This is a preprint, the final version is subject to change, of the American Mineralogist (MSA)
Cite as Authors (Year) Title. American Mineralogist, in press.
DOI: https://doi.org/10.2138/am-2018-6152

The inverse problem of unpolarised infrared spectroscopy of geological materials: Estimation from noisy random sampling of a quadratic form

Andrew Jackson[1], Robert L. Parker[2], Malcolm Sambridge[3], Catherine Constable[2] & Aaron S. Wolf[4]

1. Institute for Geophysics, ETH Zurich, Switzerland.
2. Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, USA.
3. Research School of Earth Sciences, Australian National University, Canberra Australia.
4. Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, MI.

Abstract

We address the problem of unpolarised light spectroscopy of geological materials. Using infrared radiation, the aim of this technique is to learn about the absorbing species, such as hydroxyl. The use of unoriented samples leads to the need to perform a rigorous statistical analysis, so that the three principal absorbances of the crystal can be retrieved. We present here such an analysis based on a derivation of the probability density function for a single random measurement. Previous methods for retrieval of the absorbances are shown to be suboptimal, producing biased results that are sometimes even unphysical (e.g. negative estimates for an inherently positive quantity). The mathematical structure of the problem is developed in order to use the maximum likelihood estimation method, and we show how to optimise for the three absorbance parameters. This leads to good parameter retrieval on both synthetic and real data sets.

Introduction

In the analysis of geological samples using both polarised and unpolarised infrared radiation to determine principal absorbances, there is a logical desire to implement procedures based on unpolarised light spectroscopy. In a transmission geometry, the intensity of light is measured after passage through a (possibly) birefringent crystal; when polarised light is used, the physics of the experiment is clear and allows unambiguous determination of the absorbance as the sum of the principal polarised absorbances along each of the principal axes of the absorbance indicatrix. Unfortunately, the use of unpolarised light presents a much less clear physical problem, but remains scientifically invaluable since it enables the analysis of small unoriented mineral chips that might not be otherwise measured. Regardless of the technique employed, the aim is to obtain quantitative information on absorbance, which can help constrain the presence of hydroxyl in the sample.

The unpolarised implementation is simpler than the more protracted method whereby polarised light is used, but the technique has not received widespread acceptance because of arguments that it is not possible to use it to obtain quantitative determinations of absorbance in anisotropic materials (e.g. Libowitzky & Rossman (1996); Bell et al. (2003)). However, Sambridge et al. (2008) derived, from first principles, a simple relationship between transmittance and the direction, and polarisation angle of incident light. This led to a theory for unpolarised transmittance and also allowed quantification of the conditions under which approximate formulae for unpolarised absorbance can be applied, namely that the maximum linear unpolarised absorbance should not exceed 0.15. The latter does not appear to have been widely appreciated and led some to continue to claim a controversy over which theory is correct (Withers, 2013). In fact, the primary difference remains not which theory is correct but rather which quantity should be treated in analysis, namely average unpolarised absorbance (Sambridge et al., 2008; Kovács et al., 2008) or average unpolarised transmittance (Withers, 2013); fortunately we are able to neatly sidestep this issue in our presentation, as the methods that we develop will apply equally to either of the competing approaches, as do those of Sambridge et al. (2008). Instead, the focus of our paper is concerned with how to optimally treat the data that are collected in the unpolarised scenario when unoriented samples are used. This question of optimal estimation must be answered using a correct statistical analysis that has, so far, been lacking from the literature. We begin with a description of the physics and mathematics of both polarised and unpolarised spectroscopy, and then go on to develop the statistical treatment appropriate for random unoriented measurements in unpolarised light. At the heart of our development is the aim to determine the three principal absorbances of the mineral, and we show that previous heuristic procedures for determining these quantities are suboptimal. We illustrate the use of our theory with analysis of two data sets, the first synthetic (in which case the true answer is known) and the second real (measurements on olivine from the literature).

Theory

Relative to the axes of the optical indicatrix, we let A_a,A_b and A_c signify the principal absorbances of a (possibly) birefringent crystal. In an isotropic crystal or, when looking along one of the principal directions, the transmission geometry of typical experiments, the ratio of light measured (I) to that of the source (I_0) is

$$T=\frac{I}{I_0} \tag{1}$$

and the absorption $A=-\log T$. It is well-known (Libowitzky & Rossman, 1996) that measurements of polarised light along the principal axes are sufficient to determine the absorbance through

$$A=A_{\text{tot}}=A_a+A_b+A_c \tag{2}$$

and indeed, it is sufficient to make measurements within three mutually perpendicular unoriented sections (subscripted 1,2,3) where the maximum and minimum values A^i_{max}, A^i_{min} are observed; the absorbance is then

$$A=\sum_{i=1}^{3}(A^i_{\text{max}}+A^i_{\text{min}})/2. \tag{3}$$

The difficulty of producing truly polarised light is one of the reasons for the interest in the experimental implementation of a protocol using unpolarised light.

We work in standard spherical polar coordinates where φ is the angle between the radial and the z axis, and ψ is the angle between the x-axis and the projection of r on the x-y plane. When unpolarised light is used, the absorbance measured for an incidence angle (ψ,φ) is either

$$A_{\text{unpol}}=\frac{1}{2}\left\{A_a(\cos^2\varphi\cos^2\psi+\sin^2\psi)+A_b(\cos^2\varphi\sin^2\psi+\cos^2\psi)+A_c\sin^2\varphi\right\} \tag{4}$$

or

$$T_{\text{unpol}}=\frac{1}{2}\left\{T_a(\cos^2\varphi\cos^2\psi+\sin^2\psi)+T_b(\cos^2\varphi\sin^2\psi+\cos^2\psi)+T_c\sin^2\varphi\right\} \tag{5}$$

depending whether one adopts the practice of using absorbance (Sambridge et al., 2008) or using transmission (Libowitzky & Rossman, 1996; Withers, 2013). We avoid entering that particular debate by noting that the theory that we shall develop applies equally well to either type of observation, but for concreteness we simply use the "A" notation. Should the reader wish to adopt the transmission-based "T" theory, this can be done with a trivial interchange of variable names.

Sambridge et al. (2008) noted that (4) could be written as

$$A_{\text{unpol}}=\frac{1}{2}(A_b+A_c)\sin^2\varphi\cos^2\psi+\frac{1}{2}(A_a+A_c)\sin^2\varphi\sin^2\psi+\frac{1}{2}(A_a+A_b)\cos^2\varphi \tag{6}$$

or, in Cartesian coordinates (x_1,x_2,x_3), as a quadratic form Q that is central to our paper,

$$Q(x_1,x_2,x_3)=\sum_{i=1}^{3}a_i x_i^2 \tag{7}$$

where the point (x_1,x_2,x_3) lies on the unit sphere ($|\mathbf{x}|=1$) and the semi-axes of the quadratic form are

$$a_1 = \frac{1}{2}(A_b+A_c) \tag{8}$$

$$a_2 = \frac{1}{2}(A_a+A_c) \tag{9}$$

$$a_3 = \frac{1}{2}(A_a+A_b) \quad (10)$$

which comprise the vector we will call **a**, derived from the triplet $\mathbf{A}=(A_a,A_b,A_c)$. At the heart of our paper is a methodology for the determination of (a_1,a_2,a_3) from experimental data. Having found these parameters, we can clearly then determine

