and sensor management. He also works in various areas of mathematics including harmonic analysis, representation theory, and number theory.

Du Q. Huynh is a Senior Lecturer at the School of Computer Science and Software Engineering, The University of Western Australia. She obtained her PhD in Computer Vision in 1994 at the same university. Since then, she has worked for the Australian Cooperative Research Centre for Sensor Signal and Information Processing (CSSIP) and Murdoch University. She has been a visiting scholar at Chinese University of Hong Kong, Malmö University, Gunma University, and The University of Melbourne. Dr. Huynh is currently a recipient of a Discovery Project grant and two Linkage Project grants funded by the Australian Research Council. Her research interests include shape from motion,
multiview geometry, video image processing, visual tracking, and signal processing.

Xuezhi Wang received the B. S. degree in avionics from Northwest Polytechnical University, Xian, China in 1982 and the PhD degree in Signal and Systems from the Department of Electrical & Electronic Engineering, The University of Melbourne, Australia in 2001.

He has been involved in radar and sonar signal processing research and development since 1983. He had been working with the Cooperative Research Center for sensor signal and information processing (CSSIP), Australia from 2001 to 2005. He is now a research fellow with the Melbourne Systems Laboratory, in the department of Electrical & Electronic Engineering, The University of Melbourne, Australia. He has worked as a key player on many industrial contracts in the area of target tracking, data fusion and sensor network issues for defense research agencies both in Australia and USA since 2001. His research interests are in stochastic signal processing, information theory, Bayesian estimation, data fusion and Situation assessment.

Barbara La Scala received her PhD in Systems Engineering from the Australian National University in 1994. She is a quantitative analyst with the National Australia Bank. Before then she worked with The Melbourne Systems Laboratory in the Electrical and Electronic Engineering Department at the University of Melbourne, Australia. MSL performs basic and applied research on optimization and control of distributed systems, including multi-sensor and multi-target tracking systems. She was Assistant Editor-in-Chief of the Journal of Advances in Information Fusion (2006-2008), and an Associate Editor of IEEE Transactions on Aerospace and Electronic Systems for papers on target tracking and multi-sensor systems (2003-2008).