Chapter 3

A Bayesian Approach to Determining Equations of State in the Diamond Anvil Cell

Abstract

In this work, we apply the general Bayesian statistical approach to analyzing powder diffraction data from diamond anvil cell experiments. Statistical problems over a wide range of complexities arise in this effort, and we show how the Bayesian statistical framework provides all of the tools necessary to adeptly meet these challenges. In particular, we demonstrate how this method for data analysis naturally leads to probabilistically rigorous solutions that can range from standard least-squares methods, which are fully consistent with non-Bayesian methods, to uniquely Bayesian solutions to problems that require a more sophisticated application of Bayesian priors. This proves especially useful in cases where the data do not fully constrain important physical parameters, but where outside knowledge provides the necessary information to yield reasonably well-constrained solutions. These include situations such as the use of previous studies of the composition-dependence of zero-pressure crystal volumes as well as the usage of smoothness criteria for the com-

pression evolution of axial ratios. In this chapter, we therefore demonstrate through many examples the flexibility and power of Bayesian data analysis as applied to experimental measurements in mineral physics.

3.1 Introduction

Fitting models to data is one of the most common and routines tasks in scientific investigations. It is nevertheless extremely common for model-fitting to be carried out in ad-hoc manner. In this chapter, we present the Bayesian approach to data analysis and specifically apply it to the interpretation of x-ray diffraction data. We hope to show that rather than giving a compendium of specific recipes for analyzing numerical data, Bayesian statistics instead provides a general framework for data inversion problems which transform information, in the form of noisy data, into useful knowledge. Once absorbed, this enables practitioners to develop data processing methods that are grounded in the rules of probability.

3.2 Intro to Bayesian Statistics

The Bayesian approach to statistics essentially boils down to the idea that every data analysis problem can be cast in terms of evaluating the probability of each data point given some model of reality. Any description of Bayesian statistics must start with Bayes' theorem, which is the basic foundational principle that underlies everything else:

$$\begin{gathered} prob(\vec{M}|\{d\}) \propto prob(\{d\}|\vec{M})prob(\vec{M}) \\ posterior \propto likelihood \times prior \end{gathered} \tag{3.1}$$

which can be read as: The probability of the model values $\vec{M}$ given the data, called the posterior, is proportional to the probability of the data $\{d\}$ given the model values, called the likelihood, times the initial probability of the model values to begin with, called the prior.[1]

The derivation of Bayes' theorem is deceptively simple, relying only on the rule of conditional probability $prob(A, B) = prob(A|B)prob(B)$, which states that the probability of two things both being true, $prob(A, B)$, is equal to the probability of one alone being true $prob(B)$ times the conditional probability of the other given that the first is true $prob(A|B)$. But the choice of which variable to condition on is arbitrary, so we can write an equivalent expression for $prob(B, A)$. Simple algebraic manipulation then leads directly to Bayes' theorem. The real power of Bayes' theorem is not the expression itself, which is an irrefutable expression of basic probability, but the willingness to interpret A and B as the model and data sets vectors. Once this is accepted, however it leads to countless insights—it is not unreasonable to think of it as the equivalent of Newton's second law for probability studies, where at some level it represents an assertion that if accepted has almost boundless explanatory power.

As we will refer regularly to each of these terms throughout this chapter, and they can

[1] The missing constant of proportionality is often called the evidence, but it is unimportant for parameter estimation problems and therefore we will not discuss it here.

often be confused, we will attempt to make their meanings plain. The posterior is the final answer of any statistical data analysis problem and is often quoted as a best-fit along with associated errors on the best-fit parameters. While perfectly correct, this statement of the results can be somewhat misleading. In actuality, the end result of parameter estimation is not a single answer, but rather a *probability distribution*, describing the relative probability of different combinations of parameter values after analyzing the data. In this way, the posterior summarizes our final belief about the parameter values after performing our analysis. The practice of quoting best-fit values with associated uncertainties is a concise way of conveying this information for posteriors which are approximately described by normal (Gaussian) distributions. We will discuss this point further in Section 3.3.3 when we obtain the posterior covariances for the powder diffraction analysis.

The likelihood is the most familiar of the three terms in Bayes' theorem, conveying the relative "goodness-of-fit" of the data to a particular set of model parameters. As we will show later, the most familiar form of the likelihood uses weighted least-squares or "chi-square" minimization, which seeks to minimize the difference between the data and a model. Since the prerequisite of any data analysis problem presupposes that we can generate a model for any particular set of model parameters $\vec{M}$, the likelihood then uses probability distributions for the data points $\{d\}$ to determine the probability of the data given the model values. The likelihood can therefore be thought of as asking the data, "What is the probability of this datum?"[2], and then combining the probability of every data point into an overall probability factor.

[2] These words are taken directly from John Johnson. They succinctly convey the root question that the likelihood addresses, and so we quote them here.

Lastly, the prior is simply the term that encapsulates our knowledge about the system before performing the analysis. This term is unfortunately the one that people tend to have the most difficulty with accepting, but it is actually rather straightforward. The prior simply mathematically encodes our "belief" about the model parameter values *before* the data is analyzed, just as the posterior conveys our belief *after* the data is analyzed. The prior therefore explicitly acknowledges that we usually have some previous or outside knowledge about a problem that is influencing our analysis. The most straightforward prior is just the posterior from a previous investigation, though we must be careful that we are comparing two directly analogous studies and be sure that the previous investigator properly reported their final uncertainties. Oftentimes, when lacking a directly applicable previous investigation, we might choose to use weakly informed priors, typically in the form of very wide normal distributions for each parameter. In some cases, we may choose to disregard all outside knowledge and therefore favor totally "uninformed" priors, which are wide and entirely flat. This option seems attractive to many at first, as they deem it the most "objective"; it is easy to recognize that this is the result of flawed thinking, however, since the choice of model parameterization is largely arbitrary and a flat prior for one parameterization can be a very biased prior in another. Assuming a total lack of prior knowledge—which is almost never the case—there is no "objective" method available to choose which parameterization is best. Therefore, it is always important to make informed choices about the combination of parameterization and prior, since together they directly affect the posterior. That said, we needn't worry too much about such issues given enough good quality data, since the likelihood will always win out over the prior because its affect grows as more data is added while the prior remains fixed.