$$A_a = a_2+a_3-a_1 \quad (11)$$

$$A_b = a_1+a_3-a_2 \quad (12)$$

$$A_c = a_1+a_2-a_3. \quad (13)$$

We begin with a few remarks on the nature of the quadratic form (7). Assuming, without loss of generality, that $a_3 \le a_2 \le a_1$ (implying $A_c \ge A_b \ge A_a$), we note that the maximum and minimum values that are measured at any angle can be at most (least) a_1 (a_3), so the maximum and minimum of the data set $A_{\max}^{\text{obs}}$ and $A_{\min}^{\text{obs}}$ are estimators for these values. We should be aware, however, that these estimates are biased toward less extreme values as a result of incomplete sampling. A typical dataset is shown in Figure (1), where this is apparent. To find the third unknown, Sambridge et al. (2008) made the observation that, in the case of perfect data measured uniformly over the surface of the sphere, one would find

$$E\{Q\}=A_{\text{avg}}^{\text{obs}}=\frac{1}{3}(a_1+a_2+a_3)=\frac{1}{3}(A_a+A_b+A_c) \quad (14)$$

where $E\{\}$ signifies expectation. There is a nagging doubt as to whether this intuitively appealing estimator is likely to be an efficient one in the case of small datasets, as it is unlikely that one would achieve uniform coverage of all solid angles, but it does at least serve to close the problem. One can deduce from integrations that the variance σ^2 of this estimate is

$$\sigma^2=E\{Q^2\}-E\{Q\}^2=\frac{2}{45}\left\{(a_1-a_2)^2+(a_2-a_3)^2+(a_1-a_3)^2\right\} \quad (15)$$

so that, provided the principal absorbances are not drastically different, when the data set is large, the mean is a rather precise quantity.

We now have three estimates from which to try to derive the principal absorbances: straightforward algebra shows that (Sambridge et al., 2008)

$$A_a = 3A_{\text{avg}}^{\text{obs}}-2A_{\max}^{\text{obs}} \quad (16)$$

$$A_b = 2(A_{\min}^{\text{obs}}+A_{\max}^{\text{obs}})-3A_{\text{avg}}^{\text{obs}} \quad (17)$$

$$A_c = 3A_{\text{avg}}^{\text{obs}}-2A_{\min}^{\text{obs}}. \quad (18)$$

Although these estimators may work reasonably well in some cases, their success is highly dataset-dependent, and may even produce unphysical results if either $A_{\max}^{\text{obs}}>3/2A_{\text{avg}}^{\text{obs}}$ or $A_{\text{avg}}^{\text{obs}}>2/3(A_{\min}^{\text{obs}}+A_{\max}^{\text{obs}})$, which yield meaningless negative estimates of A_a or A_b, respectively. This can occur if the anisotropy is relatively strong.

As a motivating example, we analyse a synthetic dataset of representative size for typical laboratory analysis: for random points on the sphere, we synthesize N=23 values of Q; see Table 1. For the synthesis we use the principal values $(A_a,A_b,A_c)=(1.89,45.6,80.94)$ which dictates that $(a_1,a_2,a_3)=(23.745,41.415,63.27)$. From the dataset we see that $\min(Q)=25.063$ and $\max(Q)=62.4398$, giving estimators $\hat{a}_3$ and $\hat{a}_1$. The mean of the data is 40.676 and therefore we deduce the following estimates for the principal absorbances:

$$A_a = -2.85 \tag{19}$$

$$A_b = 52.98 \tag{20}$$

$$A_c = 71.90. \tag{21}$$

The latter two estimators are good to about 10% but, unfortunately, for this small dataset, positivity of A_a is not respected, and there is no obvious fix for the estimators. In the case that $N\to\infty$ none of these shortcomings exist.

A resolution to these possible issues of bias and disrespect of positivity is to properly derive the probability density function (pdf), or equivalently the cumulative density function (cdf), for the value of an unpolarised observation taken at a random orientation. We do this in very general fashion in the next section. In the final section we apply the theory to both synthetic and real datasets, and illustrate its efficacy.

The distributions of *Q*

Here we derive closed-form expressions for the probability distribution function and cumulative distribution function of Q in (7) that are needed for our estimation procedures. Imagine contouring Q on the unit sphere. Now we ask what is the probability that a random value, say Q, is less than Q_0? Since the directional distribution is uniform, that must be the fraction of the area covered by the set of points on the sphere where $Q \le Q_0$; that is

$$F(Q_0)=P(Q\le Q_0)= \frac{1}{4\pi} \int\limits_{Q(\mathbf{r})\le Q_0} d^2\mathbf{r} \,. \tag{22}$$

The boundary of region of integration in equation (22) is the contour line $C(Q_0)$ and the integral covers values below the contour level. The function F in (22) is by definition *the cumulative distribution function* (cdf), the indefinite integral of the pdf. Figure (2) shows the level lines of Q on the sphere.

We begin by obtaining an equation for the contour lines of constant Q. We will adopt spherical polar coordinates with colatitude Θ and longitude φ and $\Theta=0$ is the point of minimum Q, so that a_3 is the smallest coefficient: recall that for definiteness we take: $a_3\le a_2\le a_1$ throughout. Close to this point, contours tend to be circles centered on the north pole, see Figure (2). We will initially assume that Q_0 is close enough to a_3 that the line $C(Q_0)$ is a smooth ring around the pole. We write (7) in polar coordinates

$$Q=a_1\cos^2\varphi\sin^2\Theta + a_2\sin^2\varphi\sin^2\Theta + a_3\cos^2\Theta \,. \tag{23}$$

Then on the contour line $Q=Q_0\ge a_3$ a little algebra shows

$$\cos\Theta=c(\varphi)= \sqrt{\frac{a_1\cos^2\varphi + a_2\sin^2\varphi - Q_0}{a_1\cos^2\varphi + a_2\sin^2\varphi - a_3}} \,. \tag{24}$$

There are two separate regions where $Q\le Q_0$, one in the northern hemisphere, the other symmetrically in the south. We will treat the northern region, and double its area. That area is given by

$$2\pi F(Q_0) = \int_0^{2\pi} d\varphi \int_0^{\cos^{-1}c(\varphi)} \sin\Theta\, d\Theta = \int_0^{2\pi} [1-c(\varphi)]\, d\varphi$$

$$= \int_0^{2\pi} \left[1 - \sqrt{\frac{a_1\cos^2\varphi + a_2\sin^2\varphi - Q_0}{a_1\cos^2\varphi + a_2\sin^2\varphi - a_3}}\right] d\varphi \tag{25}$$

and thus

This is a preprint, the final version is subject to change, of the American Mineralogist (MSA)
Cite as Authors (Year) Title. American Mineralogist, in press.
DOI: https://doi.org/10.2138/am-2018-6152

$$F(Q_0)=1-\frac{2}{\pi}\int_0^{\frac{\pi}{2}}\sqrt{\frac{a_1+a_2-2Q_0+(a_1-a_2)\cos 2\varphi}{a_1+a_2-2a_3+(a_1-a_2)\cos 2\varphi}}\,d\varphi\,. \tag{26}$$

The numerator under the square root is nonnegative on the interval $(0,2\pi)$ provided that $Q_0 \le a_2$, but not otherwise. When $Q_0 > a_2$ the level line $C(Q_0)$ forms two closed loops around the maximum principal axis, not the north pole, and then the integral in (25) must be performed by breaking the domain into subintervals. A much simpler solution is to choose a new coordinate system, with the maximum value at the north pole, resulting in a similar calculation to before. We can summarize the results: the cdf is

$$F(Q_0)=\begin{cases}1-\frac{2}{\pi}\int_0^{\frac{\pi}{2}}\sqrt{\frac{a_1+a_2-2Q_0+(a_1-a_2)\cos 2\varphi}{a_1+a_2-2a_3+(a_1-a_2)\cos 2\varphi}}\,d\varphi; & Q_0\le a_2\\ \frac{2}{\pi}\int_0^{\frac{\pi}{2}}\sqrt{\frac{2Q_0-a_2-a_3+(a_2-a_3)\cos 2\varphi}{2a_1-a_2-a_3+(a_2-a_3)\cos 2\varphi}}\,d\varphi; & Q_0>a_2\,.\end{cases} \tag{27}$$

