Multistage Analysis In Astrostatistics

a thesis presented for the degree of

Doctor of Philosophy of Imperial College London

and the

Diploma of Imperial College

by

Xixi Yu

Department of Mathematics

Imperial College London

180 Queen's Gate, London SW7 2AZ

September 2020

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Copyright

Thesis advisor: Professor David A. van Dyk Xixi Yu

Multistage Analysis In Astrostatistics

Abstract

A sequence of statistical analyses often needs to be conducted by different groups of researchers where the output of each analysis feeds into subsequent analyses. The statistical and systematic uncertainties of estimated quantities, especially high dimensional quantities, are hard to quantify and difficult to carry forward into subsequent analyses. In practice, uncertainty is often ignored and the estimated quantities are often treated as fixed and known, leading to erroneous interpretation of the data and underestimation of uncertainties.

An astrophysical example occurs when we use the spectra of the solar and stellar coronae to estimate its density, temperature, and other composition. This requires the use of the results of atomic physical experiments and calculations. In the example, the interaction of statistical and atomic uncertainties in the context of a spectral model inform the analysis of the images of solar and stellar physics. The interpretation of the spectral observations in relation to atomic data and their statistical uncertainties is necessary for deriving meaningful uncertainties on the solar and stellar coronal plasma parameters like electron density and temperature. Understanding how uncertainties in the underlying atomic physics propagates to the uncertainties in the inferred plasma parameters is an essential component of this analysis.

We propose a principled *multistage analysis* to carry forward the model-generated atomic data uncertainties and statistical uncertainties obtained from preliminary analyses to a primary analysis based on the observed spectral lines under a Bayesian framework. Besides the Bayesian methodology that considers the atomic data uncertainties as fully specified and uncorrectable (the so-called *pragmatic Bayesian* method), we allow for the observed data to update the atomic data uncertainties (the *fully Bayesian*

method). The former generally increases the uncertainties on the inferred parameters compared with models that incorporate only statistical uncertainties. In contrast, the latter reduces the uncertainties on the inferred parameters. To incorporate uncertainties into a primary analysis, we summarize a Monte Carlo sample of the atomic data that represents its statistical uncertainty by treating these samples as equally likely. We also consider a degenerated multivariate Gaussian model derived via a principal component analysis as a low dimensional summary of the uncertainty in the atomic data. Markov Chain Monte Carlo based model fitting is implemented including Multi-step Monte Carlo Gibbs Sampler and Hamiltonian Monte Carlo.

The multistage analysis is able to cope with case studies of different levels of complexity. Two-stage analysis is used to infer the plasma parameters in spectral analysis with case studies on the density-sensitive only Fe XIII lines and the temperature-sensitive only Fe XVII lines. Three-stage analysis is used on the density- and temperature-sensitive O VII lines in the X-ray regime considering one more parameter and its corresponding source of statistical uncertainties. This principled multistage analysis has the potential to be applied on complicated models with a variety of sources of uncertainties.

In loving memory of my grandma, Kong Qingzhi (1939-2018).

It is your endless love for family and thirst for knowledge that encourages and inspires me all the time.

Acknowledgments

I would like to express my sincere gratitude to my supervisor Prof. David van Dyk for the insightful guidance and continuous encouragement throughout my Ph.D research. I would like to acknowledge the generous support from the International Space Science Institute and the 'Improving the Analysis of Solar and Stellar Observations' international team, especially Dr. Vinay Kashyap (CfA), Dr. Giulio Del Zanna (Cambridge), and Dr. David Stenning (Imperial/SFU) for generous contribution and insightful discussions for my research. I would like to thank Marie Skodowska-Curie RISE Grant and European Space Agency to support my research and travel. I would also like to thank the CHIANTI project, the *Hinode* project, the *Chandra* mission, and the SIMBAD Astronomical Database for providing spectrum and atomic data sources.

Thank you Lifan Xuan, Yazhe Li, Lekha Patel, Johannes Lutzeyer, Josh Plasse, Matt Price-Williams, Jordan Noble, Jake Dunn, Elizabeth Riddle-Workman, Jack Hogan, Jure Vogrinc, and all my College friends over the years for coffee-breaks, pub time, and all the perceptive thoughts along the beer. Thank you Nazma Islam, Valentina Missaglia, and Konstantina Anastasopoulou for our very special time at CfA. Thank you my many many friends, you are always there with encouragement or listening ear. Thank you Shawn for your enthusiasm and brilliance to share my map of the world. Finally, big thanks to my parents and all my families. It is your love, encouragement, and companion walking me through any long and sometimes bumpy road.

Contents

1

Introduction and Review

Solar and stellar coronae are complex and dynamic systems. Measuring physical properties of the corona is important for understanding the processes that lead to the complex structure of corona. We aim to infer physical quantities of the solar and stellar atmosphere, e.g., density and temperature, while we only observe intensities or photon counts in several spectral lines. The inferences also rely on models for the underlying atomic physics and the atomic data serves as a bridge to enhance the connection between the inferred physical quantities of the solar and stellar atmosphere and the observed spectral lines.

In this chapter, we introduce background about the spectrum and uncertainties in atomic data, and the important role that atomic uncertainties plays in the spectral analysis in order to obtain valid scientific estimations of the physical quantities of the solar and stellar atmosphere. To account for atomic uncertainties and all other source of uncertainties, we design a principled multistage analysis which is applicable to complicated models with a variety of sources of uncertainties.

1.1 Spectrum and uncertainties in atomic data

1.1.1 Solar and stellar coronae

A corona is an aura, made up of plasma, that surrounds the Sun and other stars, which can be detected using X-ray telescopes. A spectral line is a dark or bright line in a spectrum corresponding to specific wavelengths of light absorbed or emitted by an object. It originates from specified electronic transitions between energy levels of ions after an inelastic collision of ions in the plasma. Energy differences are unique to the transition of electrons, therefore, the energy of the spectral lines can identify the ions. Moreover, the relative strength of the spectral lines can identify the density and temperature of the plasma. From a statistical perspective, the atomic physics data provide the relative likelihood of various transitions and thus the relative intensities of the spectral lines as a function of the plasma parameters, like the electron density and temperature of the plasma. Given the significant progress in atomic calculations in recent years (e.g., see Del Zanna et al. 2004, Del Zanna 2011, Del Zanna & Storey 2012), high quality atomic data can be used in the study of astrophysical and fusion plasmas to enhance the plasma parameter estimations.

An emissivity curve is one of atomic data indicating how much energy is radiated from a given level transition from unit volume of the plasma at given density and temperature values. Spectral observations of solar and stellar coronae, mostly taken in the X-ray, extreme-ultraviolet (EUV), and ultraviolet (UV) part of the spectrum, are often combined with atomic data to infer fundamental plasma parameters such as electron temperatures and densities. This information is essential for constraining models of coronal heating. Reliable and accurate atomic data is essential for interpreting and modelling accurate X-ray observations (Kallman & Palmeri 2007). In reality, instead of taking them as fixed and known, the uncertainties associated with the plasma emissivities are likely to be comparable to or larger than those from counting statistics, therefore, a proper statistical treatment should be incorporated via a proper data analysis (Yu et al. 2018).

1.1.2 Accounting for systematic uncertainty

The problem of interpreting astrophysical spectra is still complex. Ignoring systematic uncertainties may lead to erroneous interpretation of the spectral data and underestimation of uncertainties. The temptation of accounting for any systematic uncertainties, like instrument calibration uncertainty and atomic uncertainty, in the analysis with broad, ad hoc assumptions may result in the failure of accounting for obvious correlations within the analysis. For example, instead of assuming the same impact, the atomic data for strong lines resulting of transitions from the ground state are likely to be more accurate than the atomic data for weak lines influenced by many different transitions (e.g., Foster et al. 2010).

Fortunately, the availability in the reductions of computational cost in recent years has made it possible to carry out detailed statistical analysis on those complex systems (Drake et al. 2006, Lee et al. 2011, Xu et al. 2014, Yu et al. 2018). Especially, Drake et al. (2006), Lee et al. (2011), and Xu et al. (2014) have proposed quantitative methods and Bayesian approaches accounting for calibration uncertainties in a subsequent spectral analysis, which is a good direction for any future study on accounting for systematic uncertainties. The works focused on an instrument calibration product, effective area, which is the area that must be used when calculating the physical properties of sources in the sky.

Instead of being universally ignored or relying on the experience of the calibration scientists, as an initial step, Drake et al. (2006) adopted a truncated Gaussian distribution to represent the distribution of calibration uncertainties. This method relies on a large calibration sample, demands a heavy cost of computing, and requires prior knowledge of the unknown model parameters in practical implementation, while demonstrates the importance of including calibration uncertainties with respect to the accuracy of error bars for the model parameters. To address these problems, a pragmatic Bayesian method for sampling from a distribution of plausible calibration curves was presented by Lee et al. (2011). This method incorporates the calibration uncertainties into a comprehensive spectral analysis of high-energy Chan-

dra spectra while ignoring the potential information in the spectral data for narrowing the uncertainty for the calibration product. They also used a Principal Component Analysis (PCA) to efficiently quantify the uncertainty of the calibration curves of an X-ray telescope. The complex high-dimensional calibration sample is represented by a compact and concise summary via a small number of components from a PCA so that the computational efficiency can be improved significantly.

Following that work, Xu et al. (2014) enabled a principled fully Bayesian method that consistently accounts for the calibration uncertainty in the high-energy spectral analysis and updates estimation of calibration product with the help of the observed spectral data. Comparing with the pragmatic Bayesian method, the fully Bayesian method allows the data to simultaneously provide information for estimation of the source spectral parameters and for the calibration product given the spectral model. Along with valid estimates of calibration uncertainty, the spectral parameters can be estimated accurately and efficiently. As a result, to narrow the uncertainty for the calibration product, we are able to make use of information in large-count observed spectra. Based on these previous works, we now can broadly apply those general Bayesian statistical techniques in handling systematic errors and extend them to account for the uncertainties in atomic data for several main ions, which is described in the following chapters.

The accuracy of spectral observations and of the atomic data has progressed hand in hand for a long time. In the past years, thanks to large-scale computations, the atomic data calculations have improved significantly and the atomic physicists have benchmarked the atomic data for several of the main ions (e.g., see Del Zanna et al. 2004, Del Zanna 2011, Del Zanna & Storey 2012). In this thesis we consider the effect of the uncertainties in the atomic data and how these uncertainties propagate to the determination of the composition, the temperature and the density, of the plasma based on spectral observations, via case studies in several main ions.

1.1.3 The analysis we consider

We begin with a simple case that several spectral lines observed with the EUV Imaging Spectrometer (EIS) on *Hinode* satellite (EIS, Culhane et al. 2007) in the solar active region and associated with a widely used ion, Fe XIII. The EIS instrument observes many emission lines whose intensities are used to form density-sensitive ratios and is regularly used to measure electron densities from coronal iron ions (e.g., see Watanabe et al. 2009, Young et al. 2009). Figure 1.1 demonstrates a representative observation of a solar active region. The intense magnetic fields in the active region lead to the formation of $3-4$ MK plasma on the relatively short loops in the active region core (e.g., see Warren et al. 2012, Del Zanna 2013, Del Zanna & Mason 2014). It also illustrates the spectral region near 202 Å observed with EIS, dominated by the intense Fe XIII lines. Fe XIII is a particular ion that the corresponding emissivity curves are density-sensitive only, i.e., do not depend on temperature. This reduces the dimension of the parameter space in the statistical analysis and simplifies the problem greatly. By attaching reasonable uncertainties to the Fe XIII atomic data, atomic physicists developed simple models for the uncertainties in some of the atomic parameters for Fe XIII and generated realizations of the plasma emissivities that capture this uncertainty. We can then apply this ensemble of emissivity curves to analyze the density sensitive Fe XIII spectral lines (Yu et al. 2018). In short, we firstly characterize the uncertainties of the atomic data. Following that, we aim to incorporate the atomic uncertainty and to infer the plasma parameters, in particular, plasma densities, in a spectral analysis using Bayesian statistical techniques.

Fe XVII is a popular line system for temperature measurements where its density sensitivity is pretty much ignorable (Brickhouse et al. 1995). This is another simplified case with reduced dimension of the parameter space in the statistical analysis. Because of the temperature dependence, we have to consider differential emission measures (DEMs) which are a summary of the temperature structure of the coronae of star (Brickhouse et al. 2000). The ensemble of Fe XVII emissivity curves can first be generated in the same way as Fe XIII. Following that, the models and algorithms used for Fe XIII analysis above is also applicable to Fe XVII in order to incorporate the Fe XVII

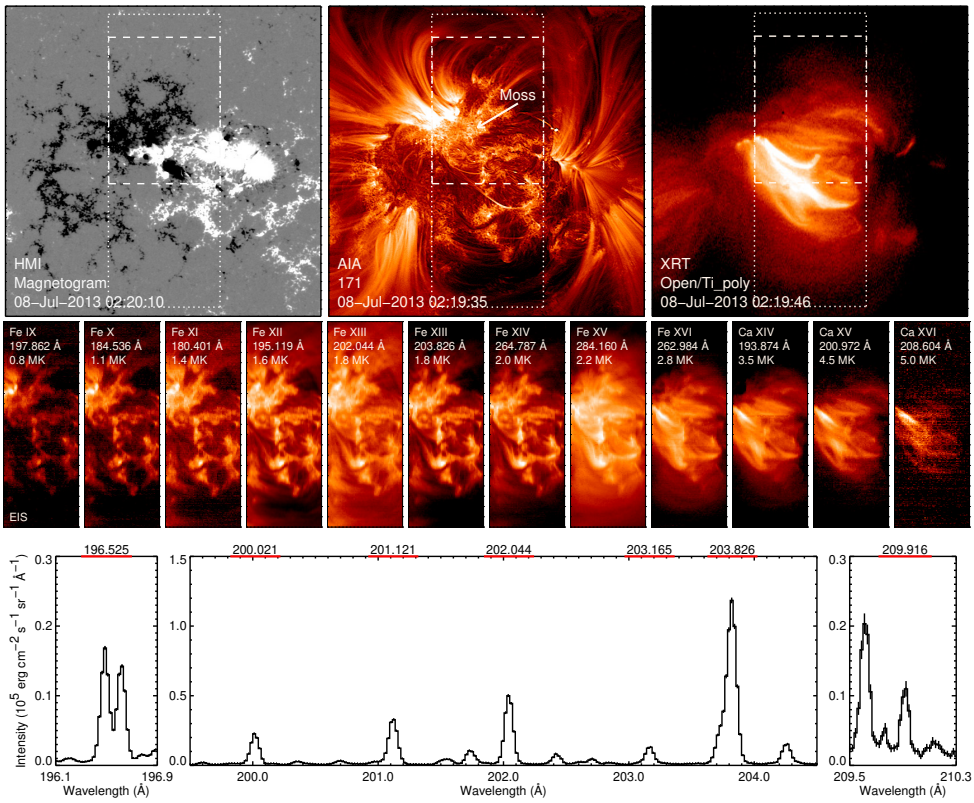

Figure 1.1: Observations of NOAA active region 11785 on 8 July 2013 near 2 UT (Universal Time). The top panels show the photospheric magnetic field measured with HMI instrument, million degree emission observed with the AIA instrument in Fe IX 171 Å channel, and high temperature loops observed with XRT instrument. The middle panels show EIS rasters in a series of emission lines that range in temperature of formation from 0.8 MK to 5.0 MK. The bottom panel shows an EIS spectrum near 202 Å from a single spatial pixel with the Fe XIII lines of interest highlighted. The EIS rasters are from an observation that began at 01:55 UT, and this field of view is indicated by the dotted lines in the top panels. The EIS full CCD spectrum is from an observation that began at 00:20 UT; this field of view is indicated by the dashed lines in the top panels.

atomic uncertainty and to infer the plasma parameters, in particular, plasma temperature.

We then move onto a more complicated case. We aim to estimate both the coronal density and temperature of Capella using the O VII line systems in the soft X-ray regime over the course of the *Chandra* mission (Mewe et al. 2001). The Capella, a quadruple star system organized in two binary pairs, is one of the brightest stellar X-ray sources in the sky. Unlike the Fe XIII and Fe XVII emissivity curves, the O VII emissivity curves are dependent on both density and temperature. Besides atomic uncertainties, density estimates must be conditional on temperature estimates as well, which comes from the output information of Fe XVII spectral analysis. In summary, the uncertainties of the O VII atomic data is characterized first. Secondly, the uncertainties of temperature is characterized from the above Fe XVII spectral analysis. Lastly, we combine measures of uncertainty due to the O VII atomic data with all kinds of statistical uncertainties and explore the effect of different emission measure distributions in order to derive meaningful uncertainties on the plasma density and temperature on the coronae of Capella.

1.1.4 Summary

In conclusion, the solar and stellar coronae are complex and dynamic systems. Instead of being measured directly, information about the physical properties of astrophysical objects can be inferred by interpreting spectral observations in relation to the atomic physics calculations. Specifically, the relative intensity of the emission lines depends on both the electron density and temperature of the emitting plasma. Emissivity curves can be used as a bridge to relate the expected intensity of the emission lines as a function of the electron density and temperature. Understanding how uncertainties in the underlying atomic physics propagates to the uncertainties in the inferred plasma parameters is an essential component of this analysis. Existing methodologies cannot use uncertainties on the atomic parameters which are not even included in the atomic physics databases.

We propose a robust and principled statistical method to deal with such problems with different complexities in this thesis. We start with two families of spectral lines, the Fe XIII spectral lines that are not sensitive to the temperature and the Fe XVII spectral lines that are not sensitive to the density. In both cases, the emissivities are in lower dimensions and as is the model parameters, simplifying the analysis greatly. Another case study on the O VII spectral lines follows to derive complete plasma parameter estimations.

1.2 Multistage analysis

Even though it is well known that the measurements of the physical properties, e.g., emissivities, have associated measurement uncertainties and it is recognized that these uncertainties can cause large systematic errors in the inferred plasma parameters of the spectral model, the uncertainties are often based on experience, with only nominal estimates used in data analysis. The astrophysicists often use the standard approach that treats the measurements as fixed and known, e.g., does not account for uncertainties in emissivities and obtains the best-fit values of the parameters. We call this approach the *standard method*. However, this method can produce biased estimates of the model parameters and can underestimate the error bars associated with these estimates significantly. Such biases and underestimations can result in inaccurate, even incorrect, interpretations of the analytical results.

There are always multiple sources of information affecting parameter estimations in complex problems. Considering only one source of information may not be good enough to obtain parameter estimations and corresponding error bars accurately. Solving complex problems is often divided into multiple stages depending on the source of information. Output information obtained in earlier stages may constrain parameter estimations in all later stages.

In an example discussed above, in a first stage, the calibration scientists learned about the calibration uncertainties with effective area curves. In a second stage, Lee et al. (2011) and Xu et al. (2014) presented general Bayesian statistical methods to sample from a distribution of plausible calibration curves and to incorporate calibration uncertainties into a comprehen-

sive spectral analysis of high-energy *Chandra* spectra. These two stages have reduced the bias in the estimation of model parameters and have improved the accuracy of the variance of these estimates. In particular, in the second stage, they also compared different statistical frameworks, the standard method, the pragmatic Bayesian method that does not allows the spectra data to provide any information for the calibration products, and the fully Bayesian method that allows the spectra data to provide information for both estimation of model parameters and the calibration products. Jones et al. (2015) developed a Bayesian statistical method that first probabilistically assigns photons to different sources. They then fitted spectral analyses and estimated the parameters describing each of the individual sources accounting for the uncertainty in the photon allocation. They presented a two-stage analysis that combines both spatial information and spectral information to separate photons among the sources and increases the accuracy with which the parameters are inferred. Blocker et al. (2013) demonstrated that carrying out multiple procedures sequentially and properly, where each procedure taking the output of the previous procedure as its input, can surpass single-procedure estimations in efficiency and robustness.

Along the same theme, here we design a robust principled method, as a common language or a standard procedure, to improve the accuracy of model parameter inference when the output from one or several preliminary analyses is required for a following primary analysis. We refer to this approach as the *multistage analysis* to carry forward the output from a *preliminary* stage to a *primary* stage. The statistical and systematic uncertainties of estimated quantities, especially high dimensional quantities, can be carried forward into subsequent analyses easily. A general two-stage analysis is specified and the detailed generic statistical methods are discussed in Section 3.2. In general, the physicists deploy a simple, but realistic, model to describe the uncertainties in some of the atomic parameters, use it to generate sets of different realizations of the plasma emissivities in a preliminary stage. We then apply this ensemble of atomic data to analyze the spectral lines observed in solar or stellar active region using a complex statistical framework in the primary stage. In summary, combining an ensemble of emissivities with observed

spectral data to account for uncertainties in atomic data fits a two-stage methodology well. Applications of a two-stage analysis via case studies in different ions are described in Chapter 4 and Chapter 5 respectively. A more complicated three-stage analysis, where more sources of uncertainties are considered, and its application are described in Chapter 6. The big picture for the application of the multistage analysis on astrophysical problems with different complexities is shown in Figure 1.2.

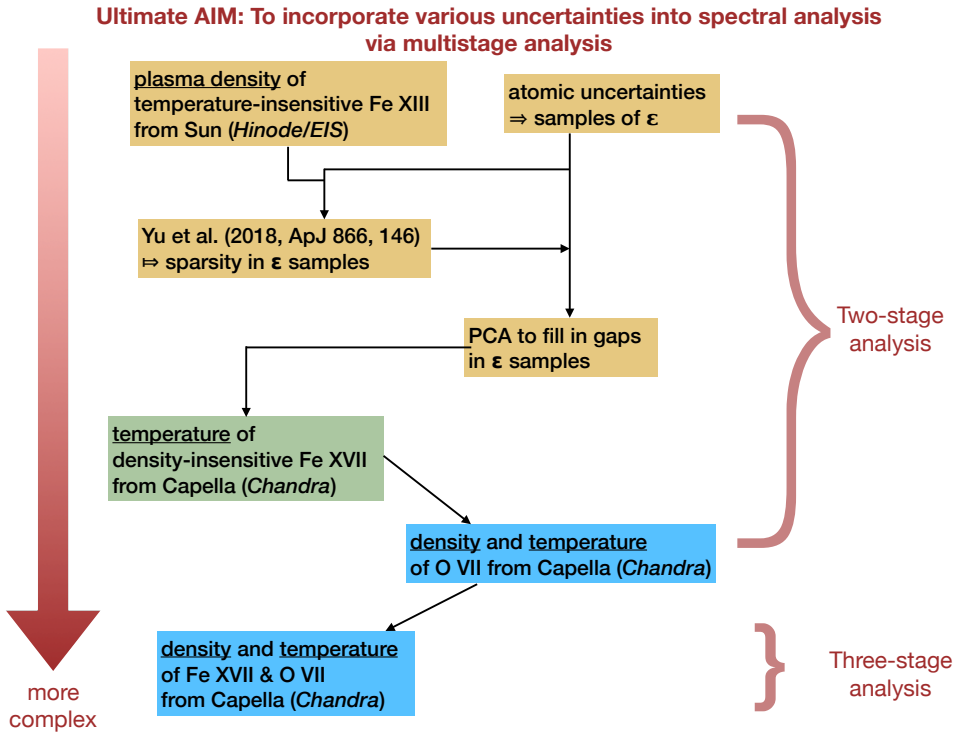

Figure 1.2: The big picture for the astrophysical problems with different complexities discussed in this thesis.

1.3 Astrophysical terminology

We have assumed that we can definitely count the observed photons emitted at discrete energies and associated with each spectral line, which is not possible in practice. If we have counts for each line of interest, we can learn about it easily. Unfortunately, photon wavelengths are recorded with errors

resulting in the difficulty in the learning process. Therefore, there are two corrections needed for the telescopes.

Effective area When the X-ray beam reaches the telescope mirror, the light is partially reflected at the surface and the rest is absorbed. The effective area determines the expected number of detected photons (i.e., how much light) from a source of constant intensity. However, effective area is strongly dependent on wavelength due to the reflectance variation of telescope mirror. The Chandra X-Ray Observatory is equipped with two gratings/detectors which split and diffract/refract the X-ray beam into different directions depending on the wavelength. Thus, enabling us to distribute photon counts from the observed wavelength to the true wavelength region by substituting spatial resolution for spectral resolution. In addition, the effective area of the telescope depends on the observation date since they have degraded over time (approx 20 years). For the same object of interest, the effective area varies significantly from the beginning to the end. Thus, observation identification need to be specified. In summary, the effective area depends on many factors including the wavelength of the beam, the grating, and the moment when the observation was taken place. Because of those dependencies, in order to estimate the spectral or the relative intensity of the lines of recorded wavelength with effective area, corrections are required (Chen et al. 2019).

The line response function As we cannot achieve perfect focus on the telescope with X-rays, the beam of X-ray spreads out when it goes through the grating/detector. Wavelength is therefore recorded with error. In this case, the line response function (LRF) needs to be taken into account to avoid an inaccurate setup of physics. Under each grating and dispersed order combination, LRF collects photons from nearby channels to obtain the intensity of wavelength. In probability, it is the probability density function or probability mass function of a conditional distribution of the recorded wavelength, w, given the true wavelength, ω, for grating g and dispersed order o, as denoted by $R_{go}(\mathrm{w}|\omega)$. In this thesis, the line response function for the HRC-S+LETGS grating combination can be approximated as (Schwartz

2014, Chandra 2019, Chen et al. 2019)

$$R_{go}(\mathrm{w}|\omega) = \frac{\Gamma(\frac{\nu+1}{2})}{\sigma\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left[1+\frac{(\mathrm{w}-\omega)^2}{\nu\sigma^2}\right]^{-\frac{\nu+1}{2}}, \tag{1.1}$$

which is of the form a generalized t-distribution with degrees of freedom $\nu \equiv 4$, location parameter ω, and scale parameter σ, and the latter two could be different for different lines of interest.

1.4 Contributions and outline

The contribution of this thesis to statistics and astrophysics can be summarized as follows. Statistically, we have proposed a principled multistage analysis, which is applicable to the problems under the setting where the output from one or several preliminary analyses is required for a following primary analysis. The statistical and systematic uncertainties of estimated quantities, especially high dimensional quantities, can be carried forward into subsequent analyses easily. Astrophysically, this robust principled method is able to account for atomic uncertainty in the spectral analysis

We aim to combine measures of uncertainty due to atomic data with statistical uncertainty in spectral data and explore the interaction of both uncertainties in a spectral analysis in order to derive meaningful uncertainties on the solar and stellar coronal plasma parameters like electron density and temperature. A critical component of this analysis is understanding how uncertainties in the underlying atomic physics in a preliminary analysis propagate to uncertainties in the inferred plasma parameters in a primary analysis. We apply the model-generated atomic data uncertainties in the emissivities of different ions, based on lab measurements and different calculation schemes, to the estimation of plasma parameters, based on the observed spectral lines of different ions.

We implement both the pragmatic and the fully Bayesian approaches to interpret the spectral observations in the context of emissivity realizations of atomic data. The former generally increases the uncertainties on the inferred

parameters compared with models that incorporate only statistical uncertainties and is a simplified model providing conservative parameter estimations. In contrast, the latter, as a more complete model, reduces the uncertainties on the inferred parameters. To incorporate uncertainties, we summarize sample of uncertainties by treating them equally likely or using a principal component analysis, a Gaussian distribution approximation, a t distribution approximation, and kernel density estimations. A general two-stage analysis, generic Bayesian statistical methods including the pragmatic and the fully Bayesian methods, and the efficient statistical computation algorithms are discussed.

With simulation studies and real data analyses conducted on different ions, Fe XIII and Fe XVII, we demonstrate that the fully Bayesian method can provide precise and accurate parameter estimations than the pragmatic Bayesian method and prove the feasibility of two-stage analysis, while the pragmatic Bayesian method, as a simplified but well-established model, is easy to be implemented and provides conservative parameter estimations with bias and large error bars. The pragmatic Bayesian method is still worth exploring because, ordinarily, when the emissivities are passed to the statisticians who have limited knowledge on the atomic physics, the model might be set up incorrectly leading to certain amount of bias and large error bars. We also work on a more complicated spectral model of O VII considering one more parameter and its corresponding source of uncertainties, and generalize our analysis into three stages. This principled multistage analysis has the potential to be applied on complicated models with a variety of sources of uncertainties. The pragmatic Bayesian method is preferred by the astrophysicists because the physical properties of the solar and stellar coronae, e.g., plasma parameters, might not take a single value but follows a distribution so that conservative estimations with larger error bars are helpful for downstream research. The fully Bayesian method is preferred by the statisticians because the parameter estimations are precise and accurate.

In Chapter 2, basic Bayesian inference and computation techniques are reviewed. In Chapter 3, we discuss a general two-stage analysis, the detailed generic Bayesian statistical methods and computation algorithms, and the

Acronym	Description
BMA	Bayesian Model Averaging
DEM	Differential emission measures
EIS	EUV Imaging Spectrometer
EUV or UV	Extreme-ultraviolet or ultraviolet
FWHM	Full width at half maximum
HMC	Hamiltonian Monte Carlo
KDE	kernel density estimation
LRF	Line response function
MAP	Maximum a posteriori
MC	Monte Carlo
MCMC	Markov chain Monte Carlo
MH	Metropolis-Hastings
MSE	mean square errors
MVN	Multivariate normal
NUTS	No-U-Turn Sampler
PCA	Principal Component Analysis
RMSE	Root mean square error
TQR	Trapezoidal quadrature rule

Table 1.1: Glossary of Acronyms used in the text.

quantification methods for uncertainties of estimated quantities. Applications of a two-stage analysis via case studies in Fe XIII and Fe XVII are described in Chapter 4 and Chapter 5 respectively. A more complicated three-stage analysis, where more sources of uncertainties are considered, and its application on Fe XVII and O VII are described in Chapter 6. In the end, a general three-stage analysis and the potential of a multistage analysis are discussed in Chapter 7. A glossary of the terms and acronyms that we use is given in Table 1.1.

2

Bayesian Inference and Statistical Computation

To account for atomic uncertainties and all other source of uncertainties in a principled multistage analysis, we adopt Bayesian methods throughout the thesis. Basic Bayesian inference and computation techniques are reviewed in this chapter.

2.1 Bayesian inference

We take a Bayesian approach in our statistical analysis because it enables us to build in the complex hierarchical dependencies engendered by uncertainties of estimated quantities, like atomic uncertainties. Such an approach offers a probability-based formalism for combining information from our prior knowledge and the current data. This requires both a *prior distribution*, which quantifies the uncertainty in the values of the unknown model parameters before the data is observed, and a *likelihood function*, which describes the distribution of the data given the model parameters. The likelihood function allows us to assess the viability of a parameter value given the observed

data under a proposed statistical model. The likelihood function is combined with the prior distribution to yield the *posterior distribution*, which quantifies the uncertainty in the values of the unknown model parameters taking account of the observed data. If we let X and ψ represent generic data and unknown model parameters, respectively, Bayes' theorem provides the posterior distribution as

$$p(\psi|X) = \frac{L(\psi|X)\ p(\psi)}{p(X)}, \tag{2.1}$$

where $L(\psi|X)$ is the likelihood of X given ψ (sometimes written as $p(X|\psi)$) and $p(\psi)$ the prior distribution of ψ. The term $p(X)$ is a normalizing constant necessary to make $p(\psi|X)$ a proper probability distribution, however, it might be difficult but unnecessary to compute. (The term $p(X)$ is sometimes referred to as the 'evidence' in the astrophysics literature.) The posterior distribution, which combines information in the data with our prior knowledge, is our primary statistical tool for deriving parameter estimates and the corresponding uncertainties. Specifically, in the multistage analysis, the prior distribution in the primary analysis is replaced with the posterior distribution from the preliminary analysis.

2.2 MCMC Bayesian computation

When a direct sampling is difficult or computationally inefficient, *Markov chain Monte Carlo* (MCMC) is a popular numerical method to use. It involves, iteratively, proposing samples from a proposal distribution and filtering the samples so that the distribution of these samples matches the target posterior distribution which is often complicated or in high dimensions. When updating iteratively, a Markov chain is formed where the conditional distribution of the new sample given the full history of the samples depends only on the most recent sample. This conditional distribution preserves the target posterior distribution. It should be designed to be easy to sample and to ensure the convergence of the sample to the target posterior distribution. When sampling long enough, the distribution of the current samples is close

enough to the target posterior distribution.

In this section, we review some basic MCMC algorithms to generate a sample from the posterior distributions under the models discussed in this thesis, including the Metropolis-Hastings and the Metropolis algorithms (Metropolis et al. 1953, Hastings 1970), the adaptive Metropolis algorithm (Haario et al. 2001, Roberts & Rosenthal 2009), Gibbs Sampler (Geman & Geman 1984), and Hamiltonian Monte Carlo (Duane et al. 1987, Neal et al. 2011, Gelman et al. 2014). We always start these algorithms with random starting values and keep sampling draws until they converge to the target posterior distribution. The draws sampled from these algorithms at early iterations may not necessarily be an appropriate representation of the target posterior distribution, which therefore are always discarded as the burn-in. Convergence diagnostics is also necessary by running multiple chains with random starting values until the chains after burn-in have mixed well and the distributions of the samples between and within chains are identical. Moreover, the computation efficiency can typically be improved by combining different algorithms when dealing with more complex models.

2.2.1 The Metropolis-Hastings and the Metropolis algorithms

The Metropolis-Hastings (MH) algorithm (e.g., Metropolis et al. 1953, Hastings 1970) is a general term for a family of Markov chain simulation methods that are typically useful for sampling from Bayesian posterior distributions. Let $p(\psi|X)$ be the probability density function of a d-dimensional target posterior distribution, using the notations in Section 2.1. A proposed ψ^* is sampled from a proposal distribution $q(\psi^*|\psi^{(t)})$ at iteration $t+1$. Calculating the acceptance ratio,

$$\rho = \frac{p(\psi^*|X)\ q(\psi^{(t)}|\psi^*)}{p(\psi^{(t)}|X)\ q(\psi^*|\psi^{(t)})},$$

we accept the proposed value and set $\psi^{(t+1)} = \psi^*$ with probability $\min(\rho, 1)$. Otherwise, we reject the proposed value and set $\psi^{(t+1)} = \psi^{(t)}$.

The Metropolis algorithm is a special case of the MH algorithm when symmetry property is attached to the proposal distribution, i.e., $q(\psi^*|\psi) = q(\psi|\psi^*)$

for all ψ^* and ψ. Then the acceptance ratio simplifies to, at iteration $t+1$,

$$\rho = \frac{p(\psi^*|X)}{p(\psi^{(t)}|X)}.$$

Thus the Metropolis algorithm save a little time in calculating the acceptance ratio ρ but otherwise have no advantages over the MH algorithm.

The choice of the proposal distribution is important for achieving rapid convergence in the MH algorithm. Finding an optimal proposal for a particular target distribution is both crucial and challenging. We focus on a symmetric random-walk Metropolis algorithm (RWM), where the proposal at iteration $t+1$ is given by $\psi^* = \psi^{(t)} + e^{(t+1)}$, where $\{e^{(t)}\}$ is independent and identically distributed (i.i.d.) from proposal increment distribution $\mathcal{N}(0, \sigma^2 I_d)$ with scaling parameter $\sigma > 0$. Too small of a value for σ will slow down the convergence of the chain, whereas too large of a value for σ will result in a higher chance of rejections for the new proposal. To prevent both the extreme cases, numerical studies have been made in identifying appropriate proposal scalings σ to optimise the resulting MCMC algorithm (see e.g., Gelman et al. 1996, Roberts et al. 1997, Roberts & Rosenthal 2001). Assuming that each component can be sampled separately from its one-dimensional distribution and the proposal increment distributions are of i.i.d normal, the optimal acceptance rate is approximately 0.234 when $d > 2$ and approximately 0.440 when $d = 1$. The corresponding optimal proposal scalings σ is typically selected manually from a fine grid on the values of possible scalings, which is often successful but generally time-consuming.

2.2.2 Adaptive Metropolis

Determining the optimal proposal scalings in the MH algorithm is often done manually and can be problematic in high dimensions. An alternative approach is adaptive MCMC, which can automatically tune the optimal proposal scalings and learn the better parameter estimation on the fly, resulting in efficient convergence. Specifically, the choice of the proposal distribution depends on the entire history to improve convergence. In this setting, as

the conditional distribution of the new sample given the full history of the samples not only depends on the most recent draw, the sequence of sample from adaptive MCMC algorithms is not forming a Markov chain anymore.

Assuming the Diminishing Adaption condition, i.e., two successive transition kernels are similar, and the Boundary Convergence condition, i.e., ergodicity of transition kernels, Roberts & Rosenthal (2007) proved the asymptotic convergence and the weak law of large numbers to guarantee the convergence in the adaptive proposal distribution to the target distribution.

We consider an important use of adaptive MCMC, the adaptive Metropolis algorithm of Haario et al. (2001) and Roberts & Rosenthal (2009). We use the same d-dimensional target posterior distribution $p(\psi|X)$. A Metropolis algorithm is performed with a multivariate normal (MVN) distribution as proposal distribution at iteration $\ell \leq L'$, whereas at iteration $\ell > L'$, a mixture of two MVN distributions is used, where $L'(\geq 2d)$ is the initial number of runs. Specifically, at iteration ℓ, we sample $\psi^* = (\psi_1^*, \cdots, \psi_d^*)$ from a proposal distribution

$$q(\psi^* \mid \psi^{(\ell-1)}) = \begin{cases} \mathcal{MVN}(\psi^* \mid \psi^{(\ell-1)}, \frac{0.1^2}{d}\Sigma_0), & \text{if } \ell \leq L', \\ (1-\beta) \cdot \mathcal{MVN}(\psi^* \mid \psi^{(\ell-1)}, \frac{2.38^2}{d}\Sigma^{(\ell)}) + & \\ \qquad \beta \cdot \mathcal{MVN}(\psi^* \mid \psi^{(\ell-1)}, \frac{0.1^2}{d}\Sigma_0), & \text{if } \ell > L', \end{cases} \tag{2.2}$$

with $\ell > L'$

$$\Sigma^{(\ell)} = \begin{cases} \begin{pmatrix} \text{var}(\{\psi_1^{(\cdot)}\}_1^{\ell-1}) & \cdots & \text{cov}(\{\psi_1^{(\cdot)}\}_1^{\ell-1}, \{\psi_d^{(\cdot)}\}_1^{\ell-1}) \\ \vdots & \ddots & \vdots \\ \text{cov}(\{\psi_d^{(\cdot)}\}_1^{\ell-1}, \{\psi_1^{(\cdot)}\}_1^{\ell-1}) & \cdots & \text{var}(\{\psi_d^{(\cdot)}\}_1^{\ell-1}) \end{pmatrix}, & \\ & \text{if } \ell \text{ is a multiple of } \mathcal{L}, \\ \Sigma^{(\ell-1)}, & \text{otherwise}, \end{cases} \tag{2.3}$$

where, if ℓ is a multiple of $\mathcal{L}$, $\Sigma^{(\ell)}$ is equal to the current empirical estimate of the variance-covariance matrix based on the previous runs, $\psi^{(1)}, \cdots, \psi^{(\ell-1)}$, so far. $\Sigma^{(\ell)}$ is used to estimate the optimal Σ of the target distribution and

is updated at every $\mathcal{L}$ draws to prevent unnecessary computing. β is a small positive constant (we take $\beta = 0.05$). Σ_0 is prior variance matrix of the unknown model parameters.

We initially use a Metropolis algorithm with a fixed MVN proposal distribution for the first L' iterations when the empirical covariance $\Sigma^{(\ell)}$ is not yet well-defined. A mixture proposal distribution is used in the following adaptation once the empirical covariance is well-defined. According to Roberts et al. (1997) and Roberts & Rosenthal (2001), it is approximately optimal to have a proposal distribution $\mathcal{MVN}(\psi^* \mid \psi^{(\ell-1)}, \frac{2.38^2}{d}\Sigma^{(\ell)})$ in a particular high-dimensional setting. The mixture with a fixed MVN, $MVN(\psi^* \mid \psi^{(\ell-1)}, \frac{0.1^2}{d}\Sigma_0)$, in proposal is a safety mechanism to avoid the adaptive Metropolis algorithm getting stuck at singular or other problematic values of $\Sigma^{(\ell)}$ (Roberts & Rosenthal 2009). Though it may require a large number of samples before the adaptation starts to improve the algorithm significantly, the adaptive Metropolis will eventually find a good algorithm in a large dimension which could never be done manually.

2.2.3 Gibbs Sampler

Another alternative approach to sampling from a high dimensional distribution is Gibbs Sampler (Geman & Geman 1984) where the conditional posterior distributions of the parameters are in good format. Instead of sampling directly from the posterior distribution itself, we iteratively generate posterior samples from the conditional posterior distribution of one of the random variables at a time or of one subset of random variables at a time given the current values of all the remaining random variables. We use the same d-dimensional target posterior distribution $p(\psi|X)$. Let $\psi = (\psi_1, \psi_2)$ where ψ_1 and ψ_2 can be either a single random variable or a subset of ψ with length d_1 and d_2 respectively, and $d = d_1 + d_2$. A general two-step Gibbs sampler proceeds as, at iteration $\ell + 1$,

Step 1: Sample $\psi_1^{(\ell+1)} \sim p(\psi_1 \mid \psi_2^{(\ell)})$,

Step 2: Sample $\psi_2^{(\ell+1)} \sim p(\psi_2 \mid \psi_1^{(\ell+1)})$.

The same mechanism is applicable to multistep Gibbs sampler where each subset of random variables is updated conditional on the current values of all the other variables, which equal the iteration $\ell+1$ values for the variables already updated and the iteration ℓ values for the others. Those draws sampled from the conditional distributions have the same distribution as if they were sampled from the target joint posterior distribution. Moreover, introducing auxiliary variables or reparameterizations can often simplify or accelerate the computations of the Gibbs sampler.

The Gibbs sampler is also a special case of the MH algorithm. To update ψ_2, we factor the target density $p(\psi_1, \psi_2) = p(\psi_1)p(\psi_2 \mid \psi_1)$ and choose $p(\psi_2 \mid \psi_1)$ as a proposal distribution. The new proposal is (ψ_1, ψ_2^*). The acceptance ratio becomes

$$\rho = \frac{p(\psi_1, \psi_2^*)p(\psi_2 \mid \psi_1)}{p(\psi_1, \psi_2)p(\psi_2^* \mid \psi_1)} = \frac{p(\psi_1)p(\psi_2^* \mid \psi_1)p(\psi_2 \mid \psi_1)}{p(\psi_1)p(\psi_2 \mid \psi_1)p(\psi_2^* \mid \psi_1)} = 1, \tag{2.4}$$

indicating the proposal is always accepted. The same rule works when updating ψ_1.

2.2.4 Hamiltonian Monte Carlo

Another efficient approach to sampling from a complicated and high dimensional distribution is Hamiltonian Monte Carlo (HMC) that introduces auxiliary momentum variables, ξ, to enable the transitions to move efficiently through the parameter space and to transform the problem of sampling from a target density distribution $p(\psi)$, typically a Bayesian posterior $p(\psi|X)$, into the problem of simulating *Hamiltonian dynamics* via the first-order gradient of the density function (Duane et al. 1987, Neal et al. 2011, Gelman et al. 2014).

The auxiliary momentum variables, ξ, typically follow a multivariate normal distribution that is independent of the model parameters, ψ,

$$\xi \sim \mathcal{MVN}(0, \Sigma),$$

where the variance-covariance matrix Σ is used to scale the target distribution.

The *Hamiltonian* function in HMC is defined as

$$\begin{aligned} H(\xi, \psi) &= -\log p(\xi) - \log p(\psi) \\ &= K(\xi) + V(\psi), \end{aligned}$$

where $K(\xi) = -\log p(\xi) = \frac{1}{2}\xi^T \Sigma^{-1} \xi$ is called the *kinetic energy* and $V(\psi) = -\log p(\psi)$ is called the *potential energy*, typically the negative log target posterior distribution on ψ.

The evolvement of the auxiliary momentum variables and the model parameters, (ξ, ψ), over time t is via *Hamilton's equations*,

$$\begin{aligned} \frac{d\psi}{dt} &= \frac{dH}{d\xi} = \frac{dK}{d\xi} = \Sigma^{-1}\xi, \\ \frac{d\xi}{dt} &= -\frac{dH}{d\psi} = -\frac{dV}{d\psi}. \end{aligned}$$

The *Leapfrog Method*, a numerical integration algorithm, is used, at each iteration, to approximate the solution to the above Hamiltonian system of differential equations by discretizing time into T discrete steps of some small stepsize, ϵ. At each step, the Leapfrog algorithm alternates a half-step update of the momentum, a full-step update of the model parameters using the new values for the momentum, and another half-step update of the momentum using the new values for the model parameters. This can be summarized as follows,

$$\begin{aligned} \xi &\leftarrow \xi - \frac{\epsilon}{2}\frac{\partial V}{\partial \psi}, \\ \psi &\leftarrow \psi + \epsilon \Sigma^{-1} \xi, \\ \xi &\leftarrow \xi - \frac{\epsilon}{2}\frac{\partial V}{\partial \psi}. \end{aligned}$$

Repeating the above three steps for the T steps, the last state, denoted by (ξ^*, ψ^*), is treated as a new proposed state at the current iteration. Ap-

plying a Metropolis algorithm and calculating the acceptance ratio, the new proposed state is accepted with probability

$$\min(\exp(H(\xi, \psi) - H(\xi^*, \psi^*)), 1),$$

otherwise, the previous state is kept.

In summary, starting with specifying an initial set of model parameter values, the detailed HMC algorithm proceeds for iteration $\ell = 2, \ldots, L$ with

Step 1: Sample a new momentum $\xi_0 \sim \mathcal{MVN}(0, \Sigma)$.

Step 2: Update the momentum and the model parameters using the Leapfrog algorithm with T discrete sub-steps and stepsize ϵ, and label the value at the Tth sub-step (last sub-step) as a new proposed state (ξ^*, ψ^*).

