Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 24 August 2011
Accepted 7 March 2012

Bayesian analysis of the evidence for minor components in crystallographic models: an alternative to the Hamilton $\mathcal{R}$ test

Anthony E. Phillips[a]‡ and Jacqueline M. Cole[a,b]*

[a]Cavendish Laboratory, University of Cambridge, J. J. Thomson Avenue, Cambridge CB3 0HE, England, and [b]Department of Chemistry, University of New Brunswick, PO Box 4400, Fredericton, NB E3B 5A3, Canada. Correspondence e-mail: jmc61@cam.ac.uk

A simple test based on Bayesian statistics for the presence of minority populations in single-crystal structure refinement is presented. The test is illustrated by analysis of photocrystallographic experiments on single crystals of ruthenium–sulfur-dioxide-based complexes. In data sets collected after irradiation with light, conventional refinements of the populations of different metastable states to values below 4% are shown to be statistically significant. The results also confirm that the photo-induced states are absent from data collected in the dark.

1. Introduction

A number of photolinkage isomerism complexes have now been studied using single-crystal photocrystallographic methods (Kovalevsky *et al.*, 2002, 2003; Bowes *et al.*, 2006; Schaniel *et al.*, 2005; Schaniel, Cormary *et al.*, 2007; Schaniel, Woike *et al.*, 2007; Phillips *et al.*, 2010). With some notable exceptions (Warren *et al.*, 2009; Cormary *et al.*, 2009), only relatively small photoconversion fractions are generally achievable; indeed, too high a metastable-state population may place the crystal under such high internal stress that it shatters (Cole, 2004). Small minority structural populations such as these pose problems in data analysis: their occupancy will be highly correlated with other variables, such as their atomic displacement parameters and parameters associated with overlapping atoms in the disordered periodic model. As a result, the accuracy of the refined population is often suspect. In extreme cases it may not be clear whether or not photo-excitation has actually been achieved.

The question of whether or not there is sufficient evidence to introduce a metastable geometry into the crystallographic model of 'light' data is a particular case of a more general data-fitting problem well known in crystallography and across the sciences. Adding a new parameter to a model cannot increase the best-fit sum-of-squares statistic, and in practice invariably decreases it, regardless of whether or not the added parameter has any physical meaning. Of course, a single sum-of-squares value is a rather crude measure of the goodness of fit of a given model, and experts in a field will recognize features indicating an unsatisfactory fit: for instance, in crystallography, large residual peaks in the Fourier difference density of a model will be suspicious no matter how good the R factor. Nonetheless, careful analysis may be required to distinguish between a genuinely good model and an overfitted one.

The classical approach to this problem is to use a likelihood ratio test: either the F test or a crystallographically convenient variation such as the $\mathcal{R}$ test (Hamilton, 1965). These tests are applicable to *nested* models, in which the null model H_0 consists of a more general model H_1 with the addition of a linear hypothesis. Suppose that H_0 contains q additional linear constraints compared to H_1, and therefore has q fewer independent parameters. Hamilton (1965), for instance, gives the example of comparing certain isotropic and anisotropic refinements, in which H_0, the isotropic model, consists of H_1, the anisotropic model, with the addition of the linear constraints that $U_{11} = U_{22} = U_{33}$ and $U_{12} = U_{13} = U_{23} = 0$ for all atoms (where U_{ij} is the atomic displacement tensor). Given observed data (structure factors) $\mathbf{F}_o^2$, consider the likelihood ratio test statistic $\Lambda = \mathcal{P}(\mathbf{F}_o^2|H_0)/\mathcal{P}(\mathbf{F}_o^2|H_1)$, where $\mathcal{P}$ denotes probability. If H_0 is true, then, as the number of observations goes to infinity, $-2\log\Lambda$ converges under suitable conditions to a χ^2 distribution with q degrees of freedom (Wilks, 1938). Thus the hypothesis H_0 can be tested by comparing the observed likelihood ratio with a suitable χ^2 distribution. A statistic which is very unlikely under this distribution (as indicated by a small p value) is interpreted as evidence against H_0.

The assumptions leading to the $\mathcal{R}$ test are not always appropriate, and several alternatives have been proposed to overcome these difficulties in particular cases. Objections to Hamilton's assumptions fall into three main categories. First, the dimension of the hypothesis or the number of degrees of freedom may not be clear. This is the case, for instance, when comparing the two possible absolute configurations of a structural model (Rogers, 1981), leading to the now standard method of representing the configuration by a refinable

‡ Current address: School of Physics and Astronomy, Queen Mary University of London, Mile End Road, London E1 4NS, England.

parameter (Flack, 1983). Second, the assumption of linearity may not be justifiable (indeed, this may be the cause of the first problem); and in some cases, such as when the hypotheses being compared differ only in the atom assignment so that the same set of variables are refined in both models, this assumption is unnecessary (Rothstein *et al.*, 1978). Third, it is now well established that F tests in general rely strongly on the normality of the residual distributions whose variances are being compared (Box, 1953, and references therein). It has been known for some time that this assumption, and the assumption that the means of the residuals should be zero, are often untrue in practice (Vacca & Kennard, 1977).