In our calculations we shall make great use of this cdf and therefore we give the analytic forms for the integrals. We begin by defining the auxiliary function G where

$$G(c,b,p)=2(c-b)\sqrt{\frac{1}{(p+1)(b+c)}}\,\Pi\left(\frac{2b}{b+c},\frac{2(b+cp)}{(b+c)(p+1)}\right) \tag{28}$$

where

$$\Pi(n,k)=\int_0^{\pi/2}\frac{d\Theta}{(1-n\sin^2\Theta)\sqrt{1-k\sin^2\Theta}} \tag{29}$$

is a complete elliptic function of the third kind. Then the cdf F is given by

$$F(Q)=\begin{cases}0 & Q\le a_3\\ 1 & Q\ge a_1\\ \frac{2}{\pi}\tan^{-1}\left(\sqrt{\frac{a_2-a_3}{a_1-a_2}}\right) & Q=a_2\\ \frac{1}{\pi\sqrt{2a_1-a_2-a_3}}G\left(-a_2-a_3+2Q;a_3-a_2;\frac{a_2-a_3}{2a_1-a_2-a_3}\right) & Q>a_2\\ 1-\frac{1}{\pi\sqrt{a_1+a_2-2a_3}}G\left(a_1+a_2-2Q;a_2-a_1;\frac{a_1-a_2}{a_1+a_2-2a_3}\right) & Q<a_2\,.\end{cases} \tag{30}$$

We also need to evaluate the pdf Φ in terms of known special functions. To find the pdf we differentiate:

$$\Phi=\frac{dF}{dQ}. \tag{31}$$

After laborious manipulations we arrive at the definitive expressions for the pdf of Q_0

This is a preprint, the final version is subject to change, of the American Mineralogist (MSA)
Cite as Authors (Year) Title. American Mineralogist, in press.
DOI: https://doi.org/10.2138/am-2018-6152

$$\Phi(Q_0)=\begin{cases} \dfrac{1}{\pi\sqrt{(a_2-a_3)(a_1-Q_0)}}\,K\left(\dfrac{(a_1-a_2)(Q_0-a_3)}{(a_2-a_3)(a_1-Q_0)}\right); & Q_0\leq a_2 \\ \dfrac{1}{\pi\sqrt{(a_1-a_2)(Q_0-a_3)}}\,K\left(\dfrac{(a_2-a_3)(a_1-Q_0)}{(a_1-a_2)(Q_0-a_3)}\right); & a_2<Q_0 \end{cases} \tag{32}$$

where $K(m)$ is the complete elliptic integral of the first kind defined by

$$K(m)=\int_0^{\frac{\pi}{2}} \frac{dx}{\sqrt{1-m\,\sin^2 x}}\,. \tag{33}$$

See, for example, Chapter 17 of Abramowitz & Stegun (1970). A typical curve is plotted in Figure (3), showing the presence of an integrable logarithmic singularity at $Q_0=a_2$.

Estimation

Exact data

The work horse of estimation is the maximum likelihood estimator (see, for example, Rice (2007)). If we assume the observations to be statistically independent then we get a likelihood function by multiplying together all the individual pdfs evaluated at the data values:

$$L(a_1,a_2,a_3)=\prod_j \Phi(Q_j)\,. \tag{34}$$

Recalling that **a** is the vector with unknowns (a_1,a_2,a_3), the maximum likelihood method requires that one view the likelihood as a function of **a** given the data, the converse to Figure (3) that plots the probability of Q_0 given values of **a**. Each of the contributing $\Phi(Q_j)$ is regular for all a_1 and a_3 but there is a singularity in each at $a_2=Q_j$. Unfortunately the presence of the logarithmic singularity at $Q=a_2$ means that for a data set consisting of N values, the likelihood possesses N singularities at which the likelihood is infinite; thus maximising the likelihood is fruitless since there is no single unique maximum. Figure (4) plots the likelihood for the test dataset given in Table 1; we assume that the values a_1 and a_3 are known and plot (34) as a function of a_2.

We reach the surprising conclusion that, *in the case of precise data*, the estimation problem cannot be solved by the method of maximum likelihood! The loss of the maximum likelihood method is a severe blow in this case, and we have consequently sought recourse to other methods based on properties of the cdf, that we do not report. The more important case of inaccurate data is instead analysed, in which it transpires that the presence of errors in the data do allow the maximum likelihood method to be implemented.

Inexact data

Given an exact value Q_0 and a measured value $Q=Q_0+e$, let $p(e)$ be the pdf of the errors. The laws of probability give that

$$p(Q|Q_0,\mathbf{a})=p(e)*p(Q_0|\mathbf{a}) \tag{35}$$

where * indicates convolution. In order to derive a concrete form for the pdf in this case, we assume that the errors on the data are uniformly distributed between ±Δ (a so-called top-hat function), so that

$$p_{\mathrm{u}}(e|\Delta)=\begin{cases} (2\Delta)^{-1} & |e|\leq\Delta \\ 0 & |e|>\Delta\,. \end{cases} \tag{36}$$

This choice of error distribution (which is a matter of mathematical convenience) now allows us to discover a pdf of the observations that is free of logarithmic singularities:

$$p(Q|\mathbf{a},\Delta)=(2\Delta)^{-1}\left[F(Q+\Delta)-F(Q-\Delta)\right] \tag{37}$$

where we recall that F is the cdf of Q. We may now plot the pdf of Q for a given choice of Δ; Figure (5) gives several examples. Note that there is now finite probability of an observation having a value in the range $a_3-\Delta$ or in $a_1+\Delta$ as a result of the incorporation of the error model (recall in everything we do we have $a_1>a_2\geq a_3$).

Equation (37) now forms the basis for creating the likelihood L for inexact data that will be used for all the subsequent calculations we present:

$$L(a_1,a_2,a_3)=\prod_j p(Q_j|\mathbf{a},\Delta). \tag{38}$$

Depending on the values of the data, the number of data N and particularly Δ, the pdf may have multiple local maxima and it is theoretically possible for two local maxima to have the same value. Such a circumstance is highly unlikely as $N\gg 1$, but one could imagine a scenario with e.g. four data in which two maxima with the same likelihood exist. The beauty of the maximum likelihood method is that it exposes such unlikely situations without the need for any special treatment.

We should remark that (37) allows other assumed pdfs for the observational noise, $p(e)$, to be treated very accurately, should the user so wish. Any symmetric distribution (e.g. a Gaussian) can be approximated by a superposition of top-hat functions:

$$p(e)=\sum_i c_i\, p_{\mathrm{u}}(e|\Delta_i) \tag{39}$$

for some coefficients c_i, and the linearity of the convolution operation means that the effect on the exact pdf can be determined by a superposition of results of the form (37). We do not pursue this avenue, and doubt whether there will be situations in which the user will have full information on the appropriate error model.

Positivity

We note from (11)–(13) that positivity of the principal absorbances generates constraints on the allowed values of **a**. Noting, as we have throughout, that $a_1\geq a_2\geq a_3$, we have the following inequalities:

$$a_1-a_2\leq a_3\leq a_2 \tag{40}$$

and

$$a_2\leq a_1\leq a_2+a_3. \tag{41}$$

The allowed region of parameter space for given a_2 is shown in Figure (6). We use this in the following way: we scan over all a_2 using $0.9\min(Q_{\mathrm{obs}})$ and $1.1\max(Q_{\mathrm{obs}})$ as boundaries; then for a given a_2 we scan over a_1 and a_3 whilst respecting the inequalities (40) and (41).