Step 3: A Metropolis accept-reject step. Set

$$\psi^{(\ell)} = \begin{cases} \psi^*, & \text{with probability} \\ & \min(\exp(H(\xi^{(\ell-1)}, \psi^{(\ell-1)}) - H(\xi^*, \psi^*)), 1), \\ \psi^{(\ell-1)}, & \text{otherwise.} \end{cases}$$

Practically there are three parameters need to be tuned to improve sampling efficiency for HMC, number of steps T, stepsize ϵ, and covariance matrix Σ, while these tuning processes are extremely difficult, time consuming, and even problematic (Neal et al. 2011, Gelman et al. 2014). If T is too small, the trace of the trajectory will be too short and the algorithm will devolve to an unwanted random walk behaviour with large statistical errors. If T is too large, the algorithm will do unnecessary work. Hoffman & Gelman (2014) presented the No-U-Turn Sampler (NUTS), an extension of HMC, that provides an automatic selection on the number of Leapfrog steps at each iteration to generate efficient transitions traversing the posterior. Rather than fixing a value for the number of steps T, in NUTS, the trajectory at each iteration moves both forwards and backwards preserving the detailed balance until a U-turn appears. Mathematically, the trajectory is terminated

when the angle between the momentum at the final step and the distance traveled by the model parameters across all the steps since the start of the current iteration is more than 90 degrees, i.e., a negative value of their dot product is obtained. This criterion essentially gets the trajectory move as far as possible at that iteration without doing unnecessary work. Standard HMC uses a fixed step size and number of steps, while NUTS figures out the step size ϵ and the number of steps T during the warm-up stage. NUTS is self-tuning and computationally efficient but has trouble in jumping from the centre to the tails if the distribution is heavy or light tailed.

3

Model Uncertainties in Atomic Data

In this chapter we describe uncertainties in atomic data and the generic methodology for combining uncertainties of estimated quantities into primary data analysis. We illustrate the concept of a two-stage analysis, and review generic Bayesian statistical methods including standard method, multiple imputation, the pragmatic and the fully Bayesian methods. There are two ways deployed to incorporate uncertainties in atomic data into our highly structured statistical model. A discrete analysis where the atomic realizations are considered individually and a continuous analysis where PCA is used to fully summarize the atomic uncertainties via a multivariate Gaussian distribution. Efficient statistical computation algorithms are also discussed.

3.1 Uncertainties in atomic data

Systematic differences in atomic data compilations have real universal consequences and affect the underlying spectral models (Mernier et al. 2020). According to Yu et al. (2018), the collisional excitation and spontaneous decay rates are two important atomic rates that affect the intensity of a spectral line. They have modified the standard CHIANTI routines (Freeland

& Handy 1998) to produce realizations of emissivities based on the uncertainties in the collisional excitation and spontaneous decay rates in order to incorporate the atomic data uncertainties.

Here is an example of generating realizations of the Fe XIII emissivities. Each rate is randomly sampled from a normal distribution within the estimated uncertainty. We then used the standard CHIANTI routine to calculate the emissivities for the spectral lines. There are a total of 1000 realizations of the Fe XIII emissivities generated for each line, as shown in Figure 3.1. The figure noticeably demonstrates how the spectral lines vary their emissivities as a function of the density.

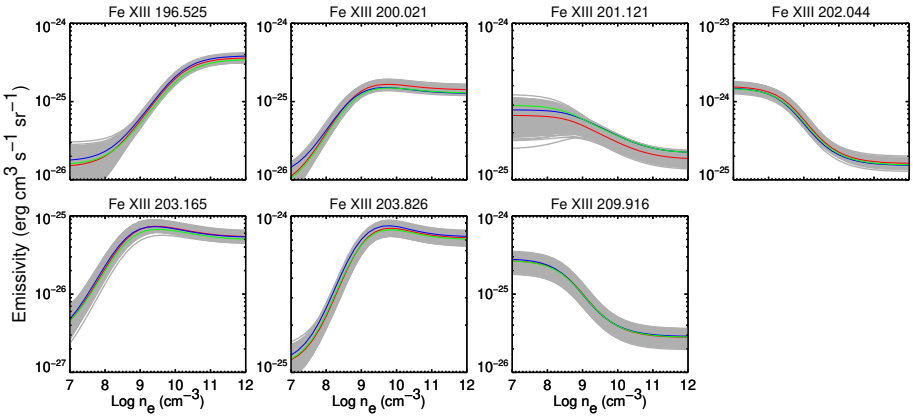

Figure 3.1: Emissivities of the seven Fe XIII lines considered in Chapter 4. The emissivities are computed assuming a temperature of 1.8 MK, the temperature of formation for Fe XIII. The grey lines represent the 1000 realizations of the CHIANTI atomic data. The red curve is the default value from CHIANTI v.8. As discussed in Chapter 4, the blue curve is identified as being most likely match to the observations (#471) and the green curve the second most probable (#368).

In summary, by attaching reasonable uncertainties to the atomic data we are able to produce realizations of the emissivities that incorporate this uncertainty. We can then use the ensemble of emissivities to characterize the uncertainties of the atomic data to infer physical parameters like plasma densities in a spectral analysis. Our methodology for combining an ensemble of emissivities with observed data to account for uncertainties in atomic data is discussed in detail in the following sections and chapters.

3.2 Two-stage analysis and generic statistical methods

Our ultimate aim is to account for the statistical and systematic uncertainties of estimated quantities and to carry them into a subsequent analysis. Specifically, the aim is to account for the effect of uncertainties in the atomic data and propagate them to the analysis of solar and stellar observations to the determination of plasma parameters. This can generally be summarized as a two-stage analysis, as described in Section 1.2. In the preliminary analysis, atomic physicists are learning about the emissivity curves, which quantify the expected intensities of spectral lines change as a function of the plasma parameters, and are using the CHIANTI routine to calculate the emissivities (see, e.g., Del Zanna & Storey 2012, Del Zanna et al. 2015, Yu et al. 2018). Statistically, this ends up with a Monte Carlo (MC) sample of the emissivities. In the primary analysis, we take those emissivity curves obtained in the preliminary analysis forward into the spectral analysis to learn about the plasma parameters. In this section we are going to propose a general framework for a two-stage analysis and compare four detailed generic methods for principled statistical inference: the standard method, multiple imputation, the pragmatic, and the fully Bayesian methods. The latter two Bayesian methods are implemented in a novel way in this thesis. Case studies on specific ions with different model complexities are used as running examples in this thesis, Fe XIII in Chapter 4, Fe XVII in Chapter 5, and Fe XVII & O VII in Chapter 6.

3.2.1 Two-stage analysis

To fix ideas, we focus attention on a general two-stage analysis problem. In the preliminary stage, the dataset, X_1, is modelled with respect to the unknown parameters, ψ_0 and ψ_1. In the primary stage, the dataset, X_2, is modelled with respect to the unknown parameters, ψ_1 and ψ_2. Note that ψ_1 represents an unknown parameter or a set of unknown parameters that is common between the two models in a two-stage analysis. The output information, typically estimates or posterior distributions, for ψ_1 in the preliminary stage is required for and will be carried forward into the primary

stage. In short, denoting the data and unknown parameter information by "data | parameter" in each stage, a two-stage analysis can be summarized by

$$\text{Preliminary stage: } X_1 \mid \psi_0,\psi_1, \tag{3.1}$$

$$\text{Primary stage: } X_2 \mid \psi_1, \psi_2. \tag{3.2}$$

As a general framework, in principle, either of the two datasets could be unavailable to different groups of researchers and the parameter estimation is based on the other dataset. We aim to fit the parameter, ψ_2, via the analysis of the observed dataset, X_2, while still accounting for the parameter, ψ_1, which can be quantified/estimated in the preliminary analysis. Practically we are focusing on the parameter estimation with respect to the dataset X_2, given the unavailability of the dataset X_1, in a two-stage analysis.

3.2.2 Standard method

Whereas our primary goal is to consider methods for joint inference of ψ_1 and ψ_2, for completeness, we compare such methods with a simplified standard method by treating ψ_1 as fixed and known, which is commonly used by the astrophysicists and statisticians. In a Bayesian analysis (e.g., see Van Dyk et al. 2001), the standard method involves estimating ψ_2 via its posterior distribution given data X_2 and a nominal estimate of ψ_1. In the context of a two-stage analysis, we first fix ψ_1 to the value that is estimated in the preliminary analysis. Conditioning on that, we fit the parameter, ψ_2, in the primary analysis. It is a common astronomical data analysis strategy that can lead to erroneous interpretation of the data. Assuming that $\psi_1 = \hat{\psi}_1$ and $\hat{\psi}_1$ is the best-fit estimator of $p(\psi_1 \mid X_1)$ from the preliminary analysis, where $p(\psi_1 \mid X_1)$ is obtained by integrating $p(\psi_0, \psi_1 \mid X_1)$ over ψ_0, ψ_2 is estimated using its posterior distribution

$$p(\psi_2 \mid X_2, \hat{\psi}_1). \tag{3.3}$$

The best-fit values of ψ_2 can be obtained via maximum a posteriori (MAP) estimate by assuming $X_2 \mid \psi_1, \psi_2$ follows a Gaussian distribution or via the

estimate that minimizes the χ^2 statistic (by comparing the expected X_2 with the observed X_2). Because this approach assumes that $\psi_1 = \hat{\psi}_1$, it does not account for the uncertainties in ψ_1. It can lead to misleading estimates of ψ_2 and can significantly underestimate the error bars associated with these estimates. Nevertheless, because this is the standard approach in practice, we treat it as a baseline in our numerical comparisons.

3.2.3 Multiple imputation

Multiple imputation is a well-established simulation based approach that is designed to handle incomplete data (Rubin 1987). MC replications of the incomplete data, called the *imputations*, are used to impute the statistical uncertainty in the missing values of the incomplete data. The application of this method depends heavily on the accessibility of the imputations. Suppose we have a MC sample of the parameter ψ_1. The sample of ψ_1 itself is not missing data, however, it plays exactly the role of the imputations and represents the statistical uncertainty of ψ_1. Therefore, we are able to apply multiple imputation method.

Given a sample of ψ_1 from $p(\psi_1 \mid X_1)$, multiple imputation can be applied straightforwardly. Suppose we have M sets of independent estimates of ψ_1 from $p(\psi_1 \mid X_1)$, denoted by $\psi_1^{(m)}$ for $m = 1, \ldots, M$ and called the *multiple imputation sample*. We apply the standard method, described in Section 3.2.2, M times, and $\hat{\psi}_1$ in Eq (3.3) is replaced with each of the M imputation samples of ψ_1. This generates M sets of MAP estimates of ψ_2 and the corresponding estimated variance-covariance matrices, denoted by $\hat{\psi}_2^{(m)}$ and $\mathrm{Var}(\hat{\psi}_2^{(m)})$, respectively, for $m = 1, \ldots, M$. Suppose that each $\hat{\psi}_2^{(m)}$ follows a multivariate normal distribution with mean ψ_2. The *multiple imputation combining rules* (e.g., Rubin 1987, Harel & Zhou 2005), as a set of simple moment calculations, can then be used for combining these estimates and their variance-covariance matrices and calculating the final fitted values and error bars based on the normal assumption above.

The overall estimate of ψ_2 is simply the average of the individual fitted values,

$$\hat{\psi}_2 = \frac{1}{M} \sum_{m=1}^{M} \hat{\psi}_2^{(m)}. \tag{3.4}$$

There are two sources of uncertainty, the within-imputation variance that would appear even if ψ_1 was known with certainty and the between-imputation variance coming from uncertainty in ψ_1 itself, to be calculated and combined into the estimated total variance. Each of the M standard analyses is applied assuming ψ_1 was known and each $\mathrm{Var}(\hat{\psi}_2^{(m)})$, therefore, is an estimated statistical uncertainty. The estimated within-imputation variance is simply the average of these individual estimates of the statistical uncertainty,

$$V = \frac{1}{M} \sum_{m=1}^{M} \mathrm{Var}(\hat{\psi}_2^{(m)}). \tag{3.5}$$

On the other hand, the between-imputation variance is estimated by thinking of how changing ψ_1 in each of the M standard analyses affects the fitted parameter ψ_2. Thus, the estimated between-imputation variance is the variance of the fitted values,

$$B = \frac{1}{M-1} \sum_{m=1}^{M} (\hat{\psi}_2^{(m)} - \hat{\psi}_2)(\hat{\psi}_2^{(m)} - \hat{\psi}_2)^\top. \tag{3.6}$$

Combining the above two sources of variance, we obtain the estimated total variance,

$$T = V + (1 + \frac{1}{M})B, \tag{3.7}$$

where the $\frac{1}{M}$ term considers the case when the number of imputations, M, is small. If M is smaller than the dimension of ψ_2, B is very noisy. T is therefore very unstable and more complicated assumptions need to be added, e.g., Li et al. (1991) assume that the between- and within-imputation variances are proportional to each other. Here, we focus attention on univariate estimates and error bars which rely on one element of $\hat{\psi}_2$ at a time and the corresponding diagonal entry of T.

The multiple imputation method is designed to give approximate error bars on each element of ψ_2 that include the effects of the imputed quantity, ψ_1 here. However, if the Gaussian assumption is inappropriate, even wrong, or a full posterior distribution on ψ_2 is expected, more detailed Bayesian methods need to be considered in the following subsections.

3.2.4 Pragmatic Bayesian method

We have discussed the standard analysis assumes that ψ_1 is fixed in Eq (3.3) and multiple imputation assumes that a sample of ψ_1 is obtained from $p(\psi_1 \mid X_1)$. Following that, ψ_2 is estimated conditional on a fixed or a sample of ψ_1. When the uncertainty in ψ_1 need to be quantified, we can not condition on a known value or a sample of ψ_1. For example, in the practical problem, the atomic uncertainty in emissivity need to be considered, we can not condition on a known value of emissivity. To eliminate this assumption, we treat ψ_1 as unknown, rather than conditioning on a fixed value. We adopt a Bayesian framework in our statistical analysis because it enables us to build in the complex hierarchical dependencies engendered by ψ_1.

We expect the information for ψ_2 comes primarily from X_2 rather than X_1, at least given ψ_1, because X_1 typically represents the large-scale dataset which we do not have access to. Mathematically, this assumption can be written as

$$p(\psi_2 \mid X_1, X_2, \psi_1) = p(\psi_2 \mid X_2, \psi_1). \tag{3.8}$$

For example, we do not have access to the large-scale dataset used by atomic physicists to generate emissivities. Practically, the universal emissivities and atomic physics data, typically what ψ_1 and X_1 represent, would not influence the electron temperature and density of the solar and stellar coronae represented by ψ_2. It is also sensible that researchers in different stages have different level of access to datasets in any particular example. Therefore, no access to X_1 is a necessary and reasonable assumption to be made throughout this thesis.

Moreover, we pragmatically assume the primary source of information for parameter ψ_1 is not the dataset X_2 but the dataset X_1. Mathematically, this assumption can be written as

$$p(\psi_1 \mid X_1, X_2) = p(\psi_1 \mid X_1). \tag{3.9}$$

Now we call the analyses under the independence assumptions in Eq (3.8) and Eq (3.9) the *pragmatic Bayesian method* (Lee et al. 2011). Therefore, when treating ψ_1 as unknown, the appropriate objective function used for probabilistic estimation and calculation of error bars under the pragmatic Bayesian method is the joint posterior distribution of all the parameters, ψ_1 and ψ_2, given all the datasets, X_1 and X_2,

$$\begin{aligned} p_{\mathrm{pB}}(\psi_1, \psi_2 \mid X_1, X_2) &= p(\psi_2 \mid X_1, X_2, \psi_1)\ p(\psi_1 \mid X_1, X_2) \\ &= p(\psi_2 \mid X_2, \psi_1)\ p(\psi_1 \mid X_1). \end{aligned} \tag{3.10}$$

The subscript pB indicates that this is the pragmatic Bayesian posterior distribution under the pragmatic assumptions in Eq (3.8) and Eq (3.9). The latter assumption indicates that the dataset X_2 and the parameter ψ_1 are independent, that is, the dataset X_2 provide no information for the uncertainty in ψ_1 in the primary stage. Under the pragmatic Bayesian model in Eq (3.10), inference for ψ_2 is based on its marginal posterior distribution

$$p_{\mathrm{pB}}(\psi_2 \mid X_1, X_2) = \int p(\psi_2 \mid X_2, \psi_1)\ p(\psi_1 \mid X_1)\ \mathrm{d}\psi_1. \tag{3.11}$$

The pragmatic Bayesian method accounts for the uncertainty of ψ_1 in a conservative manner. The assumption that $p(\psi_1 \mid X_1, X_2) = p(\psi_1 \mid X_1)$ ignores information in the dataset, X_2, that may reduce the uncertainty in ψ_1 and hence in ψ_2. Whether or not including dataset X_2 is an assumption that is questionable. We now consider methods that allow X_2 to be informative for ψ_1.

3.2.5 Fully Bayesian method

In contrast to the pragmatic Bayesian method, the *fully Bayesian method*, as described by Xu et al. (2014), incorporates the potential information in the dataset, X_2, to learn about ψ_1. The joint posterior distribution of ψ_1 and ψ_2 given X_1 and X_2 under the fully Bayesian method can be written as

$$p_{\text{fB}}(\psi_1, \psi_2 \mid X_1, X_2) = p(\psi_2 \mid X_1, X_2, \psi_1)\; p(\psi_1 \mid X_1, X_2), \tag{3.12}$$

where we use subscript fB to emphasize that this is the fully Bayeaian joint posterior. The marginal posterior distribution of ψ_2 under the fully Bayesian method is given by

$$p_{\text{fB}}(\psi_2 \mid X_1, X_2) = \int p(\psi_2 \mid X_1, X_2, \psi_1)\; p(\psi_1 \mid X_1, X_2)\; \mathrm{d}\psi_1. \tag{3.13}$$

The fully Bayesian method is a more principled approach from a statistical perspective following the principles of Bayesian analysis, while the pragmatic method makes simplifying assumptions that tend to overestimate the final uncertainty on the fitted parameters (see Xu et al. 2014). If a certain subset of ψ_1 samples are plausible before seeing the data X_2 but inconsistent with the data X_2 once it is observed, this subset should not play an important role in the primary analysis. This method allows the data X_2 to inform our choice of possible ψ_1. By investigating the pragmatic and the fully Bayesian methods, we can conclude the sensitivity of our results to the assumption that whether or not including X_2 as a source of information for ψ_1 samples.

To focus attention on the primary analysis we suppress the conditioning on X_1 for the rest of this thesis.

3.3 Quantifying the atomic uncertainties into the primary analysis

As described in Section 3.2, in the preliminary analysis, we must be able to summarize $p(\psi_1 \mid X_1)$, which is denoted simply as $p(\psi_1)$ now, and to carry

forward it into the primary analysis.

In the practical example, the objective is to quantify atomic uncertainties on emissivities, denoted by ϵ and playing the role of ψ_1 in Section 3.2, in general sense and to propose well-defined and general methods to incorporate the information on emissivities obtained from the preliminary analysis into the primary analysis in a robust principled manner. We propose a Bayesian framework, where knowledge of emissivities is quantified through a prior probability distribution. In this way, information quantified from the preliminary analysis, typically quantified by the astrophysicists in practice, can be incorporated into a coherent statistical analysis. Operationally, this involves fitting a bespoke statistical model that does not assume emissivities as known and fixed quantities, but rather incorporates its uncertainty through a prior distribution.

From a general MC Bayesian perspective, suppose we obtain an ensemble of emissivity sample, provided by the atomic physicists, as part of the preliminary analysis, denoted by

$$\mathcal{M} = \{\epsilon^{(m)}, m = 1, \ldots, M\}.$$

The ensemble can be either sampled from $p(\epsilon \mid X_1)$ or simulated with certain astrophysical strategy as described in Section 3.1. Below we formulate the priors in two ways, a discrete uniform distribution and a Gaussian distribution via principal component analysis (PCA) for incorporating this uncertainty into the primary analysis. Statistically, we treat them as two different prior distributions. The real difference between the two, in physics, is that the PCA approach provides a continuous distribution that effectively 'fills' the gaps from the emissivity samples. In Chapter 4, we describe how we go from a discrete sparse sampling to a continuous distribution and how this actually works on particular problems. Using the prior on $\psi_1(= \epsilon)$, we ultimately aim to obtain estimations of the plasma parameters, denoted by θ and playing the role of ψ_2 in Section 3.2.

3.3.1 A discrete uniform distribution

A discrete uniform prior distribution for the ensemble, $\mathcal{M}$, is straightforward, where we assume the M samples are a priori equally likely. Equivalently, we could treat the emissivity index, m, as unknown parameter and assume it has a discrete uniform distribution,

$$p(\epsilon^{(m)}) = p(m) = \frac{1}{M} \quad \text{for each } m = 1, \ldots, M. \tag{3.14}$$

This requires large storage and is computational expensive when M is large and the concern is magnified when we consider those ϵ with more complicated structure in astrophysical model. See more discussions on specific case studies in Section 4.5.5 and Section 4.5.7.

The marginal posterior distribution of θ under the pragmatic and the fully Bayesian models can then be rewritten from Eq (3.11) and Eq (3.13) as,

$$p_{\text{pB}}(\theta \mid X_2) = \sum_{m=1}^{M} p(\theta \mid X_2, m)\ p(m), \tag{3.15}$$

$$p_{\text{fB}}(\theta \mid X_2) = \sum_{m=1}^{M} p(\theta \mid X_2, m)\ p(m \mid X_2), \tag{3.16}$$

respectively, where $p(m \mid X_2)$ is the posterior probability of each of $\epsilon^{(m)}$ in $\mathcal{M}$. Recall that we are suppressing the conditioning on X_1. We consider each of these emissivity curves $\epsilon^{(m)}$ to represent a different model $p(\theta \mid X_2, m)$ and the posterior probability of each $\epsilon^{(m)}$ to represent weights for each individual model. The resulting marginal for the parameter θ under the fully Bayesian method, Eq (3.16), is a model averaging posterior. This is consistent with the so-called *Bayesian Model Averaging* (BMA) (see Roberts 1965). It is an average over all the conditional posterior distributions of θ weighted by each posterior probability of ϵ. Therefore, it allows for a direct combination of conditional posterior distributions (models) to obtain combined parameter estimates. Similarly, under the pragmatic Bayesian method, the marginal posterior, Eq (3.15), is an average over all the conditional posterior distributions of θ weighted by each prior probability of ϵ. Therefore, using a discrete

uniform prior distribution for the ensemble of emissivities can be viewed as equivalent to applying BMA with different weighting schemes. However, the BMA typically do not apply to the continuous model. We will work out a continuous one below.

3.3.2 A continuous distribution via principal component analysis

PCA is a well established linear technique for dimensionality reduction and data compression of a multivariate dataset whilst preserving as much of the relevant information as possible (Jolliffe 2002, Anderson 2003, Bishop 2006). Mathematically, PCA is defined as an orthogonal linear transformation that transforms a number of correlated variables into a few uncorrelated variables, called *principal components* (PCs), while retaining as much variability of the data as possible. The first PC defines the linear function of the original variables with the greatest variance of any projections of the data coming to lie on the first coordinate. Each succeeding component defines the linear function orthogonal to all of the previous PCs with the greatest variance on the corresponding coordinate. The aim of applying PCA is to describe and maintain as much variability of the data as possible. In practice, PCA is implemented on centred data, i.e., mean zero, by subtracting off the mean of the data before the PCA and adding it back after the analysis. The calculation of the orthogonal linear transformation is achieved by a singular value decomposition of a centred data matrix. A set of orthogonal and ordered eigenvectors corresponding to the PCs is produced, along with their eigenvalues indicating the proportion of the total variance explained by each eigenvector. Selecting the first few PCs and retaining most of the statistical properties in all of the original variables, PCA finds a subset of orthogonal eigenvectors, i.e., a subset of PCs, and their corresponding eigenvalues to effectively represent a specified percentage of the variability of a large dataset.

The ensemble of emissivities must be large enough to fully represent the atomic uncertainty in high-dimensional emissivities. There might be gaps among the sample if the given emissivity ensemble is too sparse (Yu et al.

2018). To address those problems and to quantify atomic uncertainty, in this article, we aim to use PCA to efficiently summarize and compress the ensemble of complex ϵ samples, $\mathcal{M}$, into a succinct and applicable form (Lee et al. 2011, Xu et al. 2014). The singular value decomposition is applied to a centred and scaled matrix with rows equal to the $\frac{\epsilon^{(m)}-\bar{\epsilon}}{\sigma_\epsilon}$ with arithmetic mean $\bar{\epsilon} = \frac{1}{M}\sum_{m=1}^{M}\epsilon^{(m)}$ and standard deviation σ_ϵ. We calculate the eigenvalues $(\beta_1^2, \ldots, \beta_Q^2)$, ordered as $\beta_1 \geq \beta_2 \geq \ldots, \geq \beta_Q$, and the corresponding eigenvectors / principal components, $(v_1, \ldots, v_Q)$, where Q is the number of the original variables. The proportion of the total variance of $\mathcal{M}$ explained by the qth PC, i.e., v_q, is

$$p_q = \frac{\beta_q^2}{\sum_{j=1}^{Q}\beta_j^2}. \tag{3.17}$$

Practically, a smaller number of PCs, $J \ll Q$, that sufficiently accounts for a certain proportion, e.g., 95% or 99%, of the total variance, could be used in the reconstruction. That is, very few PCs might be required to generate a new ϵ sample to high accuracy indicating the achievement of a great compression on the given ensemble of emissivities.

Assuming there is a sample from a multivariate standard normal distribution and with the PCA summary of $\mathcal{M}$ in hand, we are able to formulate and generate replicates of ϵ mimicking $\mathcal{M}$, based on the first J PCs as

$$\epsilon^{\text{rep}}(r) = \bar{\epsilon} + \sum_{j=1}^{J} r_j \ \beta_j \ v_j \tag{3.18}$$

where $r = (r_1, \ldots, r_J)$ are assumed to be independent standard normal random variables, i.e.,

$$r \sim \mathcal{MVN}(0, I_J). \tag{3.19}$$

The remaining $Q-J$ components are ignored in this representation indicating the achievement of data compression. With this approximation over the first J principle components, though we will be losing certain variability and missing certain structure of the given ensemble, it is much better than having the gaps in the given emissivity ensemble. However, if principle components are enough, it would not make significant influence. Using Eq (3.18) and

Eq (3.19), we can generate as many MC replicates from $\mathcal{M}$ as possible using only $\bar{\epsilon}$ and $(\beta_1 v_1, \ldots, \beta_J v_J)$. Details for the examples on different ion case studies are described in Section 4.3.2, Section 5.3, and Section 6.3.

3.3.3 Priors comparison

Using discrete uniform prior, we can identify which of the emissivity curves are more consistent with the observed spectral data. Because each of the emissivity curves in the ensemble is based on particular atomic physical assumptions, the atomic physics, in principle, can trace back to the assumptions associated with the most likely curves from our analysis. That is, the individual deviations used for each level and for each transition can be identified for the most likely curves. In contrast to the discrete uniform distribution, if each of the emissivity curves in the given ensemble has underlying atomic physics connected with it, using a Gaussian approximation via PCA will lose all of those connections. However, a Gaussian approximation via PCA can effectively fill in the gaps between the discrete emissivities and work on the problem of sparse sampling of the atomic data space. It requires less storage because PCA can provide a concise statistical compression when there are complex correlations in the emissivity ensemble. It can also generate emissivity samples on the fly.

3.4 Statistical algorithms

We aim to propose well-defined and general methods to incorporate the information, e.g., on emissivities ϵ, obtained from the preliminary analysis into the primary analysis in a robust principled manner. We describe two specific algorithms, two-step Monte Carlo (MC) sampler in Section 3.4.1 and Hamiltonian Monte Carlo (HMC) in Section 3.4.2, for sampling from the posterior distributions under the fully Bayesian method. The former one can also be applied under the pragmatic Bayesian method. The reason we have only one algorithm for the pragmatic Bayesian is because the model is simple and the algorithm works well, therefore, no further algorithm needs

to be tested. While for the fully Bayesian method, two-step MC sampler is very computational expensive and then HMC is considered as well.

In two-step MC sampler, an iterative Markov chain Monte Carlo (MCMC) sampler allows us to incorporate the uncertainty in ϵ directly into the fitting routine by updating ϵ and θ under certain acceptance and rejection rules, step by step, at each iteration. In this case, we can update ϵ based either solely on the information provided from the preliminary analysis (i.e., the pragmatic Bayesian method), or together with the dataset being analyzed in the primary analysis (i.e., the fully Bayesian method). Alternatively, using HMC also allows us to incorporate the uncertainty in ϵ directly into the fitting routine by updating ϵ and θ at the same time under the NUTS rule at each iteration under the fully Bayesian model.

3.4.1 Two-step Monte Carlo sampler for the pragmatic and the fully Bayesian methods

In the fully Bayesian analysis, our aim is to obtain a MC sample of (θ, ϵ) from the fully Bayesian joint posterior in Eq (3.12) and we iteratively update ϵ and θ in two separate sub-steps. We first sample ϵ according to the emissivity uncertainty, then update θ conditioning on the newly sampled ϵ. When updating θ in the second sub-step, MCMC algorithms like MH or Adaptive MH need to be applied for several inner iterations to get proper draws achieving convergence. Suppose $\epsilon^{(\ell)}$ and $\theta^{(\ell)}$ are the values of the parameters at iteration ℓ. The two-step MC sampler under the fully Bayesian model consists of the following two sub-steps:

$$\epsilon^{(\ell)} \text{ is sampled from } p(\epsilon \mid X_2) \text{ and} \tag{3.20}$$

$$\theta^{(\ell)} \text{ is iteratively sampled from } p(\theta \mid X_2, \epsilon^{(\ell)}). \tag{3.21}$$

Under the conditional independence assumption of Section 3.2.4, we can simplify the above sampler by replacing $p(\epsilon \mid X_2)$ with $p(\epsilon)$ in the first sub-step and the two-step MC sampler under the pragmatic Bayesian model is

as follow:

$$\epsilon^{(\ell)} \text{ is sampled from } p(\epsilon) \text{ and} \tag{3.22}$$

$$\theta^{(\ell)} \text{ is iteratively sampled from } p(\theta \mid X_2, \epsilon^{(\ell)}) \tag{3.23}$$

This independence assumption prevents us from estimating the posterior distribution $p(\epsilon \mid X_2)$ and simplifies the structure of the algorithm significantly.

The two-step MC sampler, as a powerful method, can be generalized to a multi-step MC sampler, which is able to deal with any level of complexity and explore interesting regions in high-dimensional parameter space. It effectively separates the complex problem of model fitting in the presence of the uncertainties in ϵ into two simpler problems: (1) the quantification of the uncertainties in emissivity independent of or conditional on the current data X_2 and (2) fitting a spectral model with a known emissivity. Either a Metropolis algorithm, a Metropolis-Hastings algorithm, a Gaussian approximation, or an adaptive Metropolis algorithm, discussed in Section 2.2, can be used to sample $\theta^{(\ell)}$ in each individual step (see Xu et al. 2014).

3.4.2 HMC for the fully Bayesian method

Another alternative method is to use the Stan* software package (Carpenter et al. 2016) to obtain MC sample of (θ, ϵ) via the HMC algorithm under the NUTS rule, discussed in Section 2.2.4, by sampling directly from its joint posterior distribution, rewritten from Eq (3.13),

$$p(\theta, \epsilon \mid X_2) = p(X_2 \mid \theta, \epsilon)\ p(\theta)\ p(\epsilon), \tag{3.24}$$

assuming that the prior distributions for θ and ϵ are independent. However, we must analytically marginalize over any discrete parameters, since Stan cannot accommodate discrete parameters.

*Stan is a probabilistic modeling language developed by Andrew Gelman and collaborators. It interfaces with the most popular data analysis languages like R, Python, etc., and is available at mc-stan.org.

4

Two-stage analysis with Fe XIII

An application of a two-stage analysis via a case study in Fe XIII is described in this chapter. The physicists deploy a simple, but realistic, model to describe the uncertainties in some of the atomic parameters for Fe XIII, and use it to generate different realizations of the plasma emissivities representing the atomic uncertainties in the preliminary stage. We then focus on the primary stage by applying this ensemble of atomic data into the analysis of the density-sensitive Fe XIII spectral lines observed in solar active region using a Bayesian framework. We have implemented the pragmatic Bayesian method that considers the atomic data uncertainties as fully specified and uncorrectable and the fully Bayesian method that allows for the observed data to update the atomic data uncertainties. The former generally increases the uncertainties on the inferred parameters compared with models that incorporate only statistical uncertainties. The latter reduces the uncertainties on the inferred parameters. To incorporate uncertainties in atomic data into our highly structured statistical model, a discrete analysis where the atomic realizations are considered individually is deployed first suggesting a couple of different realizations of emissivities are more likely than the default CHIANTI calculation and identifying areas of possible systematic problems with either

the atomic physics or the observed intensities. Following that, a continuous analysis with PCA is used to fully summarize the atomic uncertainties via a multivariate Gaussian distribution, to provide a concise statistical compression, and to mitigate the gaps among those sparse atomic realizations. The work related to discrete analysis is a review of *Incorporating Uncertainties in Atomic Data Into the Analysis of Solar and Stellar Observations: A Case Study in Fe XIII*, published by Yu et al. (2018).

4.1 Data and notation

We have randomly selected $K = 1000$ pixels from the *Hinode* Spacecraft EIS observations of a solar active region shown in Figure 1.1 for analysis. Suppose that in each of K pixels we observe the intensities of each of H spectral lines with wavelengths $\Lambda = \{\lambda_1, \ldots, \lambda_H\}$. Let $I_{k\lambda}$ be the observed intensity of the line with wavelength $\lambda \in \Lambda$ in pixel $k \in \{1, \ldots, K\}$, $\sigma_{k\lambda}$ its known standard deviation, $D_k = (I_{k\lambda_1}, \ldots, I_{k\lambda_H})$, and $\mathcal{D} = \{D_1, \ldots, D_K\}$, playing the role of X_2 in the general notation in Section 3.2.

We also have a collection of $M = 1000$ realizations of the plasma emissivities, denoted by $\mathcal{M}$,

$$\mathcal{M} = \{\epsilon_\lambda^{(m)}(\mathrm{n}_k, \mathrm{T}_k), \lambda \in \Lambda, m = 1, \ldots, M\},$$

where n_k and T_k are the electron density and temperature for pixel k and m indexes the emissivity realization (i.e., emissivity curve, $\epsilon_\lambda^{(m)}(\mathrm{n}_k, \mathrm{T}_k)$), with m=1 corresponding to the default CHIANTI emissivities. The M different realizations of the plasma emissivity curves, as shown in Figure 3.1, are simulated from a model that accounts for the uncertainty in the atomic physics calculation as described in Section 3.1.

The expected intensity of the line with wavelength λ in pixel k can be expressed as $\epsilon_\lambda(\mathrm{n}_k, \mathrm{T}_k)\mathrm{n}_k^2\mathrm{d}_k$, where d_k is the path length through the solar atmosphere for pixel k (see, for example, Mariska 1992). Let $\theta_k = (\log \mathrm{n}_k, \log \mathrm{d}_k)$ be the plasma parameters in pixel k, and $\Theta = (\theta_1, \ldots, \theta_K)$. Note that log refers to $\log_{10}$ throughout this thesis unless otherwise stated.

4.2 Statistical Model

The first step in specifying our statistical model is to construct the likelihood function. We model the intensities $I_{k\lambda}$ given ϵ, n_k, and d_k as a normal (i.e. Gaussian) distribution,

$$I_{k\lambda} \mid \epsilon, \mathrm{n}_k, \mathrm{d}_k \overset{\text{indep}}{\sim} \mathcal{N}\left(\epsilon_\lambda(\mathrm{n}_k, \mathrm{T}_k)\mathrm{n}_k^2\mathrm{d}_k,\ \sigma_{k\lambda}^2\right), \tag{4.1}$$

for $\lambda \in \Lambda$, where $\mathcal{N}(\mu, \sigma^2)$ is a normal distribution with mean μ and variance σ^2. We suppress the conditioning on the $\sigma_{k\lambda}$ throughout for notational simplicity. Thus the likelihood function of D_k given emissivity, ϵ, and plasma parameters, θ_k, is

$$\begin{aligned} L(\epsilon, \theta_k \mid D_k) &= p(D_k \mid \epsilon, \theta_k) \\ &= \prod_{h=1}^{H} \mathcal{N}\left(I_{k\lambda_h} \mid \epsilon_{\lambda_h}(\mathrm{n}_k, \mathrm{T}_k)\mathrm{n}_k^2\mathrm{d}_k,\ \sigma_{k\lambda_h}^2\right), \end{aligned} \tag{4.2}$$

since the observation independence assumption among lines of wavelengths, where $\mathcal{N}(x \mid \mu, \sigma^2)$ is the density of a normal distribution with mean μ and variance σ^2 evaluated at x.

Next, we specify the joint prior distribution on the unknown model parameters. We specify a continuous uniform distribution for $\log \mathrm{n}_k$,

$$p(\log \mathrm{n}_k) = \frac{1}{5} \qquad \text{for } 7 \leq \log \mathrm{n}_k \leq 12. \tag{4.3}$$

For ϵ, as mentioned in Section 3.3, we have two different priors, a discrete uniform distribution and a Gaussian distribution via PCA, and details are to be discussed in Section 4.3.

For $\log \mathrm{d}_k$, although there is a preference on non-informative prior, a uniform prior, $p(\log \mathrm{d}_k) \propto 1$, yields an improper posterior distribution because the likelihood converges to a positive constant as $\log \mathrm{d}_k$ goes to $-\infty$. Therefore,

we specify a Cauchy distribution for $p(\log \mathrm{d}_k)$,

$$\log \mathrm{d}_k \sim \mathrm{Cauchy}(\mathrm{center} = 9,\ \mathrm{scale} = 5), \tag{4.4}$$

which is a broad, fat-tailed distribution covering all conceivable values for the path length that we expect based on all sets of Fe XIII intensities. Such a broad prior is required because the dynamic range in the loop lengths is very large (e.g., see Berger et al. 1999, Fletcher & De Pontieu 1999). There is also a preference for loops of size 10^9cm because of the coronal scale height. The Cauchy distribution is deemed to be the best prior for this scenario, and is much superior to either a uniform distribution prior (too uninformative) or a Normal distribution prior (too informative). The Cauchy distribution prior also allows for a greater flexibility in the estimate of $\log \mathrm{n}_k$ than a Normal distribution prior.

We assume the parameters are independent a priori so that the joint prior distribution is

$$p(\epsilon, \theta_k) = p(\epsilon)\ p(\theta_k) = p(\epsilon)\ p(\log \mathrm{n}_k)\ p(\log \mathrm{d}_k). \tag{4.5}$$

Here θ is indexed by k, but ϵ is not. This reflects the fact that, although θ_k vary among the pixels, we expect the true emissivity to be an underlying physical quantity that is the same for all pixels.

We consider two ways to fit the plasma parameters, Θ, given the observed or simulated intensities, $\mathcal{D}$, while accounting for atomic uncertainty, $\mathcal{M}$. First we can analyze each single pixel separately in a sequence of pixel-by-pixel analyses. Although this may yield different estimates of ϵ, i.e., the more preferred emissivity curve among the pixels, it allows us to see if the intensities of each pixel give consistent information as to the best emissivity curve(s). Given the likelihood function in Eq (4.2) and the prior distribution in Eq (4.5), the joint posterior distribution for ϵ and θ_k under the *separate pixel-by-pixel analyses* is

$$p(\epsilon, \theta_k | D_k) = \frac{L(\epsilon, \theta_k \mid D_k)\ p(\epsilon, \theta_k)}{p(D_k)}, \tag{4.6}$$

where $p(D_k) = \int\int L(\epsilon, \theta_k \mid D_k)\ p(\epsilon, \theta_k)\ \mathrm{d}\epsilon\ \mathrm{d}\theta_k$.

Alternatively, we can simultaneously analyze the intensities from multiple pixels to arrive at an overall estimate of the most likely emissivity curve. When we consider all the K-pixel intensities together, the likelihood function of ϵ and Θ given $\mathcal{D}$, and the prior distribution of ϵ and Θ are, respectively,

$$L(\epsilon, \Theta \mid \mathcal{D}) = \prod_{k=1}^{K} L(\epsilon, \theta_k \mid D_k) \tag{4.7}$$

and

$$p(\epsilon, \Theta) = p(\epsilon)\ \prod_{k=1}^{K} p(\theta_k). \tag{4.8}$$

Thus, the joint posterior distribution of m and Θ under the *simultaneous analysis* can be expressed as

$$p(\epsilon, \Theta \mid \mathcal{D}) = \frac{L(\epsilon, \Theta \mid \mathcal{D}) p(\epsilon, \Theta)}{p(\mathcal{D})}, \tag{4.9}$$

where $p(\mathcal{D}) = \int\int L(\epsilon, \Theta \mid \mathcal{D}) p(\epsilon, \Theta)\ \mathrm{d}\epsilon\ \mathrm{d}\Theta$.

We consider both the separate pixel-by-pixel and the simultaneous analyses, and, for each, develop both the pragmatic and the fully Bayesian approaches in the following sections. Before we move on to that, we look at the two prior distributions for the ensemble of emissivities first.

4.3 The two prior distributions on emissivities

We discuss two different prior distributions for the ensemble of emissivities: a discrete uniform distribution and a Gaussian distribution via PCA.

4.3.1 A discrete uniform distribution

We specify a discrete uniform distribution for the ensemble of emissivities, $\mathcal{M}$. The prior distribution on it can be expressed exactly as in Eq (3.14). This choice of prior on m stipulates that the 1000 realizations of emissivity

curves in $\mathcal{M}$ are all a priori equally likely to be the true emissivity. As the realizations were generated by attaching reasonable uncertainties to the atomic data as described in Section 3.1, the atomic data uncertainties are contained in $p(m)$ and are thus captured by the corresponding posterior distribution. Therefore, the 1000 realizations of emissivity curves can also be considered as a sample of 1000 draws from an implicit prior distribution.

Note that we focus on methods that treat the emissivity index m as an unknown parameter, whose prior is specified above, whose posterior we estimate to determine the most likely emissivity realizations among those in $\mathcal{M}$, and whose uncertainties affect both the fit and error bars of θ_k.

4.3.2 A Gaussian distribution via principal component analysis

We specify a Gaussian prior distribution via PCA, as discussed in Section 3.3.2, to compress the given ensemble of emissivities, $\mathcal{M}$.

With the PCA representation of $\mathcal{M}$ in hand, we can easily reconstruct emissivity replicates based on the first $J(\ll Q)$ principal components via Eq (3.18) and Eq (3.19) to sufficiently account for a certain proportion of the total variance. A large amount of compression has been achieved because very few components are needed to compute the emissivity curve to high precision. For example, in the case of the given ensemble of emissivity realisations, the first 12 principal components (out of 182) can account for 95.3% of the total variance, and the first 16 principal components for 99.6%, as in Figure 4.1. Performance of the first few principal components is shown in Figure 4.2 where each principal component is orthogonal to others. After applying the PCA transformation, we are able to obtain a lower dimensional subspace where the emissivity samples are most spread along the new feature axes. Note that this approximation achieves a better reconstruction in the square root scale of the given emissivity ensemble, which is used throughout this chapter.

The complicated structure of the emissivity curves, indicating the atomic uncertainty, is illustrated in the top two panels of Figure 4.3 using the given

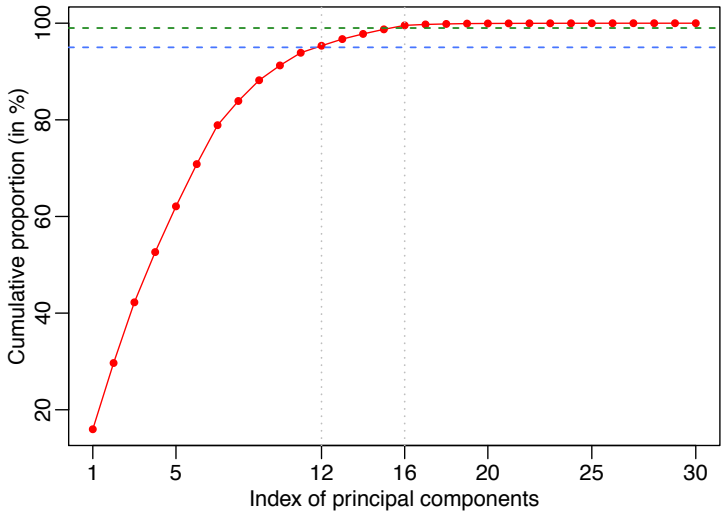

Figure 4.1: Cumulative proportion of variance explained by the first few principal components from PCA on Fe XIII emissivity curves, as in red line. The x-axis represents the number of first few principal components used and the vertical dotted grey lines correspond to the first 12 and 16 principal components. The y-axis represents the cumulative proportion of the variance explained and the horizontal dashed blue and green lines correspond to 95% and 99% respectively.