Bayes' theorem is the very heart of the inversion method, allowing us to transform data into knowledge about model parameter values. This process is not at all trivial, since the causal arrow points in the opposite direction: we start with some underlying physical process, as represented by a model, and through that process, the data is generated. In the Bayesian framework, we are required at this point to somewhat broaden our definition of "model", to include both the physical picture that dictates the "ideal" noise-free values of the data coupled with a representation of how noise or observational error are introduced to the data. Therefore, a generic expression for the observed data values is given by:

$$\{d\}^{\text{obs}} = \{y\}^{\text{phys-mod}} + \{\epsilon\}^{\text{noise-mod}} \tag{3.2}$$

where the observed data values, $\{d\}^{\text{obs}}$, result from the combination of both a physical model, generating values of $\{y\}^{\text{phys-mod}}$, and outside noise source, represented by $\{\epsilon\}^{\text{noise-mod}}$. In this way, we can see that the overall "data model" necessarily includes all terms that contribute to the measured values.[3] In fact, it is often the case that the noise model term is taken as a collection of all of the unmodeled physics in a particular analysis problem. Viewed in this way, we are given the freedom to adjust our representation of both the physical model and the error model in order to best represent the data in hand, together parameterized by the model vector $\vec{M}$.

As mentioned above, the most common method of data analysis uses "chi-square" minimization, which it is easy to show is a direct outcome of Bayes' theorem for straightforward

[3] It is usually, but not always the case, that the unpolluted signal and the noise are additive. In cases when they are not, the form of Equation 3.2 is simply adjusted to reflect their true relationship.

analysis problems. Also known as least-squares regression, chi-square minimization relies on the use of a normally distributed error model:

$$\begin{aligned} d_i^{\text{obs}} &= y^{\text{mod}}(x_i, \vec{M}) + \epsilon_i^{\text{noise}} \\ \text{with} \qquad \epsilon_i^{\text{noise}} &\sim \mathcal{N}(0, \sigma_i^2) \end{aligned} \tag{3.3}$$

where the physical model y^{mod} is a function of both an independent variable x_i and the model parameters $\vec{M}$, and the observational noise is normally distributed about the physical model value, $\sim \mathcal{N}(0, \sigma_i^2)$, with a standard deviation corresponding to the known observational errors σ_i. We can evaluate the total likelihood, $\mathcal{L}$, as simply the product of a set of normal distributions:

$$\mathcal{L} = \prod_i \mathcal{N}(d_i - y_i^{\text{mod}}, \sigma_i^2) \tag{3.4}$$

where the data points are assumed to be independent of one another, giving a total probability that is just the product of the individual probabilities for each data point. Taking the log of both sides and substituting in the expression for a normal probability distribution, we can immediately see the relation to least-squares minimization:

$$\begin{aligned} \log \mathcal{L} &= \log \left\{ \prod_i \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-0.5 \left(\frac{d_i - y_i^{\text{mod}}}{\sigma_i} \right)^2 \right] \right\} \\ &= -\frac{1}{2} \sum_i \left(\frac{d_i - y_i^{\text{mod}}}{\sigma_i} \right)^2 - \sum_i \log \sqrt{2\pi}\sigma_i \\ &= -\frac{1}{2}\chi^2 + \text{const}, \qquad \text{with} \quad \chi^2 \equiv \sum_i \left(\frac{d_i - y_i^{\text{mod}}}{\sigma_i} \right)^2 \end{aligned} \tag{3.5}$$

where we introduce here the concept of χ^2, which represents error-weighted sum-square residuals. From this equation, it is clear that minimizing χ^2 yields the maximum for the value of the likelihood, and is hence referred to as the Maximum Likelihood Estimate (MLE). If we choose to adopt wide and flat priors, then the prior is just a multiplicative constant for the posterior that combines with the second term in the log-likelihood. According to Bayes' theorem, the posterior is proportional to the product of likelihood and prior, and therefore constants have no effect on the final answer and can thus be dropped. Therefore, we can clearly see that the use of standard weighted least squares fitting is exactly equivalent to the case of fitting data with normally distributed errors and wide and flat priors. Additionally, we can easily include the effect of normally distributed priors under the same framework by merely adding on an additional "prior penalty" term for each model parameter:

$$\chi^2_{\text{tot}} = \chi^2 + \sum_j \left(\frac{M_j - M_j^{\text{prior}}}{\delta M_j^{\text{prior}}} \right)^2 \tag{3.6}$$

where M_j is the j^{th} model parameter value, and M_j^{prior} and $\delta M_j^{\text{prior}}$ give the mean and standard deviation of the prior for that parameter. We can thus easily see that this "workhorse" of frequentist statistics, which describes the more common standard statistical viewpoint, is in fact just a simple application of Bayesian statistics given a few common simplifying assumptions. By remaining in the Bayesian mindset, however, we have the flexibility to adjust this generic data model as needed to tailor the analysis to the specific properties of the data we wish to analyze.

Of course, we cannot stop at obtaining a best fit, since the posterior is actually a probability distribution, summarizing our final knowledge of the model parameters. We therefore must obtain the shape of the posterior to go along with the location in parameter space expressed by the best fit. We will find here, just as before, that the standard method is nothing more than the application of Bayes' theorem under the assumption of Normality. In this case, we must also assume that not just the data errors, but the posterior itself adopts a normal distribution. This is often a good assumption, especially in the presence of large quantities of good data, and depends on how non-linear the physical model is with respect to the model parameters. We can imagine constructing a local first-order Taylor expansion of the physical model in the region of parameter space near the best fit. As long as the physical model behaves approximately linearly over the highly probable region of parameter space, closely matching the Taylor expansion, the posterior will then be well described by a normal distribution *Sivia and Skilling* (2006)[4].

This method of estimating a normal posterior distribution is often called Optimal Estimation, since it uses function evaluations only in the local region around the best fit to approximate the posterior (*Sivia and Skilling*, 2006). In general, a normal posterior is described by a multivariate normal distribution (in multidimensional parameter space):

$$\log \mathcal{P}(\vec{M}) = -\frac{1}{2}\chi^2 + \text{const} \approx -\frac{1}{2}(\vec{M} - \vec{\mu}_M)^T \mathbf{\Sigma}_M{}^{-1}(\vec{M} - \vec{\mu}_M) + \text{const} \tag{3.7}$$

where $\log \mathcal{P}$ is the log-posterior, and $\vec{\mu}_M$ and $\mathbf{\Sigma}_M$ are the mean and covariance of the multivariate normal distribution, corresponding to the best-fit and shape of the posterior distribution for the model parameters M. As before, additive constants are unimportant for the log-posterior, since they merely represent the normalization constant. Though we

[4] It is important to recognize that such a nearly normal posterior only behaves normally over the approximately linear region of parameter space, and thus we will be unable to accurately capture the "tails" of the distribution. This is rarely an issue for parameter estimation, since we are usually interested in where the bulk of the probability lies.

must use matrix notation, since the posterior exists in multidimensional parameter space, this is merely the expression of a quadratic function in many dimensions, analogous to $\log \mathcal{P} \approx \text{const} - \frac{1}{2}(M - \mu_M)^2/\sigma^2$ in one dimension. Given the derivative properties of a quadratic, we can relate obtain the covariance matrix Σ_M by evaluating the curvature of log-posterior space:

$$\mathbf{\Sigma}_{M,ij}^{-1} \approx -\frac{\partial^2 \log \mathcal{P}}{\partial M_i \partial M_j} \tag{3.8}$$

where $\mathbf{\Sigma}_M^{-1}$ is the inverse of the covariance matrix. We will discuss how to interpret the covariance matrix in greater detail later, but for now it should be noted that it can be represented by an “confidence-ellipsoid”, which contains the most probable parameter values and has a particular size and orientation. The orientation relates the correlation that exists between different pairs of parameters, indicating how well we can independently constrain their values, and the width scales the size of the region of high confidence. This is again an example of Bayesian methods directly corresponding to standard frequentist results under certain simplifying assumptions. Throughout this chapter, we will assume roughly normal posterior distributions, as the nonlinearities in our models remain rather small over the regions of high confidence. We thus do not make use of the many useful, yet complex and time-consuming, Bayesian methods that can handle non-normal posteriors.