Convergence using a synthetic dataset

We test the theory using a synthetic data set that we constructed in the following way: data were synthesised from the model **A**=(1.89,45.6,80.94) by evaluating equation (7) for randomly generated **x**:|**x**|=1. To these data we added noise synthesised from random samples from a uniform distribution with Δ=1. We then analysed the data in two ways, firstly using the algorithm of Sambridge et al. (2008) that applies to exact data as $N\to\infty$, and secondly using the probabilistic approach described herein. The results are shown in Figure (7) when the search is performed with a resolution of 0.04 in all three directions. We see that the maximum likelihood estimate converges to the known value, with superior performance when the dataset is in the range 10−100 samples. We find that one can generally determine the total absorbance to better than 2% accuracy with 20 samples at this Δ=1 error level, which is not the case for the estimator $3A^{\mathrm{obs}}_{\mathrm{avg}}$ (equation (14)). Determination of the intermediate principal absorbance, which is generally the most difficult, can be achieved at the 5% level with a few tens of samples. The final maximum likelihood values from this N=1000 dataset are **A**=(1.86,46.3,80.4) compared to the known input (1.89,45.6,80.9). The total absorbances are 128.6 (estimated) versus 128.4 (truth).

Chips from a natural sample

Here we analyse data from natural "Pakistani" olivine samples (exact provenance unknown) reported in Kovács et al. (2008). Forty one samples were analysed with unpolarised light and the absorbances in the 3500-3630 cm^{-1} waveband normalised to 1cm thickness. As in any inverse problem, we need to assign a value for the accuracy of the measurements, and in our particular problem this plays an important role, as it modulates the pdf of the observations. It is not straightforward to assign the errors. On the one hand it can be argued that the accuracy of the theory that we use to analyse our data is good to better than 10%, and this could be the limiting factor in analysing the data (Sambridge et al., 2008). Conversely, a careful analysis of the quadrature errors that contaminate the data (sample thickness, background spectrum subtraction, radiation perfection) suggest an uncertainty of 5% (Kovács et al., 2008) or 10% (Hao et al., 2016). Our theory is appropriate for a uniform error distribution with limits $\pm\Delta$. Such a distribution has a conventional second moment (or variance) of $\frac{1}{3}\Delta^2$, or standard deviation $\Delta/\sqrt{3}$. We have assigned an error of Δ=4.3, which corresponds to a maximum error of 10% of the mean signal recorded from the olivine chips, and to a standard deviation of 6% of this mean. Unlike in our synthetic experiment where we had full knowledge of the errors, difficulties in the real experimental procedure mean it is challenging to be completely sure of the correct error attribution.

When we analyse subsets of the full 41 sample dataset, gradually increasing the sample size in number, we find convergence to the results shown as red squares in Figure (8). On the plot we show in solid lines the values suggested for the principal absorbances determined by independent polarised light analyses (Kovács et al., 2008). We see that the smallest principal absorbance is estimated to be zero, very much in line with its tiny polarised value of 1.9 (Figure (8b)). The largest principal absorbance A_c (Figure (8d) differs by about 20% from the polarised value of 80.9, which itself has a likely imprecision of about 7% based on the recorded maximum of 75.3 in a perpendicular section (Kovács et al., 2008). The intermediate axis, which is the more difficult to estimate, is determined from the 41 measurements to be very close to the unpolarised value of (Kovács et al., 2008) (Figure (8c). The total absorbance is determined to better than 10%, although it must be remarked that the "exact" theory of Sambridge et al. (2008) gives an estimate more in accord with the polarised value of 128 (Figure (8a). This is somewhat surprising as one can see that the exact theory estimates a strongly negative (and physically unrealisable) value for A_a (Figure (8b).

The maximum likelihood model is found for $\mathbf{a}$=(69.50,48.60,20.90) which corresponds to $\mathbf{A}$=(0,41.8,97.2). One can note that $a_1<\max Q$=72.7, the difference being 3.2, slightly less than our assigned Δ. One now sees that with the error model in place, $\max Q$ is now not necessarily a downward biased estimate (as in the case of assumed exact data) but is probably upward biased, because the error model can incorporate data with a value up to $a_1+\Delta$; the treatment of noisy data has significantly modified the problem. The lowered value of a_1 of the maximum likelihood model is largely due to the presence of the very small principal absorbance A_c, estimated to be zero in the model. In this instance the problem really reduces to the estimation of two parameters A_b and A_c, and as we vary the error model we observe trade-offs between their estimated values. This would not necessarily occur for other datasets in which the principal absorbances do not have such disparate values.

With the full pdf to hand we may visualise parameters of interest. In Figure (9) we show cuts through the full pdf when two of the parameters are assigned to their maximum likelihood value. The pdfs for a_2 and a_3 are simple curves from which an error on the estimate can be derived. Of particular interest, however, is the pdf for a_1, which shows the effect of the positivity constraint (41). The maximum likelihood model occurs on the upper boundary of the allowed region in Figure (6); on this boundary A_a=0. The effect of this is to strongly truncate the pdf for a_1, an effect that would not arise if one of the absorbances were not so small.

Discussion

Our results set the determination of principal absorbances from unpolarised light on a new theoretically secure footing. We have derived the exact pdf of data randomly drawn from the quadratic form (7). The cdf of the data is everywhere regular and can be used in estimation procedures that can treat exact data; such procedures cannot include the maximum likelihood method, which requires the pdf for exact data, a function that contains N singularities (where N is the number of data). When the model is extended to incorporate errors in the data, the maximum likelihood model becomes both tenable and the method of choice. We have implemented this method for

a specific error model and find good performance on a synthetic dataset. Recovery of the total absorbance is particularly good and surpasses the performance of the only previously-known algorithm based on a simple average of the data. We analyse an actual laboratory data set that is typical in its size, consisting of 41 chips of olivine with thicknesses of 210±15μm. We find that we are able to estimate the principal absorbances robustly, finding answers that are stable when the number of analysed samples $N>15$. We are able to compare the estimates to other values for the principal absorbances, themselves experimentally determined. The agreement between the two datasets is at approximately the 10% level. We feel that this helps to open up new possibilities for unpolarised analysis of difficult mineral samples.

We close by mentioning allied work by Hillier (2001), who has obtained, by a tour-de-force, the pdf of a quadratic form Q in arbitrary dimensions. In D dimensions the quadratic form contains $D-2$ singularities, and the pdf must be characterised individually in the region between each of these singularities. The application is, amongst others, to the Durbin-Watson statistic commonly used in econometrics. It is therefore possible that there are other areas of application of the methods that we have developed for the present geological context.

A computer code that implements these ideas and enables a user to analyse his or her data will be available for download at `http://www.iearth.edu.au/codes/`, or from the authors.

Implications

A mathematically rigorous estimation framework is now in place that can be used for the analysis of typical laboratory data sets (several tens of samples) consisting of unoriented unpolarised absorbance measurements from small chips. We consider it an advance in the determination of the total polarised absorbance and principal polarised absorbances from a given set of unpolarized mesurements, using as few as 15 grains even from strongly anisotropic minerals such as olivine. It can retrieve three principal absorbances in anisotropic materials and is guaranteed to always respect positivity of these quantities. This will aid the accurate determination of species such as hydroxyl in geological materials.

Acknowledgements

This work was performed while AJ was a visitor at RSES, Australian National University, and IGPP, Scripps Institution of Oceanography. He thanks both institutions for their hospitality. We thank István Kovács for supplying the olivine data and for helpful discussions and editorial comments that improved the manuscript.

References

Abramowitz, M. Stegun, I. A. 1970, *Handbook of Mathematical Functions*, Dover Publications, Inc.

Bell, Rossman, Maldener, Endisch Rauch2003bell2003 Bell, D. R., Rossman, G. R., Maldener, J., Endisch, D. Rauch, F. 2003, 'Hydroxide in olivine: A quantitative determination of the absolute amount and calibration of the IR spectrum', *Journal of Geophysical Research: Solid Earth* **108**(B2).