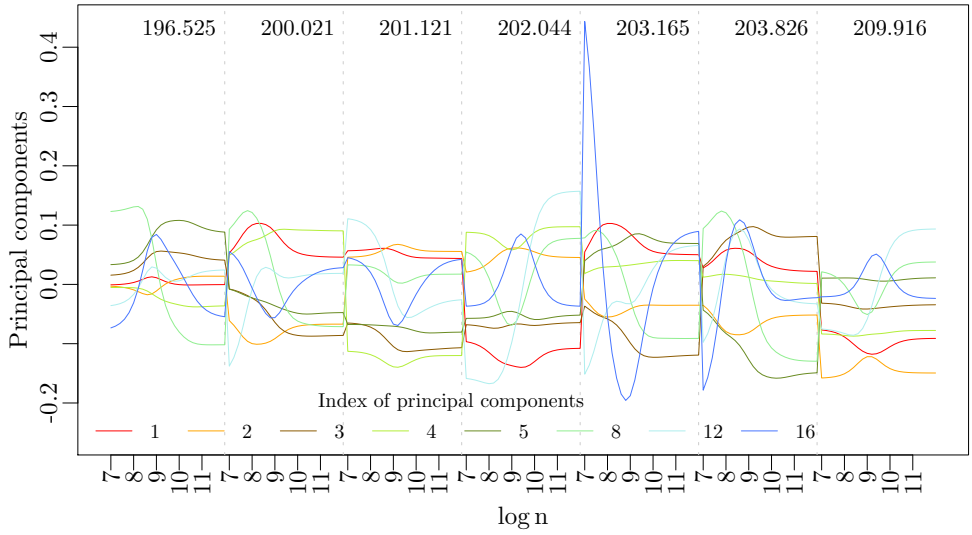

Figure 4.2: Performance of the first few major principal components from PCA on Fe XIII emissivity curves. The 1st, 2nd, 3rd, 4th, 5th, 8th. 12th, 16th principal components versus density and lines of wavelength are plotted as coloured curves. The seven horizontally-arranged sub-panels separated by vertical dotted grey lines correspond to the seven Fe XIII lines of interest. Within each sub-panel, the density $\log \mathrm{n}$ is increasing.

ensemble of 1000 emissivity realisations. A random selection of six of the $\sqrt{\epsilon^{(m)}}$ from $\mathcal{M}$ is compared with the average over all those emissivity realisations, $\sqrt{\epsilon}$, the full range, the middle 95%, and the middle 68.3% of $\mathcal{M}$ in square root space, indicating the ensemble of emissivity curves in $\mathcal{M}$ form a complex tangle that appears to defy any systematic pattern. The complexity of the uncertainty of $\mathcal{M}$ is evident. The third panel of Figure 4.3 illustrates the use of PCA compression on the emissivity realisations. We generated 1000 replicate emissivity curves, the exact same amount of emissivity curve as in the original ensemble, using Eq (3.18) and Eq (3.19) with $J = 16$. In this case, using $J = 16$ captures 99.6% of the total variation in $\mathcal{M}$, as computed with Eq (3.17). The full range, the middle 95%, and the middle 68.3% intervals of these replicates are superimposed on the corresponding intervals for the original emissivity realisations in square root scale. The correspondence between the original emissivity realisations and the PCA replicates is quite good, especially for the 68.3% intervals. Although the PCA representation cannot be perfect (e.g., it does not fully represent uncertainty overall or in certain regions) it is much better than not accounting for uncertainty at all and is able to mitigate the gaps among those sparse emissivity realizations.

4.4 General outlook on methods comparison

The ensemble set of emissivity realisations, $\mathcal{M}$, is provided by the atomic physicists and represents uncertainties in atomic data. It gives a summary for the preliminary analysis in a two-stage analysis as mentioned in Section 1.2 and Section 3.1. We aim to incorporate uncertainties in atomic data straightforward in Section 4.5, or to embed the PCA model for uncertainty in atomic data in Section 4.6, into the primary analysis via a Bayesian procedure. It simultaneously fits the model parameters and accounts for atomic uncertainty. We compare the results from different statistical methods, the standard, the pragmatic and the fully Bayesian methods, mentioned in Section 3.2.2, Section 3.2.4, Section 3.2.5, respectively. Those methods are applied to both a single-pixel (i.e., the separate pixel-by-pixel analysis) and multiple pixels (i.e., the simultaneous analysis).

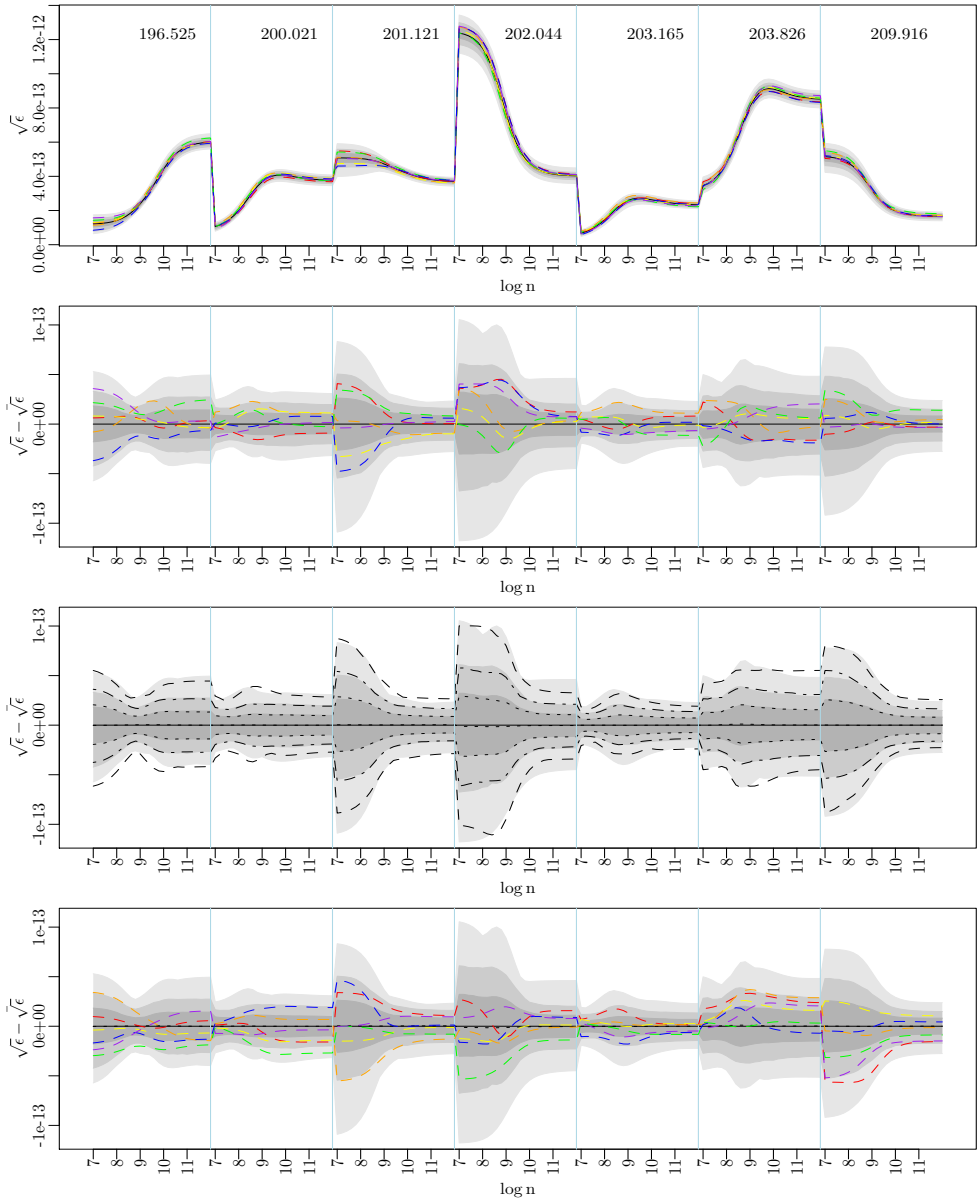

Figure 4.3: Fe XIII emissivity curves and the PCA summary. In the first panel, the light, dark, and darker grey areas cover the full range, the middle $95\%$and 68.3% of all 1000 emissivity curves in square root scale. $\sqrt{\epsilon}$ is plotted as a solid black curve. Six randomly selected curves are plotted as coloured dashed curves. Other panels are constructed in the same manner, but using $\sqrt{\epsilon^{(m)}} - \bar{\sqrt{\epsilon}}$, to magnify the structure in $\mathcal{M}$. Summarizing the emissivity samples via PCA, the dashed, dot-dash, and dotted lines, in the third panel, respectively outline intervals containing the full range, the middle 95% and 68.3% of 1000 PCA generated emissivity curves. In the last panel, six randomly selected PCA generated curves are plotted as coloured dashed curves. The seven horizontally-arranged sub-panels correspond to the seven Fe XIII lines of interest. Within each sub-panel, the density $\log \mathrm{n}$ is increasing.

4.5 Discrete prior on emissivity

In this part, we implement the standard, the pragmatic and the fully Bayesian methods applied to both a single-pixel and multiple pixels when we have a discrete uniform prior distribution on the ensemble of emissivities, given in Eq (3.14). This work has been published in Yu et al. (2018). Specifically, Section 4.5.1 develops the pragmatic and the fully Bayesian approaches to the pixel-by-pixel analyses and Section 4.5.2 describes the algorithms used to deploy these approaches. The simultaneous analysis and its algorithm are discussed in Section 4.5.3. Applications of these models and algorithms to simulated and observed intensities are discussed in Section 4.5.4 and Section 4.5.5 respectively.

4.5.1 Standard, pragmatic and fully Bayesian methods for separate pixel-by-pixel analysis

Given the 1000 equally likely emissivity realizations, we incorporate them into the standard, the pragmatic and the fully Bayesian methods discussed in Section 3.2. Treating the emissivity index m as an unknown parameter with its prior distribution in Eq (3.14), the prior independence assumption becomes

$$\begin{aligned} p(m, \theta_k) &= p(m)\ p(\theta_k) \\ &= p(m)\ p(\log \mathrm{n}_k)\ p(\log \mathrm{d}_k). \end{aligned} \tag{4.10}$$

The likelihood function of D_k given emissivity index, m, and plasma parameter, θ_k, is

$$L(\theta_k, m \mid D_k) = \prod_{h=1}^{H} \mathcal{N}\left(I_{k\lambda_h} \mid \epsilon_{\lambda_h}^{(m)}(\mathrm{n}_k, \mathrm{T}_k)\mathrm{n}_k^2\mathrm{d}_k,\ \sigma_{k\lambda_h}^2 \right). \tag{4.11}$$

Given the likelihood function in Eq (4.11) and the prior distribution in Eq (4.10), the joint posterior distribution for m and θ_k under the separate

pixel-by-pixel analyses is

$$p(m, \theta_k | D_k) = \frac{L(m, \theta_k \mid D_k)\ p(m, \theta_k)}{p(D_k)}, \tag{4.12}$$

where $p(D_k) = \sum_{m=1}^{M} \int L(m, \theta_k \mid D_k)\ p(m, \theta_k)\ \mathrm{d}\theta_k$.

Then the marginal posterior distribution $p(\theta_k \mid D_k)$ can be obtained by summing over m,

$$p(\theta_k \mid D_k) = \sum_{m=1}^{M} p(m, \theta_k \mid D_k). \tag{4.13}$$

In this way, we are able to infer θ_k accounting for uncertainties of the atomic data via the ensemble in $\mathcal{M}$. Moreover, uncertainty can be quantified with a list of the most likely emissivity realizations from $\mathcal{M}$ (or their indices, m) along with their associated posterior probabilities.

Standard method

Assuming that the emissivity curves are completely and correctly specified, $\epsilon = \epsilon^{(1)}$, and $\epsilon^{(1)}$ is the default emissivity curve, θ_k is estimated using its conditional posterior distribution given D_k and $\epsilon^{(1)}$,

$$p(\theta_k \mid D_k, m = 1). \tag{4.14}$$

The best-fit values of θ can be obtained by using χ^2 minimization or the MAP estimation. As this approach assumes that $\epsilon = \epsilon^{(1)}$, it does not account for atomic uncertainty. Both Figure 4.7 and Lee et al. (2011) illustrate that this can mislead estimates of θ_k and can significantly underestimate the error bars associated with these estimates.

Now, equipped with the ensemble set of CHIANTI emissivities corresponding to randomly selected EIS pixels, we can consider the uncertainties in the fitted plasma parameters in each case that result from both statistical fluctuations in the observed intensities and the atomic data uncertainties incorporated in the ensemble of CHIANTI emissivities.

This thesis focuses on independent analyses for randomly selected pixels. There is future potential to model spatial structure of the estimated physical quantities into the analyses. In this case, we have to consider a more complicated model where nearby pixels have similar density and similar temperature which requires Gaussian Process* for the density and temperature parameters. The model will get more complicated and is out of the scope of this thesis.

Pragmatic Bayesian method

Rather than conditioning on a fixed value, we treat $\epsilon^{(m)}$, or more precisely, its index m, as unknown. For the pragmatic Bayesian method, we assume that the observed intensities are uninformative as to the most likely emissivities. That is, we do not take into account the information in the intensities for narrowing the uncertainty in the choice of emissivity realizations. Mathematically, this assumption can be written as $p(m \mid D_k) = p(m)$, i.e., $\epsilon^{(m)}$ and D_k are independent. Thus, the pragmatic Bayesian joint posterior distribution of m and θ_k can be written from Eq (3.10) as

$$p_{\mathrm{pB}}(m, \theta_k \mid D_k) = p(\theta_k \mid m, D_k)\ p(m). \tag{4.15}$$

Under the pragmatic model in Eq (4.15), inference for θ_k is based on its marginal posterior distribution, which can be rewritten from Eq (3.10) as,

$$\begin{aligned} p_{\mathrm{pB}}(\theta_k \mid D_k) &= \sum_{m=1}^{M} p_{\mathrm{pB}}(m, \theta_k \mid D_k) \\ &= \sum_{m=1}^{M} p(\theta_k \mid D_k, m)\ p(m). \end{aligned} \tag{4.16}$$

The pragmatic Bayesian method accounts for atomic uncertainty in a conservative manner. The assumption that $p(m \mid D_k) = p(m)$ ignores information in the intensities, D_k, that may reduce uncertainty of atomic data repre-

*A Gaussian process is a stochastic process where every finite linear combination of those random variables has a multivariate normal distribution.

sented by m and hence of θ_k. We now consider methods that allow D_k to be informative for m.

Fully Bayesian method

The fully Bayesian method incorporates the potential information in the data (i.e., the intensities) to learn about $\epsilon^{(m)}$. The fully Bayesian joint posterior distribution of m and θ_k can be rewritten from Eq (3.12) as

$$p_{\mathrm{fB}}(\theta_k, m \mid D_k) = p(\theta_k \mid m, D_k)\; p(m \mid D_k), \tag{4.17}$$

and the marginal posterior distribution of θ_k is given by

$$\begin{aligned} p_{\mathrm{fB}}(\theta_k \mid D_k) &= \sum_{m=1}^{M} p_{\mathrm{fB}}(m, \theta_k \mid D_k) \\ &= \sum_{m=1}^{M} p(\theta_k \mid D_k, m)\; p(m \mid D_k), \end{aligned} \tag{4.18}$$

where each $p_{\mathrm{fB}}(\theta_k \mid D_k)$ is normalized so that $\sum_{m=1}^{M} p(m \mid D_k) = 1$.

Using Bayes' theorem and assuming each emissivity realization in $\mathcal{M}$ is equally likely, as indicated by Eq (3.14), we can directly compute the probability of each emissivity realization, m, given the data separately,

$$p(m \mid D_k) = \frac{p(D_k \mid m)}{\sum_{m=1}^{M} p(D_k \mid m)}. \tag{4.19}$$

This is the marginal posterior probability among those emissivity realizations in $\mathcal{M}$. Eq (4.19) holds because each of the m has the same prior probability (see Eq (3.14)).

The Bayesian posterior distribution in Eq (4.19) allows the observed intensities to be informative for the atomic physics, following the principles of Bayesian analysis (Xu et al. 2014). It enables us to use the intensities to determine which emissivity realizations are more or less likely and to average over (posterior) uncertainty in emissivity realizations.

4.5.2 Algorithms for the separate pixel-by-pixel analyses

Algorithms for pragmatic Bayesian in the separate pixel-by-pixel analyses

To obtain a MC sample of (m, θ_k) from the pragmatic Bayesian posterior in Eq (4.15), we first obtain a MC sample of the emissivity index, $\{m^{(1)}, \ldots, m^{(L)}\}$, from its prior distribution, Eq (3.14). For each $m^{(\ell)}$, with $\ell = 1, \ldots, L$, we can then sample $\{\theta_k^{[\ell,t]}, t = 1, \ldots, T\}$ from $p(\theta_k \mid m^{(\ell)}, D_k)$ using the MH algorithm discussed in Section 2.2.1. This requires that we specify the proposal distribution $q(\theta^*|\theta^{(t)})$. To do so, we first compute the value of θ_k that maximizes $\log p(\theta_k \mid m^{(\ell)}, D_k)$, i.e., the MAP estimates, $\hat{\theta}_k$, along with the 2×2 Hessian matrix evaluated at the mode $\hat{\theta}_k$, $H(\hat{\theta}_k)$, for each $m^{(\ell)}$. We then use $t_4\left(\theta_k \mid \hat{\theta}_k, (-H(\hat{\theta}_k))^{-1}\right)$ as the MH proposal distribution, where $t_\nu(x \mid \mu, \Sigma)$ is the density of a multivariate t distribution with ν degrees of freedom, mode μ, and scale matrix Σ, evaluated at x. This type of MH sampler is known as an independence sampler (Gilks et al. 1996). The use of Hessian matrix evaluated at the mode makes a proposal distribution that mimics the correlation structure of the target distribution, i.e., the conditional posterior distribution $p(\theta_k \mid m^{(\ell)}, D_k)$. In other words, we are sampling from a correlated proposal distribution that is already matching the target distribution. We run MH for T iterations, the last of which is taken as the MC sample corresponding to $m^{(\ell)}$, i.e., $\theta_k^{(\ell)} = \theta_k^{[\ell,T]}$.

Algorithms for fully Bayesian in the separate pixel-by-pixel analyses

In the fully Bayesian separate pixel-by-pixel analyses, our aim is to obtain a MC sample from the joint posterior distribution, Eq (4.17), and we propose three basic strategies for doing this: (i) two-step MC with MH, (ii) two-step MC with a Gaussian approximation, and (iii) HMC, as described below. Specifically, the first strategy uses the MH algorithm while the second strategy makes a Gaussian approximation to the conditional distribution of θ_k given the sampled emissivity realization m. Comparing the three strategies,

the two-step MC with MH is preferred because of the accuracy of estimates with moderate computation time, while two-step MC with a Gaussian approximation may be faster (but less accurate) and HMC can be more accurate (but slower) under certain conditions.

Implementation of two-step MC with MH for fully Bayesian in the separate pixel-by-pixel analyses

In order to implement the fully Bayesian method and to obtain a MC sample of θ_k via Eq (4.18), we first evaluate Eq (4.19) for each m where

$$p(D_k \mid m) = \int L(m, \theta_k \mid D_k)\, p(\theta_k)\, \mathrm{d}\theta_k \tag{4.20}$$

is the Bayesian evidence conditional on a given emissivity. For each sampled θ_k, we need only evaluate the likelihood for $m = 1, \ldots, M$, and then renormalize the M likelihood values by this weighted sum, which can be achieved via a two-step sampling as described in this section.

The two dimensional integral in Eq (4.20) can be evaluated numerically using the grid generated from the trapezoidal quadrature rule (TQR), which is suitable for finite domain quadrature[†]. The product-rule is also used in the construction of multivariate grids, which leads to an evenly designed grid.

The two dimensional quadrature can then be expressed as

$$\begin{aligned} &\int L(m, \theta_k \mid D_k)\, p(\theta_k)\, \mathrm{d}\theta_k \\ &\qquad = \sum_{i,j} w_{i,j}\, L(m, \log \mathrm{n}_k^{(i)}, \log \mathrm{d}_k^{(j)} \mid D_k)\, p(\log \mathrm{n}_k^{(i)}, \log \mathrm{d}_k^{(j)}) \end{aligned} \tag{4.21}$$

where nodes $(\log \mathrm{n}_k^{(i)}, \log \mathrm{d}_k^{(j)})$ and weights $(w_{i,j})$ are defined by the chosen quadrature rule[‡]. The integral range of the two parameters is $(\hat{\theta}_k - 3 \times$

[†]Package 'mvQuad' provides a collection of methods for (potentially) multivariate quadrature in R, and is available at https://cran.r-project.org/web/packages/mvQuad/.

[‡]TQR and Product-Rule are used in the construction of multivariate grids, where level = 5 is a subcommand in the grid creating commander, which represents accuracy

$\text{sdev}_k, \hat{\theta}_k + 3 \times \text{sdev}_k)$ where sdev_k is a vector of the square root of the diagonal elements in variance-covariance matrix $(-H(\hat{\theta}_k))^{-1}$.

Having evaluated Eq (4.19) at each m, we can obtain a MC sample of the emissivity index, $\{m^{(1)}, \ldots, m^{(L)}\}$. For each $m^{(\ell)}$ we sample from $p(\theta_k \mid D_k, m^{(\ell)})$ using an independence sampler exactly as described in the pragmatic Bayesian case above. For each $m^{(\ell)}$, we run the independence sampler for T iterations to obtain the MC sample corresponding to $m^{(\ell)}$, and set $\theta_k^{(\ell)} = \theta_k^{[\ell,T]}$. For Pixel k, i.e. the kth set of intensities, the detailed two-step MC with MH ($\mathcal{S}_{MH}$) proceeds for $\ell = 1, \ldots, L$ with

Step 1: Sample $m^{(\ell)} \sim p(m \mid D_k)$ via Eq (4.19).

Step 2: For $t = 1, \ldots, T-1$,

Step 2.1: Sample $\theta_k^{[\text{prop}]} \sim t_4\left(\theta_k \mid \hat{\theta}_k, (-H(\hat{\theta}_k))^{-1}\right)$ and compute

$$\rho = \frac{p(\theta_k^{[\text{prop}]} \mid D_k, m^{(\ell)})\ t_4\left(\theta_k^{[t]} \mid \hat{\theta}_k, (-H(\hat{\theta}_k))^{-1}\right)}{p(\theta_k^{[t]} \mid D_k, m^{(\ell)})\ t_4\left(\theta_k^{[\text{prop}]} \mid \hat{\theta}_k, (-H(\hat{\theta}_k))^{-1}\right)}. \quad (4.22)$$

Step 2.2: Set

$$\theta_k^{[\ell,t+1]} = \begin{cases} \theta_k^{[\text{prop}]}, & \text{with probability } \min(\rho, 1), \\ \theta_k^{[t]}, & \text{otherwise.} \end{cases} \quad (4.23)$$

Step 3: Set $\theta_k^{(\ell)} = \theta_k^{[\ell,T]}$.

For simplicity at each iteration, if the sampled emissivity index in Step 1 is the same as the previous draw, we do not need to iterate MH to sample θ_k in Step 2 since we already have a good proposal distribution for the same target distribution. Moreover, if there does exist one dominant emissivity curve, e.g., there exists m^* such that $p(m^* \mid D_k) \geq 0.9999$, we only need to sample this m^* all the time.

level, typically number of evaluation points for the parameters in each dimension.

Implementation of two-step MC with Gaussian approximation for fully Bayesian in the separate pixel-by-pixel analyses

This is an alternative method to sample $p(m, \theta_k \mid D_k)$ based on Eq (4.18) and Eq (4.19). Same as above, we can evaluate Eq (4.19) at each m and obtain a MC sample of the emissivity index, $\{m^{(1)}, \ldots, m^{(L)}\}$. For each $m^{(\ell)}$, instead of using exact MH algorithm, we can then sample from $p(\theta_k \mid D_k, m^{(\ell)})$ by considering an approximate algorithm via Gaussian approximation.

We can conduct a Gaussian approximation to $p(\theta_k \mid D_k, m^{(\ell)})$ with mean equal to the MAP estimates, $\hat{\theta}_k$, and variance-covariance matrix $(-H(\hat{\theta}_k))^{-1}$. Specifically the Gaussian approximation distribution $\mathcal{N}\left(\theta_k \mid \hat{\theta}_k, (-H(\hat{\theta}_k))^{-1}\right)$ has the same mode and curvature as the target conditional distribution $p(\theta_k \mid D_k, m^{(\ell)})$. Thus the two-step MC with Gaussian approximation ($\mathcal{S}_G$) proceeds for $\ell = 1, \ldots, L$ with

Step 1: Sample $m^{(\ell)} \sim p(m \mid D_k)$ via Eq (4.19).

Step 2: Sample $\theta_k^{(\ell)} \sim \mathcal{N}\left(\theta_k \mid \hat{\theta}_k, (-H(\hat{\theta}_k))^{-1}\right)$, where $\hat{\theta}_k$ depends on $m^{(\ell)}$.

Similar to two-step MC with MH above, if there is one dominant emissivity curve, we only need to sample this dominant one all the time.

Implementation of Hamiltonian Monte Carlo for fully Bayesian in the separate pixel-by-pixel analyses

Another alternative method to obtain a MC sample from the joint posterior distribution in Eq (4.12) via the separate analyses is to start by obtaining a sample from their marginal posterior distribution,

$$\theta_k^{(1)}, \ldots, \theta_k^{(L)} \sim p(\theta_k \mid D_k).$$

First, we rewrite

$$p(\theta_k \mid D_k) \propto L(\theta_k \mid D_k)\, p(\theta_k), \tag{4.24}$$

where

$$
\begin{aligned}
L(\theta_k \mid D_k) &= \sum_{m=1}^{M} L(m, \theta_k \mid D_k)\ p(m \mid \theta_k) \\
&= \frac{1}{M} \sum_{m=1}^{M} L(m, \theta_k \mid D_k) \\
&= \frac{1}{M} \sum_{m=1}^{M} \prod_{h=1}^{H} \mathcal{N}\left(I_{k\lambda_h} \mid \epsilon_{\lambda_h}^{(m)}(\mathrm{n}_k, \mathrm{T}_k)\mathrm{n}_k^2\mathrm{d}_k,\ \sigma_{k\lambda_h}^2 \right),
\end{aligned}
\tag{4.25}
$$

since the prior independent assumption, $p(m \mid \theta_k) = p(m) = 1/M$, and the observation independent assumption among lines of wavelengths.

Evaluating $p(\theta_k \mid D_k)$ in this way we can use the Stan package, as described in Section 3.4.2, to obtain $\{\theta_k^{(1)}, \ldots, \theta_k^{(L)}\}$ via HMC to sample directly from its marginal posterior distribution, Eq (4.24). However, we must analytically marginalize over m, via Eq (4.25), since it cannot accommodate discrete parameters.

With these MC samples $\{\theta_k^{(1)}, \ldots, \theta_k^{(L)}\}$ in hand, we can sample m from its conditional posterior distribution,

$$
\begin{aligned}
p(m \mid \theta_k^{(\ell)}, D_k) &= \frac{p(m)\ L(m, \theta_k^{(\ell)} \mid D_k)}{\sum_{\tilde{m}=1}^{M} p(\tilde{m})\ L(\tilde{m}, \theta_k^{(\ell)} \mid D_k)} \\
&= \frac{L(m, \theta_k^{(\ell)} \mid D_k)}{\sum_{\tilde{m}=1}^{M} L(\tilde{m}, \theta_k^{(\ell)} \mid D_k)},
\end{aligned}
\tag{4.26}
$$

for $\ell = 1, \ldots, L$.

Sampling multimodal posterior distributions with Stan

The samples obtained above show bimodal posterior distributions for parameters $\theta_k = (\log \mathrm{n}_k, \log \mathrm{d}_k)$ for a couple of pixel datasets. Specifically, the two modes correspond to the two different emissivity curves. The resulting relative size of the two modes does not match the actual posterior distributions indicating HMC algorithm has trouble in jumping between the modes. This multiple-mode problem may be due to an insufficient number of emis-

sivity curves because our set of emissivities sample the full uncertainty range sparsely. To solve this problem, we have experimented with adding a few strategically chosen synthetic emissivity curves to the set $\mathcal{M}$ and the augmented set of curves is denoted by $\mathcal{M}^{\text{aug}}$, where $\mathcal{M}$ is a subset of $\mathcal{M}^{\text{aug}}$, i.e., $\mathcal{M} \subset \mathcal{M}^{\text{aug}}$. These tend to connect the modes and allow HMC to jump between modes. We can then remove the samples associated with the synthetic emissivity curves to get MC samples purely from the original target.

We run the algorithm described above with $\mathcal{M}$ replaced by $\mathcal{M}^{\text{aug}}$. For each sampled value of $\theta_k^{(\ell)}$, $\ell = 1, \ldots, L$, we compute $p(m \mid \theta_k^{(\ell)}, D_k)$ for each $m \in \mathcal{M}^{\text{aug}}$, with $\mathcal{M}$ replaced by $\mathcal{M}^{\text{aug}}$ in Eq (4.26), and sample a value of m, say $m^{(\ell)}$, from it. Once we have these sample values of m, $m^{(\ell)}$, for $\ell = 1, \ldots, L$, we can then extract the samples of θ_k that correspond to the non-synthetic emissivity curves to get MC samples purely from the original target, i.e., consider the conditional posterior distribution $p(m \mid \theta_k^{(\ell)}, D_k)$ for each $m \in \mathcal{M}$.

This creative method of adding synthetic emissivity curves in HMC can be generalised to all pixel datasets. If all the multiple-mode pixels have two modes and these two modes depend on the two same emissivity curves, the same synthetic emissivity curves can be added into the original ones and the above procedure can be repeated to all pixel datasets.

4.5.3 Model and algorithms for simultaneous analysis

When we consider all the K-pixel intensities together in a simultaneous analysis using the fully Bayesian method, the likelihood function of m and Θ given $\mathcal{D}$, and the prior distribution of m and Θ are, respectively,

$$L(m, \Theta \mid \mathcal{D}) = \prod_{k=1}^{K} L(m, \theta_k \mid D_k) \tag{4.27}$$

and

$$p(m, \Theta) = p(m) \prod_{k=1}^{K} p(\theta_k). \tag{4.28}$$

Thus, the joint posterior distribution of m and Θ can be expressed as

$$p(m, \Theta \mid \mathcal{D}) = \frac{L(m, \Theta \mid \mathcal{D}) p(m, \Theta)}{p(\mathcal{D})}, \tag{4.29}$$

where $p(\mathcal{D}) = \sum_{m=1}^{M} \int L(m, \Theta \mid \mathcal{D}) p(m, \Theta) \, \mathrm{d}\Theta$. Similarly, treating m as an unknown parameter, we express the left hand side of Eq (4.29) as

$$p(m, \Theta \mid \mathcal{D}) = p(\Theta \mid \mathcal{D}, m) \; p(m \mid \mathcal{D}), \tag{4.30}$$

and we conduct statistical inference by obtaining a MC sample from this joint posterior distribution.

First we can use all the data simultaneously to obtain the marginal posterior probability of each emissivity realization m,

$$p(m \mid \mathcal{D}) = \frac{\prod_{k=1}^{K} p(D_k \mid m)}{\sum_{m=1}^{M} \prod_{k=1}^{K} p(D_k \mid m)}. \tag{4.31}$$

and sample $m^{(\ell)}$, for $\ell = 1, \ldots, L$, with weights given by the marginal posterior probabilities in Eq (4.31) so that those favoured by the data are sampled more frequently. The computation of $p(D_k \mid m)$ for each k and m is discussed in Section 4.5.2.

For each sampled m, we sample θ from its conditional posterior distribution

$$\begin{aligned} p(\Theta \mid \mathcal{D}, m) &\propto L(m, \Theta \mid \mathcal{D}) \; p(\Theta) \\ &= \prod_{k=1}^{K} L(m, \theta_k \mid D_k) \; p(\theta_k) \\ &= \prod_{k=1}^{K} \prod_{h=1}^{H} \mathcal{N}\left(I_{k\lambda_h} \mid \epsilon_{\lambda_h}^{(m)}(\mathrm{n}_k, \mathrm{T}_k) \mathrm{n}_k^2 \mathrm{d}_k, \; \sigma_{k\lambda_h}^2 \right) \times p(\log \mathrm{n}_k) \; p(\log \mathrm{d}_k). \end{aligned} \tag{4.32}$$

as these K-pixel datasets were randomly selected from the observations indicated in Section 1.1.1, so that we can safely assume conditional independence among them.

Similarly, an MH sampler is used to obtain a correlated MC sample, $\{\Theta^{[t]}, t = 1, \ldots, T\}$, from $p(\Theta \mid m^{(\ell)}, \mathcal{D})$. A $t_4\left(\theta_k \mid \hat{\theta}_k, (-H(\hat{\theta}_k))^{-1}\right)$ proposal distribution is used for each pixel independently and separately to make the computation more efficient. With those proposal distributions, we run the MH for T iterations over all the K-pixel intensities and obtain the MC sampler corresponding to $m^{(\ell)}$, $\Theta^{(\ell)} = \Theta^{[\ell,T]}$. The detailed two-step MC with MH via simultaneous analysis ($\mathcal{S}_{MH_{simul}}$) proceeds for $\ell = 1, \ldots, L$ with

Step 1: Sample $m^{(\ell)} \sim p(m \mid \mathcal{D})$ via Eq (4.31).

Step 2: Proceed for $t = 1, \ldots, T$,

Step 2.1: For each pixel $k = 1, \ldots, K$,
sample $\theta_k^{[\text{prop}]} \sim t_4\left(\theta_k \mid \hat{\theta}_k, (-H(\hat{\theta}_k))^{-1}\right)$
and set $\Theta^{[\text{prop}]} = (\theta_1^{[\text{prop}]}, \ldots, \theta_K^{[\text{prop}]})$.

Step 2.2: Compute

$$\rho = \frac{\prod_{k=1}^{K} p(\theta_k^{[\text{prop}]} \mid D_k, m^{(\ell)}) \cdot \prod_{k=1}^{K} t_4\left(\theta_k^{[t]} \mid \hat{\theta}_k, (-H(\hat{\theta}_k))^{-1}\right)}{\prod_{k=1}^{K} p(\theta_k^{[t]} \mid D_k, m^{(\ell)}) \cdot \prod_{k=1}^{K} t_4\left(\theta_k^{[\text{prop}]} \mid \hat{\theta}_k, (-H(\hat{\theta}_k))^{-1}\right)}. \tag{4.33}$$

Step 2.3: Set

$$\Theta^{[\ell,t+1]} = \begin{cases} \Theta^{[\text{prop}]}, & \text{with probability } \min(\rho, 1), \\ \Theta^{[t]}, & \text{otherwise.} \end{cases} \tag{4.34}$$

Step 3: Set $\Theta^{(\ell)} = \Theta^{[\ell,T]}$.

Similar to the separate analyses, for simplicity at each iteration, if the sampled emissivity index in Step 1 is not updated, we do not need to iterate MH to sample each θ_k in Step 2. If there does exist one dominant emissivity curve, we only need to sample the dominant all the time.

4.5.4 Application to simulated intensities

Here we illustrate both the separate pixel-by-pixel and the simultaneous analyses, mentioned in Section 4.5.1 and Section 4.5.3, with a simulated case. The $K = 1000$ simulated sets of intensities for each of $H = 7$ spectral lines are generated from known density and path lengths and from the $m = 1$ default CHIANTI emissivity, i.e., $\epsilon^{(1)}$, as in Section 3.1 and Yu et al. (2018). This will allow us, statistically, to test the ability to recover physical parameters from the Fe XIII intensities and, physically, to illustrate how the variations in the atomic data led to the variations in the inferred densities and path lengths. It allows for the comparison of the inferred values with the true values to compare the separate pixel-by-pixel analyses with the simultaneous analysis, making sure the posterior distribution is giving a reasonable estimate of the parameters.

We run the two-step MC with MH in both the separate pixel-by-pixel analyses and the simultaneous analysis described in Section 4.5.2 and Section 4.5.3. TQR and Product Rule are used in computing multivariate quadrature in Eq (4.20) and we obtain a MC sample of emissivity index via Eq (4.19) or Eq (4.31). For both analyses, 30 MH samplers, which is determined by constructing autocorrelation plots in this setting (Xu et al. 2014), are drawn for each sampled emissivity realization $m^{(\ell)}$, and the last MH sampler is taken as a MC sampler. There are 8000 MC samplers drawn in each simulation. We run the two-step MC with Gaussian approximation in the separate pixel-by-pixel analyses. Same as two-step MC with MH, TQR and Product Rule are used to obtain a MC sample of emissivity index, a Gaussian approximation is conducted to $p(\theta_k \mid D_k, m^{(\ell)})$ for each sampled $m^{(\ell)}$ and each pixel D_k, and 8000 MC samplers drawn for each pixel. We also run HMC in the separate pixel-by-pixel analyses where 5 chains are running, 4000 iterations each, and the first half of the iterations of each chain are discarded as burn-in.

The comparison of the relative posterior probability $p(m \mid D_k)$ for each emissivity index and for each pixel, for all three algorithms in separate pixel-by-pixel analyses, is shown in the top panel of Figure 4.4. The emissivity realization with index 1 occupies almost all of the probability. Similarly, in

the simultaneous analysis, the posterior probability of the emissivity realization with index 1 is nearly to one. Both analyses recover the fact that all of the simulated sets of intensities are computed from the default CHIANTI atomic data (the emissivity realization with index 1) instead of the perturbed atomic data (the emissivity realizations with other indices) as described in Yu et al. (2018).

The results in Figure 4.5 compare the fitted values and error bars using two-step MC with MH to the true value of both parameters $\log \mathrm{n}$ and $\log \mathrm{d}$ via both the separate pixel-by-pixel and the simultaneous analyses. Compared with the separate pixel-by-pixel analyses, it shows that the error bars are smaller around the true values when we use the simultaneous analysis than using one pixel dataset at a time. The results in the plots illustrate that, as more data are used in the analysis by simultaneously analyzing those pixels, incorporating the uncertainty in the atomic physics calculations results in more accurate fitted values.

Comparing the results from the separate pixel-by-pixel and the simultaneous analyses using their mean square errors (MSEs), a measure of how well the fitted values explain the given set of observations, Table 4.1 shows simultaneous analysis achieves smaller MSE values and indicates the more data we have, the smaller MSE is achieved, i.e., simultaneous analysis gives a better explanation of the given set of observations (i.e., intensities). Consistently, the three algorithms in separate pixel-by-pixel analyses achieves almost the same MSE, but all larger than that of from the simultaneous analysis.

Algorithm		MSE	
		$\log \mathrm{n}$	$\log \mathrm{d}$
Pixel-by-pixel	$\mathcal{S}_{MH}$	1.345×10^{-5}	4.936×10^{-5}
	$\mathcal{S}_{G}$	1.184×10^{-5}	4.443×10^{-5}
	$\mathcal{H}$	4.931×10^{-5}	1.936×10^{-4}
Joint	$\mathcal{S}_{MH_{joint}}$	6.748×10^{-7}	2.241×10^{-6}

Table 4.1: MSE between the fitted values and the true values for both parameters $(\log \mathrm{n}, \log \mathrm{d})$ via both separate pixel-by-pixel and simultaneous analyses in simulated data set.

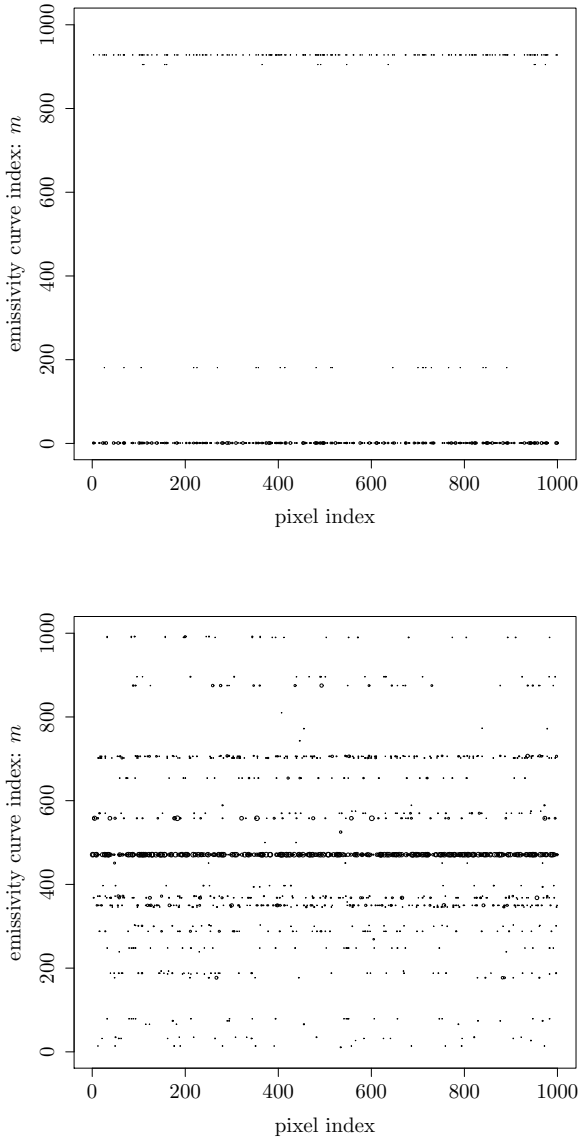

Figure 4.4: Selecting the optimal emissivity curves with separate pixel-by-pixel fully Bayesian analysis. The x-axis and the y-axis represent the index of the pixels and the index of the emissivity curves, respectively. For each pixel, the relative posterior probability is plotted along a vertical column for the emissivity indices, where index 1 represents the default CHIANTI emissivities. The size of the dots represents the relative values of the posterior probability ($p(m \mid D_k)$, for emissivity index m and pixel data set D_k) assigned to each emissivity index for a given pixel. The analyses carried out for the simulated data set (top; generated using default $m = 1$, and showing only $p(m \mid D_k) > 0.06$) and for a real data set (bottom; showing only $p(m \mid D_k) > 0.1$) are shown.

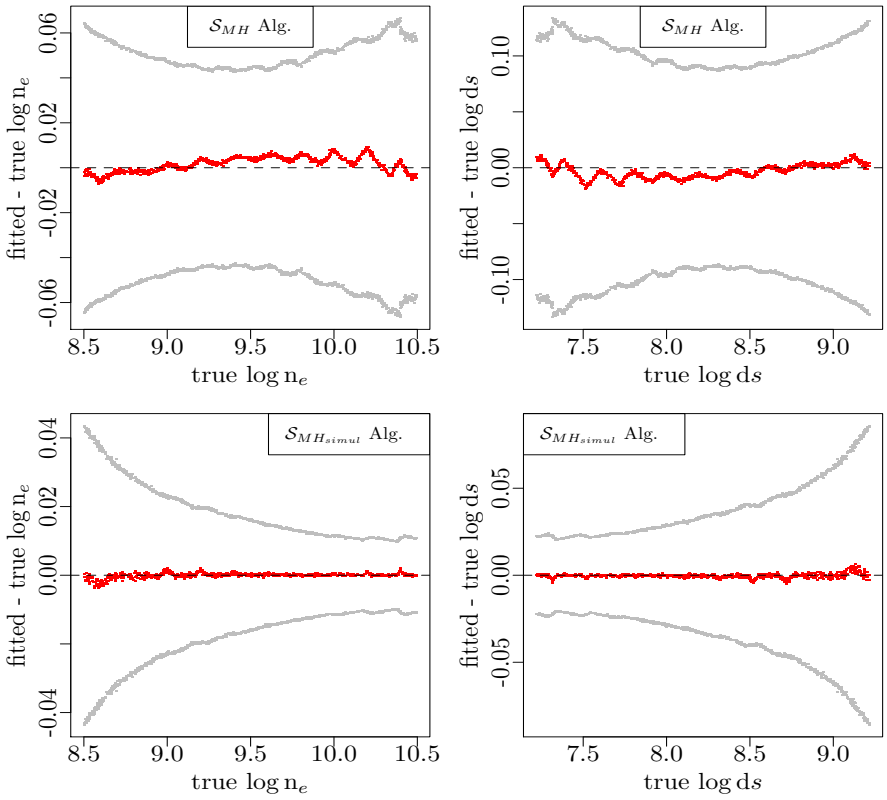

Figure 4.5: Comparison of best-fit values and actual input for the simulated dataset. The comparisons are shown for both $\log \mathrm{n}$ (left column) and $\log \mathrm{d}$ (right column). Calculations are performed using the fully Bayesian two-step MC with MH for each pixel dataset separately (top row) and for all the pixel datasets simultaneously (bottom row). The red dots represent the difference between the best-fit value and the actual input and the horizontal dashed lines represent the line of equality. The grey dots represent a vertical error of ± 1 standard deviation for the fitted values that incorporates atomic data uncertainty. Notice that the uncertainties are reduced when all the K-pixel datasets are used simultaneously.

4.5.5 Application to observed intensities

Here we demonstrate the effects of the different types of analyses by applying them to a real dataset, the EIS full-CCD observations of an active region used as an example (Yu et al. 2018). This dataset comprises sets of measured intensities of $H = 7$ spectral lines of Fe XIII in $K = 1000$ distinct, independent pixels.

Two-step MC with MH ($\mathcal{S}_{MH}$) and two-step MC with Gaussian approximation ($\mathcal{S}_G$) algorithm

Same as what is done in the simulated case, TQR and Product-Rule are used to obtain a MC sample of emissivity index first. A MH sampler is used to obtain a correlated MC sample from $p(\theta_k \mid D_k, m^{(\ell)})$ for each sampled $m^{(\ell)}$ and each pixel D_k in $\mathcal{S}_{MH}$. We have 30 MH samplers drawn for each sampled emissivity curve $m^{(\ell)}$ and the last MH sampler is taken as a MC sampler. Equivalently, a Gaussian approximation is conducted to $p(\theta_k \mid D_k, m^{(\ell)})$ for each sampled $m^{(\ell)}$ and each pixel D_k in $\mathcal{S}_G$. There are 8000 MC samplers drawn for each pixel in each algorithm.

HMC with Stan ($\mathcal{H}$)

A few strategically chosen synthetic emissivity atomic data curve replicates are added, as described in Section 4.5.2. There are 5 chains running, 4000 iterations each, and the first 2000 iterations of each chain are discarded as burn-in.

In the realistic case, once we run HMC with Stan as described in Section 4.5.2, bimodal posterior distributions are observed for several pixels. The two modes correspond to two different emissivity curves with index 471 and 368, i.e., Emis_{471} and Emis_{368}. Moreover, the relative size of the two modes does not match the actual posterior distribution as shown in the left column of Figure 4.6. Therefore, a few strategically chosen synthetic emissivity curves

are added to the original set and the augmented set is

$$\mathcal{M}^{\text{aug}} \cap \mathcal{M}^c = \{w_1 * \text{Emis}_{471} + w_2 * \text{Emis}_{368}\}$$

where $(w_1, w_2) = (0.75, 0.25)$, $(0.50, 0.50)$, and $(0.25, 0.75)$. The HMC with Stan is run once more with $\mathcal{M}$ replaced by $\mathcal{M}^{\text{aug}}$, as described in Section 4.5.2. Samples of $\theta_k^{(\ell)}$, $\ell = 1, \ldots, L$, are obtained as shown in the middle column of Figure 4.6. For each sampled value of $\theta_k^{(\ell)}$, we compute $p(m \mid \theta_k^{(\ell)}, D_k)$ for each $m \in \mathcal{M}^{\text{aug}}$, via Eq (4.26), and sample a corresponding $m^{(\ell)}$ from it. Considering the conditional posterior distribution $p(m \mid \theta_k^{(\ell)}, D_k)$ for each $m \in \mathcal{M}$, we can then extract the samples $\theta_k^{(\ell)}$ that correspond to the non-synthetic emissivity curves to get MC samples purely from the original target as shown in the right column of Figure 4.6. The actual conditional posterior distributions are also computed by direct evaluation on a fine grid of the parameters for comparison to the corresponding MC samples.