3.3 Bayesian Analysis of Powder Diffraction Data

In applying the general Bayesian approach to analyzing powder diffraction data from diamond anvil cell experiments, a wide range of statistical problems arise of varying complexity. In this and following sections, we show how the flexible Bayesian framework is particularly well suited to developing individualized analysis routines that are well suited to the task. We also demonstrate how while always remaining within the Bayesian framework, we arrive at solutions that range from standard methods, directly equivalent to the more familiar frequentist approaches, to entirely Bayesian approaches that lie outside the bounds of the standard frequentist statistical toolbox.

The primary goal of many high-pressure powder diffraction studies is to obtain the equation of state of the material of interest. This thermodynamic function describes how the volume of the crystal unit cell varies as a function of some or all of the thermodynamic state variables: pressure, temperature, and composition. Furthermore, for non-cubic crystals, it is also important to obtain the variation of the unit cell dimensions with changing environmental conditions. In order to accomplish this task, the investigator must carry out a large, many-step inversion process, that converts powder diffraction spectra into estimates of the thermodynamic properties. In this study, we develop and present a number of general data analysis techniques for high-pressure (and high-temperature) diffraction experiments. These methods are demonstrated on powder diffraction data for pure endmember and 13% Fe-bearing Mg-silicate perovskite, obtained in the study presented in Chapter 2.

3.3.1 Estimating Peak Positions and Uncertainties

We begin with the task of obtaining unit cell parameters from a set of integrated 1D diffraction spectra. These integrated spectra were obtained using the technique described in Chapter 2. As discussed in the previous chapter, due to its greater insensitivity to spectral irregularities, which often arise in high-pressure multiphase experiments, we favor the use of individual peak identification and fitting over the whole pattern refinement method. In the process of peak fitting, the spectrum is split up into different sections that each contain sample diffraction lines. These lines are then fit individually, or in small clustered groups where the peaks are seen to overlap one another. For the peak-fitting step, we employ a standard least-squares fitting approach to obtain the positions of each peak from the integrated 1D diffraction profile (e.g., *Press et al.*, 2007). As discussed in the previous section, this least-squares method is entirely consistent with the Bayesian framework under the assumptions of normally distributed errors.

Each spectral peak is assumed to be well described by a pseudo-Voigt line profile, which is generally considered the standard for most powder diffraction applications (e.g., *Toby*, 2001).

$$\begin{aligned} y_i^{\mathrm{mod}} = & \sum_j^{N_{\mathrm{line}}} A_j \left[(1-s_j) \exp\left(-\ln 2 \left(\frac{x_i - p_j}{w_j} \right)^2 \right) + \frac{s_j}{1 + \left(\frac{x_i - p_j}{w_j} \right)^2} \right] \\ & + f_{\mathrm{bkg}}(x_i, \vec{c}_{\mathrm{bkg}}) \end{aligned} \tag{3.9}$$

where the modeled intensity, y_i^{mod}, is a function of the observed inverse d-spacing values x_i [$\mathring{A}^{-1}$]. This equation represents the total intensity as a sum of contributions from a set of N_{line} profiles together with an added polynomial background function, f_{bkg}, described by a vector of roughly 3 to 5 background coefficients $\vec{c}_{\mathrm{bkg}}$. The attributes of each line profile are described by its position p_j, half-width w_j, amplitude A_j, and shape s_j (which smoothly transforms the profile between a perfect Gaussian line shape at $s_j = 0$ to a perfect Voigt line shape at $s_j = 1$).

We start the fitting process with an initial guess for the peak positions, based on the positions of a few of the most intense and easily identified lines (prior knowledge from previous investigations is also useful to obtain a reasonable initial guess). After obtaining an initial guess for the peak position model, the peaks are fit with an iterative procedure, where the identification of each additional line helps to better constrain the model, thereby increasing the ability to identify further lines. This is particularly important for data with many phases visible within each spectrum as well as data that contains a low symmetry phase like perovskite, which has a significant number of partially overlapping peaks. Once the peaks are generally identified, they are fit using the pseudo-Voigt profile model given in Equation 3.9 assuming that the error bars are given by simple photon counting statistics, which is approximately normally distributed for large photon counts. The spectrum fitting is carried using a combined user-driven and automated minimization routine written in MATLAB, finding the set of parameters that maximizes the posterior probability by minimizing chi-square. The resulting individual best-fit peak positions for the Fe-free and Fe-bearing datasets are shown in Figures 3.1 and 3.2.

Once each region has been well modeled to extract the best-fit line positions, the next challenge is to obtain errors for these line measurements, as it is those line position uncertainties that propagate into the errors on the extracted crystal volumes and unit cell dimensions (and eventually into the equation of state parameters). As there are potentially many unmodeled physical effects present in the spectra, it is not practical to use optimal estimation to obtain errors in the peak positions, and thus we seek an alternate way to assess