Hao, Xia, Tian Liu2016Hao2016315 Hao, Y.-T., Xia, Q.-K., Tian, Z.-Z. Liu, J. 2016, 'Mantle metasomatism did not modify the initial H_2O content in peridotite xenoliths from the Tianchang basalts of eastern China', *Lithos* **260**, 315 – 327.

Hillier, G. 2001, 'The density of a quadratic form in a vector uniformly distributed on the n-sphere', *Econometric Theory* **17**(1), 1–28.

Kovács, I., Hermann, J., O'Neill, H., Fitz Gerald, J., Sambridge, M. Horvath, G. 2008, 'Quantitative absorbance spectroscopy with unpolarized light: Part II. Experimental evaluation and development of a protocol for quantitative analysis of mineral IR spectra', *American Mineralogist* **93**, 765–778.

Libowitzky, E. Rossman, G. R. 1996, 'Principles of quantitative absorbance measurements in anisotropic crystals', *Physics and Chemistry of Minerals* **23**, 319–327.

Rice, J. A. 2007, *Mathematical Statistics and Data Analysis*, 3rd edn, Duxbury Press.

Sambridge, M., Fitz Gerald, J., Kovacs, I. O'Neill, H. 2008, 'Quantitative absorbance spectroscopy with unpolarized light: Part I. Physical and mathematical development', *American Mineralogist* **93**, 751–764.

Withers, A. C. 2013, 'On the use of unpolarized infrared spectroscopy for quantitative analysis of absorbing species in birefringent crystals', *American Mineralogist* **98**, 689–697.

Table and Figure captions

44.15967	49.92802	43.72898	34.96102	37.11223	48.20369	44.02242	32.64254
39.17191	39.30958	41.48838	60.79360	27.77383	62.43976	42.75821	30.67905
61.95585	25.06300	33.80495	42.13757	36.95285	29.79391	26.66335	

Table 1: The 23 synthesized Q values for the model with $(A_a, A_b, A_c) = (1.89, 45.6, 80.94)$.

Figure 1: (a) Histogram showing the frequency distribution of 600 binned synthetic data for a problem where the principal absorbances are A_a=1.89,A_b=45.6 and A_c=80.94, giving theoretical values a_1=23.75,a_2=41.42 and a_3=63.27. There is a peak in the distribution around the middle value of a_2=41.42. The maximum and minimum values observed (indicated on each plot as "max" and "min" respectively) will generally never agree with these true values for a finite dataset; here the observed values are only slightly biased from the true extrema (downwards by .09 for the maximum and upwards by .06 for the minimum). (b) Histogram showing the frequency distribution of 600 binned synthetic data for a problem where the principal absorbances are A_a=2,A_b=2 and A_c=80.94, giving a_1=41.47,a_2=41.47 and a_3=2. Note that in this case of having two principal absorbances equal, the maximum of the frequency distribution is systematically skewed towards one extremum.

Figure 2: Plotted on the sphere are the level lines of Q for the case **a**=(3,2,1). The z-axis (vertical) is surrounded by almost circular contours. In this right-handed coordinate system, the value Q=2 occurs along the y-axis (the negative y-axis is facing the reader) and defines two separate contours encircling the sphere, while the x-axis on the right is surrounded by highly elliptical contours.

Figure 3: The pdf $\Phi(Q_0)$ for **a**=(3,2.5,1). Note there exists a logarithmic singularity on this plot at $Q=a_2$ (plotted values are truncated for visual clarity). Despite the singularity introducing a narrow infinite spike in the pdf, the integral remains finite yielding a sensible probability distribution.

Figure 4: Likelihood $p(a_2|Q,a_1,a_3)$ for the test dataset with N=23 points, showing the presence of N logarithmic singularities at $a_2=Q_i$. The amplitudes of singularities are truncated for visual clarity.

Figure 5: Probability $p(Q|\mathbf{a})$ for **a**=(3,2.5,1) and a) Δ=0.01 b) Δ=0.05 and c) Δ=1. The pdf is now entirely regular.

Figure 6: For a given value of a_2, the figure shows the allowed region (shaded) of (a_1,a_3) space that satisfies the inequalities required to guarantee positivity of the principal absorbances. The upper diagonal boundary corresponds to A_a=0 and two other coefficients non-zero; the lower boundary corresponds to two coefficients being equal, A_a=A_b=a_3, the minimum value, and the right-hand vertical boundary corresponds to two coefficients A_b=A_c=a_1, the maximum value. The right-hand lower corner represents an isotropic crystal.

Figure 7: Convergence of the results for a) A_{tot}, b) A_a, c) A_b and d) A_c, as a function of the sample size in a synthetic experiment. Data were synthesized from the model **A**=(1.89,45.6,80.94) and random noise from a uniform distribution with Δ=1 was added. Red squares show the maximum likelihood estimates described in this paper, whereas blue circles show the results from the method of Sambridge et al. (2008), which assumes exact data. The horizontal lines are the known correct values.

Figure 8: Convergence of the results for a) A_{tot}, b) A_a, c) A_b and d) A_c, as a function of the sample size. Red squares show the maximum likelihood estimates described in this paper, whereas blue circles show the results from the method of Sambridge et al. (2008), which assumes exact data. The horizontal lines are the values estimated by polarised analysis and are themselves subject to errors that were not determined.

Figure 9: Conditional probability of one variable when the other two parameters are set to their maximum likelihood values. Thus $p(a_1|Q,\hat{a}_2,\hat{a}_3)$ is a slice through the joint pdf at fixed $(\hat{a}_2,\hat{a}_3)$. a) $p(a_1|Q,\hat{a}_2,\hat{a}_3)$. b) $p(a_2|Q,\hat{a}_1,\hat{a}_3)$. c) $p(a_3|Q,\hat{a}_1,\hat{a}_2)$. In the case of a_1 we see the effect of the positivity constraints: the maximum likelihood is achieved on the line a_1=a_2+a_3 (upper boundary of Figure (6)), so that A_a is estimated to be zero.