The results of posterior samples are shown in Figure 4.7 for the pixel #217. The joint posterior probability density distribution $p(\theta_k \mid D_k)$ computed using the pragmatic and the fully Bayesian methods are shown as contour plots, and marginalized 1-D posterior densities $p(\log \text{n}_k \mid D_k)$ and $p(\log \text{d}_k \mid D_k)$ are shown as curves along the corresponding axes. The estimates of $\log \text{n}_k$ and $\log \text{d}_k$ computed via the standard analysis, i.e., the χ^2 minimization of Equation (4.14), are marked with straight lines. Notice that the pragmatic Bayesian method inflates the error bars relative to the standard method as it accounts for the atomic data uncertainties. The fully Bayesian method shrinks the error bars relative to the pragmatic Bayesian method and shifts the best estimate since it selects a subset of the full range of atomic uncertainties that are consistent with the data. The standard method underestimates the uncertainties in all cases, and is shifted relative to the fully Bayesian estimate.

The comparison of the relative posterior probability $p(m \mid D_k)$ for each emissivity index and for each pixel, in separate pixel-by-pixel analyses, is shown in the bottom panel of Figure 4.4. There are two dominant emissivity re-

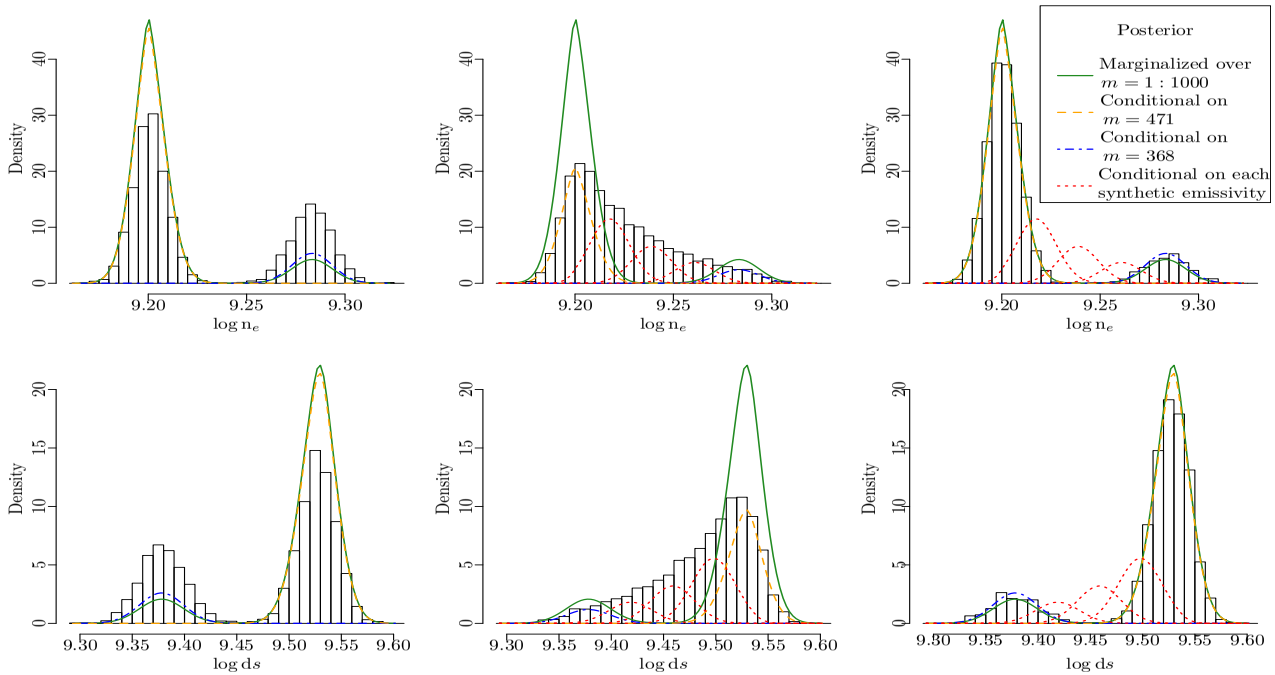

Figure 4.6: The histograms of the posterior values of $\log \mathrm{n}$ (first row) and $\log \mathrm{d}$ (second row) conditional on Pixel **593** in real observed data set via $\mathcal{H}$. The left column is conditional on all 1000 emissivity curves. The middle column is conditional on both 1000 emissivity curves and 3 synthetic emissivity curves. The right column extracts the samples $\theta_k^{(\ell)}$ that correspond to the non-synthetic emissivity curves to get MC samples purely from the original target. The green line is the density line of actual posterior distribution of parameters conditional all 1000 emissivity curves. The orange or blue lines are the density line of actual posterior distribution of parameters conditional on one of the two extracted emissivity curves, Emis_{471} or Emis_{368}. The red line is the density line of actual posterior distribution of parameters conditional on one of the three synthetic emissivity curves, normalized and standardized properly. The actual posterior distribution is computed by direct evaluation on a fine grid of the parameters.

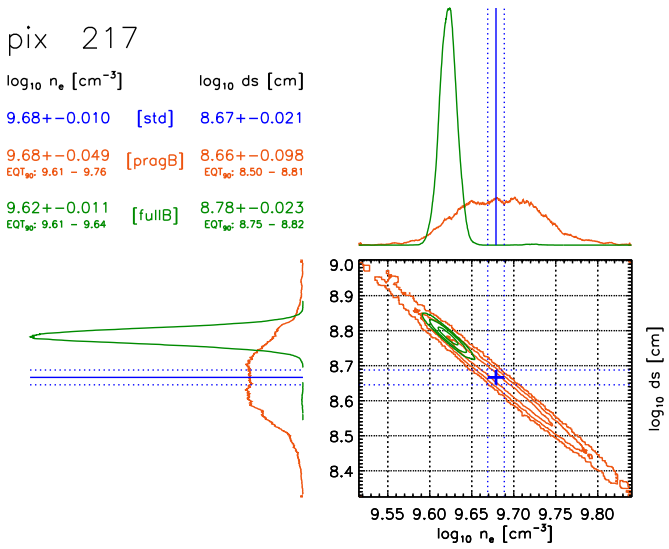

Figure 4.7: Comparisons of the inferred density $\log n_e$ and path length $\log ds$ using different methods for pixel $\#217$. The results from the different methods used are colour coded, with blue representing the standard method, red the pragmatic Bayesian method, and green the fully Bayesian method. The contour plots (with levels at $0.01\times$, $0.1\times$, and $0.5\times$ the maximum) show where the majority of the mass of the joint probability distributions of $(\log \mathrm{n}_k, \log \mathrm{d}_k)$ fall, and their marginalized distributions along each axis are shown to the top and to the left of the corresponding axis. The results from the standard analysis is shown along with the histograms as straight lines (solid for the best-fit and dotted denoting the $\pm 1\sigma$ errors on the best-fit obtained from the default CHIANTI emissivity functions), extending into the contour plot region. The best-fit value from standard analysis is also marked on the contour plot with a '$+$' sign, with the arms of the symbol corresponding to the sizes of the error bars. The standard deviations of the marginalized posterior densities, as well as the 90% equal-tail bounds for both the pragmatic and the fully Bayesian cases are listed in the legend. As expected, the density and path length are highly correlated. The standard method underestimates the uncertainties, the pragmatic Bayesian method inflates them due to atomic data uncertainties. The fully Bayesian method strikes a balance between atomic data uncertainties and how well the data are fit, shrinking the error bars relative to the pragmatic Bayesian and shifting the estimates. The full set of plots for all 1000 pixels considered here are available as a supplementary figure.

alizations which have a combined posterior probability of over 0.99 using the two-step MC and HMC. An example of the posterior probability of the two dominant emissivity realizations given Pixel 593 is shown in Table 4.2. Similarly, in the simultaneous analysis, the posterior probability of the emissivity curve with index 471 is exactly one. It indicates that the emissivity realizations reveal consistent feature of the solar atmosphere.

m	$\mathcal{S}_{MH}$ or $\mathcal{S}_G$	$\mathcal{H}$	$\mathcal{S}_{MH_{simul}}$
471	0.894	0.860	1.000
368	0.105	0.138	0.000
others	< 0.001	< 0.000	0.000

Table 4.2: The posterior probability of the two dominant emissivity realizations given Pixel 593, $p(m \mid D_{593})$, via both separate pixel-by-pixel and simultaneous analyses.

The computational time is considered in terms of (i) the elapsed time and (ii) the sum of the user and system times, which is a closer measure to real clock time. The computation time over all 1000 pixels with the two different measurements using $\mathcal{S}_{MH}$, $\mathcal{S}_G$, and $\mathcal{H}$ in the separate pixel-by-pixel analyses and using $\mathcal{S}_{MH_{simul}}$ in the simultaneous analysis, is shown in Table 4.3. The computation time for two-step MC samplers, $\mathcal{S}_{MH}$, $\mathcal{S}_G$, and $\mathcal{S}_{MH_{simul}}$ consist of both the quadrature part and the sampling part. The computation of the former part is exactly the same for both the separate pixel-by-pixel and the simultaneous analyses with a computation time of 1.2 hours for both measurements. After comparison, HMC with Stan is the most time-consuming algorithm among the three separate pixel-by-pixel analyses algorithms. Moreover, though there are 2000 parameters in the simultaneous model, the computation time using the two-step MC with MH in the simultaneous analyses is quite less than that of the separate pixel-by-pixel analyses. The computation time of the separate pixel-by-pixel analyses using the two-step MC with MH or Gaussian is significantly longer because MAP need to be calculated and Hessian matrix need to be evaluated at MAP for each pixel at each iteration and each inner iteration, which are time consuming. However, though there are 2000 parameters in the simultaneous model, the computation time using the two-step MC with MH in the simultaneous analyses is quite less than that of the separate pixel-by-pixel analyses. That

is because there is one dominant emissivity curve as in Table 4.2 and we sample the dominant all the time so that the calculation of MAP and the evaluation of Hessian at MAP for each pixel only need to be done for one time, as described in Section 4.5.3.

Alg.	Computation Time (hrs)	
	The elapsed time	The sum of user and system time
$\mathcal{S}_{MH}$	14.5	41.0
$\mathcal{S}_{G}$	8.0	20.7
$\mathcal{H}$	51.4	135.5
$\mathcal{S}_{MH_{simul}}$	6.0	6.0

Table 4.3: The computation time (in hours) over all 1000 pixels using the three algorithms, $\mathcal{S}_{MH}$, $\mathcal{S}_{G}$, or $\mathcal{H}$, in separate pixel-by-pixel analyses, with two different measures, the elapsed time (left column) or the sum of user and system times (right column).

4.5.6 Comparison of algorithms and output data analysis

Using test statistics

To obtain a MC sample of the parameters, $\log \mathrm{n}_k$ and $\log \mathrm{d}_k$, via the separate pixel-by-pixel analyses with joint posterior distribution in Eq (4.12), three algorithms, $\mathcal{S}_{G}$, $\mathcal{S}_{MH}$, and $\mathcal{H}$, are implemented for the fully Bayesian model on each of the 1000 pixel observed datasets in Section 4.5.5.

Our aim is to find which algorithm provides a more accurate simulation to the target posterior distribution and is the best to be used to make statistical inference. From a statistical point of view, we assume the HMC, which might give the best result, as the base line, and to see whether these two two-step MC samplers provide better inference or not.

The first test statistic we consider is the z-statistic, which is the difference in posterior mean between the sample values from $\mathcal{S}_{G}$ or $\mathcal{S}_{MH}$ and from $\mathcal{H}$ divided by the standard deviation of samples from $\mathcal{H}$ because HMC is assumed to be the base line, indicating how far away that estimate is from the mean in standard units, i.e.,

$$z^i_{\text{score}} = \frac{\text{mean}_{\mathcal{S}_i} - \text{mean}_{\mathcal{H}}}{\text{sd}_{\mathcal{H}}}, \text{ for } i = G \text{ or } MH. \tag{4.35}$$

Figure 4.8 shows the histograms of z-scores for both parameters, $\log \text{n}$ and $\log \text{d}$, in two comparisons, $\mathcal{S}_G$ to $\mathcal{H}$ and $\mathcal{S}_{MH}$ to $\mathcal{H}$, respectively considering all the 1000 pixels. Looking at the worst case scenarios, the most extremes we see from the comparison on the left-hand side is about 0.12 to 0.25 of standard deviation off, which corresponds to Pixel 36, 87, 302, 453, 650, and 934. The comparison on the right-hand side indicates the most extremes are about 0.15 of standard deviation off occurring at Pixel 302 and 364. The vertical lines correspond to the z-score values of these extracted pixels. This suggests that we need to look at the full posterior distributions for those extreme pixels and for the three algorithms more closely, which will be found below to get some insights.

The second test statistic to compare is the ratio of standard deviations between $\mathcal{S}_G$ or $\mathcal{S}_{MH}$ and $\mathcal{H}$, i.e.,

$$\frac{\text{sd}_{\mathcal{S}_i}}{\text{sd}_{\mathcal{H}}}, \text{ for } i = G \text{ or } MH, \tag{4.36}$$

which essentially gives the relative size of confidence intervals that we compute.

Figure 4.9 shows the histograms of the ratio of standard deviations for both parameters, $\log \text{n}$ and $\log \text{d}$, in two comparisons, $\mathcal{S}_G$ to $\mathcal{H}$ and $\mathcal{S}_{MH}$ to $\mathcal{H}$, respectively considering all the 1000 pixels. The most extremes we see from the comparison on the left-hand side corresponds to Pixel 634 and 779. The comparison on the right-hand side indicates the most extremes occurring at Pixel 396, 418, 634, and 779. The vertical lines correspond to the ratio values of these extracted pixels. An example of their posterior distributions for the three algorithms can be found below.

Parallelization

To improve the efficiency of the code, we parallelize the 1000 pixels into 20

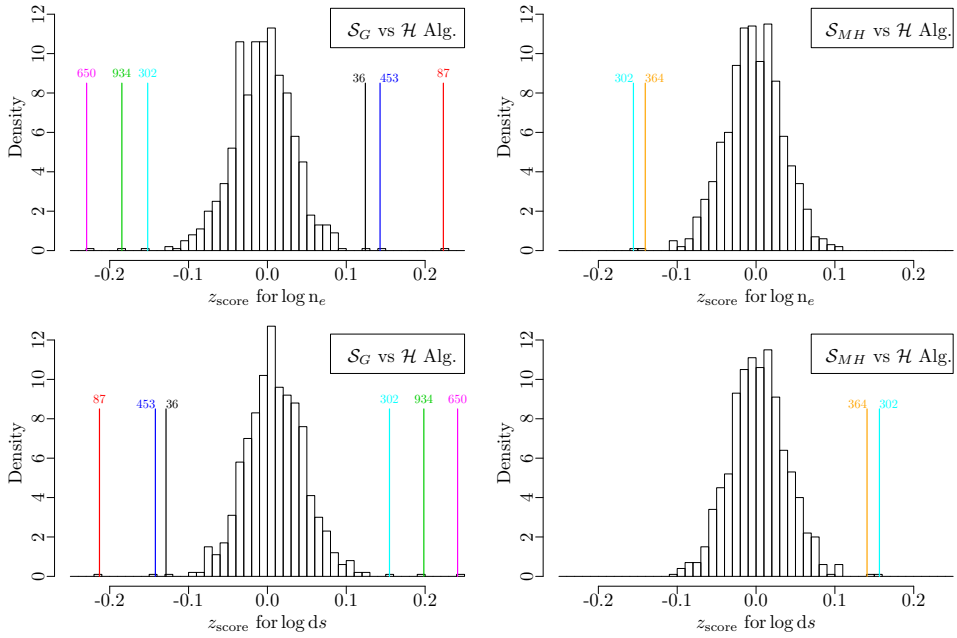

Figure 4.8: Selecting the extreme pixels via z-scores (unit difference in posterior mean) which are computed using the outputs of the three algorithms, $\mathcal{S}_G$, $\mathcal{S}_{MH}$, and $\mathcal{H}$ under fully Bayesian method and separate pixel-by-pixel analyses. The histograms represent the z-scores for both parameters, $\log \mathrm{n}$ (top row) and $\log \mathrm{d}$ (bottom row), in two comparisons (left: $\mathcal{S}_G$ to $\mathcal{H}$, right: $\mathcal{S}_{MH}$ to $\mathcal{H}$) respectively considering all the 1000 pixels. The vertical lines correspond to the values of pixel indices, top left: $650, 934, 302, 36, 453, 87$, top right: $302, 364$, bottom left: $87, 453, 36, 302, 934, 650$, bottom right: $364, 302$, from left to right.

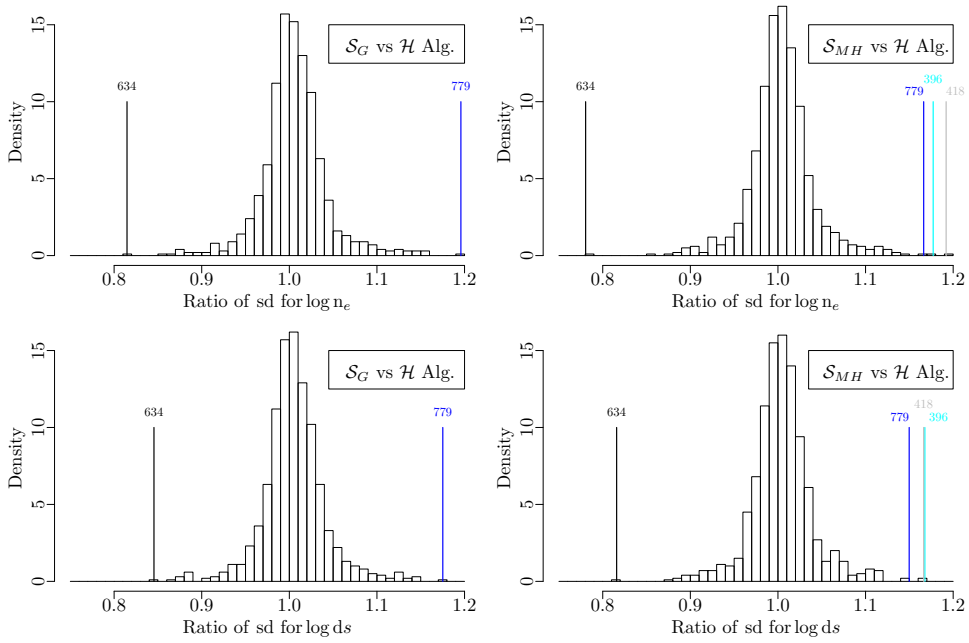

Figure 4.9: Selecting the extreme pixels via the ratio of standard deviations which are computed using the outputs of the three algorithms, $\mathcal{S}_G$, $\mathcal{S}_{MH}$, and $\mathcal{H}$ under fully Bayesian method and separate pixel-by-pixel analyses. The histograms represent the ratio of standard deviations for both parameters, $\log \mathrm{n}$ (top row) and $\log \mathrm{d}$ (bottom row), in two comparisons (left: $\mathcal{S}_G$ to $\mathcal{H}$, right: $\mathcal{S}_{MH}$ to $\mathcal{H}$) respectively considering all the 1000 pixels. The vertical lines correspond to the values of pixel indices, top left: $634, 779$, top right: $634, 779, 396, 418$, bottom left: $634, 779$, bottom right: $634, 779, 418, 396$, from left to right.

or 10 completely separate processes when pre-processing emissivities (i.e., obtaining the posterior probability of each emissivity curve) or sampling θ, for all the three algorithms. The doParallel package is used to provide a mechanism to execute foreach loops in parallel within each process, where a multi-core backend is registered and a four worker cluster (of a 64-bit 2.5 GHz CPU with 128 GB of RAM) is created and used. Specifically, in the source builds, we set the number of processors to use for the build to the number of cores on our machine we want to devote to the build, which is thirty-two. We also set the maximum number of additional R processes allowed to be run in parallel to the current R processes, which is thirty-two as well. For $\mathcal{H}$, each pixel is run with multiple cores and four pixels are run at the same time. For $\mathcal{S}_G$ or $\mathcal{S}_{MH}$, we run each pixel with a different core and thirty-two multi-core backends are used in parallel.

The posterior values of the parameters for the three algorithms and for the extracted pixels

By comparing the three algorithms using the two test statistics mentioned above, several extreme pixels are picked out from each comparison.

Figure 4.10 show the histograms of the posterior values of the parameters $\log \mathrm{n}$ and $\log \mathrm{d}$ conditional on all 1000 emissivity curves and the certain extracted pixel datasets respectively. The results of the three sampling algorithms are compared: $\mathcal{S}_G$ algorithm, $\mathcal{S}_{MH}$ algorithm, and $\mathcal{H}$ algorithm. Three more synthetic emissivity curves are conditioned when using $\mathcal{H}$ as described in Section 4.5.2.

For Pixel 364 (the left panel of Figure 4.10), which are extracted from the right column of Figure 4.8, having used the synthetic emissivity curves, it is still not good at jumping between the modes for $\mathcal{H}$ algorithm in this bimodal case.

For Pixel 396 (the middle panel of Figure 4.10), it is the histograms of $\mathcal{H}$ algorithm that does not quite get into the tail that makes the standard deviation from $\mathcal{H}$ algorithm relatively small and filters this pixel out from the right column of Figure 4.9.

Similarly, for Pixel 650 (the right panel of Figure 4.10), which are extracted from the left column of Figure 4.8, the $\mathcal{S}_G$ algorithm is not good at recovering the actual posterior with a noticeable discrepancy in the mode.

4.5.7 Discussion and conclusions

We have presented the first comprehensive treatment of atomic physics uncertainties in the analysis of solar spectra. To make this analysis tractable, we have considered the relatively simple problem of inferring the electron density and path length from a set of observed Fe XIII intensities and a simple model for the emission (see Equation (4.1)). For this work we have used observed Fe XIII intensities from the EIS spectrometer on the *Hinode* satellite. If we consider only the uncertainties due to counting statistics, we obtain very small error bars on the electron density and path length, suggesting that the parameters are very precisely determined by the observations. We have shown that the effect of atomic uncertainties on estimates of coronal properties can be substantial, and that a pathway exists to reduce this uncertainty by finding subspaces that are preferred by the data, i.e., the emissivity realizations that are supported by the observed intensities.

We have used a Bayesian framework to interpret the observed intensities in the context of the different realizations of the atomic data. A pragmatic Bayesian approach, where each realization of set of emissivities is considered to be equally likely, yields larger uncertainties in the electron density and path length than the uncertainty implied by counting statistics alone. A fully Bayesian approach, where we allow the observed intensities to update the uncertainty in the emissivity curves, reduces the uncertainties in the plasma parameters, but also suggests that certain different realizations of the atomic data are more likely than the default CHIANTI calculation. This indicates some combination of systematic errors in the atomic physics and the observed intensities.

We also have considered two basic strategies for obtaining a MC sample from the joint posterior distribution, HMC and two-step MC sampler. By comparing the histograms of the posterior values, there is definitely an issue with

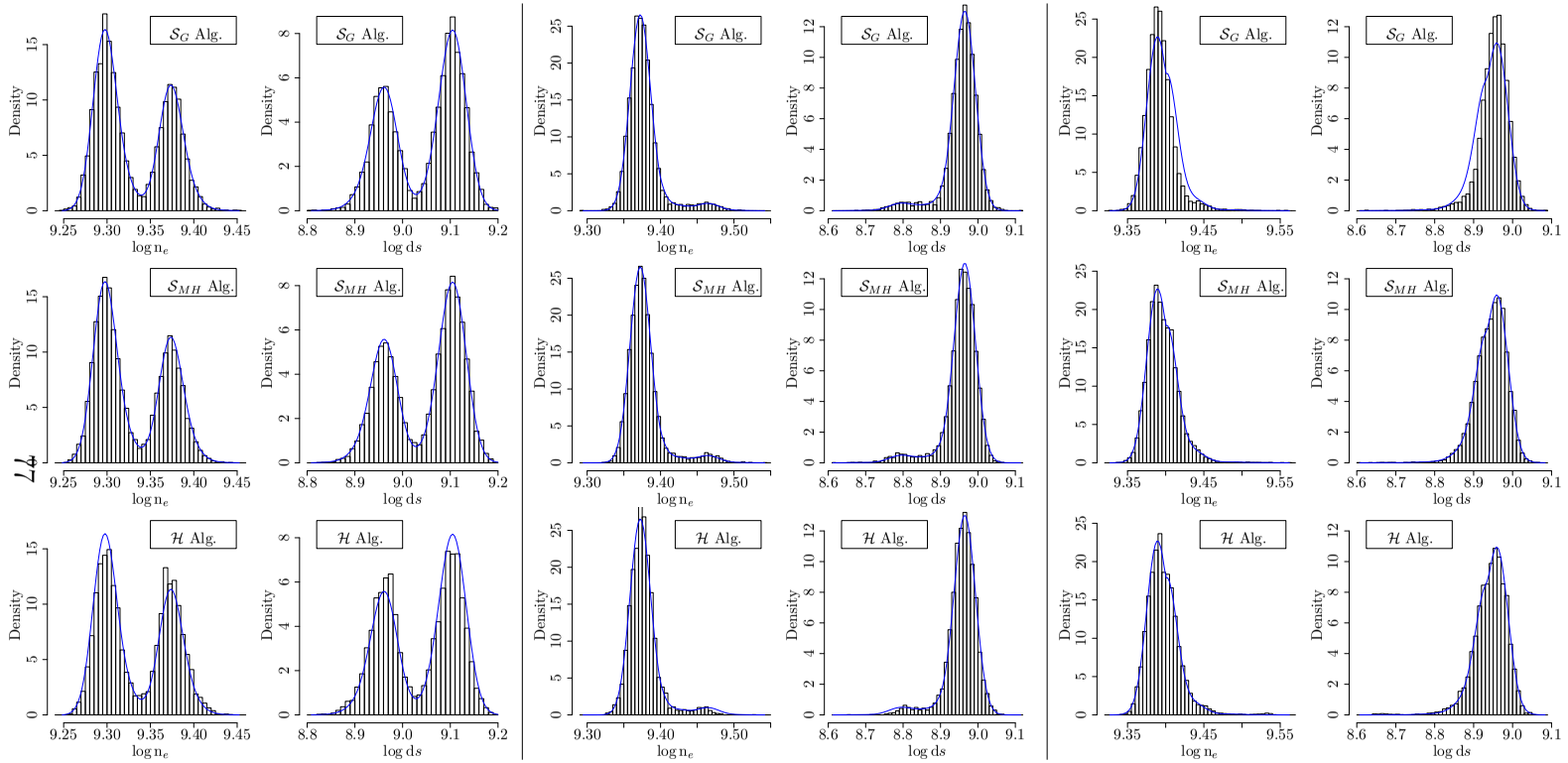

Figure 4.10: Selecting the optimal algorithm from the full posterior distribution for the selected extreme pixels under fully Bayesian method. The histograms represent the posterior distributions of $\log \mathrm{n}$ (left column of each panel) and $\log \mathrm{d}$ (right column of each panel) conditional on all 1000 emissivity curves and Pixel $\mathbf{364}$ (left panel), or $\mathbf{396}$ (medium panel), or $\mathbf{650}$ (right panel) in real observed data set respectively. The sampling algorithms compared are $\mathcal{S}_G$ (first row), $\mathcal{S}_{MH}$ (second row), and $\mathcal{H}$ (third row). The actual conditional posterior distributions, computed by direct evaluation on a fine grid of the parameters, are shown as the blue lines.

the Gaussian assumption (two-step MC with Gaussian approximation) where the MC samplers are not matching very well with the actual posterior and it is more conservative. HMC algorithm looks appropriate but occasionally does not estimate the relative size of the mode correctly, though after adding synthetic emissivity curves. For all 1000 pixels, the MC samplers generated from two-step MC with MH match the density line of actual posterior very well and this algorithm takes moderate computation time. Therefore, more accurate and significantly faster, two-step MC with MH would be the best to use to make statistical inference. An improved version of two-step MC with MH will be used in more complexed case studies in the following chapters.

In our experiment, there are $H = 7$ spectral lines with corresponding wavelengths being considered, whereas two of them are not close to others in wavelength, 196.525 and 209.916 vs 200.021-203.826 Å, and we call them extreme wavelengths. We have experimented with one of the two-mode-case pixels (Pixel 593), where the two extreme wavelengths are removed one at a time from the analysis and the three algorithms mentioned in Section 4.5.2 are repeated. Whether we consider the two extreme wavelengths or not, the resulting MC samplers have a good match to their actual posterior distributions; however, the shape of the actual posterior distribution differs dramatically when including wavelength 196.525Å compared to when it is excluded from the analysis. Because including the extreme wavelengths does not impact the ability of the MC samplers to recover the actual posterior distributions, we use the seven-wavelength dataset in all the experiments.

To keep the analysis simple, in this section, we have assumed that the observed emission can be described by a simple model with a single density, temperature, and path length. Despite its simplicity, this model reproduces the observed intensities remarkably well. However, it cannot overcome any limitations in the model used to interpret the observations from an astrophysics point of view. It would be necessary to consider more complex emission measure distributions if we seek to interpret the plasma parameters derived from the observations.

Referring back to Section 4.5, the real issues of occurring these multiple modes might be we do not have enough diverse set of emissivity curves. In this section, a more realistic and immediate way, PCA, that contains statistical compression of the ensemble of emissivity realizations as described in Section 3.3.2 is implemented, and we aim to incorporate it into the two-step MC Gibbs samplers and the HMC algorithm. Once the PCA generated emissivity is incorporated into the models, more parameters need to be considered. Therefore, a simultaneous analysis, considering multiple datasets jointly, will also be discussed.

Doing with discrete prior on emissivities as in Eq (3.14), we can point to the atomic physicists which of the emissivity curves is relevant. As each of the curves has associated with relative assumptions, it can be traced back to the individual deviations that we used for each level and for each iteration in the atomic physics. Although doing with continuous prior via PCA will lose the connection with the underlying atomic physics, it can fill the gaps between the discrete emissivities and we can work around the problem of sparse sampling of the atomic data space.

In this part, we have a Gaussian prior distribution via PCA on the ensemble of emissivities and aim to embed the PCA model for uncertainty in atomic data into a Bayesian procedure that simultaneously fits the model parameters and accounts for atomic uncertainty. With the PCA representation obtained via a Gaussian prior distribution in Section 4.3.2, we can generate a new MC emissivity replicate from $\mathcal{M}$ on the fly. In this subsection, we describe specific algorithms that incorporate atomic uncertainty into the three data analysis routines, the standard, the pragmatic and the fully Bayesian analyses. However, the incorporation of the PCA generated emissivity curves will result in an increase in the number of model parameters. To avoid overfitting the data, we consider two ways to fit all the plasma parameters, Θ, given all the observed or simulated intensities, $\mathcal{D}$, while accounting for atomic uncertainty, ϵ. First we can analyze a single dataset from each pixel separately in a sequence of pixel-by-pixel analyses and develop both the pragmatic and the

fully Bayesian methods (see Section 4.6.1). Although this may yield different estimates of ϵ, it allows us to see if the intensities of each pixel give consistent information as to the best emissivity curve(s). Here we totally have $H + 2$ model parameters versus 7 data points and the model is overfitted if $H \geq 5$. Alternatively, we can simultaneously analyze the intensities from multiple datasets, e.g., K multiple pixels, to arrive at an overall estimate of the most likely emissivity curve, where there are $H + 2K$ model parameters versus $7K$ data points, and develop both the pragmatic and the fully Bayesian methods (see Section 4.6.2).

4.6.1 Incorporating the PCA generated emissivities into a single pixel

Models and algorithms

Here we incorporate the PCA generated emissivity curve into the models and algorithms in the separate pixel-by-pixel analyses discussed in Section 4.5.1 and Section 4.5.2. Rewriting Eq (3.18), here we have $\epsilon(r) = r \cdot G + b$ with G a constant matrix and b a constant vector. Equivalently, the prior independence assumption becomes $p(\theta_k, r) = p(\theta_k)\ p(r)$ and the r follows a standard multivariate normal distribution from Section 3.3.2.

The likelihood function of D_k given emissivity, r, and plasma parameter, θ_k, is

$$L(\theta_k, r \mid D_k) = \prod_{h=1}^{H} \mathcal{N}\left(I_{k\lambda_h} \mid \epsilon_{\lambda_h}(r; \mathrm{n}_k, \mathrm{T}_k)\mathrm{n}_k^2\mathrm{d}_k,\ \sigma_{k\lambda_1}^2\right). \tag{4.37}$$

Pragmatic Bayesian model and two-step MC sampler algorithm

For the pragmatic Bayesian method, as described in Section 3.2.4, we assume that the observed intensities, D_k, are uninformative as to the PCA generated emissivities. That is, we do not take into account the information in the intensities of a single pixel for narrowing the uncertainty in the choice of emissivity realizations. Mathematically, this assumption can be written $p(r \mid D_k) = p(r)$, i.e., r and D_k are independent. Thus, the pragmatic Bayesian

joint posterior distribution of r and θ_k in the separate pixel-by-pixel analyses is

$$p(r, \theta_k \mid D_k) = p(\theta_k \mid D_k, r)\ p(r \mid D_k) = p(\theta_k \mid D_k, r)\ p(r), \tag{4.38}$$

The pragmatic Bayesian method accounts for atomic uncertainty in a conservative manner. The assumption that $p(r \mid D_k) = p(r)$ ignores information in the intensities, D_k, that may reduce uncertainty of atomic data represented by r and hence of θ_k. We next consider methods that allow D_k to be informative for r.

To obtain a Monte Carlo (MC) sample of (r, θ_k) from the pragmatic joint Bayesian posterior distribution in Eq (4.38), at iteration ℓ, we first obtain a MC sample of $r^{(\ell)}$, from its prior distribution in Eq (3.19). Conditional on $r^{(\ell)}$, we can then sample $\theta_k^{(\ell)}$ from $p(\theta_k \mid D_k, r^{(\ell)})$ using the MH algorithm, exactly same as what is done in Section 4.5.2.

Fully Bayesian model and HMC algorithm The fully Bayesian joint posterior distribution for r and θ_k conditional on a single pixel ($\#k$) dataset, i.e., under the separate pixel-by-pixel analyses, is

$$p(\theta_k, r \mid D_k) = \frac{L(\theta_k, r \mid D_k)\ p(\theta_k)\ p(r)}{p(D_k)}, \tag{4.39}$$

where normalization constant $p(D_k) = \int \int L(\theta_k, r \mid D_k)\ p(\theta_k)\ p(r)\ \mathrm{d}\theta_k \mathrm{d}r$.

Evaluating $p(\theta_k, r \mid D_k)$ in this way, we can use the Stan software package (Carpenter et al. 2016) again on a higher dimension posterior to obtain the MC sample,$\{(\theta_k^{(1)}, r^{(1)}), \ldots, (\theta_k^{(L)}, r^{(L)})\}$, via HMC by sampling directly from their posterior distribution, Eq (4.39).

Application to simulated intensities

Data simulation We simulate the intensity replicates for each pixel k and for each $\lambda \in \Lambda$ from a normal distribution

$$I_{k\lambda} \mid \epsilon, \mathrm{n}_k, \mathrm{d}_k \overset{\text{indep}}{\sim} \mathcal{N}\left(\epsilon_\lambda(\mathrm{n}_k, \mathrm{T}_k)\mathrm{n}_k^2\mathrm{d}_k,\ \sigma_{k\lambda}^2\right), \tag{4.40}$$

with the most possible emissivity curve in the ensemble $\mathcal{M}$, $\epsilon^{(471)}$, and the parameters, $\theta_k = (\log \mathrm{n}_k, \log \mathrm{d}_k) = (9.4, 9.3)$ for each k, which is the posterior mean for Pix #1 obtained in Section 4.5.

Application and output summary The results of comparing the inferred density $\log \mathrm{n}_k$ and path length $\log \mathrm{d}_k$ from all the five methods, standard, the pragmatic and the fully Bayesian methods using discrete uniform prior or Gaussian prior via PCA, discussed so far on a single pixel (#1) are shown in Figure 4.11. The joint posterior probability density distribution $p(\theta_k \mid D_k)$ computed using the pragmatic and the fully Bayesian methods using discrete uniform prior or Gaussian prior via PCA are shown as contour plots, and marginalized 1-D posterior densities $p(\log \mathrm{n}_k \mid D_k)$ and $p(\log \mathrm{d}_k \mid D_k)$ are shown as curves along the corresponding axes. The estimates of $\log \mathrm{n}_k$ and $\log \mathrm{d}_k$ computed via the standard analysis, i.e., the MAP minimization or the χ^2 minimization, are marked with straight lines. As expected, the density and path length are highly correlated. Notice that, for those using discrete uniform prior, the pragmatic Bayesian method inflates the error bars relative to the standard method as it accounts for the atomic data uncertainties. The fully Bayesian method shrinks the error bars relative to the pragmatic Bayesian method and shifts the best estimates towards the true values since it selects a subset of the full range of atomic uncertainties that are consistent with the data. The standard method underestimates the uncertainties in all cases, and is shifted away from the true values relative to the fully Bayesian estimate. For those methods using Gaussian prior via PCA, we use 7 PCs when fitting the model accounting for 78.89% of the total variance. The fully Bayesian method using Gaussian prior via PCA inflates the error bars significantly relative to that using discrete uniform prior though still shrinking the error bars slightly and shifting the best estimates slightly towards the true values relative to the pragmatic Bayesian method since it generates a subset of the whole range of atomic uncertainties that are more consistent with the simulated data. However, there is not much difference between the pragmatic and the fully Bayesian method when PCA is embedded where the latter one is supposed to select those atomic uncertainties that are consistent with the data, while the former one does not. The

reason is the datasets do not provide enough information to the selection of atomic uncertainties. Next, we are going to consider multiple pixels to make sure there is enough information provided by the intensities.

Figure 4.12 compares the prior distribution of emissivities (i.e., the given ensemble of emissivity realisations) with the two posterior distributions of emissivities respectively, non-PCA original posterior with discrete uniform prior in Eq (4.19) discussed in Section 4.5 and PCA posterior with Gaussian prior in Eq (4.39) discussed in this section. The first panel of Figure 4.12 illustrates the incorporation of PCA compression on the emissivity realisations into the HMC algorithm, where a PCA MC sample of emissivity curve is generated using Eq (3.18) with $J = 7$ at each iteration and the reconstructed emissivity curves are able to capture the structure of the original $\mathcal{M}$ nicely. The full range, the middle 95%, and the middle 68.3% intervals of these PCA posterior samples of emissivity curves are superimposed on the corresponding intervals for the given ensemble of emissivity realisations. In the second panel, seven of the $\sqrt{\epsilon^{(m)}}$ from $\mathcal{M}$ haveing top posterior probabilities with discrete uniform prior in Eq (4.19) are compared with the full range, the middle 95%, and the middle 68.3% intervals of $\mathcal{M}$. In the third panel, the same seven of the $\sqrt{\epsilon^{(m)}}$ are compared with the full range, the middle 95%, and the middle 68.3% intervals of these PCA posterior samples of emissivity curves. Those highly likely emissivity curves tend to lie in high probability density area and are well-captured when PCA is incorporated.

4.6.2 Models for incorporating the PCA generated emissivities into multiple pixels

Assuming that there are K pixel datasets and their prior distributions for r and θ_k, $k = 1, \ldots, K$, are independent, the joint posterior distribution for r and $\theta = (\theta_1, \ldots, \theta_K)$ given K-pixel datasets, $D = \{D_1, \ldots, D_K\}$, under the simultaneous analysis is

$$p(r, \theta \mid D) \propto L(r, \theta \mid D)\ p(r)\ p(\theta) \propto \prod_{k=1}^{K} L(r, \theta_k \mid D_k)\ p(\theta_k) \cdot p(r), \qquad (4.41)$$

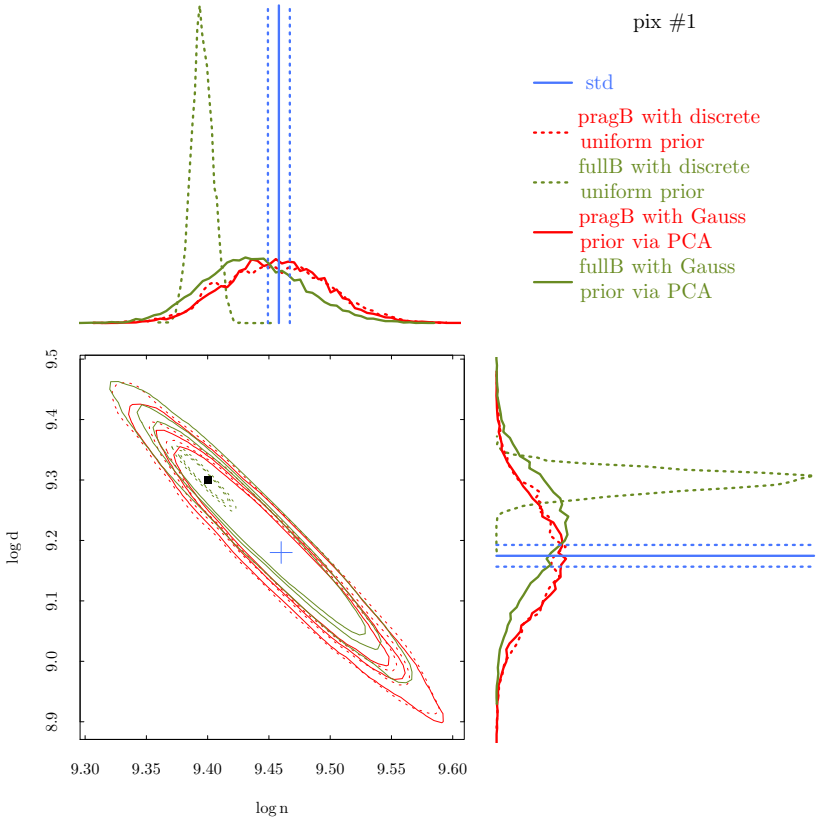

Figure 4.11: Comparisons of the inferred density $\log \mathrm{n}_k$ and path length $\log \mathrm{d}_k$ using different methods for pixel $\#1$. The results from different methods used are colour coded, with blue representing the standard method, dotted red the pragmatic Bayesian method using discrete uniform prior, dotted green the fully Bayesian method using discrete uniform prior, solid red the pragmatic Bayesian method using Gaussian prior via PCA, and solid green the fully Bayesian method using Gaussian prior via PCA. The contour plots (with levels at $0.01\times$, $0.05\times$, and $0.1\times$ the maximum) show where the majority of the mass of the joint probability distributions of $(\log \mathrm{n}_k, \log \mathrm{d}_k)$ fall, and their marginalized distributions along each axis are shown to the top and to the right of the corresponding axis. The results from the standard analysis is shown along with the histograms as straight lines (solid for the best-fit and dotted denoting the $\pm 1\sigma$ errors on the best-fit obtained from the default CHIANTI emissivity functions), extending into the contour plot region. The best-fit value from standard analysis is also marked on the contour plot with a '$+$' sign, with the arms of the symbol corresponding to the sizes of the error bars. The true values of $\log \mathrm{n}_k$ and $\log \mathrm{d}_k$ used to generate the simulated dataset is also marked on the contour plot with a square dot. As expected, the density and path length are highly correlated. The standard method underestimates the uncertainties, the pragmatic Bayesian method using discrete uniform prior inflates them due to atomic data uncertainties. The fully Bayesian method using discrete uniform prior strikes a balance between atomic data uncertainties and how well the data are fit, shrinking the error bars relative to pragmatic Bayesian using discrete uniform prior and shifting the estimates. The fully Bayesian method using Gaussian prior via PCA mitigates the gaps between the two modes and shifts the distribution.