For the purpose of analysing photocrystallographic data, a fourth objection becomes important. Among the conditions for the χ^2 distribution to be obtained is the topological requirement that the parameter space Θ_0 of H_0 must be in the interior of the parameter space Θ_1 of H_1 and that both spaces must be open (Kent, 1982; Protassov *et al.*, 2002). In other words, if the null hypothesis H_0 is a boundary case of H_1, the likelihood ratio distribution is not guaranteed to converge to a χ^2 distribution, and the F and $\mathcal{R}$ tests are no longer tractable since their null distributions are not known. In particular, in photocrystallographic experiments, among other parameters such as the locations of the atoms in the metastable state, the metastable-state occupancy fraction η is refined. Since H_0 has $\eta = 0$ (*i.e.*, no metastable state present) and η cannot be negative, the topological condition above is not met and Hamilton's $\mathcal{R}$ test should not be used.

This topological condition has recently been discussed in the astrophysics literature (Protassov *et al.*, 2002), where analysis analogous to the above applies to the possible presence of a new peak in a spectrum. However, this result does not appear to be used widely in crystallography.

As a result of the topological criterion, we must look elsewhere for a statistical test that is suitable to detect the presence of a metastable state. One promising avenue is the field of Bayesian statistics, which has been applied to many aspects of crystallography (Gilmore, 1996), and which has been suggested as a means to overcome the boundary-case problem in astrophysics (Protassov *et al.*, 2002; Trotta, 2008). Rather than calculating the probability that a null hypothesis is capable of explaining the observed data, these methods proceed by estimating the relative likelihood of two different models, given the observed data.

2. Theory

The following analysis follows the method of Gull (1988), as presented by Sivia (2006, p. 78). Label the ground-state model, in which there is no metastable state present, A, and the model including some fraction of a metastable state B. Then B has more parameters than A; we will consider first the case where there is a single such parameter θ before generalizing to multiple parameters.

We would like to estimate the ratio $\mathcal{P}(A|\mathbf{F}_o^2)/\mathcal{P}(B|\mathbf{F}_o^2)$, where $\mathbf{F}_o^2 = (F_{o1}^2, F_{o2}^2, \ldots, F_{on}^2)$ are the observed intensities. Applying Bayes' theorem to the numerator and denominator gives

$$\frac{\mathcal{P}(A|\mathbf{F}_o^2)}{\mathcal{P}(B|\mathbf{F}_o^2)} = \frac{\mathcal{P}(\mathbf{F}_o^2|A)}{\mathcal{P}(\mathbf{F}_o^2|B)} \times \frac{\mathcal{P}(A)}{\mathcal{P}(B)}. \quad (1)$$

The factor $\mathcal{P}(A)/\mathcal{P}(B)$ reflects our relative belief in the two models irrespective of (prior to considering) the data. We will adopt an agnostic position as to whether or not a metastable state has been experimentally generated by setting this ratio to 1.

To evaluate $\mathcal{P}(\mathbf{F}_o^2|B)$, we must take into account the extra variable θ:

$$\mathcal{P}(\mathbf{F}_o^2|B) = \int \mathcal{P}(\mathbf{F}_o^2|\theta, B)\mathcal{P}(\theta|B)\,\mathrm{d}\theta. \quad (2)$$

The first part of this integral is the probability of observing the data within model B for a given value of θ. We approximate this by a Gaussian distribution centred on the best-fit value θ_0 with standard deviation estimated by the fitting procedure:

$$\mathcal{P}(\mathbf{F}_o^2|\theta, B) = \mathcal{P}(\mathbf{F}_o^2|\theta_0, B)\exp\left[-\frac{(\theta-\theta_0)^2}{2\sigma(\theta)^2}\right]. \quad (3)$$