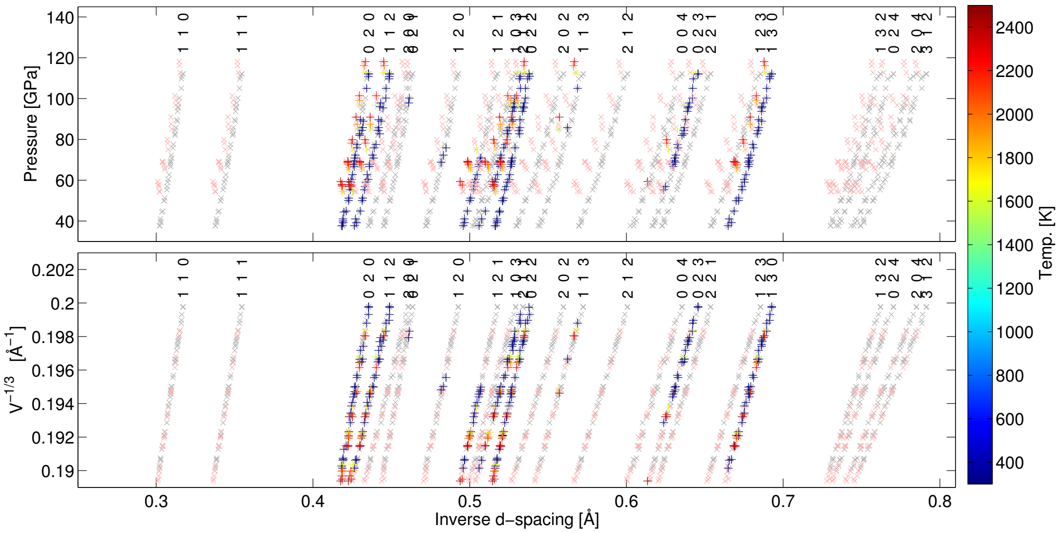

Figure 3.1: Peak positions for 0% Fe perovskite sample are plotted against pressure in the top panel and against $V^{-1/3}$ in the bottom panel. Plotting against pressure is difficult to interpret since it includes effects from the equation of state; in contrast, the purely geometric space leads to linear trends with minimal scatter and no systematic offsets due to temperature. The hkl value of each line is labeled at the top and in pale gray and pale red, the best-fit model peak positions are shown for the cold and heated data points.

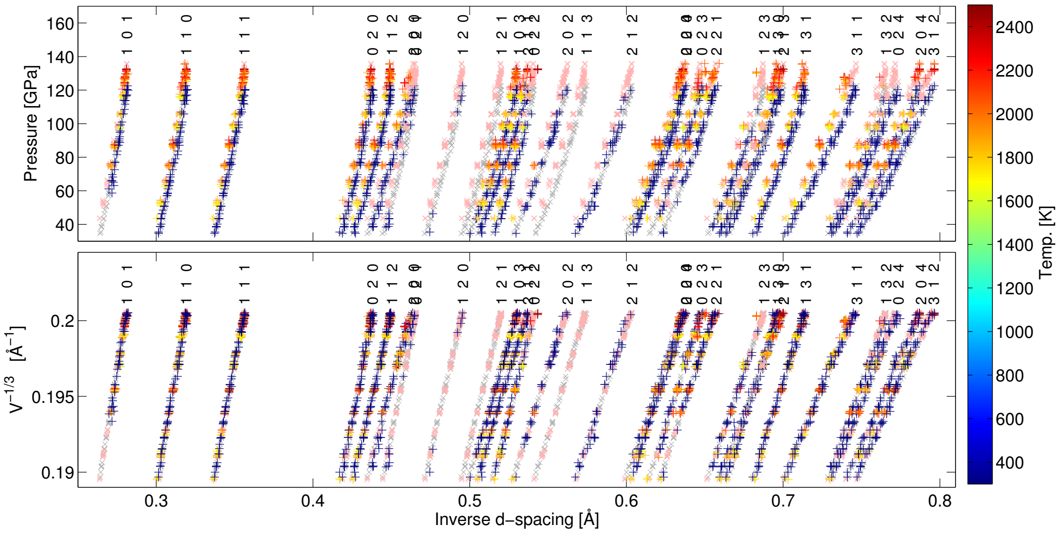

Figure 3.2: Peak positions for 13% Fe perovskite sample are plotted against pressure in the top panel and against $V^{-1/3}$ in the bottom panel. Plotting against pressure is difficult to interpret since it includes effects from the equation of state; in contrast, the purely geometric space leads to linear trends with minimal scatter and no systematic offsets due to temperature. The hkl value of each line is labeled at the top and in pale gray and pale red, the best-fit model peak positions are shown for the cold and heated data points.

reasonable uncertainties. Thinking back to the peak parameters, we note that the primary variable controlling the ability to determine the location of a peak is its width assuming the amplitude is sufficient to make it clearly visible. This reflects the fact that the peak width directly controls the relative curvature of the spectrum in the region of the peak. We therefore suppose that the line position error can be modeled as roughly proportional to the peak width, where the constant of proportionality is specific to each line. If we can identify this constant of proportionality by examining the observed scatter in the measured line positions, we will then have a straightforward way or assigning uncertainties to each position measurement.

The most obvious space in which to characterize the scatter is simply pressure vs inverse d-spacing, since the peak positions should follow a systematic trend with increasing pressure. This choice is not ideal, however, since the peak positions are sensitive to both pressure and temperature, where thermal pressure causes the heated data points to “lift” up off of the 300K curves, making it impossible to define a single curve that the data points are scattering about. This effect is plainly visible in the upper panels of Figures 3.1 and 3.2. In addition, by plotting against pressure, whose value is uncertain, we inadvertently add an additional source of scatter. In searching for a better alternative, we note that the evolution of the relative axial ratios is primarily a geometric phenomenon, and is thus only a function of volume. The planar spacings within the crystal, which are determined by the crystal geometry, should also therefore be nearly independent of temperature. Furthermore, since the cube root of the volume roughly sets the overall length-scale for the crystal, inverse d-spacings for each line should be roughly linear in $V^{-1/3}$. The lower panels of Figures 3.2

and 3.1 show the evolution of the line positions with changing inverse crystal length-scale, where the symbols are color coded by temperature. The inverse length scale can be thought of as a proxy for pressure, where the line positions are found to be conveniently nearly linear in this variable and independent of temperature.

We can therefore use the observed scatter relative to a best-fit linear trend as a way to obtain the line position uncertainty constants. For simple normally distributed data, it is well known that the sample standard deviation is a good estimate of the width of the normal distribution—it is in fact the Maximum Likelihood Estimator. This familiar result comes from the fact that the expectation value of the error-weighted square residuals is simply equal to one for a normal distribution,$\langle(\Delta x_i/\sigma_i)^2\rangle = 1$. Thus we can trivially determine the approximate error on each measured position, σ_i, by substituting in our assumption that the error is proportional to the line width and solving for the proportionality constant yielding:

$$\sigma_i = \alpha w_i, \quad \text{where} \quad \alpha \approx \sqrt{\langle(\Delta x_i/w_i)^2\rangle} \tag{3.10}$$

Using this method, we can easily obtain accurate estimations of the uncertainties of each peak position measurement, completing the first stage of the statistical analysis.

3.3.2 Estimating Unit Cell Parameters and Accounting for Misidentified Lines

After each spectrum is reduced into a list of measured peak positions and uncertainties, the next step in the analysis is to fit these peak lists with a crystal reflection model. The details of this method are given in the previous chapter. Here it is only important to note that for orthogonal perovskite, we have a model that is a function of three parameters, the lengths of the three crystal axes a, b, and c. These three parameters must then be fit to the 10 to 25 measured peak positions.