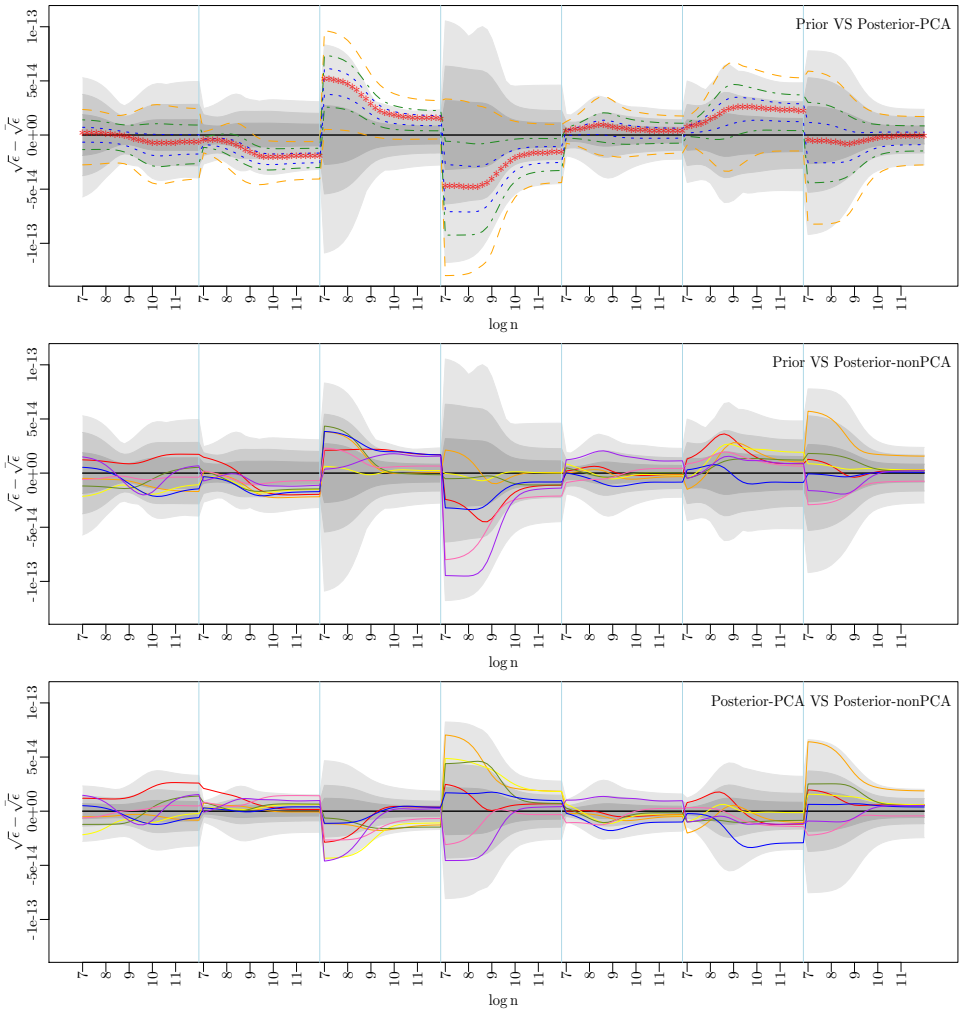

Figure 4.12: Comparison of the prior distribution and the posterior distribution for emissivities. The top panel illustrates the incorporation of PCA compression on the emissivity realisations into the HMC algorithm, where a PCA MC sample of emissivity curve is generated using Eq (3.18) with $J = 7$ at each iteration. The dashed, dot-dash, and dotted lines, respectively, superimpose the full range, the middle 95%, and the middle 68.3% intervals of these PCA posterior samples of emissivity curves on the corresponding intervals for the given the ensemble of emissivity realisations, plotted in light, dark, and darker grey areas. Posterior mean of the PCA generated emissivity curves is plotted as red star line and compared with the average over all given emissivity realisations, $\sqrt{\epsilon}$, that is plotted as a solid black line. Seven of the $\sqrt{\epsilon^{(m)}}$ from $\mathcal{M}$ with the top posterior probabilities in Eq (4.19) is plotted as coloured solid lines and compared with the full range, the middle 95%, and the middle 68.3% intervals of $\mathcal{M}$ (the middle panel), or compared with that of PCA generated posterior samples of emissivity curves (the bottom panel), plotted in light, dark, and darker grey areas respectively.

where we have omitted the denominator since it is a normalizing constant determined by the numerator. Under the model in Eq (4.41), inference for θ is based on its conditional posterior distribution

$$p(\theta \mid r, D) \propto L(r, \theta \mid D)\ p(\theta) \propto \prod_{k=1}^{K} L(r, \theta_k \mid D_k)\ p(\theta_k), \tag{4.42}$$

and inference for r is based on its conditional posterior distribution

$$p(r \mid \theta, D) \propto L(r, \theta \mid D)\ p(r) \propto \prod_{k=1}^{K} L(r, \theta_k \mid D_k) \cdot p(r). \tag{4.43}$$

Pragmatic Bayesian method For the pragmatic Bayesian method, similar to the separate pixel-by-pixel analyses, we assume that the observed intensities are uninformative as to the PCA generated emissivities. That is, we do not take into account the information in the intensities for narrowing the uncertainty in the choice of emissivity realizations. Mathematically, this assumption can be written $p(r \mid D) = p(r)$, i.e., r and D are independent. Thus, the pragmatic Bayesian joint posterior distribution of r and θ is

$$p(r, \theta \mid D) = p(\theta \mid D, r)\ p(r \mid D) = p(\theta \mid D, r)\ p(r), \tag{4.44}$$

The pragmatic Bayesian method accounts for atomic uncertainty in a conservative manner. The assumption that $p(r \mid D) = p(r)$ ignores information in the intensities, D, that may reduce uncertainty of atomic data represented by r and hence of θ. We next consider methods that allow D to be informative for r.

Fully Bayesian method In contrast to the pragmatic Bayesian method, the fully Bayesian method, as described by Xu et al. (2014), incorporates the potential information in the data (i.e., the intensities) to learn about r and the corresponding emissivity. The fully Bayesian joint posterior distribution of r and θ is given in Eq (4.41) and the conditional posterior distributions of θ and r is given in Eq (4.42) and Eq (4.43) respectively. The Bayesian posterior

distribution in Eq (4.43) allows the observed intensities to be informative for the atomic physics, following the principles of Bayesian analysis. It enables us to use the intensities to determine the more likely PCA generated emissivities.

4.6.3 Algorithms for incorporating the PCA generated emissivities into multiple pixels

Here instead of considering all the pixels together and dealing with super high-dimensional problem, we consider multiple pixels by starting from a relative low dimension. We apply the same algorithms, as what is done to a single pixel in Section 4.6.1, with multiple pixels to incorporate the PCA generated emissivity curve into the pragmatic and the fully Bayesian models.

Algorithms for the pragmatic Bayesian with multiple datasets

Under the pragmatic Bayesian model, we aim to construct MC sampler to account for emissivity uncertainties under the assumption that the observed intensities carry little information as to the PCA generated emissivity curve using the two-step MC sampler. At iteration ℓ, we iteratively update $r^{(\ell)}$ and $\theta^{(\ell)}$ by sampling them from Eq (3.19) and Eq (4.42). Sampling of $r^{(\ell)}$ is essentially random from the entire space. For each sampled $r^{(\ell)}$, updating $\theta^{(\ell)}$ is exactly same as what is done in the non-PCA model in Section 4.5.3. The Metropolis Hastings (MH) algorithm is used to sample $\theta^{(\ell)}$ from $p(\theta|r^{(\ell)}, D)$. For each sampled $r^{(\ell)}$, we use $t_4\left(\theta_k \mid \hat{\theta}_k, (-H(\hat{\theta}_k))^{-1}\right)$ proposal distribution, where $\hat{\theta}_k$ is the maximum a posteriori (MAP) estimates for $\log p(\theta_k \mid r^{(\ell)}, D_k)$ along with the 2×2 Hessian matrix evaluated at the mode $\hat{\theta}_k$, $H(\hat{\theta}_k)$. It is used for each pixel independently and separately to make the computation more efficient. With this proposal distribution, we run the MH for T iterations over the K-pixel intensities and obtain the MC sampler corresponding to $r^{(\ell)}$, $\theta^{(\ell)} = \theta^{[\ell,T]}$. The detailed two-step MC Sampler with MH under the pragmatic Bayesian model via simultaneous analysis proceeds for iteration $\ell = 1, \ldots, L$ with

Step 1: Sample $r^{(\ell)} \sim \mathcal{N}(0, I)$, set $\epsilon^{(\ell)} = \bar{\epsilon} + \sum_{j=1}^{J} r_j^{(\ell)}\ \beta_j\ v_j$.

Step 2: Proceed for inner interation $t = 1, \ldots, T-1$, and for each pixel $k = 1, \ldots, K$, sample $\theta_k^{[\text{prop}]} \sim t_4\left(\theta_k \mid \hat{\theta}_k, (-H(\hat{\theta}_k))^{-1}\right)$, compute

$$\rho = \frac{p(\theta_k^{[\text{prop}]} \mid D_k, r^{(\ell)})\ t_4\left(\theta_k^{[t]} \mid \hat{\theta}_k, (-H(\hat{\theta}_k))^{-1}\right)}{p(\theta_k^{[t]} \mid D_k, r^{(\ell)})\ t_4\left(\theta_k^{[\text{prop}]} \mid \hat{\theta}_k, (-H(\hat{\theta}_k))^{-1}\right)}, \tag{4.45}$$

set

$$\theta_k^{[\ell,t+1]} = \begin{cases} \theta_k^{[\text{prop}]}, & \text{with probability } \min(\rho, 1), \\ \theta_k^{[t]}, & \text{otherwise.} \end{cases} \tag{4.46}$$

Then set $\theta^{(\ell)} = \theta^{[\ell,T]}$.

Algorithms for the fully Bayesian with multiple datasets

Two-step Gibbs sampler Under the fully Bayesian model, we aim to construct MC sampler to account for emissivity uncertainties allowing the observed intensities to be informative for the PCA generated emissivity curve using the two-step Gibbs sampler. We iteratively update $r^{(\ell)}$ and $\theta^{(\ell)}$ by sampling them from Eq (4.43) and Eq (4.42). With this conditional distribution, sampling of $r^{(\ell)}$ does not depend on the multivariate integral, used in non-PCA case (i.e., the discrete prior case), in Section 4.5.2 anymore. The MH algorithm could be used to update $r^{(\ell)}$ as long as accepting reasonably and the symmetric property of the standard normal density makes the computation more efficient. For each sampled $r^{(\ell)}$, updating $\theta^{(\ell)}$ is exactly same as what is done under the pragmatic Bayesian model above.

The two-step MC Gibbs Sampler with MH under the fully Bayesian model via simultaneous analysis proceeds for iteration $\ell = 1, \ldots, L$ with

Step 1: Sample $r^{[\text{prop}]} \sim \mathcal{N}(0, I)$, set $\epsilon^{[\text{prop}]} = \bar{\epsilon} + \sum_{j=1}^{J} r_j^{[\text{prop}]}\ \beta_j\ v_j$, compute

$$\rho_r = \frac{p(r^{[\text{prop}]} \mid D, \theta^{(\ell-1)})}{p(r^{(\ell-1)} \mid D, \theta^{(\ell-1)})}, \tag{4.47}$$

and set

$$r^{(\ell)} = \begin{cases} r^{[\text{prop}]}, & \text{with probability } \min(\rho_r, 1), \\ r^{(\ell-1)}, & \text{otherwise.} \end{cases} \tag{4.48}$$

Step 2: Proceed for inner interation $t = 1, \ldots, T-1$, and for each pixel $k = 1, \ldots, K$, sample $\theta_k^{[\text{prop}]} \sim t_4\left(\theta_k \mid \hat{\theta}_k, (-H(\hat{\theta}_k))^{-1}\right)$, compute

$$\rho_\theta = \frac{p(\theta_k^{[\text{prop}]} \mid D_k, r^{(\ell)})\ t_4\left(\theta_k^{[t]} \mid \hat{\theta}_k, (-H(\hat{\theta}_k))^{-1}\right)}{p(\theta_k^{[t]} \mid D_k, r^{(\ell)})\ t_4\left(\theta_k^{[\text{prop}]} \mid \hat{\theta}_k, (-H(\hat{\theta}_k))^{-1}\right)}, \tag{4.49}$$

and set

$$\theta_k^{[\ell,t+1]} = \begin{cases} \theta_k^{[\text{prop}]}, & \text{with probability } \min(\rho_\theta, 1), \\ \theta_k^{[t]}, & \text{otherwise.} \end{cases} \tag{4.50}$$

Then set $\theta^{(\ell)} = \theta^{[\ell,T]}$.

Hamiltonian Monte Carlo Under the fully Bayesian model, we incorporate the PCA generated emissivity curve into the HMC algorithm discussed in Section 3.4.2. With the prior independence assumption $p(\theta, r) = \prod_{k=1}^{K} p(\theta)\ p(r)$ and the r follows a standard multivariate normal distribution from Section 3.3.2, the joint posterior distribution for r and θ conditional on K pixel datasets is given in Eq (4.41). Evaluating $p(\theta, r \mid D)$ in this way, we can use the Stan software package again on a higher dimension posterior to obtain $\{(\theta^{(1)}, r^{(1)}), \ldots, (\theta^{(L)}, r^{(L)})\}$ via HMC by sampling directly from their joint posterior distribution.

4.6.4 Application to simulation study on multiple pixels

To illustrate the advantage of the fully Bayesian method over the pragmatic Bayesian method in the multiple-pixel case, we compare their performance in a simulation study and give detailed results under the simulation setting.

Data simulation Suppose the 471st emissivity in $\mathcal{M}$ has the PCA transformation: $\epsilon^{(471)} = \bar{\epsilon} + \sum_{j=1}^{182} r_j^{(471)}\ \beta_j\ v_j$ via Eq (3.18). We construct a new emissivity by using only the first three PCs, $\epsilon^{(471)*} = \bar{\epsilon} + \sum_{j=1}^{J} r_j^{(471)}\ \beta_j\ v_j$ and $J = 3$. The first $J = 3$ PCs accounting for 42.23% of the total variance. We will also only use the first $J = 3$ PCs when fitting both the pragmatic and the fully Bayesian model in Section 4.6.2 so that we are simulating and fitting under the *same* model. We simulate $R = 200$ intensity replicates for each pixel k and for each $\lambda \in \Lambda$ from a normal distribution

$$I_{k\lambda} \mid \epsilon, \mathrm{n}_k, \mathrm{d}_k \overset{\text{indep}}{\sim} \mathcal{N}\left(\epsilon_\lambda(\mathrm{n}_k, \mathrm{T}_k)\mathrm{n}_k^2\mathrm{d}_k,\ \sigma_{k\lambda}^2\right), \tag{4.51}$$

with $\epsilon = \epsilon^{(471)*}$ the new generated emissivity and the parameters, $\theta_k = (\log \mathrm{n}_k, \log \mathrm{d}_k) = (9.4, 9.3)$ for $k = 1, \ldots, K$, which is the posterior mean for Pix #1 obtained in Section 4.5. Here we pick the first two pixels, #1 and #2, and let $K = 2$. Note that each of the pixels is simulated with the same values of the parameters, but different replicates have different simulated data.

Application and output summary We run the two-step MC sampler for the pragmatic Bayesian model, and both the two-step MC Gibbs sampler and the HMC algorithms for the fully Bayesian model, mentioned in Section 4.6.3.

Using the two-step MC sampler in the pragmatic Bayesian model, $T = 30$ MH samplers, determined by constructing autocorrelation plots in this setting (Xu et al. 2014), are drawn for the plasma parameters given each PCA generated emissivity, and the last MH sampler is taken as an MC sampler. There are 1000 MC samplers drawn in each simulation. Similarly, using the two-step Gibbs sampler in the fully Bayesian model, $T = 30$ MH samplers are drawn for the plasma parameters and for each PCA generated emissivity from theirs marginal posterior distributions respectively, the last MH sampler is taken as an MC sampler, and 1000 MC samplers are drawn in each simulation. For using HMC in the fully Bayesian model, there are 4 chains running, 4000 iterations each, and the first half of the iterations of each chain are discarded as burn-in.

Figure 4.13 shows the posterior means, the 68% and 95% intervals of the posterior samples for each parameter conditional on the two simulated pixels, #1 and #2, over the first 30 replicates under the pragmatic Bayesian model (row 2), and under the fully Bayesian model using two-step MC Gibbs sampler (row 3), and using the HMC (row 4). It also includes the MAP, the 68% and 95% intervals of the parameters using the standard method (row 1). All these results are also compared with the true values of those parameters, used to simulate the intensity replicates, marked as the vertical lines. Although the error bars from the standard method are the smallest in this simulation study, this method misses the true value of θ and the estimate is biased. This is not unusual when the emissivity is misspecified. As in this case, the data was generated under the emissivity curve $\epsilon^{(471)*}$, i.e., one that is nonetheless plausible given the atomic uncertainty, instead of the default emissivity curve $\epsilon^{(1)}$. The pragmatic Bayesian method accounts for atomic uncertainty by iteratively and randomly generating emissivity curves across the whole parameter space, resulting in bias and much larger error bars, while a proper coverage that captures the true value of θ compared to that of the standard method. The fully Bayesian method accounts for atomic uncertainty by iteratively generating and selecting those emissivity curves that are consistent with the simulated data. The resulting error bars are only slightly larger than those produced with a fixed emissivity curve and significantly smaller than that from the pragmatic Bayesian method. At the same time, the fitted values for θ have shifted enough that the error bars still capture the true value. This example clearly illustrates the benefits of the fully Bayesian method: the estimates of the parameters get less biased, have smaller error bars, and achieve a proper coverage of the true value of the parameter.

Table 4.4 shows the frequency results over all of the 200 replicates. The bias is dominant in the pragmatic Bayesian model, while the error becomes dominant in the fully Bayesian model. For the pragmatic Bayesian model, the posterior samples are biased, the 68% intervals are significantly under coverage, the 95% intervals are significantly over coverage. For the fully Bayesian model using both algorithms, the estimates of the parameter get

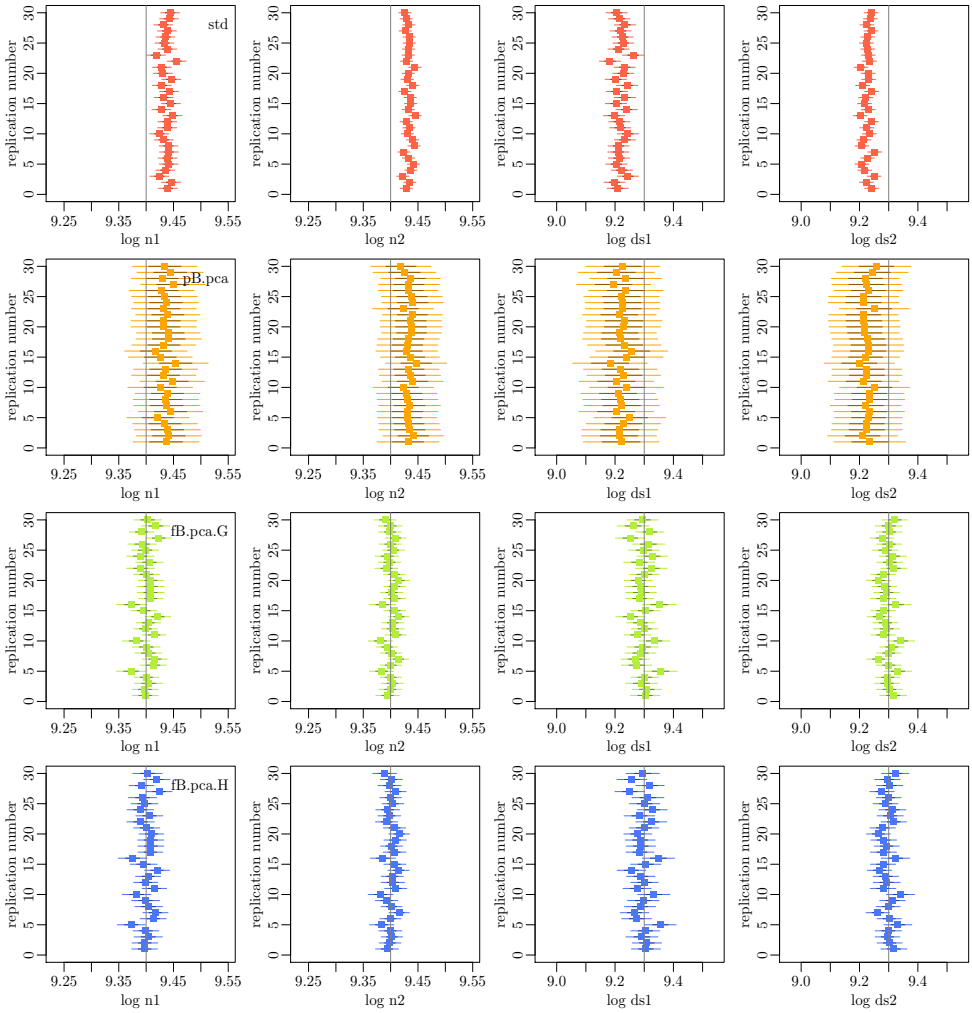

Figure 4.13: The posterior means, the 68% and 95% intervals of the posterior samples for each parameter in the simulation study with $J = 3$ principal components and $K = 2$ pixels over the first 30 replicates under the pragmatic Bayesian model (row 2), and under the fully Bayesian model using two-step MC Gibbs sampler (row 3), and using the HMC (row 4). The MAP, the 68% and 95% intervals of the parameters using the standard method (row 1).

less biased and the 68% intervals are a little bit over coverage. Moreover, the smaller bias and root mean square error (RMSE, the square root of the mean square error, thus, the average distance of the observed data from the fitted value) are obtained, the coverage is getting close to the true coverage, and the smaller coverage length.

The doParallel package is used to execute for each loops in parallel within each process and 32 processors are used. The computation times, measured by the sum of user and system times, of implementing two-step MC sampler in the pragmatic Bayesian model and implementing two-step MC Gibbs sampler and HMC in the fully Bayesian model are 8.0, 31.4, and 5.7 minutes respectively. Comparing the two algorithms for the fully Bayesian model, the latter is significantly faster.

For the two algorithms under the fully Bayesian model, using HMC is significantly faster but the tuning step in it is complex and time consuming. On the other hand, using two-step Gibbs sampler is itself time consuming, but the tuning step is very simple.

		Bias	RMSE	Coverage of 68% interval	Coverage of 95% interval	Average length of 95% interval
prag Bayes	$\log n_1$	0.0366	0.0376	0.225	1	0.1187
	$\log n_2$	0.0345	0.0350	0.155	1	0.1114
	$\log d_1$	-0.0780	0.0817	0.205	1	0.2556
	$\log d_2$	-0.0752	0.0763	0.115	1	0.2405
fully Bayes Gibbs	$\log n_1$	0.0023	0.0121	0.68	0.95	0.0483
	$\log n_2$	0.0019	0.0096	0.73	0.97	0.0405
	$\log d_1$	-0.0053	0.0260	0.67	0.94	0.1026
	$\log d_2$	-0.0045	0.0207	0.72	0.975	0.0869
fully Bayes HMC	$\log n_1$	0.0024	0.0120	0.675	0.95	0.0481
	$\log n_2$	0.0019	0.0096	0.72	0.975	0.0404
	$\log d_1$	-0.0054	0.0259	0.65	0.945	0.1023
	$\log d_2$	-0.0046	0.0207	0.72	0.965	0.0867

Table 4.4: Summary of the bias, RMSE, and the coverage levels of the posterior estimates for each parameter in the simulation study with $J = 3$ principal components and $K = 2$ pixels over all 200 replicates, under the pragmatic Bayesian model (row 1), and under the fully Bayesian model using two-step MC Gibbs sampler (row 2), and using the HMC (row 3).

4.6.5 Conclusions and discussion

We use PCA to summarize the complex structure of the atomic data, e.g., emissivity ensemble, via a lower dimensional Gaussian approximation. This continuous method can fill in the gaps between the discrete emissivities, work on the problem of sparse sampling of the atomic data space, and require less storage. We are able to embed the PCA model for atomic data uncertainty, a high dimensional quantity, into a Bayesian procedure that simultaneously fits the model parameters and accounts for atomic uncertainty in the primary stage of a two-stage analysis. We demonstrate the advantage of incorporating the PCA model into a fully Bayesian method. Compared to incorporating the PCA model into a pragmatic Bayesian method, it gives estimates with smaller bias and narrower error bars. We are going to apply this method to more complicated case studies in the following two chapters.

5

Two-stage analysis with Fe XVII

Another application of a two-stage analysis via a case study in Fe XVII is described in this chapter. Fe XVII is a popular line system for temperature measurements where its density sensitivity is pretty much ignorable. Therefore, the models and algorithms used for Fe XIII analysis in Chapter 4 are also applicable to Fe XVII working on the temperature information instead of the density information. Similarly, we focus on the primary stage by applying the ensemble of atomic data to analyze the temperature-sensitive Fe XVII spectral lines observed in a stellar active region with the *Chandra* X-ray Observatory using a Bayesian framework. As we have proved that the Gaussian approximation via PCA model for atomic data uncertainty is good at dealing with sparse sampling and mitigating gaps of sparse samples in Section 4.6, we use PCA to efficiently represent the ensemble of sparse Fe XVII emissivity realizations. Both the pragmatic Bayesian method where the observed data is not informative to the choice of the atomic uncertainties and the fully Bayesian method where the observed data is allowed to update the atomic uncertainties are implemented. The ultimate aim is to estimate the plasma parameters, e.g. temperature and volume, in the spectral analysis while incorporating atomic uncertainties. A summarization of those estimations will

also be used as a prior knowledge for a further stage in the next chapter. Therefore, the whole analysis on Fe XVII in this chapter will be treated as a preliminary analysis and the estimation information will then be carried forward into a further primary analysis in a three-stage analysis.

5.1 Data and the astrophysical model

Instead of observing the Fe XIII intensities from a solar active region with an EIS observation on *Hinode* Spacecraft in Chapter 4, the Fe XVII and O VII photon counts, to be discussed in this chapter and next chapter, are observed from Capella with the *Chandra* X-ray Observatory. The photon counts are observed directly from the observatory while the intensities are pre-processed data. Therefore, we introduce a Poisson likelihood function for the observed photon count data (Arnaud et al. 2011) compared to a Gaussian likelihood for the observed intensity data.

In this section we discuss the general expressions for notations and equations applicable to different ions, Fe XVII discussed in this chapter and O VII in Chapter 6. In Section 5.1.1 we introduce the spectral data and the general statistical distributions that photon counts are modelled as. In Section 5.1.2, details of a general spectral model and the corresponding physical parts are discussed. Upper subscripts ^{Fe} and ^{O} are used to represent Fe XVII and O VII respectively, from Section 5.2 onwards until the end of this thesis.

5.1.1 Spectral data and its distribution

Suppose we have observed the *source* counts, contaminated with the background in an exposure, and also have observed the pure *background* counts in another exposure from each channel of the detector. Each channel is associated with one of the H spectral lines corresponding to the set of *recorded* wavelengths $\mathcal{W} = \{\mathrm{w}_h, h = 1, \ldots, H\}$. Let $Y(\mathrm{w})$ and $Z(\mathrm{w})$ be the observed photon counts from the source and the background exposure in channel $\mathrm{w} \in \mathcal{W}$, and $\mathcal{D} = \{(Y(\mathrm{w}), Z(\mathrm{w})), \mathrm{w} \in \mathcal{W}\}$.

We also have a collection of M realizations of the *plasma emissivities*, denoted by $\mathcal{M} = \{\epsilon^{(m)}(\omega; \log \mathrm{n}, \log \mathrm{T}), \omega \in \Omega, m = 1, \ldots, M\}$, where $\Omega = \{\omega_l, l = 1, \cdots, L\}$ is the set of *true* wavelengths, i.e., lines of interest. Note that each of the true wavelength ω has its own range of recorded wavelength w.

Let $\theta = (\theta_S, \theta_B)$ be the plasma parameters, where θ_S and θ_B are sets of parameters from the source and the background exposures respectively.

To fix ideas, we specify a general spectral analysis problem applicable to many different ions, e.g., Fe XVII and O VII, in this subsection. According to Arnaud et al. (2011), the observed source and background photon counts, Y and Z in channel $\mathrm{w} \in \mathcal{W}$, given θ_S and (or) θ_B, are modelled as independent Poisson distributions, respectively,

$$Y(\mathrm{w}) \mid \theta_S,\ \theta_B \overset{\text{indep}}{\sim} \text{Poisson}\big(s(\mathrm{w}; \theta_S) + \kappa(\mathrm{w}) + b(\mathrm{w})\big), \tag{5.1}$$

$$Z(\mathrm{w}) \mid \theta_B \overset{\text{indep}}{\sim} \text{Poisson}\big(\eta \cdot b(\mathrm{w})\big), \tag{5.2}$$

where Poisson(λ) is a Poisson distribution with parameter λ, $s(\mathrm{w}; \theta_S)$ is the *expected source counts* in channel w, $\kappa(\mathrm{w})$ is the *expected source continuum counts* in channel w, $b(\mathrm{w})$ is the *expected background counts* in channel w, $\theta_B = \{b(\mathrm{w}), \mathrm{w} \in \mathcal{W}\}$, and η is the area to exposure time ratio of source and background (given as a constant). To simplify the problem and to prevent overfitting, we have assumed 1) the shape of the expected continuum counts is fixed, i.e., $\kappa(\mathrm{w})$ is known in advance for each channel w; 2) the expected background count is the same for every channel, i.e., $b(\mathrm{w}) = \theta_B$ for any w. The expected photon count is physically equal to the intensity, which is equivalent to the sum of the source counts, the background counts, and the continuum counts. Different from the simplified model discussed in Chapter 4 and Yu et al. (2018), where we have ignored the background counts, the continuum counts, and the temperature dependency of the emissivities, we are now considering a more general spectral model that contains almost all the physical parts. Detailed formulas for the physical parts are discussed in Section 5.1.2.

In astrophysics, we are always counting photons and there is always a Poisson model. In Chapter 4 and Yu et al. (2018), for simplicity, we have approxi-

mated the Poisson model with a Gaussian approximation because a Poisson tends to be a Gaussian as its mean becomes sufficiently large, e.g., when there are a large amount of photon counts observed, via the central limit theorem (Le Cam 1986). In this Chapter, we address the above general spectral model with both an observed and a simulated sets of photon counts, a more general dataset than that of Chapter 4, and consider the pragmatic and the fully Bayesian models under the same kind of simplified assumptions.

5.1.2 The astrophysical spectral model

Suppose the expected source counts s in channel w from a model with a set of parameters θ_S in a given observation I that has grating $g \in \{\mathrm{HEG}, \mathrm{MEG}, \mathrm{LEG}\}$, order $o = \pm 1$, and over all the true wavelengths ω (Chandra 2019, Chen et al. 2019),

$$s_{Igo}(\mathrm{w}; \theta_S) = \sum_{l=1}^{L} R_{go}(\mathrm{w}|\omega_l) \cdot (A_{Igo}(\omega_l) \cdot f(\omega_l; \theta_S)) \cdot \mathcal{T}_I \tag{5.3}$$

where $R_{go}(\mathrm{w}|\omega)$ is the line response function, $A_{Igo}(\omega)$ is the effective area at wavelength ω, $f(\omega; \theta_S)$ is the photon flux at telescope at wavelength ω, and $\mathcal{T}_I$ is observation duration.

If the photon counts for each line of interest are observed directly, it can be learned about straightforwardly. However, as we only observe photon counts over channels, there are two corrections, line response function and effective area, need to be applied to the instruments and the gratings on the telescope, as discussed in Section 1.3. The line response function used in this chapter and Chapter 6 is expressed in Eq (1.1). The effective area is given as a function of wavelength which interpolates linearly on a given set of wavelength grid.

Because of the temperature dependence, we have to consider DEMs as discussed in Section 1.1. The DEM is usually expressed as

$$g(\mathrm{n}, \mathrm{T}(\vec{\boldsymbol{\alpha}})) = \mathrm{n}^2(\mathrm{T}) \, \frac{dV(\mathrm{T})}{d \log \mathrm{T}} \,, \tag{5.4}$$

where $\vec{\boldsymbol{\alpha}}$ includes some shape factors that determine the shape of the DEM, V is the temperature volume, and $\frac{dV(\mathrm{T})}{d\log \mathrm{T}}$ is the temperature volume differentiation performed over a reasonable temperature interval. Note that the volume emission measure is used instead of the path length emission measure in the previous chapter. Specifically, we consider the delta volume in this article,

$$\frac{dV(\mathrm{T})}{d\log \mathrm{T}} \equiv \alpha_0 \ \cdot \ \delta(\log \mathrm{T} - \alpha_1). \tag{5.5}$$

The flux at source is the product of the emissivity ϵ and the DEM, and it can be written from Eq (5.4) and Eq (5.5) as, at true wavelength ω,

$$\begin{aligned} f_s(\omega;\theta_S) &= \int \epsilon(\omega;\mathrm{n},\mathrm{T},r)\ g(\mathrm{n},\mathrm{T}(\vec{\boldsymbol{\alpha}}))\ d\log \mathrm{T} \\ &= \int \epsilon(\omega;\mathrm{n},\mathrm{T},r)\ \mathrm{n}^2\ \frac{dV(\mathrm{T})}{d\log \mathrm{T}}\ d\log \mathrm{T} \\ &= \epsilon(\omega;\log \mathrm{n},\alpha_1,r)\ \cdot\ \mathrm{n}^2\ \cdot\ \alpha_0\,, \end{aligned} \tag{5.6}$$

where $\theta_S = (\log \mathrm{n}, \log \alpha_0 = \log V, \alpha_1 = \log \mathrm{T}, r)$, and the log scale is used to stretch out the values and exaggerate the difference.

Therefore, convolving the flux with the effective area and the line response function and considering all other factors for uncertainties (Dere et al. 1997, Kashyap & Drake 1998) including ion abundance (Z), ionization fraction (i), the ratio of n_H and n_e ($\frac{n_H}{n_e}$), distance factor (d), and energy-to-photon conversion factor ($\frac{hc}{\omega}$), the expected source counts in channel w, in a given observation I that has grating g, order o, and over all the true wavelengths ω, can be expanded from Eq (5.3) and Eq (5.6) into

$$\begin{aligned} s_{Igo}(\mathrm{w};\theta_S) = \sum_{l=1}^{L}\ & R_{go}(\mathrm{w}|\omega_l)\cdot\big(A_{Igo}(\omega_l)\mathcal{T}_I\big)\cdot \\ & \frac{\epsilon(\omega_l;\log \mathrm{n},\alpha_1,r)}{\mathrm{n}}\cdot\frac{\omega_l}{hc}\cdot Z^{Fe}\cdot i^{Fe}(\alpha_1)\cdot\frac{n_H}{n_e}(\alpha_1)\cdot\frac{\alpha_0\cdot\mathrm{n}^2}{4\pi d^2}. \end{aligned} \tag{5.7}$$

Specifically, we substitute ^{O} for ^{Fe} to Z and i in the O VII analysis. The emissivities, ϵs, for the Fe XVII lines are independent of the density $\log \mathrm{n}$,

while those for the O VII lines are dependent on both the density $\log \mathrm{n}$ and the temperature $\log \mathrm{T} = \alpha_1$.

Incorporating the observation duration into the effective area and summing over all observations for which we have the same set of grating and order $\{g, o\}$, we have

$$A_{go}(\omega) = \sum_{I|go} A_{Igo}(\omega)\mathcal{T}_I \tag{5.8}$$

and

$$s_{go}(\mathrm{w}; \theta_S) = \sum_{I|go} s_{Igo}(\mathrm{w}; \theta_S). \tag{5.9}$$

Therefore, the expected source counts in channel w, grating g, order o, and over all the true wavelengths ω is

$$\begin{aligned} s_{go}(\mathrm{w}; \theta_S) = \sum_{l=1}^{L} & R_{go}(\mathrm{w}|\omega_l) \cdot A_{go}(\omega_l) \cdot \\ & \frac{\epsilon(\omega_l; \log \mathrm{n}, \alpha_1, r)}{\mathrm{n}} \cdot \frac{\omega_l}{hc} \cdot Z^{Fe} \cdot i^{Fe}(\alpha_1) \cdot \frac{n_H}{n_e}(\alpha_1) \cdot \frac{\alpha_0 \cdot \mathrm{n}^2}{4\pi d^2} \end{aligned} \tag{5.10}$$

$$\begin{aligned} = \sum_{l=1}^{L} \Bigg(& R_{go}(\mathrm{w}|\omega_l) \cdot A_{go}(\omega_l) \cdot \epsilon(\omega_l; \log \mathrm{n}, \alpha_1, r) \cdot i^{Fe}(\alpha_1) \cdot \\ & \frac{n_H}{n_e}(\alpha_1) \cdot \frac{\omega_l}{hc} \Bigg) \cdot Z^{Fe} \cdot \frac{1}{\mathrm{n}} \cdot \mathrm{n}^2 \cdot \alpha_0 \cdot \frac{1}{4\pi d^2}. \end{aligned} \tag{5.11}$$

Moreover, the expected continuum counts κ in channel w, grating g, and order o, is expressed as

$$\begin{aligned} \kappa_{go}(\mathrm{w}) &= \sum_{k=1}^{K} \delta(\mathrm{w} - \omega_k) \cdot \big(\sum_{I|go} A_{Igo}(\omega_k)\mathcal{T}_I \big) \cdot c(\omega_k, \alpha_1) \cdot \frac{1}{10^{23}} \cdot \mathrm{n}^2 \cdot \alpha_0 \cdot \frac{1}{4\pi d^2} \\ &= A_{go}(\mathrm{w}) \cdot c(\mathrm{w}, \alpha_1) \cdot \frac{1}{10^{23}} \cdot \mathrm{n}^2 \cdot \alpha_0 \cdot \frac{1}{4\pi d^2} \end{aligned} \tag{5.12}$$

where

$$\delta(\mathrm{w} - \omega_k) = \begin{cases} 1, & \text{if } \mathrm{w} = \omega_k \\ 0, & \text{otherwise,} \end{cases}$$

c is the given set of continuum counts, $\{\omega_k\}$ is a superset and spans the same range as $\{\mathrm{w}_k\}$, and $4\pi d^2 = 2.07e + 40$. There is just an extra factor of $1e23$ attached to the continuum counts for numerical purposes, which accounts for the difference from the distance modulus. Adding the continuum to the model could lead to a better model where the expected counts would match the observed counts better and generate more accurate estimations of the model parameters.

Thus, the *expected source and continuum counts* in channel w, grating g, and order o, is

$$\begin{aligned}\lambda_{SC}(\mathrm{w}, \theta_S) &= s_g(\mathrm{w}, \theta_S) + \kappa_g(\mathrm{w}) \\ &= \Bigg\{ \sum_{l=1}^{L} \Big(R_{go}(\mathrm{w}|\omega_l) \cdot A_{go}(\omega_l) \cdot \epsilon(\omega_l; \log \mathrm{n}, \alpha_1, r) \cdot \frac{Z^{Fe}}{\mathrm{n}} \cdot i^{Fe}(\alpha_1) \cdot \\ &\qquad \frac{n_H}{n_e}(\alpha_1) \cdot \frac{\omega_l}{hc} \Big) \cdot A_{go}(\mathrm{w}) \cdot c(\mathrm{w}, \alpha_1) \cdot \frac{1}{10^{23}} \Bigg\} \cdot \mathrm{n}^2 \cdot \alpha_0 \cdot \frac{1}{4\pi d^2}.\end{aligned} \tag{5.13}$$

Specifically, $\log \mathrm{n}$ is fixed as a known value in the Fe XVII spectral analysis, since the Fe XVII lines are not sensitive to it, with corresponding $\theta_S = (\log \alpha_0 = \log V, \alpha_1 = \log \mathrm{T}, r)$ in this chapter, while it is treated as an unknown parameter in the O VII spectral analysis with corresponding $\theta_S = (\log \mathrm{n}, \log \alpha_0 = \log V, \alpha_1 = \log \mathrm{T}, r)$ in Chapter 6.

5.2 Fe XVII Statistical model

Here we specify the detailed Bayesian model for the Fe XVII spectral analysis in this primary stage.

We model the source and the background counts given the source and (or) the background parameters as independent Poisson distributions respectively, for

each $\mathrm{w} \in \mathcal{W}^{Fe}$,

$$Y^{Fe}(\mathrm{w}) \mid r^{Fe}, \log\alpha_0, \alpha_1, \theta_B^{Fe} \overset{\text{indep}}{\sim} \text{Poisson}\left(s^{Fe}(\mathrm{w}; r^{Fe}, \log\alpha_0, \alpha_1) + \kappa^{Fe}(\mathrm{w}) + \theta_B^{Fe}\right), \tag{5.14}$$

$$Z^{Fe}(\mathrm{w}) \mid \theta_B^{Fe} \overset{\text{indep}}{\sim} \text{Poisson}\left(\eta \cdot \theta_B^{Fe}\right). \tag{5.15}$$

Thus, the likelihood function of Y^{Fe} and Z^{Fe} given r^{Fe}, $\log\alpha_0$, α_1, and θ_B^{Fe} is

$$\begin{aligned}
&L(r^{Fe}, \log\alpha_0, \alpha_1, \theta_B^{Fe} \mid Y^{Fe}, Z^{Fe}) \\
&= p(Y^{Fe}, Z^{Fe} \mid r^{Fe}, \log\alpha_0, \alpha_1, \theta_B^{Fe}) \\
&= p(Y^{Fe} \mid r^{Fe}, \log\alpha_0, \alpha_1, \theta_B^{Fe}) \cdot p(Z^{Fe} \mid \theta_B^{Fe}) \\
&= \prod_{\mathrm{w} \in \mathcal{W}^{Fe}} p(Y^{Fe}(\mathrm{w}) \mid r^{Fe}, \log\alpha_0, \alpha_1, \theta_B^{Fe}) \cdot p(Z^{Fe}(\mathrm{w}) \mid \theta_B^{Fe})
\end{aligned} \tag{5.16}$$

Next, we specify the joint prior distribution on the unknown model parameters. We put continuous uniform distributions for $\log_{10}\alpha_0$ and α_1 respectively, a multivariate standard normal distribution for r^{Fe}, and a Gamma distribution for θ_B^{Fe},

$$p(\log_{10}\alpha_0) = \frac{1}{4} \quad \text{for } 30 \le \log_{10}\alpha_0 \le 34, \tag{5.17}$$

$$p(\alpha_1) = \frac{1}{2.1} \quad \text{for } 5.8 \le \alpha_1 \le 7.9, \tag{5.18}$$

$$r^{Fe} \sim \text{MVN}(0, I), \tag{5.19}$$

$$\theta_B^{Fe} \sim \text{Gamma}(\text{shape} = a_1, \text{rate} = a_2), \tag{5.20}$$

where $a_1 = 0.5$ and $a_2 = 2$. A Gamma distribution for θ_B^{Fe} is a conjugate prior for a Poisson likelihood. It covers all conceivable background ranges encountered in the dataset.

Those parameters are a priori independent so that the joint prior distribution is

$$p(r^{Fe}, \log\alpha_0, \alpha_1, \theta_B^{Fe}) = p(r^{Fe})\ p(\log\alpha_0)\ p(\alpha_1)\ p(\theta_B^{Fe}). \tag{5.21}$$

Given the likelihood function and the prior distribution defined in Eq (5.16) and Eq (5.21), the complete joint posterior distribution of r^{Fe}, $\log \alpha_0$, α_1, and θ_B^{Fe} given Y^{Fe} and Z^{Fe} in the Fe XVII spectral analysis is

$$\begin{aligned} & p(r^{Fe}, \log \alpha_0, \alpha_1, \theta_B^{Fe} \mid Y^{Fe}, Z^{Fe}) \\ \propto\ & p(Y^{Fe} \mid r^{Fe}, \log \alpha_0, \alpha_1, \theta_B^{Fe}) \cdot p(Z^{Fe} \mid \theta_B^{Fe}) \cdot \\ & p(r^{Fe}) p(\log \alpha_0) p(\alpha_1)\ p(\theta_B^{Fe}). \end{aligned} \tag{5.22}$$

The marginal posterior distribution $p(\log \alpha_0, \alpha_1 \mid Y^{Fe}, Z^{Fe})$ can be obtained by integrating out r^{Fe} and θ_B^{Fe},

$$p(\log \alpha_0, \alpha_1 \mid Y^{Fe}, Z^{Fe}) = \int \int p(r^{Fe}, \log \alpha_0, \alpha_1, \theta_B^{Fe} \mid Y^{Fe}, Z^{Fe}) dr^{Fe} d\theta_B^{Fe}. \tag{5.23}$$

Implicitly, the observed source counts, $Y^{Fe} = \{Y^{Fe}(\mathrm{w})\}$, are made up of the counts from the source exposure due to the source only, $Y_S^{Fe} = \{Y_S^{Fe}(\mathrm{w})\}$, and the counts due to the background only, $Y_B^{Fe} = \{Y_B^{Fe}(\mathrm{w})\}$, i.e., $Y^{Fe}(\mathrm{w}) = Y_S^{Fe}(\mathrm{w}) + Y_B^{Fe}(\mathrm{w})$, for any $\mathrm{w} \in \mathcal{W}^{Fe}$. Though it is impossible to observe Y_S^{Fe} and Y_B^{Fe} explicitly, we treat them as missing data. The method of data augmentation is applied to simplify our statistical analysis and to make the Bayesian inference computationally efficient (van Dyk 2003). We assume $Y_S^{Fe}(\mathrm{w})$ and $Y_B^{Fe}(\mathrm{w})$ follow independent Poisson distributions with intensity $s^{Fe}(\mathrm{w}; r^{Fe}, \log \alpha_0, \alpha_1) + \kappa^{Fe}(\mathrm{w})$ and θ_B^{Fe} for each $\mathrm{w} \in \mathcal{W}^{Fe}$,

$$\begin{aligned} Y_S^{Fe}(\mathrm{w}) \mid r^{Fe}, \log \alpha_0, \alpha_1 &\overset{\text{indep}}{\sim} \text{Poisson}\left(s^{Fe}(\mathrm{w}; r^{Fe}, \log \alpha_0, \alpha_1) + \kappa^{Fe}(\mathrm{w})\right), \\ Y_B^{Fe}(\mathrm{w}) \mid \theta_B^{Fe} &\overset{\text{indep}}{\sim} \text{Poisson}\left(\theta_B^{Fe}\right). \end{aligned}$$

It is easy to estimate the missing data and the model parameters under the above reformulated models. In particular, if the model parameters are known, the conditional distribution of $Y_B^{Fe}(\mathrm{w})$ given $Y^{Fe}(\mathrm{w})$ for any $\mathrm{w} \in \mathcal{W}^{Fe}$ can

be computed using the Bayes' Theorem,

$$
\begin{aligned}
& p(Y_B^{Fe}(\mathrm{w}) \mid Y^{Fe}(\mathrm{w}), r^{Fe}, \log \alpha_0, \alpha_1, \theta_B^{Fe}) \\
= & \frac{p(Y^{Fe}(\mathrm{w}) \mid Y_B^{Fe}(\mathrm{w}), r^{Fe}, \log \alpha_0, \alpha_1, \theta_B^{Fe}) \cdot p(Y_B^{Fe}(\mathrm{w}) \mid \theta_B^{Fe})}{p(Y^{Fe}(\mathrm{w}) \mid r^{Fe}, \log \alpha_0, \alpha_1, \theta_B^{Fe})} \\
= & \binom{Y^{Fe}(\mathrm{w})}{Y_B^{Fe}(\mathrm{w})} \cdot \Big(\frac{\theta_B^{Fe}}{s^{Fe}(\mathrm{w}; r^{Fe}, \log \alpha_0, \alpha_1) + \kappa^{Fe}(\mathrm{w}) + \theta_B^{Fe}} \Big)^{Y_B^{Fe}(\mathrm{w})} \times \\
& \Big(\frac{s^{Fe}(\mathrm{w}; r^{Fe}, \log \alpha_0, \alpha_1) + \kappa^{Fe}(\mathrm{w})}{s^{Fe}(\mathrm{w}; r^{Fe}, \log \alpha_0, \alpha_1) + \kappa^{Fe}(\mathrm{w}) + \theta_B^{Fe}} \Big)^{Y^{Fe}(\mathrm{w}) - Y_B^{Fe}(\mathrm{w})}.
\end{aligned}
$$

That is in the form of the probability mass function of a binomial distribution,

$$
\begin{aligned}
& p(Y_B^{Fe}(\mathrm{w}) \mid Y^{Fe}(\mathrm{w}), r^{Fe}, \log \alpha_0, \alpha_1, \theta_B^{Fe}) \\
\overset{\text{indep}}{\sim} & \text{Binomial} \big(Y^{Fe}(\mathrm{w}), \frac{\theta_B^{Fe}}{s^{Fe}(\mathrm{w}; r^{Fe}, \log \alpha_0, \alpha_1) + \kappa^{Fe}(\mathrm{w}) + \theta_B^{Fe}} \big),
\end{aligned} \tag{5.24}
$$

where $X \sim \text{Binomial}(n, p)$ indicates that X follows a binomial distribution with n independent experiments each with probability p. Therefore, given $Y^{Fe}(\mathrm{w})$, $Y_B^{Fe}(\mathrm{w})$ follows a binomial distribution for any $\mathrm{w} \in \mathcal{W}^{Fe}$.