The second part of the integral in equation (2) is the prior likelihood of the parameters θ, again given model B. For simplicity we take this as constant over some range $\theta_{\min} \leq \theta \leq \theta_{\max}$ and zero elsewhere. Provided that θ_0 lies comfortably within this range, the Gaussian in equation (3) will be very close to zero when evaluated outside this range. Accordingly, rather than integrating from $\theta_{\min}$ to $\theta_{\max}$ in equation (2), we can integrate over the entire real line to give

$$\mathcal{P}(\mathbf{F}_o^2|B) = \frac{\mathcal{P}(\mathbf{F}_o^2|\theta_0, B)}{\theta_{\max}-\theta_{\min}} \int_{-\infty}^{\infty} \exp\left[-\frac{(\theta-\theta_0)^2}{2\sigma(\theta)^2}\right]\mathrm{d}\theta \quad (4)$$

$$= (2\pi)^{1/2}\sigma(\theta)\frac{\mathcal{P}(\mathbf{F}_o^2|\theta_0, B)}{\Delta_{\text{prior}}\theta}, \quad (5)$$

where $\Delta_{\text{prior}}\theta = \theta_{\max} - \theta_{\min}$. Generalizing to the case of multiple parameters $\boldsymbol{\theta} = (\theta^1, \theta^2, \ldots)$ gives

$$\mathcal{P}(\mathbf{F}_o^2|B) = \mathcal{P}(\mathbf{F}_o^2|\boldsymbol{\theta}_0, B)\prod_i \frac{(2\pi)^{1/2}\sigma(\theta^i)}{\Delta_{\text{prior}}\theta^i}. \quad (6)$$

One particular parameter that must be included in model B is the occupancy η of the metastable state. We treat this separately rather than including it in the vector $\boldsymbol{\theta}$ with the other parameters. This is because in marginal cases, where the η value obtained from least-squares fitting may well be comparable to its estimated standard deviation $\sigma(\eta)$, the Gaussian corresponding to equation (3) may be truncated by the requirement that $\eta \geq 0$. In this case, the marginalization integral can be explicitly evaluated:

$$\mathcal{P}(\mathbf{F}_o^2|\boldsymbol{\theta}_0, B) = \frac{\mathcal{P}(\mathbf{F}_o^2|\eta_0, \boldsymbol{\theta}_0, B)}{\eta_{\max}-0} \int_0^{\eta_{\max}} \exp\left[-\frac{(\eta-\eta_0)^2}{2\sigma(\eta)^2}\right]\mathrm{d}\eta. \quad (7)$$

Now, considering the standard assumptions for least-squares analysis (independent data and Gaussian noise), we can express the probabilities of observing data $\mathbf{F}_o^2$ under our two models in terms of the least-squares residuals of these models. The notation on the left-hand side is intended to reflect that, once particular values of the parameters have been specified, these equations are identical for the two models. F_o and F_c refer, respectively, to the observed and calculated (from the fitted model) structure factors

$$\mathcal{P}\left(\left.\begin{matrix}A\\ \mathbf{F}_o^2\end{matrix}\right|\eta_0, \boldsymbol{\theta}_0, B\right) = \prod_i^n \frac{1}{(2\pi)^{1/2}\sigma(F_{oj}^2)} \exp\left[-\sum_j^n \frac{(F_{oj}^2 - F_{cj})^2}{2\sigma(F_{oj}^2)^2}\right]. \quad (8)$$

Note that we have deliberately *not* included restraints in these equations: we treat them as an integral part of their respective models.

We take the σ values in equation (8) from the weighting scheme generated in the least-squares analysis. As an example, we consider the frequently used scheme from *SHELX* (Sheldrick, 2008):

$$\frac{1}{\sigma(F_o^2)^2} = w = \frac{1}{\sigma'(F_o^2)^2 + (aP)^2 + bP}, \quad (9)$$

where w is the weight used by *SHELX*; σ is the equivalent standard deviation, used in the following analysis; σ' is the standard deviation estimated during data collection; a and b are chosen to achieve a flat analysis of variance (*i.e.*, a χ^2 value roughly independent of intensity or resolution); and $P = \frac{1}{3}\max\{0, F_o^2\} + \frac{2}{3}F_c^2$.

Combining equations (1) and (6)–(8) gives the likelihood ratio we would like to evaluate. For convenience, we will actually calculate the logarithm of this ratio:

$$\begin{aligned}\ln\left[\frac{\mathcal{P}(A|\mathbf{F}_o^2)}{\mathcal{P}(B|\mathbf{F}_o^2)}\right] = &-\sum_j^n \left[\frac{(F_{oj}^2 - F_{cj}^2)^2}{2\sigma(F_{oj}^2)^2} + \ln\sigma(F_{oj}^2)\right]_A \\ &+ \sum_j^n \left[\frac{(F_{oj}^2 - F_{cj}^2)^2}{2\sigma(F_{oj}^2)^2} + \ln\sigma(F_{oj}^2)\right]_B \\ &+ \sum_i^{p_B - p_A} \ln\left[\frac{\Delta_{\rm prior}\theta^i}{(2\pi)^{1/2}\sigma(\theta^i)}\right] \\ &+ \ln\left\{\frac{\eta_{\max}}{\int_0^{\eta_{\max}} \exp\left[-(\eta - \eta_0)^2/2\sigma(\eta)^2\right] d\eta}\right\}. \quad (10)\end{aligned}$$