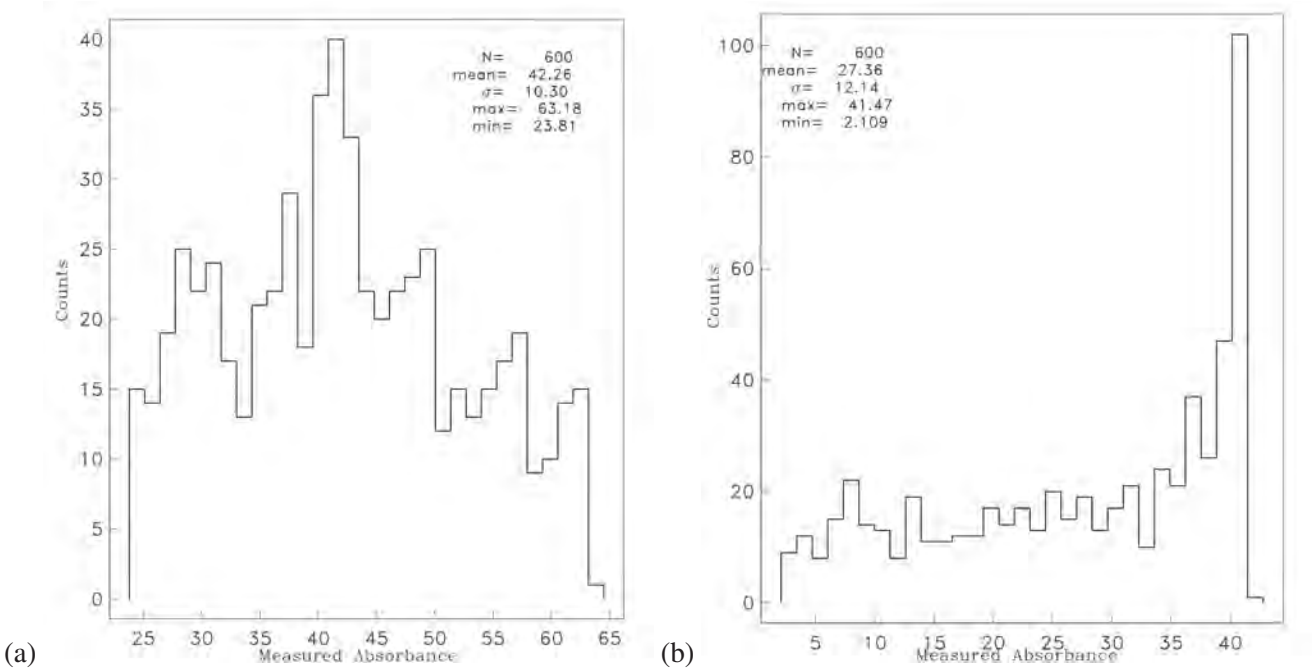

Figure 1: (a) Histogram showing the frequency distribution of 600 binned synthetic data for a problem where the principal absorbances are $A_a = 1.89, A_b = 45.6$ and $A_c = 80.94$, giving theoretical values $a_1 = 23.75, a_2 = 41.42$ and $a_3 = 63.27$. There is a peak in the distribution around the middle value of $a_2 = 41.42$. The maximum and minimum values observed (indicated on each plot as "max" and "min" respectively) will generally never agree with these true values for a finite dataset; here the observed values are only slightly biased from the true extrema (downwards by .09 for the maximum and upwards by .06 for the minimum). (b) Histogram showing the frequency distribution of 600 binned synthetic data for a problem where the principal absorbances are $A_a = 2, A_b = 2$ and $A_c = 80.94$, giving $a_1 = 41.47, a_2 = 41.47$ and $a_3 = 2$. Note that in this case of having two principal absorbances equal, the maximum of the frequency distribution is systematically skewed towards one extremum.

44.15967	49.92802	43.72898	34.96102	37.11223	48.20369	44.02242	32.64254
39.17191	39.30958	41.48838	60.79360	27.77383	62.43976	42.75821	30.67905
61.95585	25.06300	33.80495	42.13757	36.95285	29.79391	26.66335	

Table 1: The 23 synthesized Q values for the model with $(A_a, A_b, A_c) = (1.89, 45.6, 80.94)$.

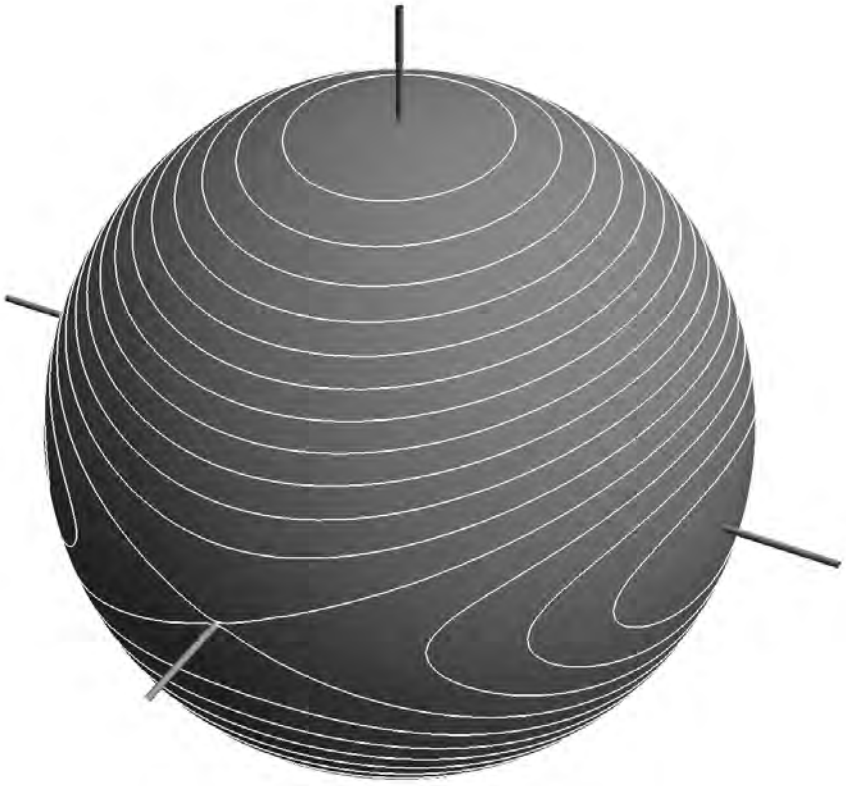

Figure 2: Plotted on the sphere are the level lines of Q for the case $\mathbf{a} = (3, 2, 1)$. The z-axis (vertical) is surrounded by almost circular contours. In this right-handed coordinate system, the value $Q = 2$ occurs along the y-axis (the negative y-axis is facing the reader) and defines two separate contours encircling the sphere, while the x-axis on the right is surrounded by highly elliptical contours.

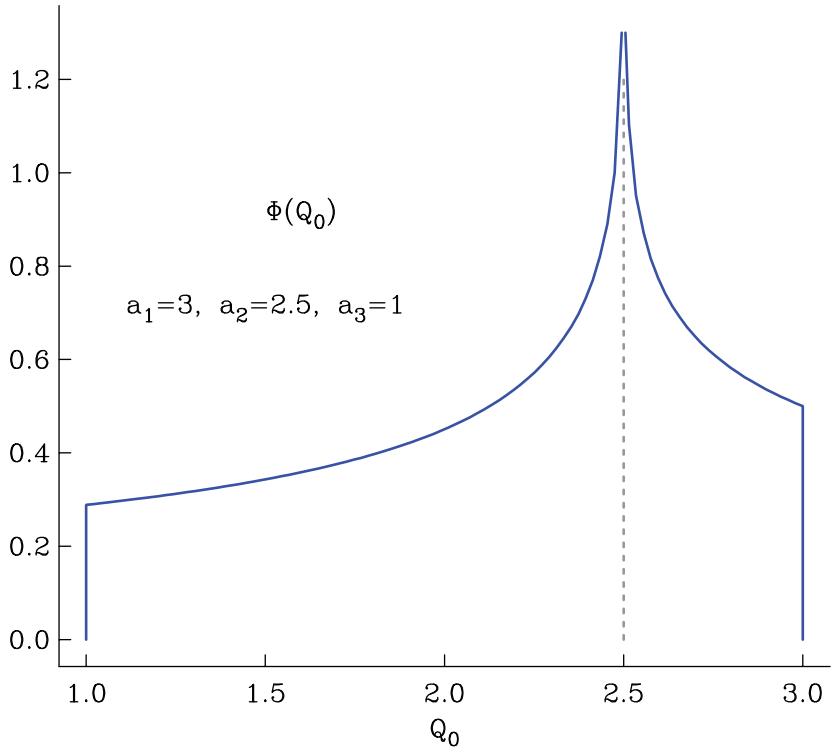

Figure 3: The pdf $\Phi(Q_0)$ for $\mathbf{a} = (3, 2.5, 1)$. Note there exists a logarithmic singularity on this plot at $Q = a_2$ (plotted values are truncated for visual clarity). Despite the singularity introducing a narrow infinite spike in the pdf, the integral remains finite yielding a sensible probability distribution.