Moreover, if θ_B^{Fe} follows a Gamma prior as in Eq (5.20), the marginal posterior of θ_B^{Fe} given Y_B^{Fe} and Z^{Fe} can also be computed using the Bayes' Theorem,

$$
\begin{aligned}
p(\theta_B^{Fe} \mid Z^{Fe}, Y_B^{Fe}) & \propto \prod_{h=1}^{H} \big(p(Z^{Fe}(\mathrm{w}_h) \mid \theta_B^{Fe}) \cdot p(Y_B^{Fe}(\mathrm{w}_h) \mid \theta_B^{Fe}) \big) \; p(\theta_B^{Fe}) \\
& \propto e^{-\theta_B^{Fe} \left(H(\eta+1) + a_2 \right)} \; \theta_B^{Fe \left(\sum_{h=1}^{H} Z^{Fe}(\mathrm{w}) + \sum_{h=1}^{H} Y_B^{Fe}(\mathrm{w}) + a_1 - 1 \right)},
\end{aligned}
$$

That is in the form of the probability density function of a Gamma distribution,

$$
\theta_B^{Fe} \mid Z^{Fe}, Y_B^{Fe} \sim \text{Gamma}(\sum_{h=1}^{H} Z^{Fe}(\mathrm{w}) + \sum_{h=1}^{H} Y_B^{Fe}(\mathrm{w}) + a_1, H(\eta+1) + a_2). \tag{5.25}
$$

Therefore, given Y_B^{Fe} and Z^{Fe}, θ_B^{Fe} follows a Gamma distribution.

Pragmatic Bayesian method

For the pragmatic Bayesian method in Section 3.2, we assume that the observed counts are uninformative as to the most likely emissivities. That is, we do not take into account the information in the counts for narrowing the uncertainty in the choice of emissivity realizations. Mathematically, this assumption can be written as,

$$p(r^{Fe} \mid Y^{Fe}, Z^{Fe}) = p(r^{Fe}) \tag{5.26}$$

i.e., r^{Fe} and $\mathcal{D}^{Fe} = \{Y^{Fe}, Z^{Fe}\}$ are independent. Thus, the pragmatic Bayesian joint posterior distribution of r^{Fe} and all other parameters Y_B^{Fe}, θ_B^{Fe}, $\log \alpha_0, \alpha_1$ given Y^{Fe} and Z^{Fe} can be written as

$$\begin{aligned}
&p(Y_B^{Fe}, \theta_B^{Fe}, \log \alpha_0, \alpha_1, r^{Fe} \mid Y^{Fe}, Z^{Fe}) \\
&= p(Y_B^{Fe}, \theta_B^{Fe}, \log \alpha_0, \alpha_1 \mid r^{Fe}, Y^{Fe}, Z^{Fe})\ p(r^{Fe} \mid Y^{Fe}, Z^{Fe}) && (5.27)\\
&= p(Y_B^{Fe}, \theta_B^{Fe}, \log \alpha_0, \alpha_1 \mid r^{Fe}, Y^{Fe}, Z^{Fe})\ p(r^{Fe}). && (5.28)
\end{aligned}$$

Inference for Y_B^{Fe} is based on its conditional posterior distribution,

$$\begin{aligned}
&p(Y_B^{Fe} \mid Y^{Fe}, Z^{Fe}, \theta_B^{Fe}, \log \alpha_0, \alpha_1, r^{Fe}) \\
&= p(Y^{Fe} \mid Y_B^{Fe}, \theta_B^{Fe}, \log \alpha_0, \alpha_1, r^{Fe})\ p(Y_B^{Fe} \mid \theta_B^{Fe}) \\
&= p(Y_B^{Fe} \mid Y^{Fe}, \theta_B^{Fe}, \log \alpha_0, \alpha_1, r^{Fe}), && (5.29)
\end{aligned}$$

following a binomial distribution as in Eq (5.24). Based on the assumption in Eq (5.26), inference for $\theta_B^{Fe}, \log \alpha_0, \alpha_1$, and r^{Fe} is based on their conditional posterior or prior distributions,

$$\begin{aligned}
&p(\theta_B^{Fe}, \log \alpha_0, \alpha_1, r^{Fe} \mid Y^{Fe}, Z^{Fe}, Y_B^{Fe}) \\
&\propto p(\theta_B^{Fe} \mid Z^{Fe}, Y_B^{Fe})\ p(\log \alpha_0, \alpha_1, r^{Fe} \mid Y_S^{Fe}) && (5.30)\\
&\propto p(\theta_B^{Fe} \mid Z^{Fe}, Y_B^{Fe})\ p(\log \alpha_0, \alpha_1 \mid Y_S^{Fe})\ p(r^{Fe}), && (5.31)
\end{aligned}$$

where θ_B^{Fe} given Z^{Fe} and Y_B^{Fe} follows a Gamma distribution as in Eq (5.25) and $Y_S^{Fe} = Y^{Fe} - Y_B^{Fe}$.

The pragmatic Bayesian method accounts for atomic uncertainty in a conservative manner. The assumption in Eq (5.26) ignores information in the counts, Y^{Fe} and Z^{Fe}, that may reduce uncertainty of atomic data represented by r^{Fe} and hence of all other parameters. We now consider methods that allow Y^{Fe} and Z^{Fe} to be informative for r^{Fe}.

Fully Bayesian method

The fully Bayesian method incorporates the potential information in the counts to learn about emissivities. The fully Bayesian joint posterior distribution of $Y_B^{Fe}, \theta_B^{Fe}, \log\alpha_0, \alpha_1, r^{Fe}$ given Y^{Fe} and Z^{Fe} is shown in Eq (5.27). Inference for Y_B^{Fe} is based on its conditional posterior distribution in Eq (5.29), following a binomial distribution as in Eq (5.24). Inference for $\theta_B^{Fe}, \log\alpha_0, \alpha_1$, and r^{Fe} is based on their conditional posterior distributions in Eq (5.30). Alternatively, the fully Bayesian joint posterior distribution of $\theta_B^{Fe}, \log\alpha_0, \alpha_1, r^{Fe}$ given Y^{Fe} and Z^{Fe} is shown in Eq (5.22), which is equal to integrating Y_B^{Fe} out from Eq (5.27).

The fully Bayesian posterior distribution allows the observed counts to be informative for the atomic physics, following the principles of Bayesian analysis. It enables us to use the counts to determine the more likely PCA generated emissivities.

MC samples of those model parameters can be obtained from their complete joint posterior distribution, Eq (5.22), via HMC algorithm directly, see Section 5.4.2. Alternatively, an iterative strategy can be implemented on the conditional distributions in Eq (5.29) and Eq (5.30) that, iteratively and separately, updates the missing background data given the model parameters and the model parameters given the sampled missing background data, see Section 5.4.2. The MC posterior sample of $\log\alpha_0$ and α_1 can be summarized with an analytical distribution to approximate the above marginal posterior distribution $p(\log\alpha_0, \alpha_1 \mid Y^{Fe}, Z^{Fe})$, see more in Section 5.7. The approximated distribution obtained in this preliminary analysis on Fe XVII spectrum will be carried forward into the primary analysis on O VII spectrum as the joint prior distribution of $(\log\alpha_0, \alpha_1)$, see more details in Chapter 6.

5.3 Gaussian prior on Fe XVII emissivities with PCA

A Gaussian prior distribution via PCA, as discussed in Section 3.3.2, is used to compress the given ensemble of emissivities, $\mathcal{M}^{Fe}$, for the fourteen Fe XVII lines.

Similar to what is done on Fe XIII in Section 4.3.2, we can easily reconstruct emissivity replicates based on the first $J = 14$ principal components (out of $Q = 308$) via Eq (3.18) and Eq (3.19) capturing 97.65% ($> 95\%$) of the total variance, as computed with Eq (3.17). A large amount of compression has been achieved because very few components are needed to compute the emissivity curve to high precision. This approximation also achieves a better reconstruction in the square root scale of the given emissivity ensemble, which is used throughout this chapter.

The complicated structure of the Fe XVII emissivities, indicating the atomic uncertainty, is illustrated in the top two panels of Figure 5.1 using the given ensemble of 1000 Fe XVII emissivity realisations. A random selection of six of the $\sqrt{\epsilon^{(m)}}$ from $\mathcal{M}^{Fe}$ is compared with the average over all those emissivity realisations, $\bar{\sqrt{\epsilon}}$, the full range, the middle 95%, and the middle 68.3% of $\mathcal{M}^{Fe}$ in square root space, indicating the ensemble of emissivity curves in $\mathcal{M}^{Fe}$ form a complex tangle that appears to defy any systematic pattern. The complexity of the uncertainty of $\mathcal{M}^{Fe}$ is evident. The bottom panel of Figure 5.1 illustrates the use of PCA compression on the Fe XVII emissivity realisations. We have generated 1000 replicate emissivity curves, the exactly same amount of emissivity curve as in the original ensemble, using Eq (3.18) and Eq (3.19) with $J = 14$. The full range, the middle 95%, and the middle 68.3% intervals of these replicates are superimposed on the corresponding intervals for the original emissivity realisations in square root scale. The correspondence between the original emissivity realisations and the PCA replicates is quite good, especially for the 68.3% intervals.

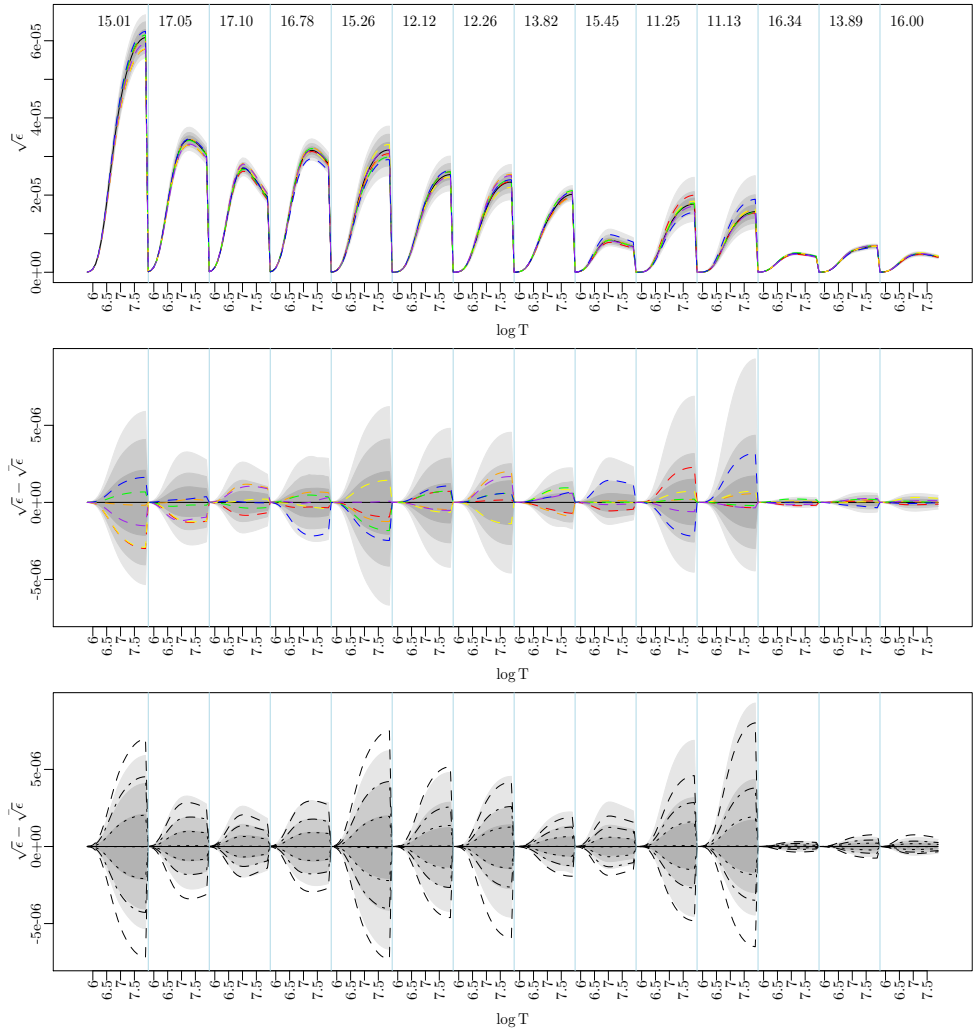

Figure 5.1: Fe XVII emissivity samples and their summary using Gaussian approximation via PCA. In the first panel, the light, dark, and darker grey areas cover the full range, the middle 95% and 68.3% of all 1000 emissivity curves in square root space. $\sqrt{\epsilon}$ is plotted as a solid black curve. Six randomly selected curves are plotted as coloured dashed curves. Other panels are constructed in the same manner, but using $\sqrt{\epsilon^{(m)}} - \bar{\sqrt{\epsilon}}$, to magnify the structure in $\mathcal{M}^{Fe}$. Summarizing the emissivity samples using PCA with $J = 14$, the dashed, dot-dash, and dotted lines, in the third panel, respectively outline intervals containing the full range, the middle 95%, and the middle 68.3% of 1000 PCA replicates of the emissivity curves. The fourteen horizontally-arranged sub-panels correspond to the fourteen Fe XVII lines of interest. Within each sub-panel, the temperature $\log \mathrm{T}$ is increasing.

5.4 Algorithms for the Fe XVII analysis

5.4.1 Algorithm for the pragmatic Bayesian

Under the assumption that the observed counts carry little information as to the uncertainty of emissivities, we aim to construct MC samples of those model parameters from the pragmatic Bayesian target posterior distribution in Eq (5.28) using a four-step Gibbs sampler. We iteratively update $\{Y_B^{Fe}(\mathrm{w})\}$, θ_B^{Fe}, r^{Fe}, and $(\log\alpha_0, \alpha_1)$ by sampling them from the corresponding conditional posterior distributions or prior distribution in Eq (5.29) and Eq (5.31).

At iteration ℓ, we first obtain an MC sample of $\{Y_B^{Fe}(\mathrm{w})^{(\ell)}\}$ from its conditional posterior distribution given the current states of all other parameters, $p(Y_B^{Fe} \mid Y^{Fe}, \theta_B^{Fe(\ell-1)}, r^{Fe(\ell-1)}, \log\alpha_0{}^{(\ell-1)}, \alpha_1^{(\ell-1)})$, which follows a binomial distribution as in Eq (5.24). Secondly, given $\{Y_B^{Fe}(\mathrm{w})^{(\ell)}\}$, we can sample $\theta_B^{Fe(\ell)}$ from $p(\theta_B^{Fe} \mid Z^{Fe}, Y_B^{Fe(\ell)})$ following a Gamma distribution as in Eq (5.25). Thirdly, under the pragmatic Bayesian assumption, sampling of $r^{Fe(\ell)}$ is essentially random from the entire space, i.e., from its prior distribution in Eq (5.19). Fourthly, given $Y_S^{Fe(\ell)} = Y^{Fe} - Y_B^{Fe(\ell)}$, we can sample $(\log\alpha_0{}^{(\ell)}, \alpha_1^{(\ell)})$ from $p(\log\alpha_0, \alpha_1 \mid Y_S^{Fe(\ell)}, r^{Fe(\ell)}, \log\alpha_0{}^{(\ell-1)}, \alpha_1^{(\ell-1)})$ using the adaptive Metropolis algorithm, as discussed in Section 2.2.2. This requires us to specify a new proposal distribution on the fly,

$$
\begin{aligned}
&q(\log\alpha_0, \alpha_1 \mid \log\alpha_0{}^{(\ell-1)}, \alpha_1^{(\ell-1)}) = \\
&\begin{cases}
\mathcal{MVN}\Big((\log\alpha_0, \alpha_1) \mid (\log\alpha_0{}^{(\ell-1)}, \alpha_1^{(\ell-1)}), \frac{0.1^2}{d}\Sigma_0\Big), & \text{if } \ell \le L', \\
(1-\beta)\cdot\mathcal{MVN}\Big((\log\alpha_0, \alpha_1) \mid (\log\alpha_0{}^{(\ell-1)}, \alpha_1^{(\ell-1)}), \frac{2.38^2}{d}\Sigma^{(\ell)}\Big) + & \\
\qquad \beta\cdot\mathcal{MVN}\Big((\log\alpha_0, \alpha_1) \mid (\log\alpha_0{}^{(\ell-1)}, \alpha_1^{(\ell-1)}), \frac{0.1^2}{d}\Sigma_0\Big), & \text{if } \ell > L',
\end{cases}
\end{aligned}
\tag{5.32}
$$

with $\ell > L'$

$$\Sigma^{(\ell)} = \begin{cases} \begin{pmatrix} \text{var}\big(\{\log {\alpha_0}^{(\cdot)}\}_1^{\ell-1}\big) & \text{cov}\big(\{\log {\alpha_0}^{(\cdot)}\}_1^{\ell-1}, \{\alpha_1^{(\cdot)}\}_1^{\ell-1}\big) \\ \text{cov}\big(\{\alpha_1^{(\cdot)}\}_1^{\ell-1}, \{\log {\alpha_0}^{(\cdot)}\}_1^{\ell-1}\big) & \text{var}\big(\{\alpha_1^{(\cdot)}\}_1^{\ell-1}\big) \end{pmatrix}, \\ \qquad\qquad \text{if } \ell \text{ is a multiple of } \mathcal{L}, \\ \Sigma^{(\ell-1)}, \qquad\qquad \text{otherwise,} \end{cases} \tag{5.33}$$

and Σ_0 is a diagonal matrix with prior variances of $\log \alpha_0$ and α_1 as diagonal entries.

The detailed four-step MC sampler under the pragmatic Bayesian model proceeds for iteration $\ell = 1, \ldots, L$ with

Step 1: Sample ${Y_B^{Fe}(\mathrm{w})}^{(\ell)} \mid Y^{Fe}(\mathrm{w}), {\log \alpha_0}^{(\ell-1)}, \alpha_1^{(\ell-1)}, {\theta_B^{Fe}}^{(\ell-1)}, {r^{Fe}}^{(\ell-1)}$
$\sim \text{Binomial}\big(Y^{Fe}(\mathrm{w}), \frac{{\theta_B^{Fe}}^{(\ell-1)}}{{\theta_B^{Fe}}^{(\ell-1)} + \lambda_{SC}(\mathrm{w}; {\log \alpha_0}^{(\ell-1)}, \alpha_1^{(\ell-1)}, {r^{Fe}}^{(\ell-1)})}\big)$
for each $\mathrm{w} \in \mathcal{W}$.

Step 2: Sample ${\theta_B^{Fe}}^{(\ell)} \mid Z^{Fe}, {Y_B^{Fe}}^{(\ell)}$
$\sim \text{Gamma}\big(\sum_{j=1}^{M} Z^{Fe}(\mathrm{w}_j) + \sum_{j=1}^{M} {Y_B^{Fe}(\mathrm{w}_j)}^{(\ell)} + a_1, M(\eta+1) + a_2\big)$,
where $\theta_B^{Fe} \sim \text{Gamma}(a_1, a_2)$ a conjugate prior.

Step 3: Sample ${r^{Fe}}^{(\ell)} \sim \mathcal{MVN}(0, I)$.

Step 4: Sample $(\log {\alpha_0}^{[\text{prop}]}, \alpha_1^{[\text{prop}]})$ from $q(\log \alpha_0, \alpha_1 \mid {\log \alpha_0}^{(\ell-1)}, \alpha_1^{(\ell-1)})$ in Eq (5.32), compute

$$\rho = \frac{p(\log {\alpha_0}^{[\text{prop}]}, \alpha_1^{[\text{prop}]} \mid {Y_S^{Fe}}^{(\ell)})}{p(\log {\alpha_0}^{(\ell-1)}, \alpha_1^{(\ell-1)} \mid {Y_S^{Fe}}^{(\ell)})}, \tag{5.34}$$

and set

$$(\log {\alpha_0}^{(\ell)}, \alpha_1^{(\ell)}) = \begin{cases} (\log {\alpha_0}^{[\text{prop}]}, \alpha_1^{[\text{prop}]}), & \text{with probability } \min(\rho, 1), \\ (\log {\alpha_0}^{(\ell-1)}, \alpha_1^{(\ell-1)}), & \text{otherwise,} \end{cases} \tag{5.35}$$

5.4.2 Algorithms for the fully Bayesian

Under the fully Bayesian model, our aim is to construct an MC sampler of those model parameters from the fully Bayesian target posterior distribution in Eq (5.22) or Eq (5.27) using two basic strategies: a four-step Gibbs sampler and HMC.

Four-step Gibbs Sampler

We construct an MC sampler of those model parameters from the fully Bayesian target posterior distribution in Eq (5.27) using a four-step Gibbs sampler. We iteratively update $\{Y_B^{Fe}(\mathrm{w})\}$, θ_B^{Fe}, r^{Fe}, and $(\log \alpha_0, \alpha_1)$ by sampling them from the corresponding conditional posterior distributions or prior distribution in Eq (5.29) and Eq (5.30).

At iteration ℓ, the sampling process of $\{Y_B^{Fe}(\mathrm{w})^{(\ell)}\}$ in Step 1, $\theta_B^{Fe(\ell)}$ in Step 2, and $(\log \alpha_0{}^{(\ell)}, \alpha_1^{(\ell)})$ in Step 4 is exactly same as what is done in Section 5.4.1. Allowing the observed counts to be informative for the uncertainty of emissivities, in Step 3, we can sample $r^{Fe(\ell)}$ from $p(r^{Fe} \mid Y_S^{Fe(\ell)}, r^{Fe(\ell-1)}, \log \alpha_0{}^{(\ell-1)}, \alpha_1^{(\ell-1)})$ using the adaptive Metropolis algorithm, as discussed in Section 2.2.2. We need to specify a new proposal distribution on the fly,

$$
q(r^{Fe} \mid r^{Fe(\ell-1)}) = \begin{cases} \mathcal{MVN}\Big(r^{Fe} \mid r^{Fe(\ell-1)}, \frac{0.1^2}{d}\Sigma_0\Big), & \text{if } \ell \le L', \\ (1-\beta)\cdot \mathcal{MVN}\Big(r^{Fe} \mid r^{Fe(\ell-1)}, \frac{2.38^2}{d}\Sigma^{(\ell)}\Big) + & \\ \qquad \beta \cdot \mathcal{MVN}\Big(r^{Fe} \mid r^{Fe(\ell-1)}, \frac{0.1^2}{d}\Sigma_0\Big), & \text{if } \ell > L', \end{cases} \tag{5.36}
$$

with $\ell > L'$

$$\Sigma^{(\ell)} = \begin{cases} \begin{pmatrix} \operatorname{var}\left(\{r_1^{Fe(\cdot)}\}_1^{\ell-1}\right) & \cdots & \operatorname{cov}\left(\{r_1^{Fe(\cdot)}\}_1^{\ell-1}, \{r_J^{Fe(\cdot)}\}_1^{\ell-1}\right) \\ \vdots & \ddots & \vdots \\ \operatorname{cov}\left(\{r_J^{Fe(\cdot)}\}_1^{\ell-1}, \{r_1^{Fe(\cdot)}\}_1^{\ell-1}\right) & \cdots & \operatorname{var}\left(r_J^{Fe(\cdot)}\}_1^{\ell-1}\right) \end{pmatrix} \\ \qquad\qquad \text{if } \ell \text{ is a multiple of } \mathcal{L}, \\ \Sigma^{(\ell-1)}, \qquad\qquad \text{otherwise,} \end{cases} \tag{5.37}$$

and Σ_0 is a diagonal matrix with prior variances of r^{Fe} as diagonal entries.

The detailed four-step Gibbs sampler under the fully Bayesian model proceeds for iteration $\ell = 1, \ldots, L$ with

Step 1: Sample $Y_B^{Fe}(\mathrm{w})^{(\ell)} \mid Y^{Fe}(\mathrm{w}), \log \alpha_0{}^{(\ell-1)}, \alpha_1^{(\ell-1)}, \theta_B^{Fe(\ell-1)}, r^{Fe(\ell-1)}$
$\sim$ Binomial $\left(Y^{Fe}(\mathrm{w}), \frac{\theta_B^{Fe(\ell-1)}}{\theta_B^{Fe(\ell-1)} + \lambda_{SC}(\mathrm{w}; \log \alpha_0{}^{(\ell-1)}, \alpha_1^{(\ell-1)}, r^{Fe(\ell-1)})}\right)$
for each $\mathrm{w} \in \mathcal{W}$.

Step 2: Sample $\theta_B^{Fe(\ell)} \mid Z^{Fe}, Y_B^{Fe(\ell)}$
$\sim$ Gamma $\left(\sum_{j=1}^M Z^{Fe}(\mathrm{w}_j) + \sum_{j=1}^M Y_B^{Fe}(\mathrm{w}_j)^{(\ell)} + a_1, M(\eta+1) + a_2\right)$,
where $\theta_B^{Fe} \sim \text{Gamma}(a_1, a_2)$ a conjugate prior.

Step 3: Sample $r^{Fe[\text{prop}]}$ from $q(r^{Fe} \mid r^{Fe(\ell-1)})$ in Eq (5.36), compute

$$\rho_1 = \frac{p(r^{Fe[\text{prop}]} \mid Y_S^{Fe(\ell)})}{p(r^{Fe(\ell-1)}) \mid Y_S^{Fe(\ell)})}, \tag{5.38}$$

and set

$$r^{Fe(\ell)} = \begin{cases} r^{Fe[\text{prop}]}, & \text{with probability } \min(\rho_1, 1), \\ r^{Fe(\ell-1)}, & \text{otherwise.} \end{cases} \tag{5.39}$$

Step 4: Sample $(\log \alpha_0{}^{[\text{prop}]}, \alpha_1^{[\text{prop}]})$ from $q(\log \alpha_0, \alpha_1 \mid \log \alpha_0{}^{(\ell-1)}, \alpha_1^{(\ell-1)})$ in Eq (5.32), compute

$$\rho_2 = \frac{p(\log \alpha_0{}^{[\text{prop}]}, \alpha_1^{[\text{prop}]} \mid Y_S^{Fe(\ell)})}{p(\log \alpha_0{}^{(\ell-1)}, \alpha_1^{(\ell-1)} \mid Y_S^{Fe(\ell)})}, \tag{5.40}$$

and set

$$(\log \alpha_0{}^{(\ell)}, \alpha_1^{(\ell)}) = \begin{cases} (\log \alpha_0{}^{[\text{prop}]}, \alpha_1^{[\text{prop}]}), & \text{with probability } \min(\rho_2, 1), \\ (\log \alpha_0{}^{(\ell-1)}, \alpha_1^{(\ell-1)}), & \text{otherwise.} \end{cases} \tag{5.41}$$

Hamiltonian MC

We construct an MC sampler of those model parameters from the fully Bayesian target posterior distribution in Eq (5.22) with the likelihood function in Eq (5.16) the prior independence assumption in Eq (5.21). We can then obtain sample of those model parameters from this higher dimension posterior via HMC by sampling directly from their joint posterior distribution.

5.5 Application to Fe XVII simulation studies

We generate sets of the source and the background counts from reasonable and known values of all the model parameters. The exact same models and algorithms are applied to this simulated datasets as to the real observed datasets in the next section. This, as a crucial approach, will allow us to test the ability of our programming to recover physical parameters from the Fe XVII photon counts.

We assume values of $\log \alpha_0 = 31.17$, $\alpha_1 = 6.75$, $\theta_B^{Fe} = 11.11$, and $J = 7$ principle components $r^{Fe} = (6.15, 5.38, -4.45, -7.20, 8.49, 3.54, 0.81)$, capturing 55% of the total variance of the emissivities for fourteen Fe XVII lines as computed with Eq (3.17), are used to construct a new emissivity via a simple linear combination in Eq (3.18). Those values are obtained from an initial try to explore the reasonable range of those parameters. In this setting, we obtain the expected values of the Poisson parameter in Eq (5.14) and Eq (5.15). We simulate 30 sets of source and background photon counts for each channel from the Poisson distributions in Eq (5.14) and Eq (5.15).

We use the Stan software package to obtain sample of those model parameters via HMC by sampling directly from their joint posterior distribution in Eq (5.22) given each of the above simulated sets of photon counts. For each simulated set of photon counts, there are 6000 iterations sampled with random initials and the first 5000 iterations are discarded as burn-in.

The true values of the parameters can be recovered. The density plots in Figure 5.2 summarise the posterior samples of those model parameters across different simulated datasets and indicate the shape of the posterior distributions of those model parameters are more like Gaussian. The density lines of those posterior samples are able to cover the true values of the corresponding parameters. The 68% credible intervals and the 95% credible intervals of the posterior samples for each parameter over the 30 simulated datasets in Figure 5.3 indicate a proper coverage as well. These credible intervals of the posterior samples are proved to be confidence intervals because they have certain coverage in this simulation study. The expected photon counts in the source exposure, $s^{Fe}(\mathrm{w}; r^{Fe}, \log\alpha_0, \alpha_1)+\kappa^{Fe}(\mathrm{w})+\theta_B^{Fe}$, for each $\mathrm{w} \in \mathcal{W}^{Fe}$, can be evaluated from the posterior means of samples for those model parameters for each simulated dataset. The expected photon counts match the simulated counts well, as in Figure 5.4, because of the ability of our programming to recover the Fe XVII parameter and the photon counts. Almost 95% of the standardised residuals of the expected photon counts, in Figure 5.5, fall in $(-2, 2)$ without any dependency indicating the model can fit the data well.

5.6 Application to Fe XVII observed counts

Here we demonstrate the effects of the different types of models by applying them to an observed dataset from grating MEG and positive order, denoted by 'amp', which comprises sets of observed counts from the source and the background exposures in 1400 channels. We also have a collection of $M = 1000$ emissivity realisations for 14 spectral lines of Fe XVII.

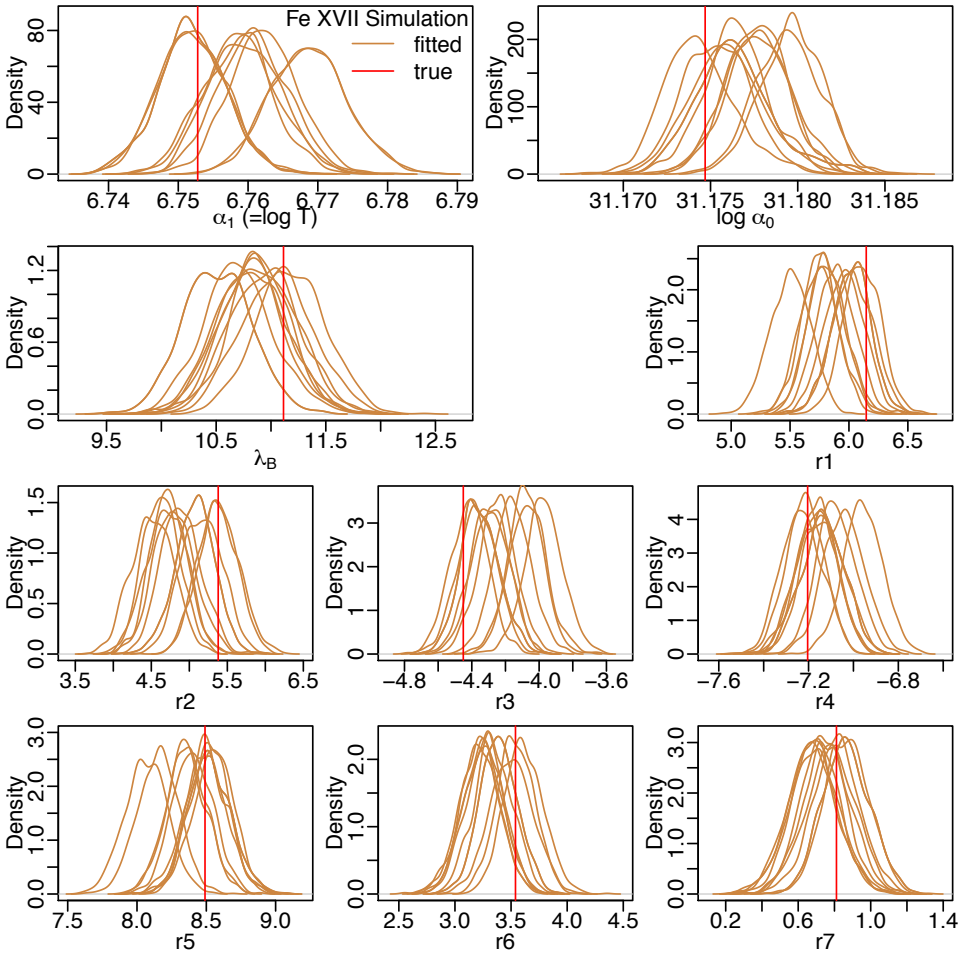

Figure 5.2: Density plots for the posterior samples of the model parameters across different simulated datasets in the Fe XVII simulation study. They all are compared with the corresponding true values of the parameters marked as red vertical lines.

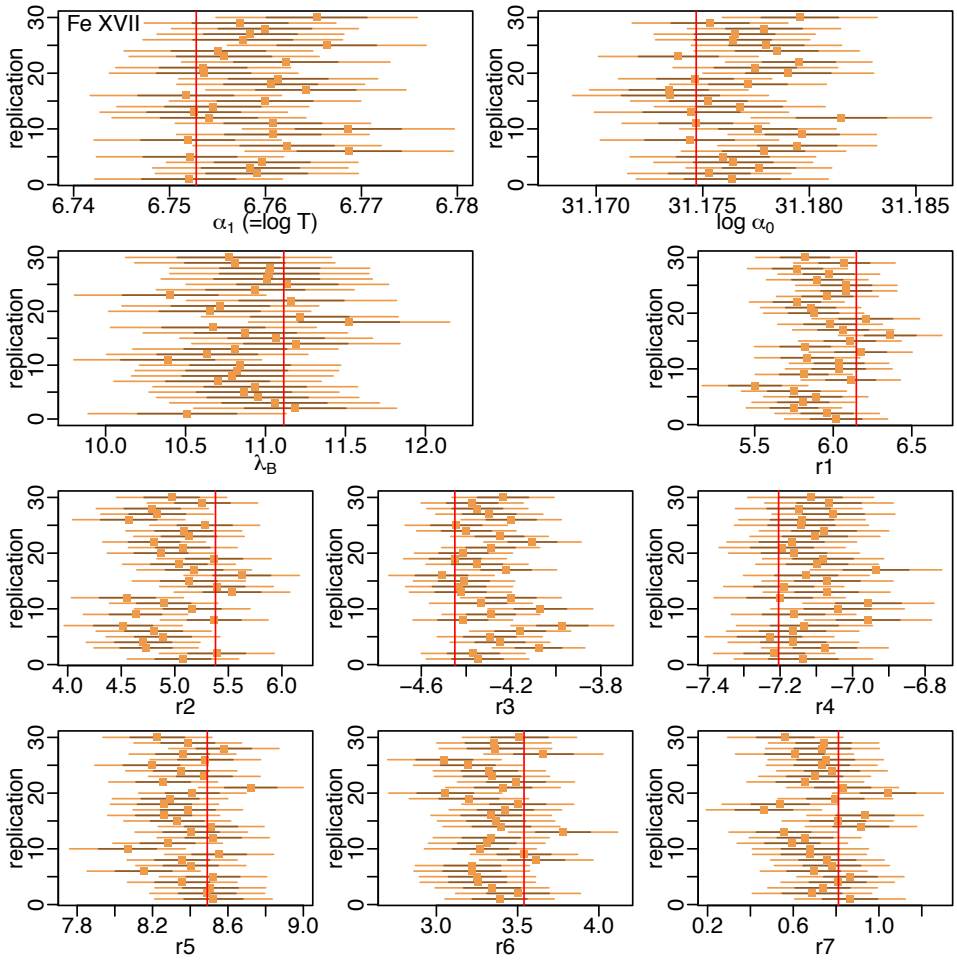

Figure 5.3: The posterior means (brown square dots), the 68% (dark brown horizontal lines) and the 95% (light brown horizontal lines) credible intervals over each of the 30 Fe XVII simulated datasets. They are compared with the true values of the corresponding parameters, marked as red vertical lines.

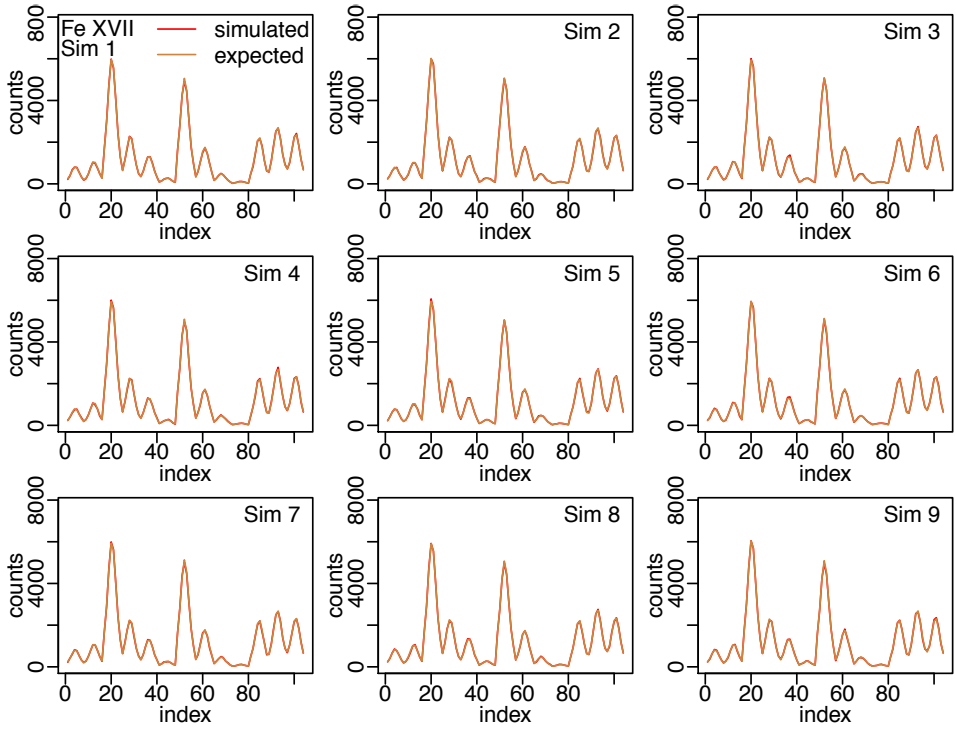

Figure 5.4: The comparison of the expected photon counts, in brown, and the simulated counts, in red, along the indices of the filtered wavelength for the first 9 replicates in the primary Fe XVII simulation study. No obvious difference between the expected and the simulated photon counts because our programming is able to recover the Fe XVII parameter and the photon counts.

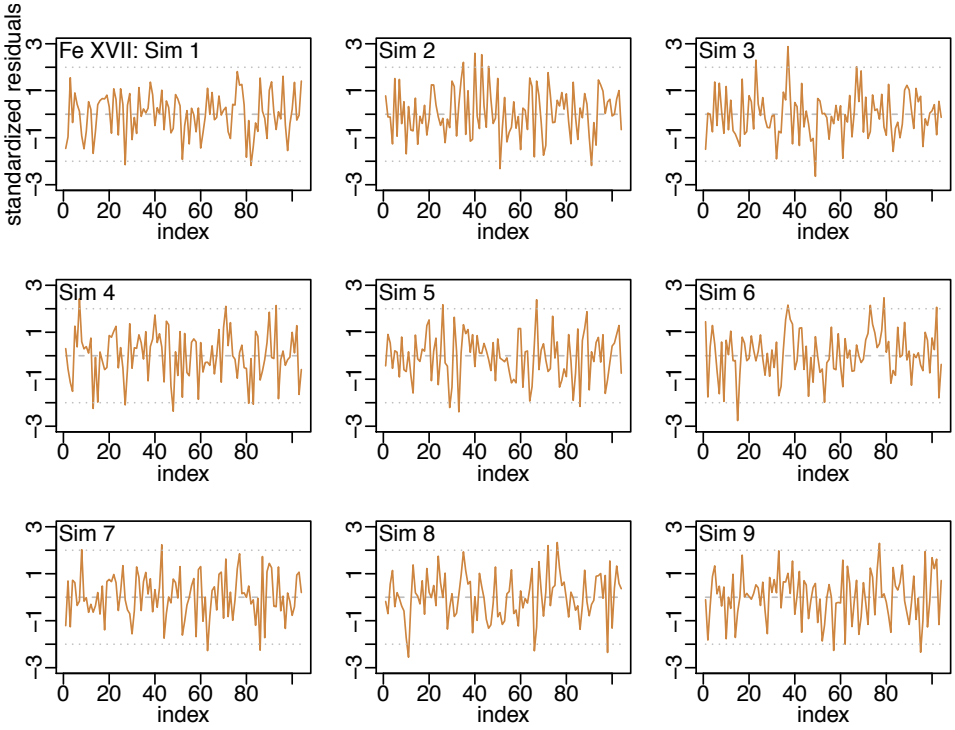

Figure 5.5: The standardised residuals of the expected source photon counts, along the indices of the filtered wavelength, for the first 9 simulated datasets in the Fe XVII simulation study. The standardised residual is the ratio of the difference between the simulated counts and the expected counts to the square root of the expected counts since the counts follow Poisson distributions.

5.6.1 Preprocessing

Filtering the wavelength, we focus on the observed counts of a subset of channels based on where the lines of interest are, typically the adjacent 9 or 10 channels on both sides of a line. Therefore, there could be gaps in $\mathcal{W}^{Fe}$.

As a calibration problem, the model wavelength location, i.e., the recorded wavelength, may be slightly displaced from the measured wavelength of the data resulting in mismatches in the spectrum profile. By shifting the wavelength of the model for the particular lines by a few channels, the expected counts could match the observed counts better. By comparing the observed counts with the expected counts at around 12.124 Å with corresponding wavelength channel indices at around 20, we see the peak of the expected data shifted by about two channels in the left panel of Figure 5.6. If those channels, that correspond to 12.124 Å line, are shifted by two channels, the shifted expected counts will match the observed counts well. We can refit the model with this shifted dataset where only those channels that correspond to 12.124 Å line are shifted. The right panel of Figure 5.6 indicates a better match between the observed counts and the refitted expected counts. We have obtained a better estimate of the flux. There is an offset between the wavelength of the data and the wavelength of the model. Unfortunately, as a calibration issue, this offset is not a constant shift. It is wavelength dependent and is highly random all over the place and across the gratings, evidenced in this preprocessing analysis. Therefore, shifting adjustment for each of the wavelengths is necessary to match their location individually, to make their profiles line up with the data, and to make sure the model is fitted properly.

Some lines of interest might be contaminated and mixed with other lines from different ions. For example, the 16.004 Å line of Fe XVII is weak itself and there is an O VII line mixed up with it (Del Zanna 2011). Incorporating this line may affect the accuracy of our results. Therefore, we will ignore any such lines having contamination. As a result, there are in total only $H = 100$ sub-channels considered in this Fe XVII analysis.

According to the right panel of Figure 5.6, there again are mismatches of line

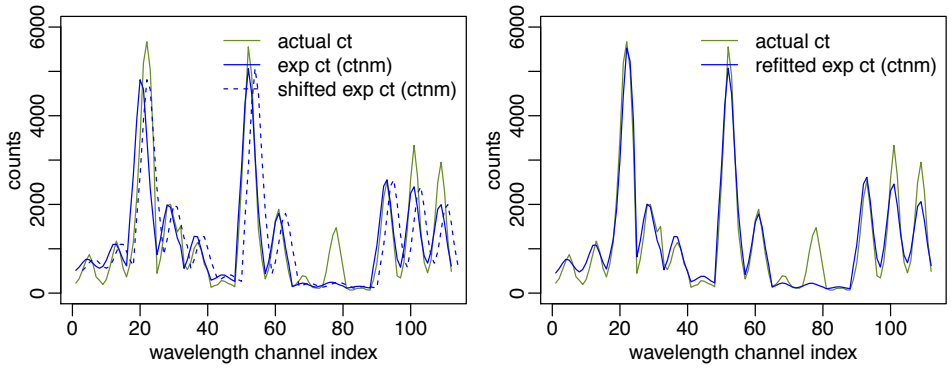

Figure 5.6: The shifting process of the Fe XVII spectrum when shifting the wavelength of the model. In the left panel, the observed counts, in solid green line, is compared with the expected counts, in solid blue line. At around 12.124 Å with corresponding wavelength channel indices at around 20, we see the peak of the expected data shifted by about two channels. If those channels are shifted by two channels, the shifted expected counts, in dashed blue line, match the observed counts well at around 12.124 Å. We then refit the model with the shifted dataset where only those channels that correspond to 12.124 Å line are shifted. The observed counts, in solid green line, is compared with the expected counts after shifting, in solid blue line, in the right panel.

locations and line width, e.g., the expected counts do not reach the peak and the trough of the observed counts for the last three lines with wavelength channel indices > 80. Moreover, some of the wavelength we have in the data are not exact. Though LRF is supposed to be constant, we might need to make small adjustments to its width and move the centre a little bit left and right for each of the lines to get spectrum profiles measurably broader or narrower and get them line up with the data properly, from an instrumental point of view. We adjust the locations and the width of LRF, which is in the form of a t-distribution as in Eq (1.1), for each line. The residual sum of squares (RSS) is the sum of the squares of the deviations of expected data from actual data. As an optimality criterion in parameter selection, it measures the variation of the photon counts which is not accounted for by the fitted model. It also plays the role of a loss function in this analysis. A small RSS indicates a tight fit of the model to the photon counts. The best-fit width, i.e., the scale parameter $\hat{\sigma}$, is chosen from a fine grid for each individual line by minimising the RSS. The best-fit centre, i.e., the location parameter $\hat{\omega}$, is chosen from another fine grid for each individual line based on the corresponding best-fit width. A summary of the best-fit width

ω (Å)	$\hat{\sigma}$	$\hat{\omega}$ (Å)
11.129	0.0090	11.131
11.250	0.0090	11.253
12.124	0.0105	12.122
12.264	0.0105	12.267
13.825	0.0120	13.827
13.890	0.0115	13.891
15.013	0.0090	15.013
15.262	0.0090	15.262
15.453	0.0135	15.454
16.336	0.0110	16.337
16.776	0.0085	16.777
17.051	0.0080	17.052
17.096	0.0075	17.097

Table 5.1: A summary of the best-fit width, $\hat{\sigma}$, and the adjusted centres, $\hat{\omega}$, for the Fe XVII LRF.

and the adjusted centres is shown in Table 5.1 and is used throughout this chapter. Moreover, the full width at half maximum (FWHM), a parameter commonly used to describe the width of a 'bump' on a curve, is $2*\sqrt{2^{\frac{2}{\nu+1}}-1}*\sqrt{\nu\hat{\sigma}^2}$ falling in $(0.017, 0.028)$ with the best-fit width values in Table 5.1. They are consistent with the FWHM for amp dataset used in this section, approximately 0.023, in Canizares et al. (2005).