At this stage the likelihood ratio can already be evaluated directly; however, we will simplify this expression slightly for ease of calculation and discussion. First, we introduce the expression for the reduced goodness of fit S reported by *SHELX*:

$$S = \left[\frac{1}{n-p}\sum_j^n \frac{(F_{oj}^2 - F_{cj}^2)^2}{\sigma(F_{oj}^2)^2}\right]^{1/2}, \quad (11)$$

where n is the number of observations, p the number of parameters and S the (non-restrained) goodness of fit at the optimized values of all parameters.

Second, we note that it is tempting to cancel the $\ln\sigma$ terms from the first two sums in equation (10). However, this would not be valid since, although the uncertainties reported by the integration software are identical for both models, the σ values used in model refinement depend *via* equation (9) on P and hence on the calculated structure factors F_{cj}^2, which will depend on the particular model being refined.

We then break equation (10) into the sum of three terms, thus writing

$$\ln\left[\frac{\mathcal{P}(A|\mathbf{F}_o^2)}{\mathcal{P}(B|\mathbf{F}_o^2)}\right] = X + W + O, \quad (12a)$$

$$X = -\frac{1}{2}[(n - p_A)S_A^2 - (n - p_B)S_B^2], \quad (12b)$$

$$W = -\sum_j^n [\ln\sigma(F_{oj}^2)_A - \ln\sigma(F_{oj}^2)_B], \quad (12c)$$

$$\begin{aligned}O = &\sum_i^{p_B - p_A} \ln\left[\frac{\Delta_{\rm prior}\theta^i}{(2\pi)^{1/2}\sigma(\theta^i)}\right] \\ &+ \ln\left\{\frac{\eta_{\max}}{\int_0^{\eta_{\max}} \exp\left[-(\eta - \eta_0)^2/2\sigma(\eta)^2\right] d\eta}\right\}. \quad (12d)\end{aligned}$$

The 'fit' term X reflects the relative goodness-of-fit values for the models. The model with more parameters, B, will be favoured by this term since these parameters allow a smaller S value.

The 'weighting' term W penalizes models where the weighting scheme has, on average, increased the uncertainty of the observed data. Before actually evaluating this term, it is not obvious which model it will favour.

Finally, the 'Ockham factor' O acts to penalize B for introducing more parameters. Which model is preferable depends on the relative magnitude of the three terms. If the sum of these terms is positive, model A is more probable and the crystal is probably entirely in its original ground state; if it is negative, then B is more probable and there is good evidence that some fractional occupancy of a metastable state has been generated.

These terms behave qualitatively as expected. A low S value for either model favours that model. Counterintuitively, it appears at first that decreasing p for a particular model will count against that model, but in practice any such change in a good model will be compensated for by a decrease in S: the $(n - p)$ factor is simply compensating for the 'reduction' performed in calculating S in the first place. A large prior range for the parameters in B will count against it: the less we are prepared to say about any parameter, the more we are simply using it as a 'fudge factor'. On the other hand, a *small* uncertainty in the refined value of any parameter counts against B: if the value of a parameter is not well known prior to the experiment, more evidence is required to claim that, as a result of the experiment, it is now known to a high level of precision than to claim that it has only approximately been measured. Either model is penalized if it requires substantial

numbers of reflections to be downweighted to achieve the flat analysis of variance that is the goal of the weighting procedure.[1]

3. Three examples: evidence for metastable geometries in Ru–SO_2-based complexes

To exemplify this analysis, we have applied it to several published studies of complexes in the $[Ru(SO_2)(NH_3)_4\mathbf{X}]\mathbf{Y}$ family, where the *trans* ligand **X** and counterion **Y** vary. In these materials, partial occupancies of two different metastable SO_2 coordination geometries, often denoted MS1 and MS2, can be induced by irradiation with light at low temperatures, although the original ground state invariably remains the major structural component even after irradiation (Kovalevsky *et al.*, 2002, 2003; Bowes *et al.*, 2006; Phillips *et al.*, 2010). In the ground state (GS), the SO_2 ligand binds to the Ru^{II} ion through the S atom; in MS2, it binds in side-on fashion through both the S and an O atom; while in MS1, the least stable of these states, it binds end-on, through an O atom. In the following examples, we consider the cases first where the metastable states are necessarily absent, because the crystal has been kept in the dark, and where a metastable state is clearly present, before proceeding to more complex data sets.