If every line in a powder diffraction spectrum is correctly identified, we could simply apply the standard least-squares approach given by Equation 3.5. Unfortunately, this is extremely unlikely when fitting lower symmetry phases like perovskite—which have many potentially visible and overlapping reflections—or when many different phases are present in the diamond anvil cell resulting in overlapping peaks—which frustrate accurate peak identification and fitting. Since the primary criterion used for peak identification is that the observed line appears at a position reasonably close to where it is expected, we must concede that misidentified peaks are bound to exist within our peak position list. To decrease the incidence of misidentified peaks, we should first examine the linear peak position maps in Figures 3.1 and 3.2. In addition to plotting the measured peak positions in dark colors, we also show the best-fit model positions for each line in pale gray and red for the cold and hot spectra. By showing the measured and modeled positions together in this well-behaved linear space, it is often easy to pick out strong outliers that are otherwise difficult to identify. We can then go back to the original spectrum to see if there was a simple and obvious mix-up between neighboring peaks, which are readily rectified. In other cases, it can be plainly seen that not just one, but all of the lines identified with a particular peak deviate systematically from the modeled position, indicating that the observed peak must belong to a different mineral phase. Worst of all identification errors are the lines that coincidentally lie near an expected peak position, but never deviate sufficiently far to draw attention to the mistake. These lines will therefore remain within the fitted data set, quietly and systematically biasing the measurements. In order to address this issue, we use a simple Bayesian mixture model approach which is robust against moderate degrees of contamination by peak identification errors.

A Bayesian Mixture Model is a general statistical tool that is useful in analyzing “polluted” datasets, where there are a number of different data models from which each data point can be generated. We start with the general data model description given by Equation 3.3. In writing that simple expression, we implicitly assumed that every data point was generated by the same physical process. When our dataset contains data from a variety of sources, such as correctly identified perovskite peaks and misidentified peaks, that expression is no longer appropriate. We therefore need to generalize Equation (3.5) to account for some fraction of misidentified peaks, which can be done by adding an additional term to the likelihood that accounts for the different possible sources of each data point. To simplify this process, we assume that each position measurement is drawn at random from one of two possible populations: either it is properly identified and drawn from the true sample line population or it is misidentified and draw from a population of confused lines. Just as before, the sample line population is represented merely by a normal distribution about the measured value. Based on how lines are identified, we expect that the confused line population is reasonably modeled by a simple flat distribution centered on the expected position value with a width of Δp, where the value of the width corresponds to how close an observed line must be to the expected position in order to be counted as an identified peak. We estimate a reasonable value for the width of $\Delta p \approx 0.02$, which is roughly a few times larger the typical uncertainties on line position determined from Equation (3.10). Therefore, the total likelihood for each data point is just a "mixture", or a weighted average, of these two distributions (*Sivia and Skilling*, 2006): $\mathcal{L} = \prod_i \left((1-f) N(p_i - p_i^{\text{mod}}, \sigma_i) + \frac{f}{\Delta p} \right)$, where f is just the expected fraction of the data points that are incorrectly identified. When there are no misidentified peaks, $f = 0$ and we recover the standard least-squares approach described above. Taking the log of both sides as before, we obtain the final expression for the Bayesian mixture model:

$$\begin{aligned}
\log \mathcal{L} &= \log \left\{ \prod_i \left(\frac{(1-f)}{\sqrt{2\pi}\sigma_i} \exp \left[-0.5 \left(\frac{p_i - p_i^{\text{mod}}}{\sigma_i} \right)^2 \right] + \frac{f}{\Delta p} \right) \right\} \\
&= \sum_i \log \left(\frac{(1-f)}{\sqrt{2\pi}\sigma_i} \exp \left[-0.5 \left(\frac{p_i - p_i^{\text{mod}}}{\sigma_i} \right)^2 \right] + \frac{f}{\Delta p} \right) \qquad (3.11)
\end{aligned}$$

In order to use this expression, we need only choose reasonable values for the width of the line identification window and the line fraction like $f \approx 0.1$. In truth, we can generally use values as large as $f = 0.5$ with the only downside being that we will increase the uncertainty in our determined best-fit values. In deriving this generalization of the least-squares method in so few steps, we can see the general power behind the Bayesian framework, which reduces everything to the evaluation of probabilities.

We apply this Bayesian mixture model to obtain reasonable estimates for the unit cell dimensions of the perovskite samples for each spectrum. The results of the crystal model fits are summarized in Figure 3.3, showing the variation of the normalized axial ratios with compression. These normalized axial ratios, defined in the figure caption, represent a simple scaling of the unit cell parameters by the average crystal length scale $V^{-1/3}$. Just as before, we plot these against inverse length scale, rather than pressure, since it reveals the purely geometric behavior of the crystal, rendering the plots essentially independent of temperature. In the left panel of the figure are the results for the 13% Fe-bearing sample. These are obtained using a direct application of the mixture model described above. Notice that the normalized axial ratios appear to evolve nearly linearly with compression, and show relatively small amplitude scatter that remains roughly constant over the compression range. In contrast, the modeling results for the 0% Fe sample is shown in the right panel of the figure. In light "plus" symbols, we show the modeling results using the same method as for the Fe-bearing sample. As can be clearly seen at the left end of the figure, the fitted axial ratios diverge from their expected trends rather dramatically at for a number of spectra, especially at low compressions. Unfortunately, the Fe-free sample has a rather limited number of fitted peak positions, due to poorer spectrum quality resulting from small sample size. While there are enough peaks to fit a model, the reduced number is significantly more susceptible to the deleterious effects of peak misidentification. Examination of the peak position map in Figure 3.1 does not show any obvious trends indicating poor peak identification, however. In order to deal with the issue, we therefore turn to the use of priors, which can provide useful outside constraints to any statistical analysis problem.

Though we have presented peak identification and crystal modeling in two separate sections, in truth they exists large overlap between the two. As alluded to earlier, peak identification and fitting is inherently an iterative process, where peak identification improves as the model is refined with the addition of each new peak. At the same time, however, the addition of an incorrect peak at such an early stage can be quite detrimental