Most of the above mismatches in the spectrum profiles come from the calibrations. Assuming the simple form of LRF for each individual line, the estimations now seem to be unbiased and no trend. There could still be other factors or more complicated LRF which have not gone into the analysis, however, they would not affect the accuracy of our results much.

5.6.2 Application to Fe XVII observed photon counts

Based on the preprocessing work in Section 5.6.1, there are in total only $H = 100$ sub-channels considered for $L = 13$ spectral lines in this Fe XVII analysis. We illustrate the algorithms for both the pragmatic and the fully

Bayesian models mentioned in Section 5.4. Note that a simple case where the continuum counts are ignored in the astrophysical model is tested initially using the same data. However, extremely large standardised residuals are obtained due to the fact that the model is less specified. Therefore, in this application section, we focus on an improved model where the continuum photon counts are specified.

Four-step MC Sampler for the pragmatic Bayesian

There are 2000000 MC iterations drawn when implementing the four-step Gibbs Sampler for the pragmatic Bayesian where the first $L' = 1000$ are used to trigger the sampler and the following first half is discarded as burn-in. The first 7 principle components are considered accounting for about 55% of the total variability of the given emissivities. Specifically, in Step 4, the adaptive Metropolis algorithm is used to sample $(\log \alpha_0, \alpha_1)$ from $p(\log \alpha_0, \alpha_1 \mid Y_S^{Fe})$. The empirical estimate of the variance-covariance matrix is updated at every $\mathcal{L} = 50$ iterations to prevent unnecessary computing. A small $\beta = 0.05$ is used in the mixture proposal distribution. The prior variance matrix of the variables is $\Sigma_0 = \text{diag}(\text{var}(\log \alpha_0), \text{var}(\alpha_1))$.

Four-step Gibbs Sampler for the fully Bayesian

There are 200000 MC iterations drawn when implementing the four-step Gibbs Sampler for the fully Bayesian where the first $L' = 1000$ are used to trigger the sampler and the following first half is discarded as burn-in. The first 7 principle components are considered as well. The adaptive Metropolis algorithm is used to sample r^{Fe} from $p(r^{Fe} \mid Y_S^{Fe})$ in Step 3 and to sample $(\log \alpha_0, \alpha_1)$ from $p(\log \alpha_0, \alpha_1 \mid Y_S^{Fe})$ in Step 4. The empirical estimate of the variance-covariance matrices in both steps are updated at every $\mathcal{L} = 50$ iterations to prevent unnecessary computing. A small $\beta = 0.05$ is used in the mixture proposal distributions in both steps. The prior variance matrix of the variables is $\Sigma_0 = I_7$ in step 3 and $\Sigma_0 = \text{diag}(\text{var}(\log \alpha_0), \text{var}(\alpha_1))$ in step 4.

HMC for the fully Bayesian

Using the same 7 principle components, we implement the Stan software package to obtain sample of those model parameters via HMC by sampling directly from their joint posterior distribution in Eq (5.22). There are 4 chains running, 6000 iterations sampled each, with random initials, and the first 5000 iterations of each chain are discarded as burn-in.

Comparison of algorithms and output data analysis

Under the fully Bayesian model, there are two algorithms, the four-step Gibbs Sampler and the HMC, used for obtaining the posterior samples of those model parameters. There are good matches in density distributions of the posterior samples for each individual parameter from the two algorithms, the four-step Gibbs Sampler and the HMC, and also in their corresponding quantiles, as indicated by the density plots and the quantile-quantile plots in Figure 5.7. Therefore, both algorithms are giving reasonable samples. Demanding less sample in achieving convergence and significantly faster, HMC is able to provide a good representation of the posterior distribution and we will therefore use it from now on.

Figure 5.8 compares the posterior samples of those model parameters in Fe XVII, $(\alpha_1, \log \alpha_0, \theta_B^{Fe}, r^{Fe})$, under the pragmatic and the fully Bayesian models. Notice that incorporating the uncertainties of emissivities under the pragmatic Bayesian model introduces large error bars, while the fully Bayesian model is shrinking the error bars relative to the pragmatic Bayesian model. For the posterior samples of α_1 and $\log \alpha_0$, there are gaps between the fully Bayesian model and the pragmatic Bayesian model indicating that they are sensitive to the choice of the emissivities and the data is more informative about the emissivities than the prior knowledge. As we use a Gaussian approximation to the prior distribution of emissivities, which has the support in the whole space, the PCA decomposition of the emissivities can be broader than the actual ensemble but with the same standard deviation. As the posterior samples of some of the principal component variables under the

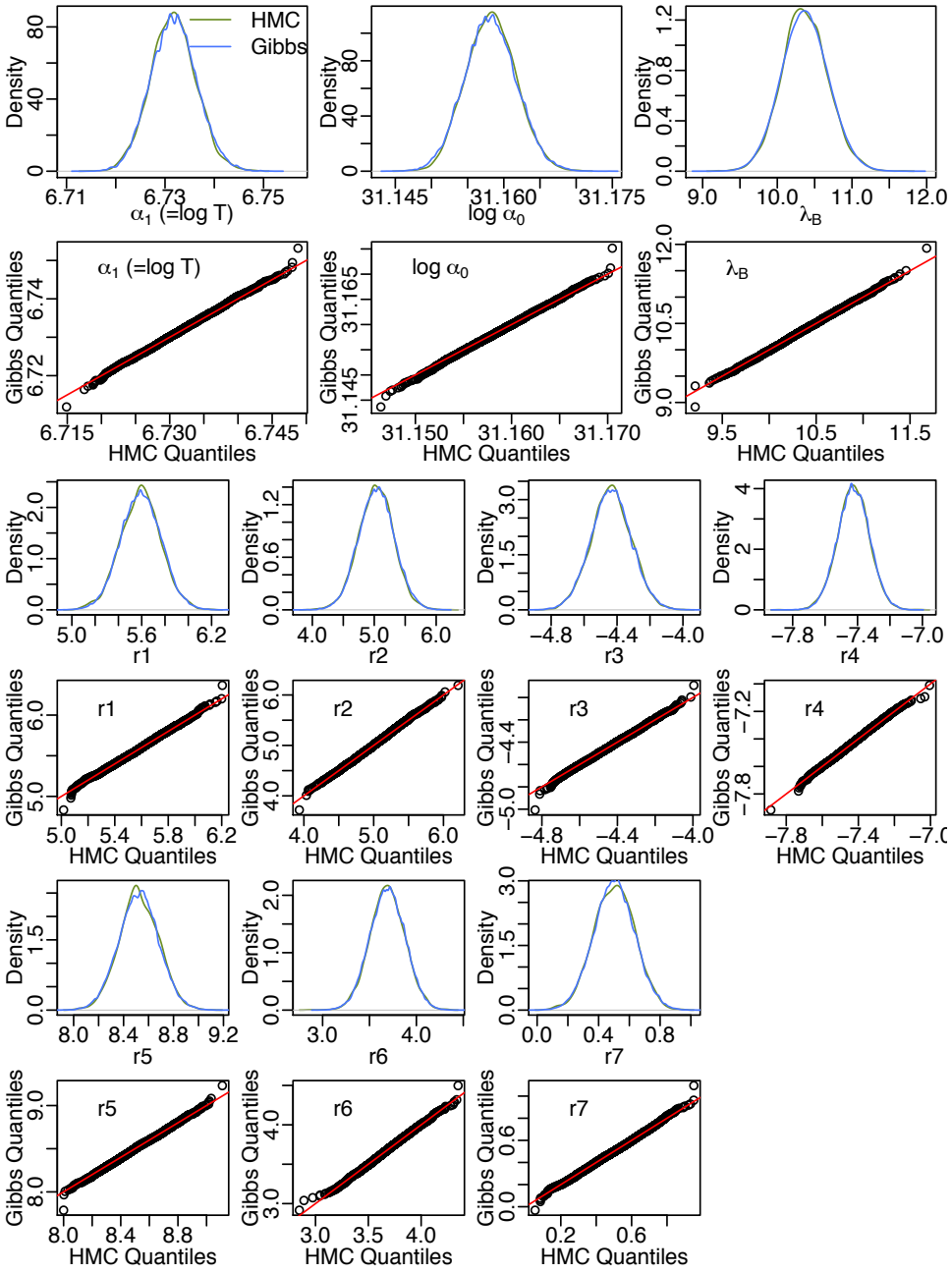

Figure 5.7: Comparison of the posterior samples of those model parameters in Fe XVII, $(\alpha_1, \log \alpha_0, \theta_B^{Fe}, r^{Fe})$, from both algorithms, the four-step Gibbs Sampler (blue) and the HMC (green), under the fully Bayesian model via the the density plots and the quantile-quantile plots. Identity lines are marked as red lines.

fully Bayesian model (green lines in the bottom two rows of Figure 5.8 are up to eight standard deviations away from the center of the prior distribution, it is possible to see the best-fit atomic emissivities for the thirteen Fe XVII lines considered in this work, which are preferred by the data, are quite far away from the default emissivity values from the CHIANTI, as in Figure 5.9. The preferred atomic emissivities for 15.013 Å or 12.124 Å could be smaller by about 30% or larger by about 100% than its default emissivity values, while that for other lines are very close to their default values.

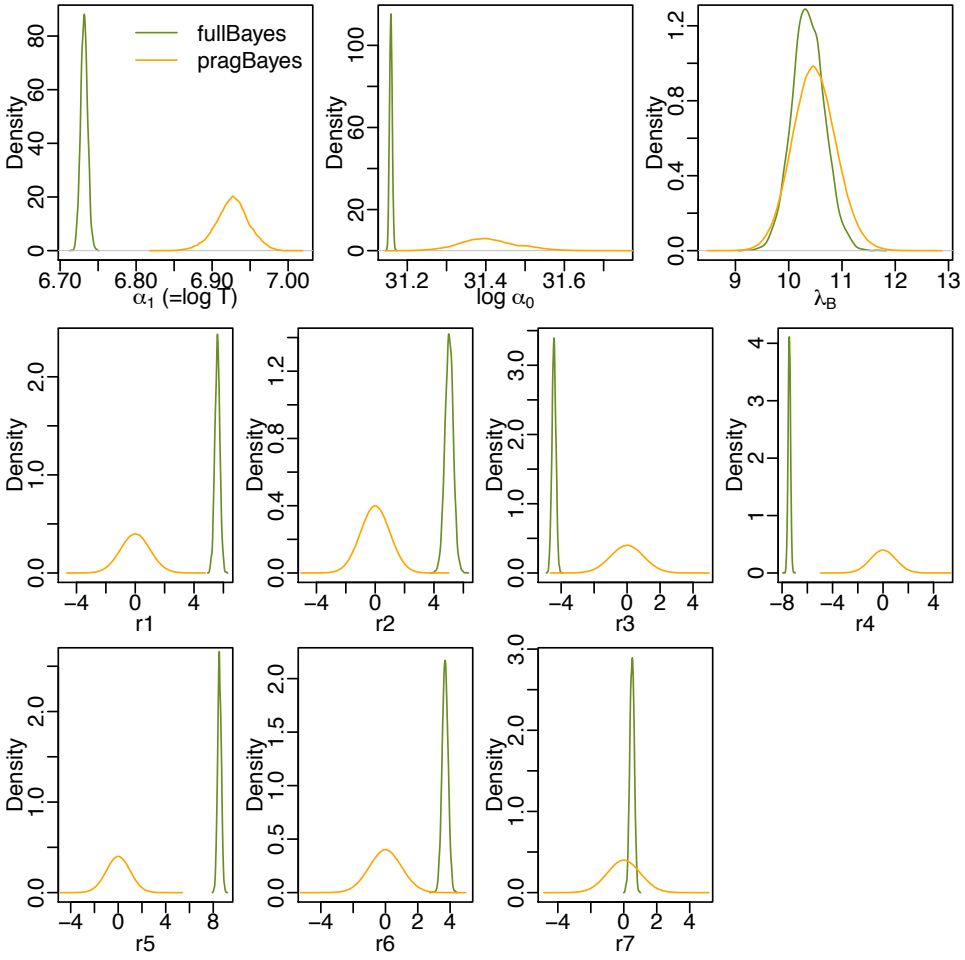

Figure 5.8: Comparison of the posterior samples of those model parameters in Fe XVII case study, $(\alpha_1, \log \alpha_0, \theta_B^{Fe}, r^{Fe})$, under the pragmatic (amber) and the fully (green) Bayesian models.

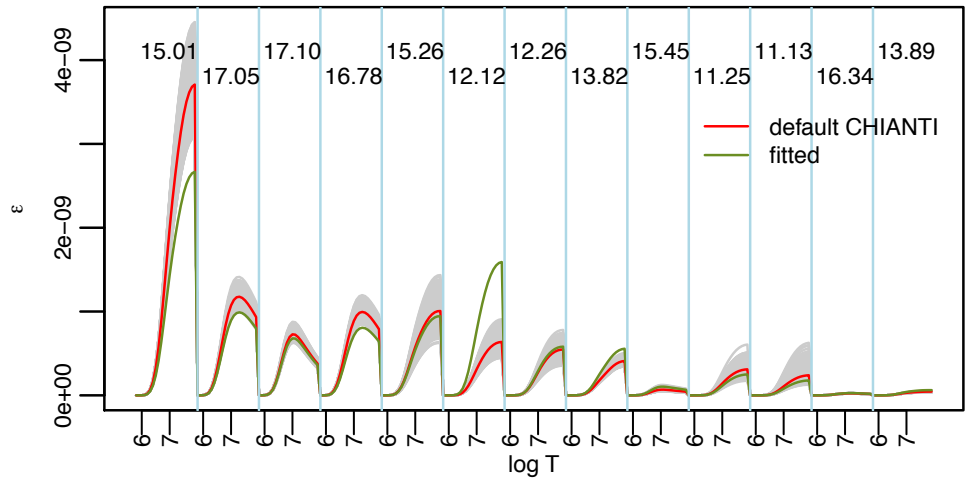

Figure 5.9: Emissivities of the thirteen Fe XVII lines considered in this work. The grey lines represent the 1000 realizations of the CHIANTI atomic data. The red curve is the default value from the CHIANTI. The green curve is the best emissivities preferred by the data under the fully Bayesian model.

The expected photon counts in the source exposure can be evaluated from the posterior means of those model parameters sampled from HMC under the fully Bayesian model. There is a good match when comparing the expected photon counts with the observed photon counts in the source exposure along the wavelength, as seen in the upper panel of Figure 5.10. The middle panel is a zoom in version, which is constructed in the same way while removing the gaps in the wavelength, to look at the detailed features of the spectrum profiles for the lines of interest. The light blue dashed curve represents the contribution of the continuum counts in the spectrum, which is about 51 counts per channel on average and there are more contribution of the continuum at shorter wavelength. Most of the standardised residuals of the expected photon counts, in the lower panel of Figure 5.10, fall in 95% confidence interval, $(-2, 2)$, without any dependency indicating the fully Bayesian model can fit the data very well. There are still complex atomic or calibration factors having been ignored in this analysis, which lead to some relatively large standardised residuals up to ± 5.

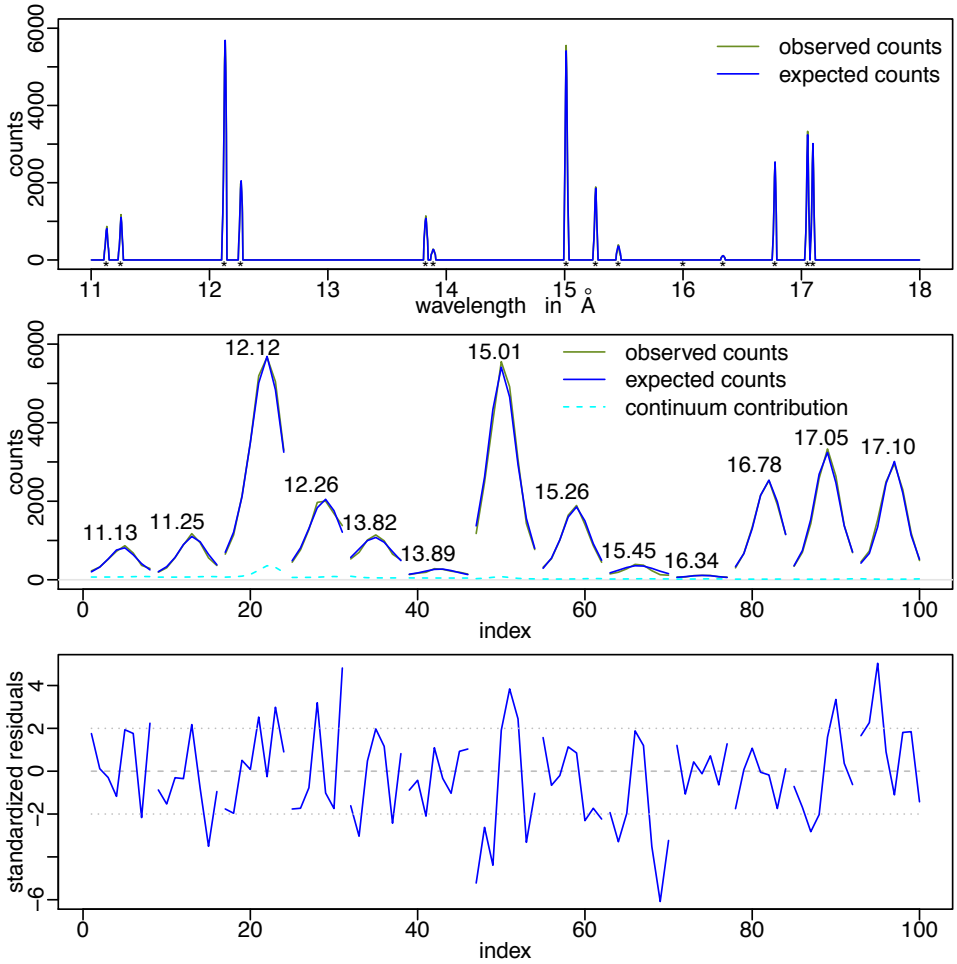

Figure 5.10: Comparison of the expected spectrum profiles under the fully Bayesian model with the observed Fe XVII spectrum profiles. The upper panel compares the expected photon counts, marked as blue curve, with the observed photon counts, marked as green curve, in the source exposure along the wavelength (Å). The lines of interest are marked as '*'. The middle panel is a zoom in version focusing on the areas covered by the lines of interest. It is constructed in the same way while removing the gaps in the wavelength to compare the spectrum profiles closely and x-axis is the indices of model wavelength. The wavelengths of interest are marked on top of each line. The light blue curve represents the contribution of the continuum counts in the spectrum. The standardized residuals, which is the ratio of the difference between the observed counts and the expected counts to the square root of the expected counts since the counts follow Poisson distributions, are plotted along the indices of model wavelength in the lower panel.

5.7 Posterior summary

We have obtained the parameter estimations $\log \alpha_0 = 31.158 \pm 0.003$ and $\alpha_1 = 6.732 \pm 0.005$, via HMC under the fully Bayesian model in Section 5.6. They can be summarized with a variety of analytical distributions, denoted by $\hat{p}(\log \alpha_0, \alpha_1 \mid Y^{Fe}, Z^{Fe})$, to approximate the above complex marginal posterior distribution $p(\log \alpha_0, \alpha_1 \mid Y^{Fe}, Z^{Fe})$ in Eq (6.18). We have tested different analytical distributions including a multivariate Gaussian distribution, a multivariate t-distribution, and kernel density estimation (KDE).

The contour plots, in Figure 5.11, show where the majority 68% and 95% of the mass of the approximated joint probability distributions of $(\log \alpha_0, \alpha_1)$ fall, using different approximation distributions: KDE, a two dimensional Gaussian distribution, and a multivariate t-distribution. They are compared with the MC posterior sample of $\log \alpha_0$ and α_1 obtained via HMC. The 68% contours for the three different approximation distributions are generally similar. While the 95% contour for the t-distribution approximation is much wider than the other two indicating there are heavy tails in it. Figure 5.12 compares the HMC sample quantiles with the three different approximation model quantiles for each parameter. Both lines for the KDE and the Gaussian approximation fall very closely to the identity line indicating a high consistency with the posterior samples. However, the tails for the t-distribution approximation deviate from the identity line significantly indicating heavy tails again. As a result, the t-distribution has a fatter tail and is designed to be conservative. Though the KDE is as good as the Gaussian approximation, the evaluation over the KDE is quite complex and time consuming. Therefore, we choose the Gaussian approximation to summarize the posterior samples.

Ultimately, the approximated distribution, $\hat{p}(\log \alpha_0, \alpha_1 \mid Y^{Fe}, Z^{Fe})$, obtained in this preliminary analysis on Fe XVII will be carried forward into the primary analysis on O VII model as the joint prior distribution of $(\log \alpha_0, \alpha_1)$. There might be different systematic errors among different datasets. We do not want to be constrained about how we carry information from one analysis to another. Therefore, as a conservative manner, we will also use a multi-

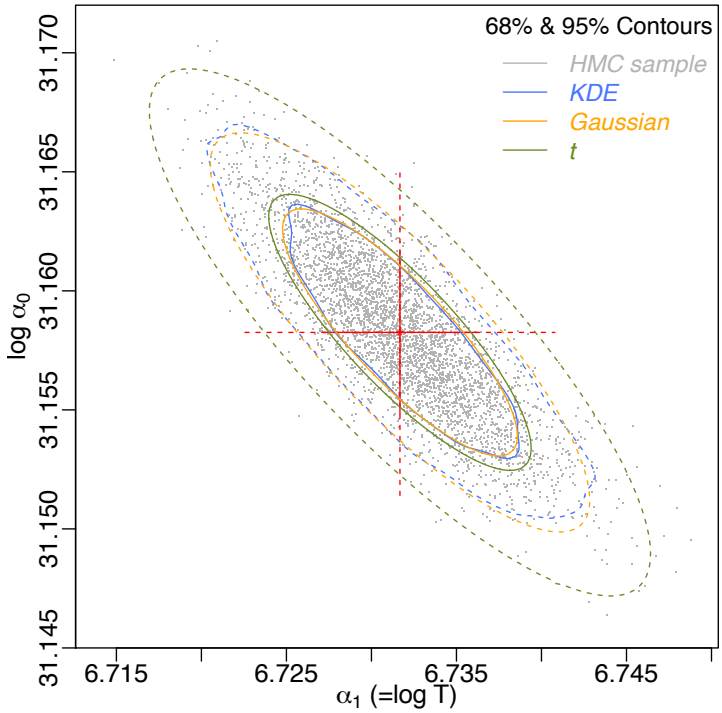

Figure 5.11: Comparison of the approximation distributions for Fe XVII posterior sample under the fully Bayesian model using HMC via contour plots. There are three approximation distributions for the joint probability distributions of $(\log \alpha_0, \alpha_1)$, with blue representing KDE, amber a two dimensional Gaussian distribution, and green a multivariate t-distribution. The solid contours contain the majority 68% of the mass of the approximated joint probability distributions and the dashed contours contain the 95% of them. They are compared with the MC posterior sample of $\log \alpha_0$ and α_1 obtained via HMC, grey dots for posterior samples, red solid lines denoting the $\pm 1\sigma$, and red dashed lines denoting the $\pm 2\sigma$ of the posterior samples.

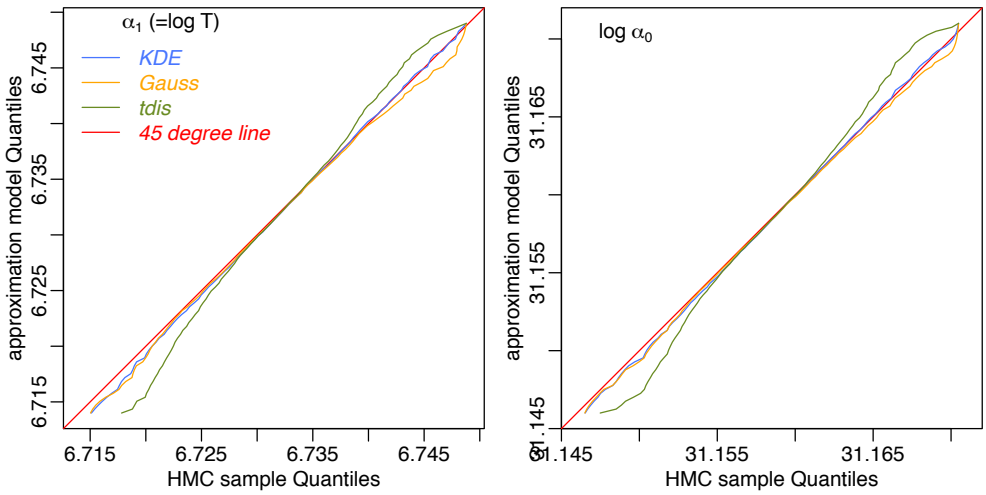

Figure 5.12: Comparison of the approximation distributions for Fe XVII posterior sample under the fully Bayesian model using HMC via quantile-quantile plots. There are three approximation distributions for the joint probability distributions of $(\log \alpha_0, \alpha_1)$, with blue representing KDE, amber a two dimensional Gaussian distribution, and green a multivariate t-distribution. A red identity line is added as a benchmark.

variate t-distribution as one of the approximation distributions. More details will be discussed in Chapter 6.

5.8 Conclusion and discussion

We have used a Bayesian framework to interpret the observed Fe XVII photon counts in terms of a given ensemble of atomic data. A fully Bayesian model, where we allow the observed photon counts to update the uncertainty in the emissivities, reduces the uncertainties in the plasma parameters, compared to that of from a pragmatic Bayesian model, where the uncertainty in the emissivities are independent of the observed photon counts. Reasonable temperature and volume values are obtained after careful preprocessing and model fitting works in this preliminary stage. The approximated distributions of these two parameters, where the Gaussian approximation matches the sample well and the t-distribution approximation is designed to be conservative, will be carried forward into the primary analysis on O VII in the next chapter.

There are still complex atomic or other systematic factors having been ignored in this analysis leading to some relatively large standardised residuals. We have adjusted the location parameters for each line in the preprocessing step to reduce the standardised residuals and improve the model fitting. We could implement this location adjustment analysis onto multiple datasets and look at the ensemble of the adjustments. If we notice that the best-fit location parameters for each line across multiple datasets were systematically different from the true value, it may point to a problem in the atomic data and in the listed wavelength lines. If they were distributed on either side as positive or negative deviations, it may point to an instrumental calibration issue.

6

Three-stage analysis with Fe XVII and O VII

More complicated than the previous two chapters, we now focus on a three-stage analysis and its application. The posterior distributions obtained in the preliminary analyses in the first two stages are carried forward into the primary analysis in the following stage. We consider, in the primary analysis, both the pragmatic method where the observed data is not informative to the choice of the atomic uncertainties and the fully Bayesian method where the observed data is allowed to update the atomic uncertainties.

In contrast to those based on Fe XIII line ratios in Chapter 4 where temperature sensitivity is ignorable and on Fe XVII line ratios in Chapter 5 where density sensitivity is ignorable, both the temperature sensitivity and the density sensitivity are not ignorable for O VII line systems in the soft X-ray regime. Thus, besides atomic uncertainties, parameter uncertainties from both density and temperature need to be considered as well. This situation serves as a bridge from the problem considering only density or temperature estimation to solving the full emissivity curve estimation problem. Based on what we have done in Chapter 4, the basic statistical issue becomes that the

number of parameters increases to include a temperature parameter, where either a delta function or a pre-specified shape can be set up as its prior distribution as what is done in Chapter 5. Overall there are three places where different uncertainties are to be incorporated: uncertainties of atomic data, uncertainties for a subset of plasma parameters, and uncertainties for all other plasma parameters. The emissivity ensemble obtained by the atomic physicists represents the uncertainties of atomic data, which is playing the role of one of the preliminary analyses. The analysis on the Fe XVII spectral lines in Chapter 5, as another preliminary analysis, provides uncertainties for a subset of plasma parameters including a priori temperature information. The former two sources of uncertainties are then taken forward into the final primary analysis on O VII spectral lines to measure the density parameter. The ultimate goal is to develop Bayesian methods to determine the thermal structures of solar and stellar coronae using spectral line intensity or photon count measurements that cover a large temperature interval and have sufficient density sensitivity to allow realistic density estimation.

6.1 Data and the astrophysical model

The He-like O VII is a popular line system for density measurements in the X-ray regime, but unlike the Fe XIII and the Fe XVII lines, its curves have both the density and the temperature dependencies. The O VII wavelength range does not overlap Fe XVII lines.

We use the same notations and equations for the data and the astrophysical spectrum model of Fe XVII analysis in Section 5.1.2 to the O VII analysis in this Chapter with an upper subscript ^{O}. Moreover, we treat $\log n$ as a random variable instead of a fixed constant as before. As both of the Fe XVII and the O VII datasets are observed from the same plasma, we could assume they share the same temperature.

6.2 O VII Statistical model

6.2.1 Combining Fe XVII and O VII analyses to estimate plasma parameters

We simplify the problem by abstracting the notations. Let $\theta^O = (r^O, \theta_B^O, \log \mathrm{n})$ and $\theta^{Fe} = (r^{Fe}, \theta_B^{Fe})$ be the parameters that only exist in the O VII and the Fe XVII analyses, and $\theta^{O,Fe} = (\log \alpha_0, \alpha_1)$ be the common parameters in both of the O VII and the Fe XVII analyses. The source and the background parameters are mingled together for simplicity.

Besides the emissivity ensemble obtained by the atomic physicists via a dataset $\mathcal{D}^{\text{atomic}}$, we preliminarily focus attention on Fe XVII spectral analysis, in Chapter 5, where the preliminary source of information for the temperature ($\alpha_1 = \log \mathrm{T}$) and volume ($\log \alpha_0$) is $\mathcal{D}^{Fe}$. We then take forward the temperature and volume information obtained, together with the atomic uncertainties of O VII, into the primary analysis regarding $\mathcal{D}^O$ to get the density information and to update the temperature information meanwhile. These can be abstracted as a special three-stage analysis and can then be summarized as below,

$$\text{Stage 1 (preliminary):} \quad \mathcal{D}^{\text{atomic}} \mid \text{atomic part of } \theta^O, \tag{6.1}$$

$$\text{Stage 2 (preliminary):} \quad \mathcal{D}^{Fe} \mid \theta^{Fe}, \theta^{O,Fe}, \tag{6.2}$$

$$\text{Stage 3 (primary):} \quad \mathcal{D}^{O} \mid \theta^{O}, \theta^{O,Fe}. \tag{6.3}$$

Similar to the two-stage analysis, we focus on the parameter estimation with respect to both of the datasets, $\mathcal{D}^{Fe}$ and $\mathcal{D}^O$, given the unavailability of the dataset $\mathcal{D}^{\text{atomic}}$, in this three-stage analysis.

The photon counts of different ions are observed independently from the telescope. Mathematically, it means the two datasets, $\mathcal{D}^O$ and $\mathcal{D}^{Fe}$, are conditionally independent given their own parameters. Thus, the likelihood function of $\mathcal{D}^O$ and $\mathcal{D}^{Fe}$ given θ^O, θ^{Fe}, and $\theta^{O,Fe}$ is

$$p(\mathcal{D}^O, \mathcal{D}^{Fe} \mid \theta^O, \theta^{Fe}, \theta^{O,Fe}) = p(\mathcal{D}^O \mid \theta^O, \theta^{O,Fe}) \cdot p(\mathcal{D}^{Fe} \mid \theta^{Fe}, \theta^{O,Fe}), \tag{6.4}$$

where $p(\mathcal{D}^O \mid \theta^O, \theta^{O,Fe})$ and $p(\mathcal{D}^{Fe} \mid \theta^{Fe}, \theta^{O,Fe})$ are the likelihood functions of the O VII and the Fe XVII spectral analyses, respectively, under Poisson distributions coming from Eq (5.1) and Eq (5.2).

Next, we assume all parameters are a priori independent, so that the joint priori distribution is

$$p(\theta^O, \theta^{Fe}, \theta^{O,Fe}) = p(\theta^O)p(\theta^{Fe})p(\theta^{O,Fe}). \tag{6.5}$$

Given the likelihood function in Eq (6.4) and the prior independence assumption in Eq (6.5), the joint posterior distribution for all the parameters given the whole datasets is

$$\begin{aligned} & p(\theta^O, \theta^{Fe}, \theta^{O,Fe} \mid \mathcal{D}^O, \mathcal{D}^{Fe}) \\ = & \frac{p(\mathcal{D}^O \mid \theta^O, \theta^{O,Fe}) \cdot p(\mathcal{D}^{Fe} \mid \theta^{Fe}, \theta^{O,Fe}) \cdot p(\theta^O)p(\theta^{Fe})p(\theta^{O,Fe})}{p(\mathcal{D}^O \mid \mathcal{D}^{Fe})\ p(\mathcal{D}^{Fe})}. \end{aligned} \tag{6.6}$$

Regrouping the terms of Eq (6.6), the joint posterior distribution can be rewritten as,

$$p(\theta^O, \theta^{Fe}, \theta^{O,Fe} \mid \mathcal{D}^O, \mathcal{D}^{Fe}) = \frac{p(\mathcal{D}^O \mid \theta^O, \theta^{O,Fe}) \cdot p(\theta^O)}{p(\mathcal{D}^O \mid \mathcal{D}^{Fe})} \cdot p(\theta^{Fe}, \theta^{O,Fe} \mid \mathcal{D}^{Fe}), \tag{6.7}$$

where $p(\theta^{Fe}, \theta^{O,Fe} \mid \mathcal{D}^{Fe})$ is the target posterior distribution of the Fe XVII spectral analysis in Stage 2 to infer $(\theta^{Fe}, \theta^{O,Fe})$, which has already been discussed in Chapter 5.

Moreover, there is a conditional independence that comes from the model, where the information for θ^O comes from $\mathcal{D}^O$ rather than $\mathcal{D}^{Fe}$. This can be

expressed mathematically as

$$\begin{aligned}
&p(\theta^O \mid \mathcal{D}^O, \mathcal{D}^{Fe}, \theta^{O,Fe}, \theta^{Fe}) \\
&= \frac{p(\theta^O, \mathcal{D}^{Fe} \mid \mathcal{D}^O, \theta^{O,Fe}, \theta^{Fe})}{p(\mathcal{D}^{Fe} \mid \mathcal{D}^O, \theta^{O,Fe}, \theta^{Fe})} \\
&= \frac{p(\mathcal{D}^{Fe} \mid \theta^O, \mathcal{D}^O, \theta^{O,Fe}, \theta^{Fe}) \cdot p(\theta^O \mid \mathcal{D}^O, \theta^{O,Fe}, \theta^{Fe})}{p(\mathcal{D}^{Fe} \mid \mathcal{D}^O, \theta^{O,Fe}, \theta^{Fe})} \\
&= p(\theta^O \mid \mathcal{D}^O, \theta^{O,Fe}, \theta^{Fe}), \qquad (6.8)
\end{aligned}$$

i.e., θ^O and $\mathcal{D}^{Fe}$ are independent. The last equality in Eq (6.8) is due to the fact that $\mathcal{D}^{Fe}$ does not depend on θ^O.

The inference for $(\theta^O, \theta^{O,Fe})$ is based on its marginal posterior distribution by integrating θ^{Fe} out from Eq (6.7),

$$p(\theta^O, \theta^{O,Fe} \mid \mathcal{D}^O, \mathcal{D}^{Fe}) = \frac{p(\mathcal{D}^O \mid \theta^O, \theta^{O,Fe}) \cdot p(\theta^O)}{p(\mathcal{D}^O \mid \mathcal{D}^{Fe})} \cdot p(\theta^{O,Fe} \mid \mathcal{D}^{Fe}). \qquad (6.9)$$

Overall there are three sources of uncertainties, atomic uncertainty in O VII emissivities and parameter uncertainty in Fe XVII model in the two preliminary analyses (Stage 1 and Stage 2), and parameter uncertainty in O VII model in the primary analysis (Stage 3). We have fit the Fe XVII spectral model as one of the preliminary analyses and have obtained an approximated distribution, $\hat{p}(\theta^{O,Fe} \mid \mathcal{D}^{Fe})$, for the marginal posterior distribution $p(\theta^{O,Fe} \mid \mathcal{D}^{Fe})$ in Section 5.7. We can then carry forward this approximated marginal posterior distribution, together with the O VII emissivity uncertainties, into the primary analysis which has target distribution $p(\theta^O, \theta^{O,Fe} \mid \mathcal{D}^O, \mathcal{D}^{Fe})$ in Eq (6.9) and θ^O is learned through the full datasets, $\mathcal{D}^{Fe}$ and $\mathcal{D}^O$, see more details in Section 6.2.2. That is, the approximated marginal posterior distribution from the preliminary analysis for Fe XVII is used as a prior distribution in the primary analysis for O VII. Four-step Gibbs Sampler and HMC algorithms are deployed to obtain sample of those parameters, see Section 6.4.

6.2.2 O VII STATISTICAL MODELS IN PRIMARY STAGE

Here we specify the detailed Bayesian model for the O VII spectral analysis in the primary stage.

We model the source and the background counts given the source and (or) the background parameters as independent Poisson distributions respectively, for each $\mathrm{w} \in \mathcal{W}^O$,

$$Y^O(\mathrm{w}) \mid \log \alpha_0, \alpha_1, r^O, \theta_B^O, \log \mathrm{n} \overset{\text{indep}}{\sim} \text{Poisson}\left(s^O(\mathrm{w}; \log \alpha_0, \alpha_1, r^O, \log \mathrm{n}) + \kappa^O(\mathrm{w}) + \theta_B^O\right), \tag{6.10}$$

$$Z^O(\mathrm{w}) \mid \theta_B^O \overset{\text{indep}}{\sim} \text{Poisson}\left(\eta \cdot \theta_B^O\right). \tag{6.11}$$

Thus, the likelihood function of $\mathcal{D}^O = \{(Y^O(\mathrm{w}), Z^O(\mathrm{w})), \mathrm{w} \in \mathcal{W}^O\}$ given $\log \alpha_0$, α_1, r^O, θ_B^O, and $\log \mathrm{n}$ is

$$\begin{aligned}
& p(\mathcal{D}^O \mid \theta^O, \theta^{O,Fe}) \\
&= L(\log \alpha_0, \alpha_1, r^O, \theta_B^O, \log \mathrm{n} \mid Y^O, Z^O) \\
&= p(Y^O, Z^O \mid \log \alpha_0, \alpha_1, r^O, \theta_B^O, \log \mathrm{n}) \\
&= p(Y^O \mid \log \alpha_0, \alpha_1, r^O, \theta_B^O, \log \mathrm{n}) \cdot p(Z^O \mid \theta_B^O) \\
&= \prod_{\mathrm{w} \in \mathcal{W}^O} p(Y^O(\mathrm{w}) \mid \log \alpha_0, \alpha_1, r^O, \theta_B^O, \log \mathrm{n}) \cdot p(Z^O(\mathrm{w}) \mid \theta_B^O).
\end{aligned} \tag{6.12}$$

Next, we specify the joint prior distribution on the unknown model parameters. A multivariate standard normal distribution is used as the prior distribution for r^O, a Gamma distribution for θ_B^O, and a continuous uniform distribution for $\log_{10} \mathrm{n}$,

$$r^O \sim \text{MVN}(0, I), \tag{6.13}$$

$$\theta_B^O \sim \text{Gamma}(\text{shape} = a_1, \text{rate} = a_2), \tag{6.14}$$

$$p(\log_{10} \mathrm{n}) = \frac{1}{7} \quad \text{for } 8 \le \log_{10} \mathrm{n} \le 13, \tag{6.15}$$

where $a_1 = 0.5$ and $a_2 = 2$. A Gamma distribution for θ_B^O is a conjugate prior for a Poisson likelihood. It covers all conceivable background ranges

encountered in the dataset.

The approximated Gaussian distribution and the approximated t-distribution obtained in Section 5.7 are used as the prior distribution for $(\log_{10}\alpha_0,\, \alpha_1)$,

$$\log\alpha_0, \alpha_1 \sim \hat{p}(\log\alpha_0, \alpha_1 \mid Y^{Fe}, Z^{Fe}), \tag{6.16}$$

Except $\log\alpha_0$ and α_1, all other parameters are a priori independent so that the joint prior distribution is

$$\begin{aligned} & p(\theta^O) \cdot p(\theta^{O,Fe} \mid \mathcal{D}^{Fe}) \\ =\, & p(r^O, \theta_B^O, \log \mathrm{n}) \cdot p(\log\alpha_0, \alpha_1 \mid Y^{Fe}, Z^{Fe}) \\ =\, & p(r^O)\; p(\theta_B^O)\; p(\log \mathrm{n})\; \hat{p}(\log\alpha_0, \alpha_1 \mid Y^{Fe}, Z^{Fe}). \end{aligned} \tag{6.17}$$

Given the likelihood function in Eq (6.12) and the prior distribution in Eq (6.17), and expanded from Eq (6.9), the joint posterior distribution in the primary stage is

$$\begin{aligned} & p(\log\alpha_0, \alpha_1, r^O, \theta_B^O, \log \mathrm{n} \mid Y^O, Z^O, Y^{Fe}, Z^{Fe}) \\ \propto\, & p(Y^O \mid \log\alpha_0, \alpha_1, r^O, \theta_B^O, \log \mathrm{n}) \cdot p(Z^O \mid \theta_B^O) \times \\ & \quad p(r^O) p(\theta_B^O) p(\log \mathrm{n}) \hat{p}(\log\alpha_0, \alpha_1 \mid Y^{Fe}, Z^{Fe}). \end{aligned} \tag{6.18}$$

Similar to the Fe XVII analysis, the observed source counts, $Y^O = \{Y^O(\mathrm{w})\}$, are implicitly made up of the counts from the source exposure due to the source only, $Y_S^O = \{Y_S^O(\mathrm{w})\}$, and the counts due to the background only, $Y_B^O = \{Y_B^O(\mathrm{w})\}$, i.e., $Y^O(\mathrm{w}) = Y_S^O(\mathrm{w}) + Y_B^O(\mathrm{w})$, for any $\mathrm{w} \in \mathcal{W}^O$. Treating them as missing data again, we assume $Y_S^O(\mathrm{w})$ and $Y_B^O(\mathrm{w})$ follow independent Poisson distributions with intensity $s^O(\mathrm{w}; \log\alpha_0, \alpha_1, r^O, \log \mathrm{n}) + \kappa^O(\mathrm{w})$ and θ_B^O for each $\mathrm{w} \in \mathcal{W}^O$,

$$\begin{aligned} Y_S^O(\mathrm{w}) \mid \log\alpha_0, \alpha_1, r^O, \log \mathrm{n} &\overset{\text{indep}}{\sim} \text{Poisson}\left(s^O(\mathrm{w}; \log\alpha_0, \alpha_1, r^O, \log \mathrm{n}) + \kappa^O(\mathrm{w})\right), \\ Y_B^O(\mathrm{w}) \mid \theta_B^O &\overset{\text{indep}}{\sim} \text{Poisson}\left(\theta_B^O\right). \end{aligned}$$

The missing data and all the model parameters can be estimated easily under

the above reformulated models. The conditional distribution of $Y_B^O(\mathrm{w})$ given all the model parameters and $Y^O(\mathrm{w})$ for any $\mathrm{w} \in \mathcal{W}^O$ can be computed using the Bayes' Theorem,

$$\begin{aligned}
&p(Y_B^O(\mathrm{w}) \mid Y^O(\mathrm{w}), \log \alpha_0, \alpha_1, r^O, \theta_B^O, \log \mathrm{n}) \\
&= \frac{p(Y^O(\mathrm{w}) \mid Y_B^O(\mathrm{w}), \log \alpha_0, \alpha_1, r^O, \theta_B^O, \log \mathrm{n}) \cdot p(Y_B^O(\mathrm{w}) \mid \theta_B^O)}{p(Y^O(\mathrm{w}) \mid \log \alpha_0, \alpha_1, r^O, \theta_B^O, \log \mathrm{n})} \\
&= \binom{Y^O(\mathrm{w})}{Y_B^O(\mathrm{w})} \cdot \Big(\frac{\theta_B^O}{s^O(\mathrm{w}; \log \alpha_0, \alpha_1, r^O, \log \mathrm{n}) + \kappa^O(\mathrm{w}) + \theta_B^O} \Big)^{Y_B^O(\mathrm{w})} \times \\
&\qquad \Big(\frac{s^O(\mathrm{w}; \log \alpha_0, \alpha_1, r^O, \log \mathrm{n}) + \kappa^O(\mathrm{w})}{s^O(\mathrm{w}; \log \alpha_0, \alpha_1, r^O, \log \mathrm{n}) + \kappa^O(\mathrm{w}) + \theta_B^O} \Big)^{Y^O(\mathrm{w}) - Y_B^O(\mathrm{w})}.
\end{aligned}$$

That is in the form of the probability mass function of a binomial distribution,

$$\begin{aligned}
&p(Y_B^O(\mathrm{w}) \mid Y^O(\mathrm{w}), \log \alpha_0, \alpha_1, r^O, \theta_B^O, \log \mathrm{n}) \\
&\overset{\text{indep}}{\sim} \text{Binomial}\big(Y^O(\mathrm{w}), \frac{\theta_B^O}{s^O(\mathrm{w}; \log \alpha_0, \alpha_1, r^O, \log \mathrm{n}) + \kappa^O(\mathrm{w}) + \theta_B^O}\big). \qquad (6.19)
\end{aligned}$$

Moreover, if θ_B^O follows a Gamma prior as in Eq (6.14), the marginal posterior of θ_B^O given Y_B^O and Z^O can also be computed using the Bayes' Theorem,

$$\begin{aligned}
p(\theta_B^O \mid Z^O, Y_B^O) &\propto \prod_{h=1}^{H} \big(p(Z^O(\mathrm{w}_h) \mid \theta_B^O) \cdot p(Y_B^O(\mathrm{w}_h) \mid \theta_B^O) \big) \; p(\theta_B^O) \\
&\propto e^{-\theta_B^O \left(H(\eta+1)+a_2 \right)} \; \theta_B^{O \left(\sum_{h=1}^H Z^O(\mathrm{w}) + \sum_{h=1}^H Y_B^O(\mathrm{w}) + a_1 - 1 \right)},
\end{aligned}$$

That is in the form of the probability density function of a Gamma distribution,

$$\theta_B^O \mid Z^O, Y_B^O \sim \text{Gamma}(\sum_{h=1}^{H} Z^O(\mathrm{w}) + \sum_{h=1}^{H} Y_B^O(\mathrm{w}) + a_1, H(\eta+1) + a_2). \qquad (6.20)$$

6.2.3 The effect of photon counts on emissivities in primary stage

Here we are going to consider the effect of photon counts on emissivities. According to the previous work in Chapter 4 and Chapter 5, whether the spectral observation is considered for estimating the emissivity parameters or not can influence the estimation of the plasma parameters significantly. In this section, referring back to the introduction on statistical methods in Section 3.2, we consider both the pragmatic Bayesian method where we do not learn about the emissivity choices from the information in the data and the fully Bayesian method where we do consider the potential information in the data to learn about the emissivities.