3.1. Evidence for MS2 before and after irradiation at 100 K

The first example consists of two data sets collected from the same crystal of aquatetraammine(sulfur dioxide)-ruthenium(II) (±)-camphorsulfonate, $[Ru(SO_2)(NH_3)_4(H_2O)](C_{10}H_{15}SO_3)_2$, (1). At 100 K, this compound is known to have a single metastable state of MS2 geometry (Phillips *et al.*, 2010). A 'dark' data set was collected after the crystal had been cooled to 100 K in the dark; the crystal was then irradiated with an Xe lamp for 2 h at 100 K, after which a 'light' data set was collected using the same data-collection strategy as the first. Refinement of all parameters against the 'light' data with *SHELX* (Sheldrick, 2008) indicated the presence of the metastable state at a fraction of only $\eta = 3.6\ (5)\%$. We will use the method described above to check whether this is real or an artefact of an overmodelled data set.

In order to evaluate equation (12*d*) we need to decide on a suitable $\Delta_{\rm prior}\theta^i$ for each extra parameter θ^i associated with the metastable-state model. There are 28 such parameters: three for the position and six for the anisotropic displacement parameter of each of the three atoms in the SO_2 unit, plus the photoconversion fraction η. We set the prior upper limit on photoconversion to $\eta_{\rm max} = 0.5$ (otherwise we would be looking for traces of the ground state against a background of the metastable one) and set $\Delta\theta$ to 1 Å for atomic positions and 0.05 Å^2 for displacement parameters. This choice of 'prior' reflects the fact that we know approximately where to expect the metastable-state atoms but are not confident of their precise electron distribution.

Table 1
Comparison of the ground-only (*A*) and ground-plus-metastable (*B*) models for a 'dark' and 'light' data set collected from compound (1).

	'Dark' data		'Light' data	
	A	*B*	*A*	*B*
n	7495	7495	7488	7488
p	470	498	470	498
S	1.049	1.049	1.056	1.046
η_0		0.0088		0.0364
$\sigma(\eta_0)$		0.0044		0.0051
X		−13.4		−88.2
W		−2.4		−17.9
O		21.8		31.1
Total		6.0		−75.0

The results of this analysis are shown in Table 1. As expected, in both cases the 'fit' term *X* favours model *B*, while the 'Ockham factor' *O* favours model *A*. Note that, because of the definition of the reduced goodness of fit, an equal *S* value in a model with more parameters indicates a better fit. In the case of the 'dark' data, the Ockham factor outweighs the fit term, indicating that it is roughly $\exp(X + W + O) \simeq 400$ times more probable that there is no metastable state present than that 0.9% photoconversion has genuinely been achieved. However, in the case of the 'light' data, despite the relatively small (< 4%) photoconversion fraction achieved, this analysis shows that it is vastly (calculated to be 32 orders of magnitude) more probable that the metastable state was indeed generated than that the crystal remained entirely in the ground state. Moreover, these results are robust to any sensible variation in the prior limits used: varying these to favour model *A* demonstrates, for instance, that the 'light' data do contain enough information to deduce the presence and geometry of the photo-induced state even if this were not previously known. This is plausible since the geometry of these metastable states was first unambiguously characterized by precisely this sort of photocrystallographic experiment.

3.2. Evidence for MS1 after irradiation at 100 K

Having seen that the test confirms crystallographic intuition in cases where it is relatively clear that this is correct, a more complex situation may be considered. It has recently been shown that the O-bound metastable state, MS1, can be generated in coexistence with MS2 and GS in compound (1) at 100 K (Phillips *et al.*, 2012). This raises the question of whether the published synchrotron data for this and the related compound tetraammine(isonicotinamide)(sulfur dioxide)-ruthenium(II) tosylate, $[Ru(SO_2)(NH_3)_4(C_6H_6N_2O)](C_7H_7SO_3)_2$, (2) (Phillips *et al.*, 2010), also contain evidence for small populations of MS1. The same analysis was therefore applied. In this case, model *A* contains both GS and MS2 while model *B* also includes MS1. Compound (2) has two crystallographically independent excitation centres of which we only

[1] A spreadsheet template for facilitating the Bayesian calculation presented here is available from the IUCr electronic archives (Reference: SC5046). Services for accessing this file are described at the back of the journal.

Table 2
Comparison of GS + MS2 models without (*A*) and with MS1 (*B*), for 'light' data collected from compounds (1) and (2).

	Compound (1)		Compound (2)	
	A	*B*	*A*	*B*
n	9304	9304	16762	16762
p	532	561	973	1030
S	0.806	0.802	1.141	1.115
η_0(MS1)		0.031		0.041†
$\sigma(\eta_0)$		0.004		0.003
X		−38.9		−495.1
W		−6.8		−19.7
O		29.5		93.5
Total		−16.2		−421.3

† Two geometries, with populations equal within experimental error.

considered possible MS1 excitation at one [labelled `Ru51` in Phillips *et al.* (2010)].