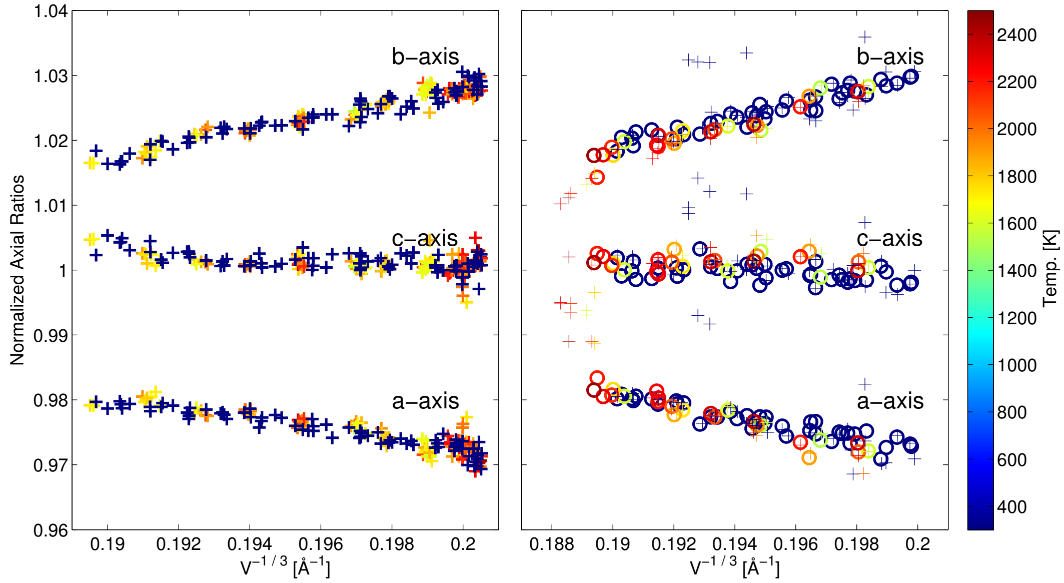

Figure 3.3: The normalized axial ratios are plotted in a purely geometric space against $V^{-1/3}$, where 13% Fe-bearing sample is on the left and the Fe-free sample on the right. For the 13% Fe sample, the data appears just as it should, being nearly linear, without temperature-dependent systematics, and with minimal scatter. The same unconstrained fit to the peak positions of the 0% Fe sample shows strong deviations from expected physically reasonable behavior, as shown in the "+" symbols. We therefore re-perform the analysis constraining the axial parameters by applying an appropriate prior, thereby suppressing this unphysical behavior, as show in the open circle symbols.

when using the standard least-squares approach. The Bayesian mixture model significantly reduces the effect of the misidentified peaks on the overall fit, making it useful both for obtaining final estimates as well as early on in the peak identification and fitting process.

As discussed initially in Section 3.2, it is common to use uninformed, or flat priors, when in the data dominated regime. As just discussed, however, the Fe-free data set does not contain enough visible spectral lines to yield fits that conform with our prior knowledge about the system. We know that the large deviations from linear behavior exhibited in the best-fit values at low pressure are not realistic, and therefore must result from an inability of the data to properly constrain all the crystal dimensions simultaneously. We can reasonably aide the fitting process that imposes the prior of linear behavior onto the crystal model. To do this, we simply fit the initial retrieved axial ratios with a line using a standard robust polynomial-fitting routine in MATLAB (this method also uses a Bayesian technique, but we will not discuss the details here). In doing so, we can estimate the expected linear trend of each of the normalized axial ratios. The prior is then constructed as a normal distribution centered on the robust-fitted line, whose width is chosen to be sufficiently small as to constrain the retrieved axial ratios. Given the scale on the y-axis, it is not unreasonable to expect that we should need to constrain the fit to variations of about half a percent from the linear trend. The results from that prior-constrained fit are shown in think circles over-plotted in the same figure. As is clearly visible, the problem has been fixed using this outside constraint. Also, we can see that the axial values are still able to vary from one spectrum to the next. This reflects how the prior can act as additional pseudo-measurement, affecting the fitted results but without totally eliminating the freedom necessary for the model to best explain the data. Though we do not show it here, it should come as no surprise that the unphysical variations in unit cell dimensions also induced large offsets in the apparent crystal volumes. Upon applying this prior-based method for constraining the crystal dimensions, all of these large deviations were seen to disappear, also providing useful verification for the approach.

3.3.3 Obtaining Unbiased Estimates and Uncertainties for Equation of State parameters

The last step of the inversion process is to obtain the equation of state (EOS) from the measurements of the crystal volumes at each pressure and temperature. The EOS expresses the dependence of crystal volume on pressure and temperature. This is captured using a set of parameters, including the zero-pressure volume, V_0, the resistance to compression given by the bulk modulus at zero pressure K_0, the pressure dependence of the bulk modulus K_0', as well as parameters that relate to the thermal properties of the crystal. While the details of the equation of the state model are given in Chapter 2, we focus here on the more general aspects of the fitting procedure and interpretation of the resulting posterior distributions.

To start with, we should begin with a discussion of priors for the parameter values. While we might wish to have enough data such that we needn't rely on any outside information, this is very often not the case. In fact, it is oftentimes not merely a question of the quantity or quality of the data, but an issue of the diversity of the data. A physical model for any complex system always relies on information about many aspects of that system's behavior. We are limited, however, in what types of data we can collect, and how directly that data speaks to different properties of the system. There are always properties that are poorly constrained, no matter how much data we collect of a particular type, and therefore we must rely on outside or prior information to constrain those values. The equation of state of perovskite is a perfect example of this. In thermodynamic equilibrium, perovskite is stable over nearly the entire range of lower mantle pressures and temperatures, but it becomes unstable at low pressures, (transforming back to pyroxene, a stable lower pressure phase of the same composition). This means that within its stability field, it is impossible to directly measure zero-pressure properties for perovskite, since the transition occurs between around 25 GPa. Unfortunately, the equation of state models rely explicitly on the zero-pressure properties. Because of this, it is generally not practical to collect enough high-pressure measurements to provide strong independent constraints on the zero-pressure properties. This makes intercomparison of the results of different studies particularly difficult, since there exist strong correlations between the zero-pressure parameters, especially the values of the bulk modulus K_0 and its pressure derivative K_0' (*Angel*, 2000; *Bass et al.*, 1981).

A typical, but systematically biasing, approach is to fix some subset of the parameters to particular values, and then fit the remaining parameters as normal. This is a logical, but probabilistically incorrect step to take. A common example of this is, for data that constrain the slope but not the curvature of the P-V compression curve is to fix the value of K_0' to 4. From a physics standpoint, there is nothing special about a value of 4—rather its special nature is related to the particular details of a commonly used equation of state. If the desire is merely to estimate the volume over the range of the data, there is nothing particularly wrong with this approach, since we have already supposed that the curvature is so small in the data region as to be non-noticeable. If, as is usually the case, we would like to take our derived equation of state and then extrapolate to higher pressures, perhaps the core-mantle boundary, then the practice of fixing parameters is a serious mistake. It will severely (and artificially) limit the allowable range of volumes predicted outside the data region. This is because uncertainties in K_0' spill over into uncertainties in other correlated parameters, and such correlated uncertainties are entirely ignored when parameters are fixed—we obtain unrealistically narrow uncertainty regions that bear little relation to our true knowledge about the system given the data. By ignoring the known trade-offs between K_0 and K_0' (the slope and the curvature), we are doomed to wildly underestimate the uncertainties resulting from extrapolation.