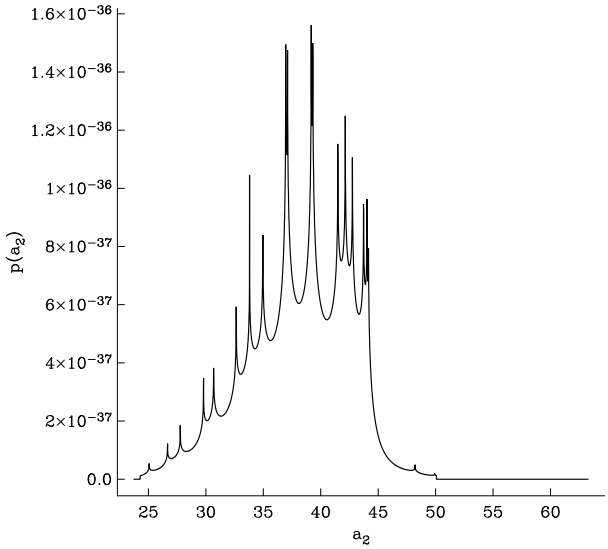

Figure 4: Likelihood $p(a_2|Q, a_1, a_3)$ for the test dataset with $N = 23$ points, showing the presence of N logarithmic singularities at $a_2 = Q_i$. The amplitudes of singularities are truncated for visual clarity.

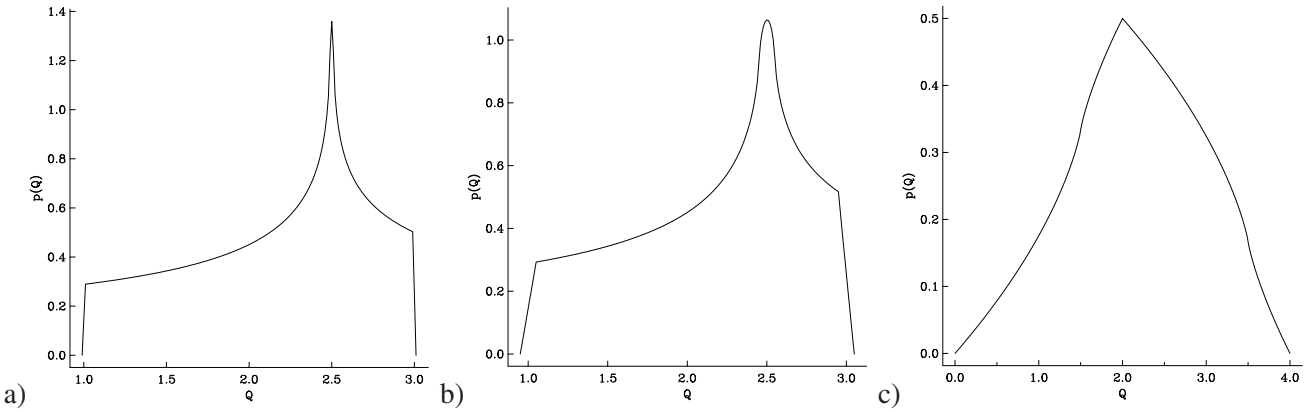

Figure 5: Probability $p(Q|\mathbf{a})$ for $\mathbf{a} = (3, 2.5, 1)$ and a) $\Delta = 0.01$ b) $\Delta = 0.05$ and c) $\Delta = 1$. The pdf is now entirely regular.

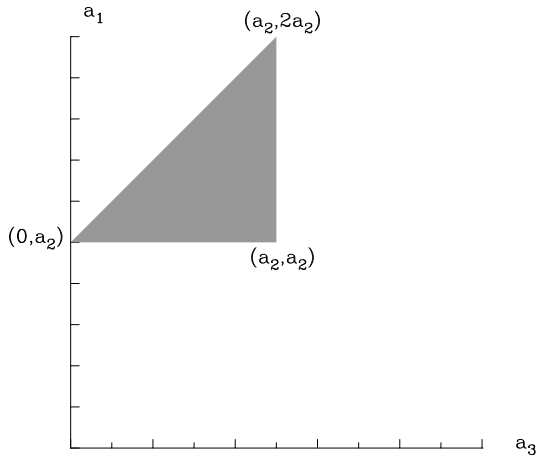

Figure 6: For a given value of a_2, the figure shows the allowed region (shaded) of (a_1, a_3) space that satisfies the inequalities required to guarantee positivity of the principal absorbances. The upper diagonal boundary corresponds to $A_a = 0$ and two other coefficients non-zero; the lower boundary corresponds to two coefficients being equal, $A_a = A_b = a_3$, the minimum value, and the right-hand vertical boundary corresponds to two coefficients $A_b = A_c = a_1$, the maximum value. The right-hand lower corner represents an isotropic crystal.

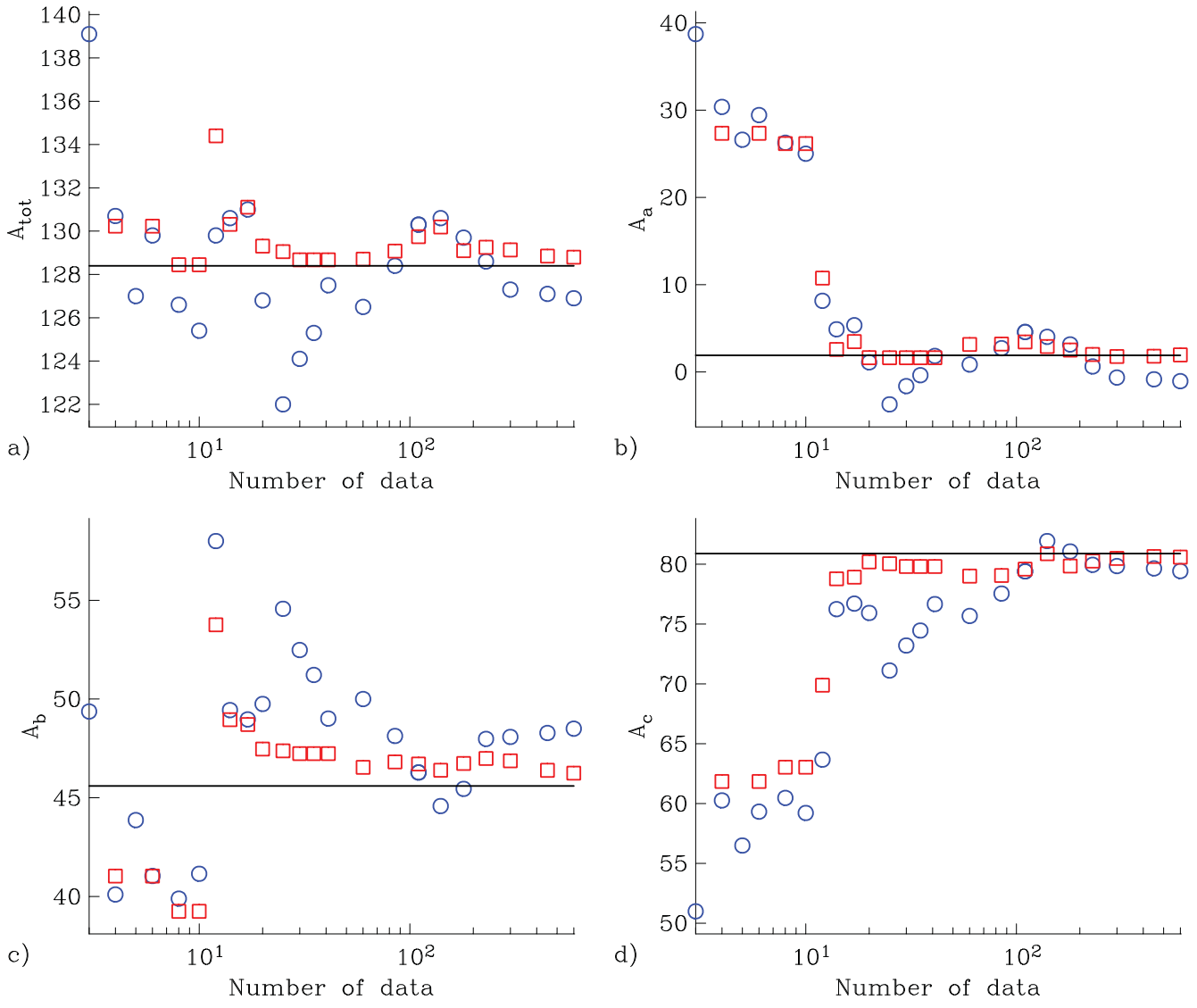

Figure 7: Convergence of the results for a) A_{tot}, b) A_a, c) A_b and d) A_c, as a function of the sample size in a synthetic experiment. Data were synthesized from the model $\mathbf{A} = (1.89, 45.6, 80.94)$ and random noise from a uniform distribution with $\Delta = 1$ was added. Red squares show the maximum likelihood estimates described in this paper, whereas blue circles show the results from the method of Sambridge et al. [2008], which assumes exact data. The horizontal lines are the known correct values.