The same priors were used for the different types of parameters as in the previous example. For compound (1), model *B* required 29 new parameters: the same 28 as in the previous example, plus a further occupancy parameter for the ground state (although note that the extra state in model *B*, to which these extra parameters refer, is now MS1 rather than MS2 in the previous example). This new occupancy parameter reflects the fact that, in *SHELX*, two parameters can be constrained to sum to unity while three or more parameters must be restrained to a constant sum. The inclusion of this extra restraint, however, ensures that the extraneous parameter does not bias the test in favour of model *A*. For compound (2), model *B* required 57 new parameters to account for MS1: three position parameters plus six atomic displacement parameters for each of six atoms (*i.e.*, two crystallographically independent MS1 geometries) plus three occupancy parameters (for the GS, MS1 and MS2 states). Table 2 shows that the resulting evidence is overwhelmingly in favour of the models that include MS1.

3.3. Evidence for MS1 decay at 120 K after irradiation at 100 K

Finally, we consider a set of data collected in order to monitor the decay of MS1 over several hours (Phillips *et al.*, 2012). A single crystal of compound (1) was cooled in the dark to 100 K, where it was irradiated with light as described in §3.1. A full data collection revealed that populations of 13.8% MS1 and 11.5% MS2 had been generated. The temperature was then raised to 120 K, where the crystal was held in the dark while repeated collections of the same set of frames were performed. The 100 K geometries of each ion were imported as rigid bodies into the 120 K refinement, taking small variations in the cell parameters into account. The approximation entailed in this procedure is justifiable since the small change in temperature will affect the state occupancies far more than the molecular geometries. Refinement of the fractional occupancies while holding the other model parameters constant revealed that MS1 decays to MS2 over a period of several hours at this temperature, and allowed the associated kinetic parameters to be estimated.

We again apply the analysis described above to evaluate the evidence for a nonzero MS1 population in each of these data sets. Here, as in §3.2, model *A* comprises GS and MS2 while model *B* also includes MS1. Note that because the atomic positions of MS1 were not refined during the decay process, in this analysis we must consider these values as being fixed at their 100 K values in model *B* rather than being parameters of this model.

As before, we take the prior range for U_{ij} parameters, $\Delta_{\rm prior}U_{ij}$, to be 0.05 Å^2, and set $\eta_{\rm max} = 0.5$ (although the latter could be reduced in this case of decay). The results of the model comparison are presented in Table 3. These show that the probability ratio $\mathcal{P}(A|\mathbf{F}_{\rm o}^2)/\mathcal{P}(B|\mathbf{F}_{\rm o}^2)$ increases with time as the fraction η of MS1 decreases, as one might expect. However, even at the end of the analysis, there is still clear evidence that some MS1 persists.

4. Discussion and conclusions

In these examples, the results of the proposed test accord with crystallographic intuition. When the sample has not been exposed to light, the test shows no evidence of the metastable state; but it concludes that there is good evidence for photoconversion in the 'light' data sets despite the low refined populations.

Despite the theoretical reasons not to use the $\mathcal{R}$ test in this case, it is interesting to consider the conclusions at which it would arrive. In fact, applying the $\mathcal{R}$ test to the examples in §§3.1 and 3.2 produces equivalent results to the analysis presented here: in the 'dark' data, there is no evidence at the 5% significance level for any metastable state, whereas in each set of 'light' data, there is evidence for the relevant metastable state at this level of significance. Nonetheless, it is reassuring to be able to place these empirical conclusions on a sturdy theoretical footing. The example in §3.3 is more subtle: while the $\mathcal{R}$ test shows evidence for the metastable state in the first seven successive data collections after irradiating the crystal, runs 8 and 9 marginally fail the test at the 5% significance level. This is not a contradiction: it means that the likelihood that the metastable state is absent is not negligibly small, but that it remains more likely still that the metastable state persists.