Luckily, this challenge is very simple to address in Bayesian statistics, since it merely reflects the desire on the part of the investigator to use outside knowledge or *prior information* to help constrain the parameter estimation. While fixing parameters is generally a poor idea for the reasons discussed above, applying a probability density function that encodes a prior belief about the parameter values is a simple and explicit way to properly constrain the fitting parameters without paying the heavy costs associated with fixing parameters. The simplest informative prior is a normal distribution for each model parameter, which therefore suggests the range of reasonable parameter values without insisting on a particular value. For example, we might replace the assumption that $K_0' = 4$ with instead a weakly informed prior of $K'0 = 4 \pm 1$. In order to incorporate this information, we turn to Bayes' theorem with says the that posterior probability is proportional to the product of the likelihood and the prior (since independent probabilities multiply). Therefore the log-posterior probability is just the sum of the two independent pieces $\log p = \log \mathcal{L} + \log \Pi$, where a simple normally distributed prior adds a squared residual term for each parameter

$$\log \Pi = -\frac{1}{2} \sum_i \left(\frac{M_i - \bar{M}_i}{\sigma_{M_i}} \right)^2 \tag{3.12}$$

where $\bar{M}_i \pm \sigma_{M_i}$ is the prior information about the i^{th} model parameter M_i. Based on its mathematical form, it is clear that such a normal prior distribution acts as an "apparent" measurement of each model parameter.

For our study of perovskite, the issue of priors on the equation of state parameters arises when considering the zero-pressure volume. While it is not possible to measure the equilibrium value of V_0, since it does not exist, perovskite is metastable at ambient pressure and temperature. The kinetics of the backward transition are so slow that the transformation is inhibited, Thus, there are ambient measurements of metastable perovskite. It is an open question as to how relevant the metastable behavior is to the behavior within the stability field. Many phases undergo softening and changes in their axial ratios as they approach a phase transition. It is usually assumed that these differences are negligible and thus we can make use of metastable measurements. Presuming that is the case, we are faced with a perfect opportunity to make use of priors.

In the absence of zero-pressure measurements of the sample volume, the proper way to constrain the parameter values is to impose a reasonable composition-dependent prior for V_0. It has long been known that many crystals that exist as solid solutions show nearly linear variation of the volume as a function of composition. *Kudoh et al.* (1990) showed that the measurements of metastable perovskite volume are roughly linear in ferrous iron composition. We use a similar analysis of the previous ambient pressure perovskite volume measurements, shown in Figure 3.4, to obtain our prior on V_0 as a function of composition. Based on the large observed scatter in that figure, as compared to the small individual measurement errors, it is clear that there is some source of additional scatter inducing large sample-to-sample variation. We therefore fit these data with a straight line, also allowing for an additional scatter term. The solid line shows the best-fit linear trend, while the dashed lines show the 68% confidence intervals, which are dominated by the intrinsic scatter term. We can summarize these results by representing them as a normal distribution whose mean is linear in composition and whose standard deviation is nearly independent of composition.

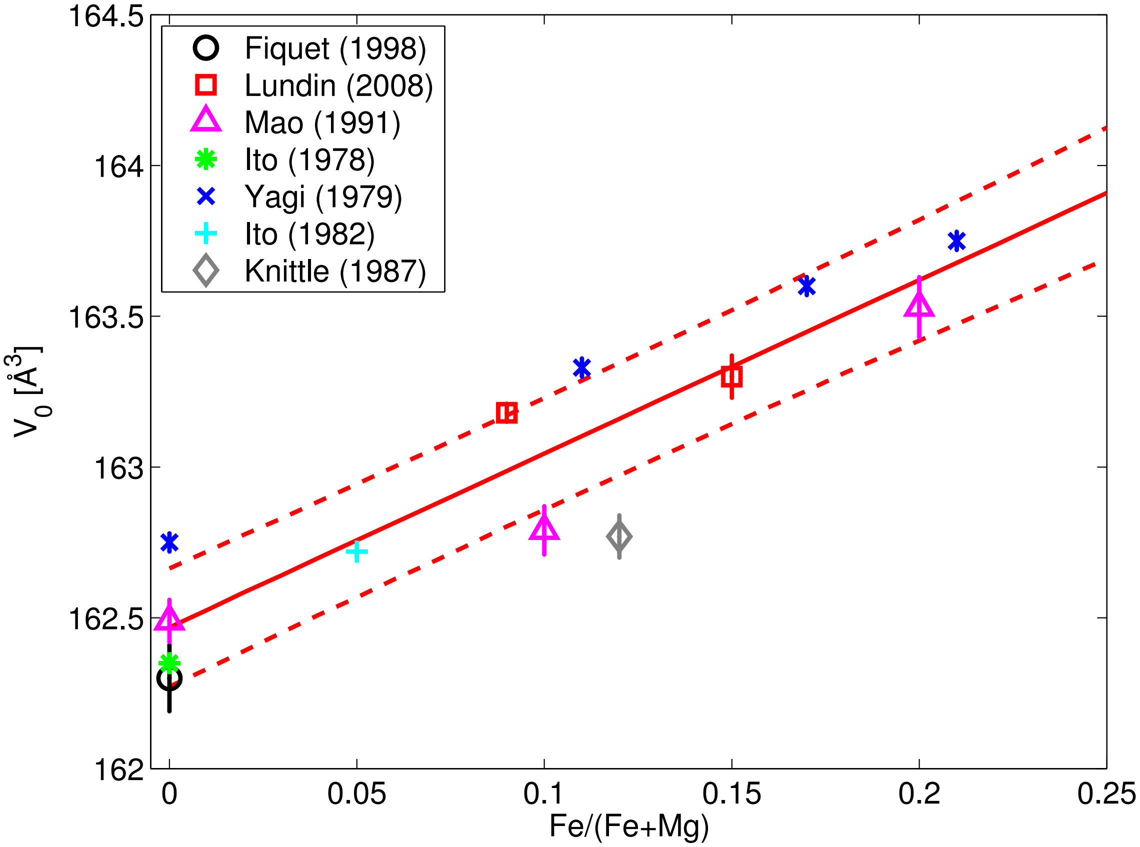

Figure 3.4: Examination of previous zero-pressure volume measurements as a function of iron content, using a linear fit combined with an intrinsic scatter due to sample-to-sample variation. 68% contours are shown.

We therefore use this curve, sampled at 0% and 13% Fe to provide the prior on $V0$ for our fit. As far as the priors on the other parameters, they are much less important and unnecessary, as the combination of the data and the V_0 prior are more than adequate to provide constraints on the parameters values.