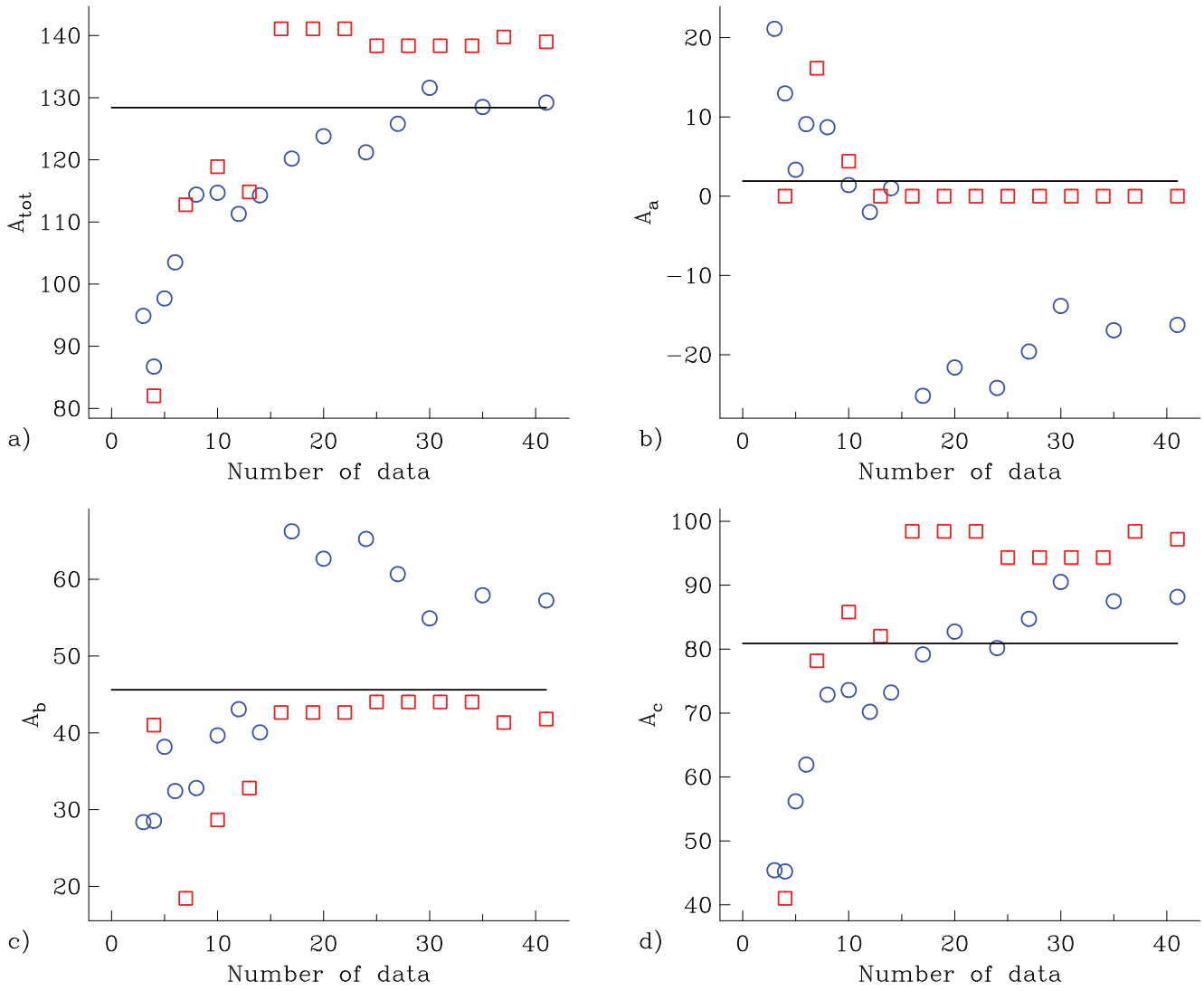

Figure 8: Convergence of the results for a) $A_{\rm tot}$, b) A_a, c) A_b and d) A_c, as a function of the sample size. Red squares show the maximum likelihood estimates described in this paper, whereas blue circles show the results from the method of Sambridge et al. [2008], which assumes exact data. The horizontal lines are the values estimated by polarised analysis and are themselves subject to errors that were not determined.

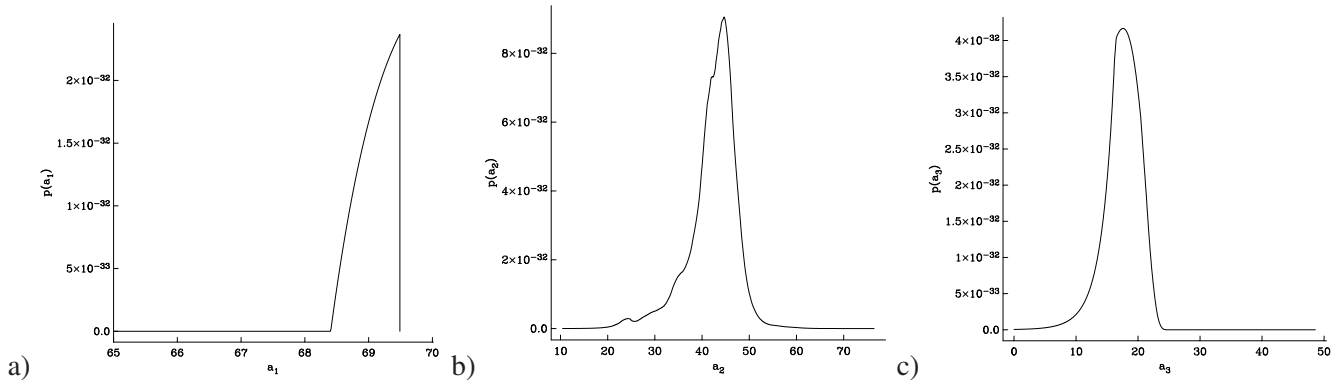

Figure 9: Conditional probability of one variable when the other two parameters are set to their maximum likelihood values. Thus $p(a_1|Q, \hat{a}_2, \hat{a}_3)$ is a slice through the joint pdf at fixed $(\hat{a}_2, \hat{a}_3)$. a) $p(a_1|Q, \hat{a}_2, \hat{a}_3)$. b) $p(a_2|Q, \hat{a}_1, \hat{a}_3)$. c) $p(a_3|Q, \hat{a}_1, \hat{a}_2)$. In the case of a_1 we see the effect of the positivity constraints: the maximum likelihood is achieved on the line $a_1 = a_2 + a_3$ (upper boundary of Figure (6)), so that A_a is estimated to be zero.