The difference between the results of the Bayesian and *F*-test approaches should not be taken as an indication that the statistical apparatus involved has any greater intrinsic power. Rather, the value of the approach presented here is that it allows a different question to be asked, which in some circumstances will be more relevant. The Hamilton $\mathcal{R}$ test asks whether, if the null hypothesis were true (*e.g.*, no metastable state were in fact present), there would be a substantial chance of observing a likelihood ratio at least as favourable to the alternative hypothesis as the one actually determined in the experiment. The Bayesian approach presented here asks for the relative likelihood of two models, given the experimentally determined data. One limitation of the latter approach is that

Table 3
Comparison of models without (A) and with MS1 (B), for 'light' data collected from compound (1) as a function of time at 120 K.

	t (h)								
	3.28	4.90	6.08	7.70	8.88	10.50	11.68	13.30	14.48
n	6326	6320	6286	6261	6259	6316	6280	6335	6338
p_A	369	369	369	369	369	369	369	369	369
S_A	1.217	1.226	1.234	1.205	1.197	1.211	1.214	1.220	1.214
p_B	389	389	389	389	389	389	389	389	389
S_B	1.202	1.218	1.227	1.199	1.193	1.208	1.212	1.218	1.212
η_0(MS1)	0.072	0.057	0.053	0.051	0.041	0.038	0.034	0.035	0.035
$\sigma(\eta_0)$	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
X	−125.0	−75.5	−67.5	−57.9	−38.4	−33.7	−27.5	−28.0	−29.1
W	−15.3	−10.2	−12.0	−9.9	−7.6	−7.1	−6.9	−7.6	−7.2
O	16.3	15.6	15.5	15.3	15.1	14.9	14.8	14.8	14.9
Total	−124.0	−70.1	−64.0	−52.4	−30.9	−25.8	−19.6	−20.7	−21.4

it leaves the possibility open that a third model might be more likely than either of the two being compared; so it will be most convincing in cases, such as the photocrystallographic examples presented here, when there is strong external evidence that only two (or, generally, a few) given models are worth considering.

We further note that, due to the deliberately simple formulation of the Bayesian approach that we employ, the numerical values obtained for the likelihood ratios are relatively crude estimates. Thus, for example, while it is clear that the ratios reported in Table 2 overwhelmingly favour the 'light' models, we cannot expect to take literally the apparent result that these models are tens or even hundreds of orders of magnitude more likely than the 'dark' alternatives. We describe below more sophisticated approaches that would enable a more accurate estimate of this ratio. For practical purposes, however, even a rough estimate has been demonstrated here to be sufficient to indicate the likely presence or absence of a particular state.

It is a practical advantage of the analysis presented here, as with Bayesian methods in general, that it is easy to include information about our prior expectations concerning the refined parameters. The priors for the extra parameters $\boldsymbol{\theta}$ in model B, here taken as uniform for simplicity, can easily be made more informative if this is appropriate. In particular, Gaussian priors $\mathcal{P}(\theta|B)$ would continue to allow $\mathcal{P}(\mathbf{F}_o^2|B)$ to be represented in closed form, as an analogue to equation (5). A further advantage of this analysis is the ease with which it can be extended to allow simultaneous comparison of three or more models.

In fact, it is not strictly necessary to have a closed form for $\mathcal{P}(\mathbf{F}_o^2|B)$ if the relevant integral can be evaluated numerically. In particular, there is no actual need to make the approximation of integrating over the entire real line to give equation (5). However, this approximation does not make an important difference to the Ockham factor [equation (12d)] in the examples considered. Indeed, even in the case of the 'dark' data in §3.1, where η_0 is very close to zero, making this approximation would change O by only 0.02.

This test may prove most useful in marginal situations where very low populations make it necessary to use very strong restraints to stabilize the refinement. In such situations, more informative prior distributions of the parameters along the lines discussed above may still enable these populations to be meaningfully described. For instance, in the examples discussed above, little prior information about the expected U_{ij} values was used. If these can be estimated – perhaps, with due precaution, from lower temperatures where higher photogenerated populations can more readily be achieved – this prior knowledge may help to detect remnant traces of the photogenerated states at higher temperatures.

It should be noted, however, that the 'very strong restraints' referred to above in practice make the refinement more brutal than typical cases. The background noise in the Fourier difference density makes refinements unstable at populations below about 3% if only reasonably soft restraints are placed on atomic displacement parameters, and none on bond lengths or angles (Brayshaw *et al.*, 2010; Phillips *et al.*, 2012). If this method is to be useful in detecting populations below this threshold, it will need to be carefully demonstrated that the restraints involved act only to prevent convergence to physically implausible minima rather than forcing an unjustified model upon the data.

This in turn will probably require a full Bayesian analysis of the relevant model. The comparison proposed here uses features from both Bayesian and conventional frequentist analysis, in that the likelihood function $\mathcal{P}(\mathbf{F}_o^2|\theta)$ is approximated based on the results of a standard least-squares crystallographic refinement [equation (3)]. This makes the test very easy to use: one of its greatest strengths is that only straightforward calculations from the F_c values output by any standard crystallographic software are required. However, this approach neglects the correlation between parameters in the refinement, which will be significant; furthermore, it relies on the potentially crude Gaussian approximation [equation (3)] for the likelihood of each parameter.