With a reasonable prior in hand, we can then fit both the ambient temperature and laser-heated data to determine the equations of state of the two perovskite samples. The details of that fitting procedure were already discussed in Chapter 2, and so we will focus here on interpreting the resulting posteriors. After obtaining the best-fit, we used the general Bayesian approach of optimal estimation, calculating the covariance matrices from the curvature of the posterior in log-space, as expressed by Equations 3.7 and 3.8. These results can be directly visualized using a "stair-step" plot, see Figure 3.5, which shows the correlated uncertainty bounds on each pair of parameters by plotting every parameter against every other parameter. The nested 68% and 95% confidence ellipses are shown in the central region for each parameter pair, with the 0% and 13% Fe samples in black and red lines, respectively. These plots show 2-D marginalized confidence regions, which means that the variations in all other parameters (not the 2 being plotted) are integrated over. This means that these are *projections* from a five-dimensional space down into a two-dimensional space. Note that they are not slices, which would happen when certain parameters are fixed, rather than marginalized over. The one-dimensional marginal probability distributions for each parameter are shown on the edges of the figure. These correspond directly to the best-fit and 1-D uncertainties on each parameter. The subfigures are color coded with cold parameters shaded blue and hot parameters shaded red. They are also organized such that correlations amongst cold parameters are at the top, the correlations between the two hot parameters are at the bottom right, and the cross-correlations are shaded gray in the lower central region. From these plots, it is clear that the correlations are strong between purely hot and purely cold parameters, but nearly zero for mixtures. Additionally, it is clear that the two different compositions have differences in their equations of state that are statistically discernible (though not necessarily all physically relevant, see Discussion in Chapter 2), where there is little to no overlap in the 2-D 95% confidence regions for any of the parameter pairs. A final thing that must be pointed out is the ability of the stair-step plot to reveal significant parameter combination differences that are poorly visible when looking at one dimension at a time, but plain to see in two dimensions. The best example of this is shown by the correlation between K_0 and K_0', which initially appear largely consistent for the two compositions, when looking at the 1D probability distributions, but can be seen to be over inconsistent with one another at over the 3-σ level when viewed in two dimensions.

These results should then be compared in exactly the same way with previous investigations of perovskite equation of state. Unfortunately, this task is made rather difficult for two reasons. Firstly, none of the previous investigations report the full covariance matrix (or equivalent correlation matrix) for their equation of state parameters. Second, and more important, is the fact that all previous investigations fixed a number of important equation of state parameters in order to improve the apparent constraints on the parameters. Unfortunately, this results in strongly underestimating the parameter uncertainties as discussed in detail above. We are forced, therefore, to settle for the more qualitative comparisons already presented in Chapter 2.

3.4 Conclusions

In this chapter, we have used the data analysis problems that arise in powder diffraction and equation of state studies to demonstrate the power and flexibility of the Bayesian Statistics. We have demonstrated how “standard” statistical techniques, such as least-squares regression, are in fact fully consistent the Bayesian framework. We also presented a number of novel analysis methods that show how Bayesian thinking can allow for the development bespoke statistical procedures that do not suffer from the use of ad-hoc assertions, but are

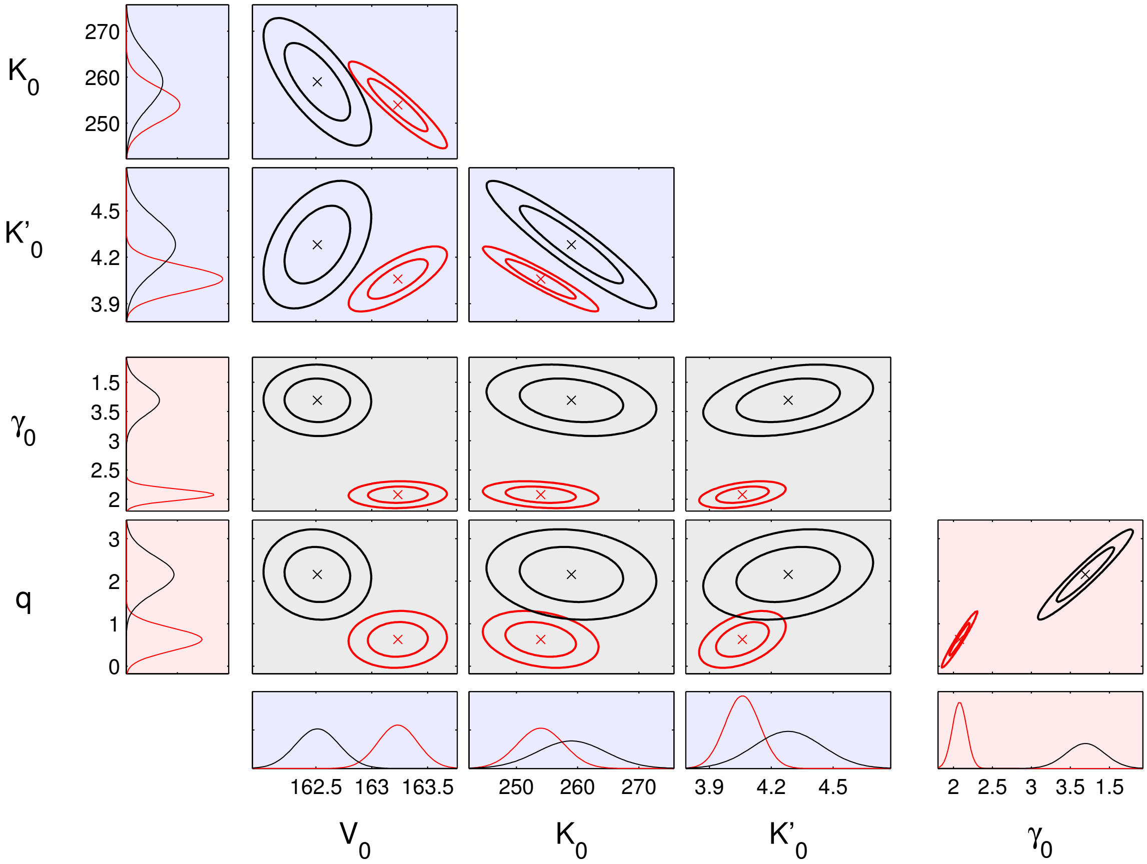

Figure 3.5: Two-dimensional posterior plots show the correlations between the different equation of state parameter uncertainties. The central plots show the nested 68% and 95% confidence regions for the 0% Fe sample in black and the 13% Fe sample in red. The marginalized one-dimensional probability distributions are shown for each parameter along the edges. Additionally, the plots are organized with the cold EOS parameters shown with a blue background, the hot parameters with a red background, and the cross-correlations with a gray background. With this scheme, it is clear that the cross-correlations between cold and hot parameters are rather small, nearly horizontal confidence ellipses, whereas the correlations between hot or cold parameters are significant, as indicated by highly angled confidence bounds.

rather directly rooted in probability theory. It is our hope that this text will help to show readers that there is no “black-magic” in Bayesian techniques, and adopting them does not require the rejection of previous statistical understandings. In short, we hope to convert others to the Bayesian way of thinking, by way of recognizing that all