The problem of correlation could be partially ameliorated by incorporating the least-squares correlation matrix. However, in order to obtain the full posterior parameter distribution, which would give the truest picture of the conclusions that can legitimately be drawn from a given data set, a purely Bayesian refinement process would be needed. In contrast to the simple calculations presented here, this approach would require the use of Markov chain Monte Carlo

or similar methods to sample the multidimensional posterior distribution. Such calculations, however, are becoming mainstream in many scientific applications; their application to single-crystal refinements is a promising avenue for further investigation.

AEP thanks Trinity College, Cambridge, for an External Research Studentship, Steve Gull (Cavendish Laboratory, Cambridge) and Lyle Gurrin (School of Population Health, University of Melbourne) for helpful discussions. JMC thanks the Royal Society for a University Research Fellowship and the University of New Brunswick for the UNB Vice-Chancellor's Research Chair.

References

Bowes, K. F., Cole, J. M., Husheer, S. L. G., Raithby, P. R., Savarese, T. L., Sparkes, H. A., Teat, S. J. & Warren, J. E. (2006). *Chem. Commun.* pp. 2448–2450.
Box, G. E. P. (1953). *Biometrika*, **40**, 318–335.
Brayshaw, S. K., Knight, J. W., Raithby, P. R., Savarese, T. L., Schiffers, S., Teat, S. J., Warren, J. E. & Warren, M. R. (2010). *J. Appl. Cryst.* **43**, 337–340.
Cole, J. M. (2004). *Chem. Soc. Rev.* **33**, 501–513.
Cormary, B., Malfant, I., Buron-Le Cointe, M., Toupet, L., Delley, B., Schaniel, D., Mockus, N., Woike, T., Fejfarová, K., Petříček, V. & Dušek, M. (2009). *Acta Cryst.* B**65**, 612–623.
Flack, H. D. (1983). *Acta Cryst.* A**39**, 876–881.
Gilmore, C. J. (1996). *Acta Cryst.* A**52**, 561–589.
Gull, S. (1988). *Maximum-Entropy and Bayesian Methods in Science and Engineering*, edited by G. J. Erickson & C. R. Smith, Vol. 1, pp. 53–74. Dordrecht: Kluwer Academic Publishers.
Hamilton, W. C. (1965). *Acta Cryst.* **18**, 502–510.
Kent, J. T. (1982). *Biometrika*, **69**, 19–27.
Kovalevsky, A. Y., Bagley, K. A., Cole, J. M. & Coppens, P. (2003). *Inorg. Chem.* **42**, 140–147.
Kovalevsky, A. Y., Bagley, K. A. & Coppens, P. (2002). *J. Am. Chem. Soc.* **124**, 9241–9248.
Phillips, A. E., Cole, J. M., d'Almeida, T. & Low, K. S. (2010). *Phys. Rev. B*, **82**, 155118.
Phillips, A. E., Cole, J. M., d'Almeida, T. & Low, K. S. (2012). *Inorg. Chem.* **51**, 1204–1206.
Protassov, R., van Dyk, D. A., Connors, A., Kashyap, V. L. & Siemiginowska, A. (2002). *Astrophys. J.* **571**, 545–559.
Rogers, D. (1981). *Acta Cryst.* A**37**, 734–741.
Rothstein, S. M., Richardson, M. F. & Bell, W. D. (1978). *Acta Cryst.* A**34**, 969–974.
Schaniel, D., Cormary, B., Malfant, I., Valade, L., Woike, T., Delley, B., Kramer, K. W. & Güdel, H.-U. (2007). *Phys. Chem. Chem. Phys.* **9**, 3717–3724.
Schaniel, D., Woike, T., Delley, B., Biner, D., Kramer, K. W. & Güdel, H.-U. (2007). *Phys. Chem. Chem. Phys.* **9**, 5149–5157.
Schaniel, D., Woike, T., Schefer, J. & Petříček, V. (2005). *Phys. Rev. B*, **71**, 174112.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Sivia, D. S. (2006). *Data Analysis: a Bayesian Tutorial*, 2nd ed. New York: Oxford University Press,
Trotta, R. (2008). *Contemp. Phys.* **49**, 71–104.
Vacca, A. & Kennard, C. H. L. (1977). *Acta Cryst.* B**33**, 3271.
Warren, M. R., Brayshaw, S. K., Johnson, A. L., Schiffers, S., Raithby, P. R., Easun, T. L., George, M. W., Warren, J. E. & Teat, S. J. (2009). *Angew. Chem. Int. Ed.* **48**, 5711–5714.
Wilks, S. S. (1938). *Ann. Math. Stat.* **9**, 60–62.