Pragmatic Bayesian method on emissivities

As described in Section 3.2, the pragmatic Bayesian method in the primary stage works under the assumption that the Fe XVII and the O VII photon counts do not have the information for narrowing the uncertainty of the O VII emissivities. Mathematically, this can be expressed as,

$$p(r^O \mid Y^O, Z^O, \mathcal{D}^{Fe}) = p(r^O) \tag{6.21}$$

i.e., r^O and $(Y^O, Z^O, \mathcal{D}^{Fe})$ are independent. Thus, the pragmatic Bayesian joint posterior distribution of r^O and all other parameters $Y_B^O, \log \alpha_0, \alpha_1, \theta_B^O$, $\log \mathrm{n}$ given (Y^O, Z^O) and $\mathcal{D}^{Fe}$ in the primary stage is

$$\begin{aligned} & p(Y_B^O, \log \alpha_0, \alpha_1, r^O, \theta_B^O, \log \mathrm{n} \mid Y^O, Z^O, \mathcal{D}^{Fe}) \\ &= p(Y_B^O, \log \alpha_0, \alpha_1, \theta_B^O, \log \mathrm{n} \mid r^O, Y^O, Z^O, \mathcal{D}^{Fe}) \cdot p(r^O \mid Y^O, Z^O, \mathcal{D}^{Fe}) \qquad (6.22) \\ &= p(Y_B^O, \log \alpha_0, \alpha_1, \theta_B^O, \log \mathrm{n} \mid r^O, Y^O, Z^O, \mathcal{D}^{Fe}) \cdot p(r^O). \qquad (6.23) \end{aligned}$$

Inference for Y_B^O is based on its conditional posterior distribution,

$$\begin{aligned} & p(Y_B^O \mid Y^O, Z^O, \mathcal{D}^{Fe}, \log \alpha_0, \alpha_1, r^O, \theta_B^O, \log \mathrm{n}) \\ = & \, p(Y^O \mid Y_B^O, \log \alpha_0, \alpha_1, r^O, \theta_B^O, \log \mathrm{n}) \; p(Y_B^O \mid \theta_B^O) \\ = & \, p(Y_B^O \mid Y^O, \log \alpha_0, \alpha_1, r^O, \theta_B^O, \log \mathrm{n}), \end{aligned} \tag{6.24}$$

following a binomial distribution as in Eq (6.19). Based on the assumption in Eq (6.21), inference for $\log \alpha_0, \alpha_1, \theta_B^O, \log \mathrm{n}$, and r^O is based on their conditional posterior or prior distributions,

$$\begin{aligned} & p(\log \alpha_0, \alpha_1, r^O, \theta_B^O, \log \mathrm{n} \mid Y^O, Z^O, Y_B^O, \mathcal{D}^{Fe}) \\ \propto & \, p(\theta_B^O \mid Z^O, Y_B^O) \; p(\log \alpha_0, \alpha_1, r^O, \log \mathrm{n} \mid Y_S^O, \mathcal{D}^{Fe}), \\ \propto & \, p(\theta_B^O \mid Z^O, Y_B^O) \; p(\log \alpha_0, \alpha_1, \log \mathrm{n} \mid Y_S^O, \mathcal{D}^{Fe}) \; p(r^O), \end{aligned} \tag{6.25}$$

where θ_B^O given Z^O and Y_B^O follows a Gamma distribution as in Eq (6.20) and $Y_S^O = Y^O - Y_B^O$.

The pragmatic Bayesian method incorporates atomic uncertainty in a conservative manner. The assumption in Eq (6.21) ignores the potential information in the Fe XVII and the O VII photon counts, $\mathcal{D}^{Fe}$ and $\mathcal{D}^O$, that may reduce uncertainty of atomic data represented by r^O and hence of those plasma parameters. We next consider methods that allow both $\mathcal{D}^{Fe}$ and $\mathcal{D}^O$ to be informative for r^O.

Fully Bayesian method on emissivities

In contrast to the pragmatic Bayesian method, the fully Bayesian method, as described in Section 3.2, eliminates the independence assumption in Eq (6.21) and incorporates the potential information in $\mathcal{D}^{Fe}$ and $\mathcal{D}^O$ to learn about r^O. The fully Bayesian joint posterior distribution of $\log \alpha_0, \alpha_1, r^O, \theta_B^O, \log \mathrm{n}$ given $\mathcal{D}^O$ and $\mathcal{D}^{Fe}$ in the primary stage is given in Eq (6.18), which is equivalent to integrating Y_B^O out from Eq (6.22).

The major difference between the pragmatic Bayesian method and the fully Bayesian method is if the information in the data is accounted for narrowing

the uncertainty of atomic data.

6.2.4 Alternative O VII statistical model with no information from Fe XVII analysis

We assume that the Fe XVII photon counts are uninformative to the estimation of the common parameters. That is, we do not take into account the information in the Fe XVII photon counts for narrowing the uncertainty of the common parameters. Mathematically, this assumption can be written as

$$p(\log \alpha_0, \alpha_1 \mid \mathcal{D}^{Fe}) = p(\log \alpha_0, \alpha_1) \tag{6.26}$$

i.e., $(\log \alpha_0, \alpha_1)$ and $\mathcal{D}^{Fe}$ are independent.

Instead of using the approximated distribution obtained from Fe XVII analysis, we can put two independent continuous uniform distributions as the prior distributions for $\log \alpha_0$ and α_1 individually,

$$p(\log_{10} \alpha_0) = \frac{1}{4} \qquad \text{for } 26 \leq \log_{10} \alpha_0 \leq 38, \tag{6.27}$$

$$p(\alpha_1) = \frac{1}{2.2} \qquad \text{for } 5.8 \leq \alpha_1 \leq 8, \tag{6.28}$$

which means no information from the preliminary analysis on Fe XVII will be carried forward into the primary analysis on O VII, i.e., Stage 2 in Section 6.2.1 is not implemented. A three-stage analysis degrades to a two-stage analysis. In this case, all the parameters are a priori independent so that the alternative joint prior distribution is

$$p(\theta^O, \theta^{O,Fe}) = p(r^O)\ p(\theta_B^O)\ p(\log \mathrm{n})\ p(\log \alpha_0)\ p(\alpha_1). \tag{6.29}$$

Given the likelihood function in Eq (6.12) and the alternative prior distribution in Eq (6.29), the alternative joint posterior distribution in the primary

stage becomes

$$
\begin{aligned}
& p(\log \alpha_0, \alpha_1, r^O, \theta_B^O, \log \mathrm{n} \mid Y^O, Z^O) \\
\propto\ & p(Y^O \mid \log \alpha_0, \alpha_1, r^O, \theta_B^O, \log \mathrm{n}) \cdot p(Z^O \mid \theta_B^O) \times \\
& p(r^O)\ p(\theta_B^O)\ p(\log \mathrm{n})\ p(\log \alpha_0)\ p(\alpha_1). \qquad (6.30)
\end{aligned}
$$

Due to the fact that there are common parameters in the last two stages, there are two main differences when comparing the joint posterior distribution in the primary stage in Eq (6.18) with its alternative one in Eq (6.30) under the assumption in Eq (6.26): 1) $\mathcal{D}^{Fe}$ becomes informative for $\theta^{O,Fe}$, and 2) $\mathcal{D}^{Fe}$ becomes informative for θ^O. The latter one is not necessary because the distribution of $\mathcal{D}^{Fe}$ does not depend on θ^O, as in Eq (6.10), Eq (6.11), and Eq (6.4), and, mathematically, it will not make any difference as well.

6.3 Gaussian prior on O VII emissivities with PCA

A Gaussian prior distribution via PCA, as discussed in Section 3.3.2, is used to compress the given ensemble of emissivities $\mathcal{M}^O$ for the seven O VII lines. We reconstruct emissivity replicates based on the first $J = 7$ principal components (out of $Q = 8211$) via Eq (3.18) and Eq (3.19) capturing 97% of the total variance, as computed with Eq (3.17). A significant amount of compression has been achieved because very few components are needed to compute the emissivity curve to a high precision. This approximation also achieves a better reconstruction in the $\log_e$ scale of the given emissivity ensemble, which is computationally efficient and is used throughout this chapter.

The complicated structure of the O VII emissivities, indicating the atomic uncertainty, is illustrated, one panel per line, in Figure 6.1 using the given ensemble of 10,000 O VII emissivity realisations. The complexity of the uncertainty of $\mathcal{M}^O$ is evident. We have generated 10,000 replicate emissivity curves, the exactly same amount of emissivity curve as in the original ensemble, using PCA compression in Eq (3.18) and Eq (3.19) with $J = 7$. The full range, the middle 95%, and the middle 68.3% intervals of these replicates is superimposed on the corresponding intervals for the original emissivity re-

alisations in $\log_e$ scale. The correspondence between the original emissivity realisations and the PCA replicates is quite good, especially for the 68.3% intervals.

6.4 Algorithms for the O VII analysis

We have considered two types of prior distributions for $(\log_{10}\alpha_0, \alpha_1)$, the approximated distributions, e.g., the approximated Gaussian distribution and the approximated t-distribution, and an alternative one with independent continuous uniform distributions individually, leading to two types of different posterior distributions, as described in Section 6.2. We have also considered the effect of photon counts on emissivities targeting on the posterior distributions with the former prior via the pragmatic and the fully Bayesian methods. Now we are working on the algorithms under the pragmatic and the fully Bayesian methods, and under the alternative model.

6.4.1 Algorithm for the pragmatic Bayesian method on emissivities

Under the assumption that the data in the primary stage carry little information as to the uncertainty of emissivities in the primary stage, we aim to construct MC sampler of those model parameters from the pragmatic Bayesian target posterior distribution in Eq (6.23) using a four-step Gibbs sampler. We iteratively update $\{Y_B^O(\mathrm{w})\}$, θ_B^O, r^O, and $(\log\alpha_0, \alpha_1, \log\mathrm{n})$ by sampling them from the corresponding conditional posterior distributions or prior distribution in Eq (6.24) and Eq (6.25).

At iteration ℓ, we first obtain an MC sample of $\{Y_B^O(\mathrm{w})^{(\ell)}\}$ from its conditional posterior distribution given the current states of all other parameters, $p(Y_B^O \mid Y^O, \theta_B^{O(\ell-1)}, r^{O(\ell-1)}, \log\alpha_0{}^{(\ell-1)}, \alpha_1^{(\ell-1)}, \log\mathrm{n}^{(\ell-1)})$, which follows a binomial distribution as in Eq (6.19). Secondly, given $\{Y_B^O(\mathrm{w})^{(\ell)}\}$, we can sample $\theta_B^{O(\ell)}$ from $p(\theta_B^O \mid Z^O, Y_B^{O(\ell)})$ following a Gamma distribution as in Eq (6.20). Thirdly, under the pragmatic Bayesian assumption, sampling of $r^{O(\ell)}$ is essentially random from the entire space, i.e., from its prior dis-

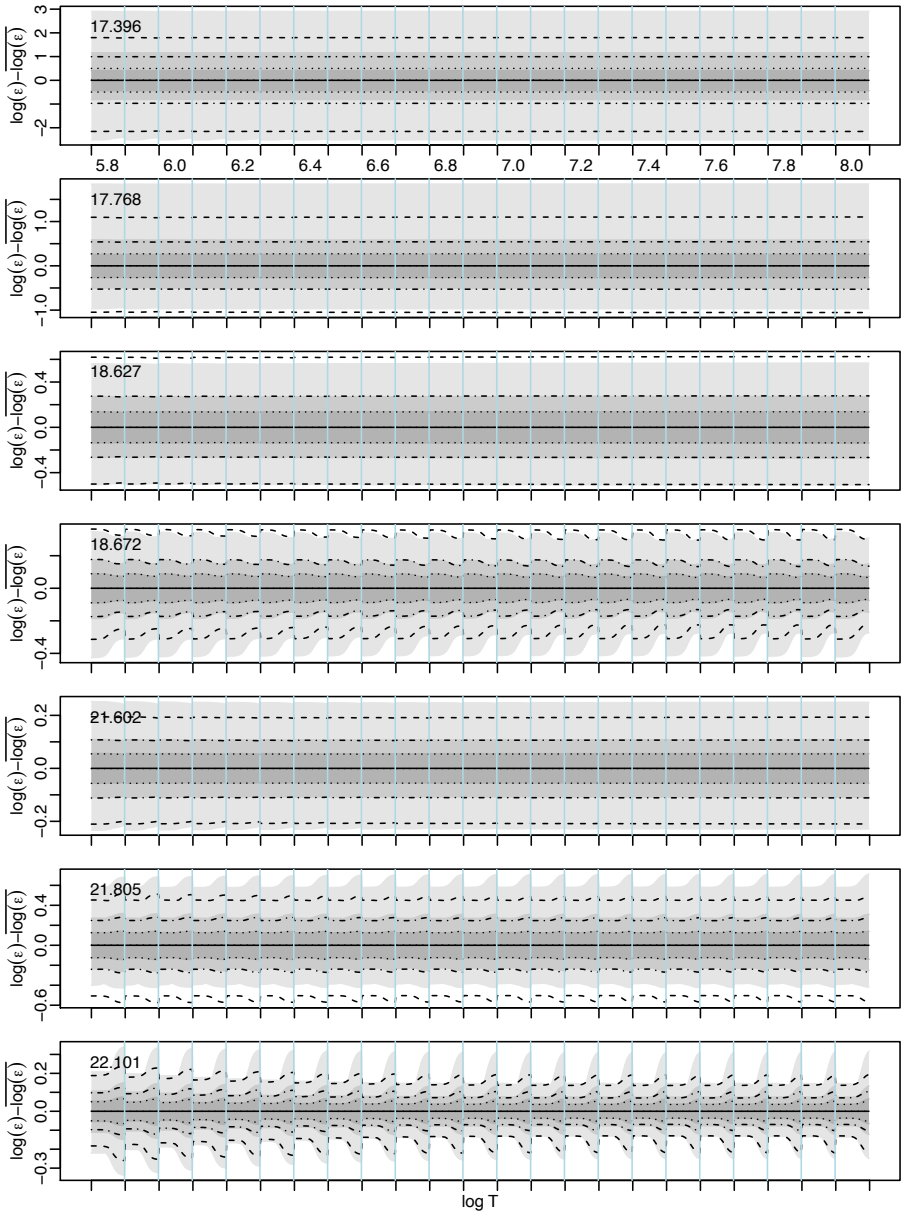

Figure 6.1: Summary of the emissivity samples of the seven O VII lines via PCA and summary of the PCA generated replicates in $\log_e$ scale. One panel for each line, the mean over all of the emissivity realisations, $\overline{\log}\,\epsilon$, is subtracted off to magnify the structure in $\mathcal{M}^O$. Zero line is plotted as a solid black curve. The light, dark, and darker grey areas cover the full range, the middle 95% and 68.3% of all $10,000$ emissivity curves in $\log_e$ scale. The dashed, dot-dash, and dotted lines respectively outline intervals containing the full range, the middle 95% and 68.3% of $10,000$ PCA generated replicates of the emissivity curves. In each panel, the horizontally-arranged sub-panels correspond to each value of the equally-spaced temperature grid, $\log \mathrm{T} \in (5.8, 8.0)$, and the temperature is increasing from left to right. Within each sub-panel, for a given temperature, the density, $\log \mathrm{n} \in (8.0, 13.0)$, is increasing.

tribution in Eq (6.13). Fourthly, given $Y_S^{O(\ell)} = Y^O - Y_B^{O(\ell)}$, we can sample $(\log \alpha_0{}^{(\ell)}, \alpha_1^{(\ell)}, \log \mathrm{n}^{(\ell)})$ from $p(\log \alpha_0, \alpha_1, \log \mathrm{n} \mid Y_S^{O(\ell)}, r^{O(\ell)}, \log \alpha_0{}^{(\ell-1)}, \alpha_1^{(\ell-1)}, \log \mathrm{n}^{(\ell-1)})$ by using the adaptive Metropolis algorithm, as discussed in Section 2.2.2, on a three-dimension space. A new proposal distribution is required to be specified on the fly,

$$
\begin{aligned}
&q(\log \alpha_0, \alpha_1, \log \mathrm{n} \mid \log \alpha_0{}^{(\ell-1)}, \alpha_1^{(\ell-1)}, \log \mathrm{n}^{(\ell-1)}) = \\
&\begin{cases}
\mathcal{MVN}\Big((\log \alpha_0, \alpha_1, \log \mathrm{n}) \mid (\log \alpha_0{}^{(\ell-1)}, \alpha_1^{(\ell-1)}, \log \mathrm{n}^{(\ell-1)}), \frac{0.1^2}{d}\Sigma_0\Big), \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \ell \le L', \\
(1-\beta)\cdot \mathcal{MVN}\Big((\log \alpha_0, \alpha_1, \log \mathrm{n}) \mid (\log \alpha_0{}^{(\ell-1)}, \alpha_1^{(\ell-1)}, \log \mathrm{n}^{(\ell-1)}), \\
\qquad\qquad \frac{2.38^2}{d}\Sigma^{(\ell)}\Big) + \\
\qquad \beta \cdot \mathcal{MVN}\Big((\log \alpha_0, \alpha_1, \log \mathrm{n}) \mid (\log \alpha_0{}^{(\ell-1)}, \alpha_1^{(\ell-1)}, \log \mathrm{n}^{(\ell-1)}), \\
\qquad\qquad \frac{0.1^2}{d}\Sigma_0\Big), \qquad\qquad\qquad\qquad\qquad\qquad \text{if } \ell > L',
\end{cases}
\end{aligned}
\tag{6.31}
$$

when $\ell > L'$, $\Sigma^{(\ell)}$ is set to a variance-covariance matrix based on $\big(\{\log \alpha_0{}^{(\cdot)}\}_1^{\ell-1}, \{\alpha_1^{(\cdot)}\}_1^{\ell-1}, \{\log \mathrm{n}^{(\cdot)}\}_1^{\ell-1}\big)$ if ℓ is a multiple of $\mathcal{L}$, set to $\Sigma^{(\ell-1)}$, otherwise; and Σ_0 is a diagonal matrix with prior variances of $\log \alpha_0$, α_1, and $\log \mathrm{n}$ as diagonal entries.

The detailed four-step MC sampler under the pragmatic Bayesian model proceeds for iteration $\ell = 1, \ldots, L$ with

Step 1: Sample $Y_B^O(\mathrm{w})^{(\ell)} \mid Y^O(\mathrm{w}), \log \alpha_0{}^{(\ell-1)}, \alpha_1^{(\ell-1)}, \theta_B^{O(\ell-1)}, r^{O(\ell-1)}, \log \mathrm{n}^{(\ell-1)}$,
$\sim$ Binomial $\big(Y^O(\mathrm{w}), \frac{\theta_B^{O(\ell-1)}}{\theta_B^{O(\ell-1)} + \lambda_{SC}(\mathrm{w}; \log \alpha_0{}^{(\ell-1)}, \alpha_1^{(\ell-1)}, r^{O(\ell-1)}, \log \mathrm{n}^{(\ell-1)})}\big)$
for each $\mathrm{w} \in \mathcal{W}^O$.

Step 2: Sample $\theta_B^{O(\ell)} \mid Z^O, Y_B^{O(\ell)}$
$\sim$ Gamma $\big(\sum_{j=1}^M Z^O(\mathrm{w}_j) + \sum_{j=1}^M Y_B^O(\mathrm{w}_j)^{(\ell)} + a_1, M(\eta+1) + a_2\big)$,
where $\theta_B^O \sim$ Gamma(a_1, a_2) a conjugate prior.

Step 3: Sample $r^{O(\ell)} \sim \mathcal{MVN}(0, I)$.

Step 4: Sample $(\log \alpha_0{}^{[\mathrm{prop}]}, \alpha_1^{[\mathrm{prop}]}, \log \mathrm{n}^{[\mathrm{prop}]})$ from
$q(\log \alpha_0, \alpha_1, \log \mathrm{n} \mid \log \alpha_0{}^{(\ell-1)}, \alpha_1^{(\ell-1)}, \log \mathrm{n}^{(\ell-1)})$ in Eq (6.31),

compute

$$\rho = \frac{p(\log {\alpha_0}^{[\text{prop}]}, \alpha_1^{[\text{prop}]}, \log \text{n}^{[\text{prop}]} \mid Y_S^{Fe(\ell)})}{p(\log {\alpha_0}^{(\ell-1)}, \alpha_1^{(\ell-1)}, \log \text{n}^{(\ell-1)} \mid Y_S^{Fe(\ell)})}, \tag{6.32}$$

and set

$$(\log {\alpha_0}^{(\ell)}, \alpha_1^{(\ell)}, \log \text{n}^{(\ell)}) = \begin{cases} (\log {\alpha_0}^{[\text{prop}]}, \alpha_1^{[\text{prop}]}, \log \text{n}^{[\text{prop}]}), & \text{with probability } \min(\rho, 1), \\ (\log {\alpha_0}^{(\ell-1)}, \alpha_1^{(\ell-1)}, \log \text{n}^{(\ell-1)}), & \text{otherwise.} \end{cases} \tag{6.33}$$

6.4.2 Algorithm for the fully Bayesian method on emissivities

We construct an MC sampler of those model parameters from the fully Bayesian target posterior distribution in Eq (6.18) with the likelihood function in Eq (6.12) the prior in Eq (6.17). We can then obtain sample of those model parameters from this higher dimension posterior via HMC by sampling directly from their joint posterior distribution.

6.4.3 Algorithm for the alternative model with no information from the preliminary analysis

The alternative target posterior distribution in Eq (6.30), with the likelihood function in Eq (6.12) the prior independence assumption in Eq (6.29), is used to construct an MC sampler of those model parameters. Sample of those model parameters can then be obtained via HMC by sampling directly from this joint posterior distribution.

6.5 Application to O VII simulation studies

We have proved that our programming is able to recover physical parameters from the Fe XVII photon counts in the preliminary stage via a simulation

study in Section 5.5, where a t-distribution prior of $(\log \alpha_0, \alpha_1)$ can be statistically approximated from the posterior sample based on each simulated Fe XVII dataset. Now, we generate sets of the source and the background counts for O VII from reasonable and known values of all the model parameters. Specifically, the values for the temperature and the volume are assumed to be the same in both stages. Simulating and fitting under the same model will allow us to test the ability of our programming to recover physical parameters from both the Fe XVII and the O VII photon counts in a three-stage analysis.

We assume values of $\log \alpha_0 = 31.17$, $\alpha_1 = 6.75$, $\log \mathrm{n} = 9.43$, $\theta_B^O = 0.93$, and $J = 6$ principle components $r^O = (-1.18, 0.61, 2.36, 0.77, 2.92, -0.77)$, capturing 97% of the total variance of the emissivities for six O VII lines as computed with Eq (3.17), are used to construct a new emissivity via a simple linear combination in Eq (3.18). Those values are obtained from an initial fit to explore the reasonable range of those parameters. The values for the former two are exactly the same as those in the Fe XVII simulation studies in the preliminary stage. In this setting, we obtain the expected values of the Poisson parameter in Eq (6.10) and Eq (6.11). We simulate 30 sets of source and background O VII photon counts for each channel from the Poisson distributions in Eq (6.10) and Eq (6.11) independently. The t-distribution approximation to $p(\log \alpha_0, \alpha_1 \mid Y^{Fe}, Z^{Fe})$ from the posterior samples of $\log \alpha_0$ and α_1 given each simulated Fe XVII dataset in the preliminary stage in Section 5.5 is used as a prior distribution for $(\log \alpha_0, \alpha_1)$ in the primary stage together with each simulated O VII dataset.

We use the Stan software package to obtain sample of those model parameters via HMC by sampling directly from their joint posterior distribution in Eq (6.18) given each of the above simulated sets of O VII photon counts. For each simulated set of photon counts, there are 4 chains sampled 4000 iterations each starting with random initials and the first 3000 iterations are discarded as burn-in. To make sure that the convergence is achieved, multiple chains are applied.

Figure 6.2 shows the posterior means, the 68% and the 95% credible intervals of the posterior samples for each parameter over the 30 simulated O VII

datasets, compared with the true values of the corresponding parameters. We note that almost all the estimates of those parameters are considerably biased. Considering more data might be helpful to reduce this bias.

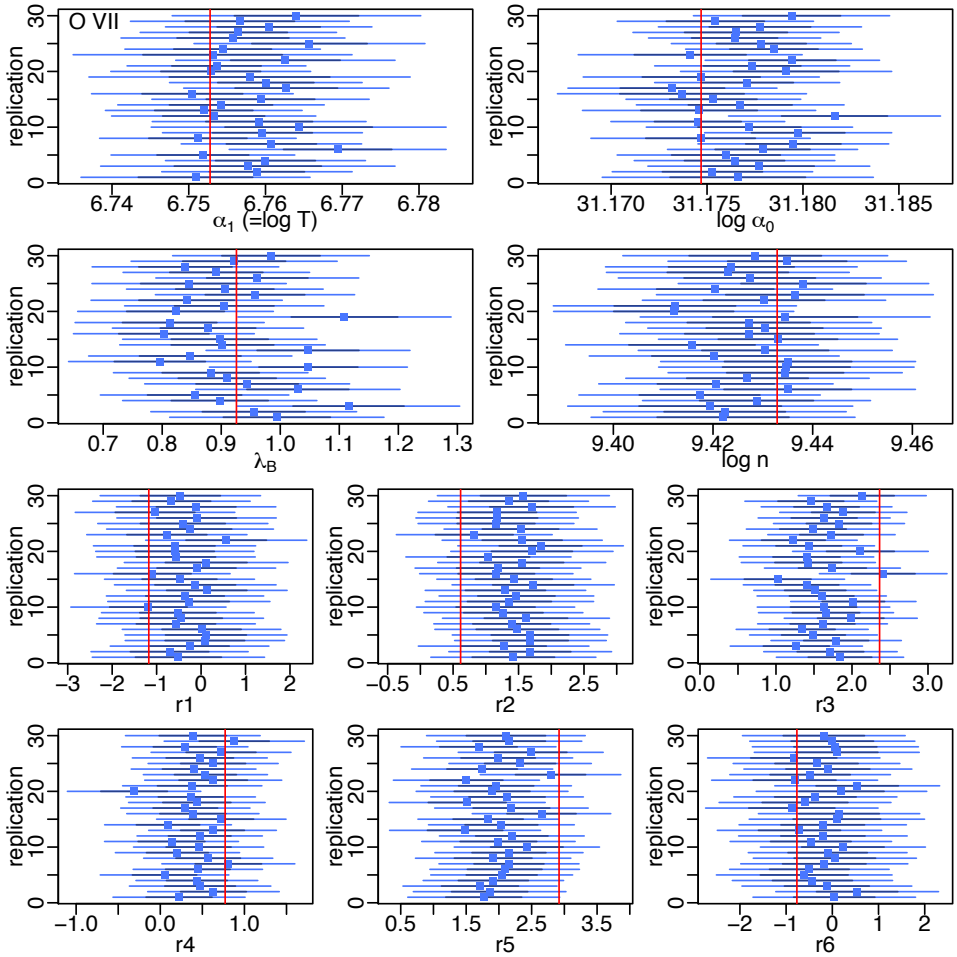

Figure 6.2: Summary of the posterior samples for the model parameters in the primary O VII simulation study based on the observed exposure. For each parameter, there are posterior means (blue square dots), the 68% (dark blue horizontal lines) and the 95% (light blue horizontal lines) credible intervals over each of the 30 simulated datasets. They are compared with the true values of the corresponding parameters, marked as red vertical lines.

We simulate more data by simply considering multiples of the original observed exposure time. We do a complete replication and consider the same multiples of the observed exposure time for both the Fe XVII preliminary and the O VII primary simulation studies. Specifically, with the true values of

the parameters, we compute the expected counts in each channel as before, multiply them with a certain number, and sample the dataset from Poisson distributions with these multipled expected values independently.

We simulate another 30 simulated O VII datasets based on 75 multiples of the observed exposure and fit the models again. Figure 6.3 is constructed in the same ways and same scales as Figure 6.2 with the new fitted results. It shows the estimates of those parameters are shifted and are less biased, and the corresponding credible intervals get smaller compared with that from the original observed exposure time. These credible intervals of the posterior samples are proved to be confidence intervals because they have certain coverage in this simulation study. The true values of the those parameters can also be recovered more accurately. Besides the recovery, those histograms in Figure 6.4 show the shape of the posterior distributions of those model parameters are more like Gaussian. The expected photon counts in the source exposure, $s^O(\mathrm{w}; \log \alpha_0, \alpha_1, r^O, \log \mathrm{n}) + \kappa^O(\mathrm{w}) + \theta_B^O$, for each $\mathrm{w} \in \mathcal{W}^O$, can be evaluated from the posterior means of samples for those model parameters for each simulated dataset. The profile of the expected photon counts match that of the simulated counts perfectly as in Figure 6.5. Almost 95% of the standardised residuals of the expected photon counts, in Figure 6.6, fall in $(-2, 2)$ without any dependency. The standardised residuals look like independent white noise as in Figure 6.7. All these indicate the model can fit the data very well.

We have considered several larger simulated datasets based on $5, 10, 25, 50,$ and 75 multiples of the observed exposure, and each of them has 30 replicates. When fitting the models with the datasets for different multiples of the observed exposure, we start the chains at same initial values of the parameters and the samples all together migrate over to the same direction no matter how large the simulated data sets are. This indicates the convergence is achieved as well.

Figure 6.8 show the bias of the average of the posterior means over all of the simulated replicates and the corresponding average standard deviations in frequency versus the multiples of the observed exposure for all the parameters in the O VII primary simulation study. The bias of the parameter estimates

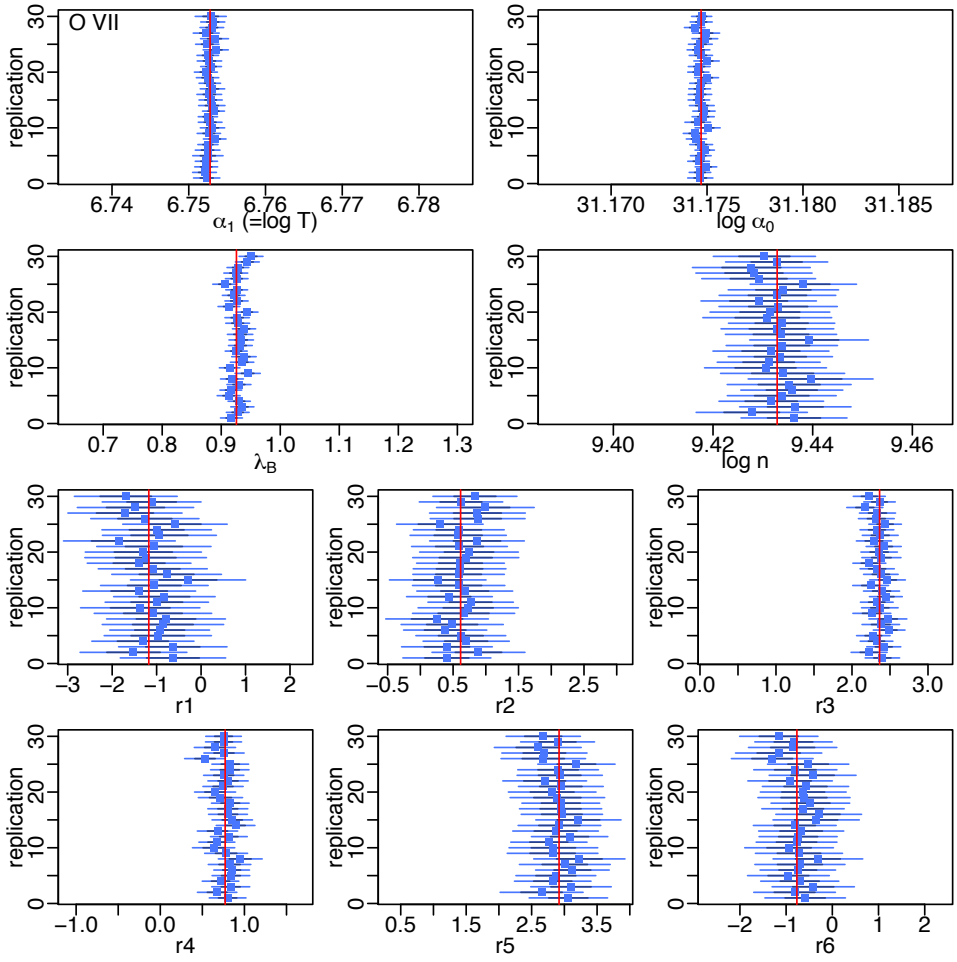

Figure 6.3: Summary of the posterior samples for the model parameters in the primary O VII simulation study based on 75 multiples of the observed exposure. The posterior means (blue square dots), the 68% (dark blue horizontal lines) and the 95% (light blue horizontal lines) credible intervals over each of the 30 simulated datasets for each parameter are plotted in the same scale as in Figure 6.2 for the purpose of comparison. The true values of the corresponding parameters are marked as red vertical lines.

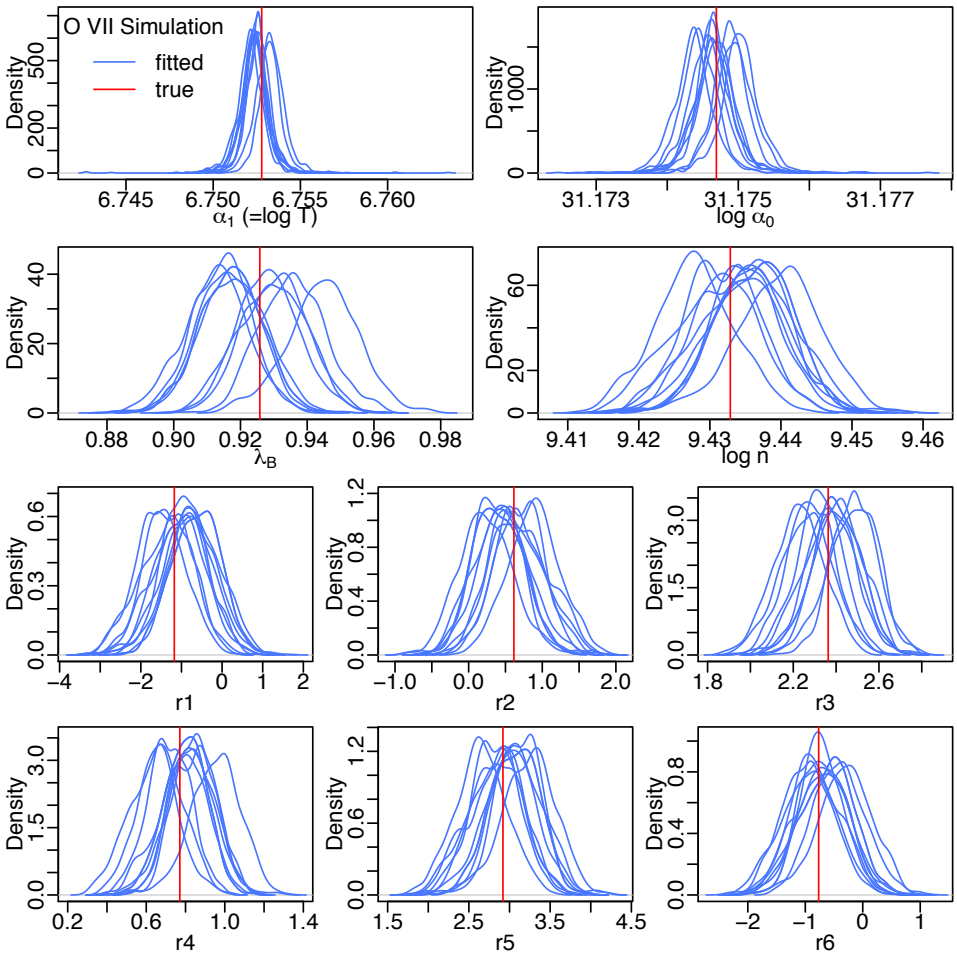

Figure 6.4: Density plots for the posterior samples of the model parameters across different simulated datasets in the primary O VII simulation study based on 75 multiples of the observed exposure. They all are compared with the corresponding true values of the parameters marked as red vertical lines.

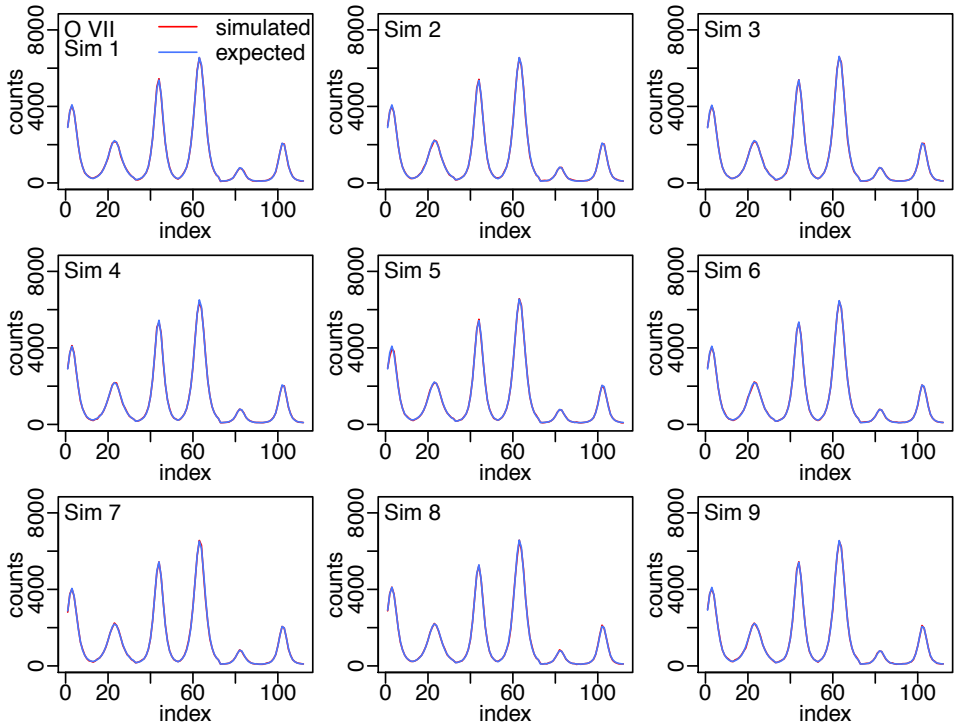

Figure 6.5: The comparison of the expected photon counts, in blue, and the simulated counts, in red, along the indices of the filtered wavelength for the first 9 replicates in the primary O VII simulation study based on 75 multiples of the observed exposure.

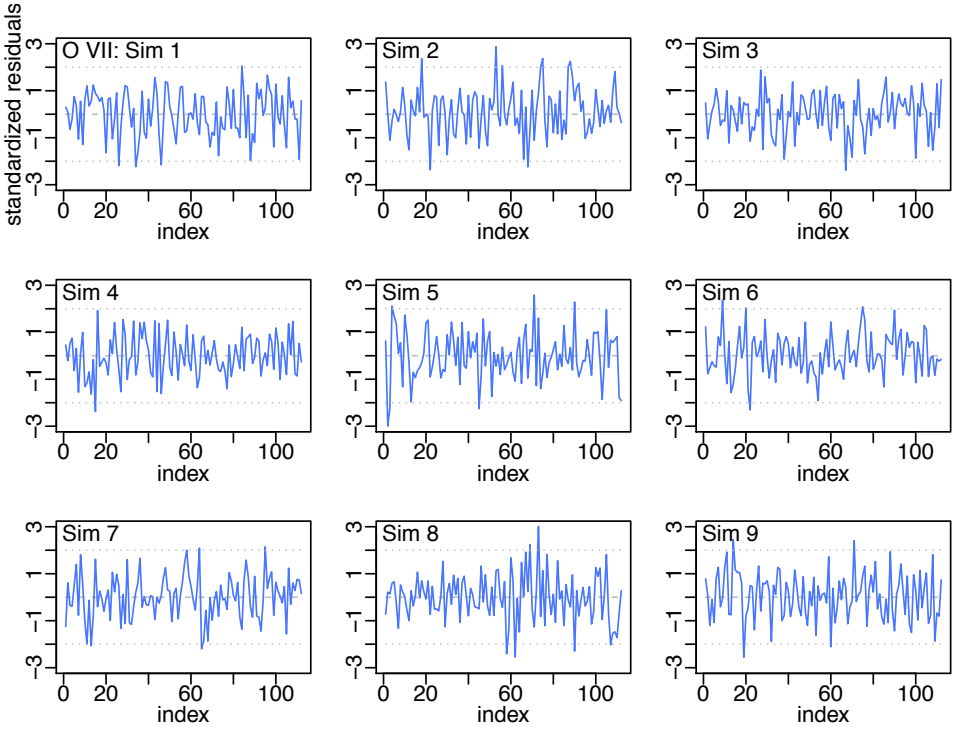

Figure 6.6: The standardised residuals of the expected source photon counts, along the indices of the filtered wavelength, for the first 9 simulated datasets in the primary O VII simulation study based on 75 multiples of the observed exposure. The standardised residual is the ratio of the difference between the simulated counts and the expected counts to the square root of the expected counts since the counts follow Poisson distributions.

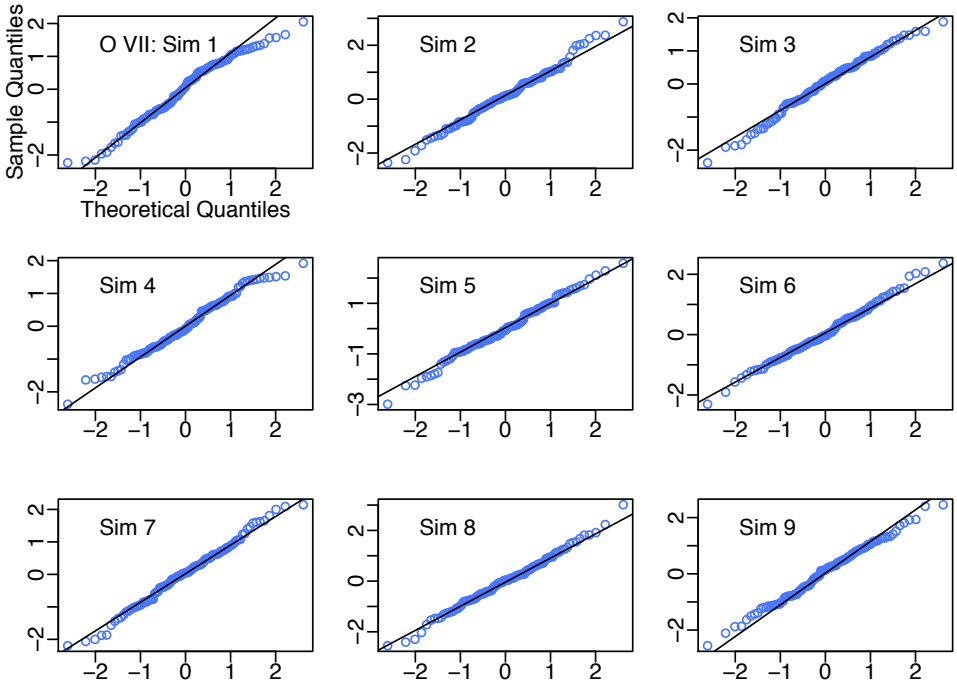

Figure 6.7: Comparison of the quantiles of the standardised residuals of the expected source photon counts versus a standard normal distribution. The results for the first 9 simulated datasets in the primary O VII simulation study based on 75 multiples of the observed exposure are presented. An identity line is added as a benchmark. The standardised residuals fall randomly around the identity lines indicating they are white noise.

decreases and the corresponding average standard deviation gets smaller as the sample size increases. The standard deviation for the bias estimate gets smaller as we get more data as well. The error starts to dominate the bias if we include about 10 or more multiples of the observed exposure for the O VII primary simulation study, i.e., 10 times more data.

Figure 6.9 show the empirical coverage rates of the 68% and the 95% confidence intervals of the posterior samples and the corresponding error bars versus the multiples of the observed exposure for all the parameters in the O VII primary simulation study. The number of 68% or 95% confidence intervals of the posterior samples covering the true values follows a Binomial distribution. The exact Binomial confidence interval is used as the error bar since it is feasible when the probability of success is equal to 0 or 1 (Clopper & Pearson 1934). The empirical coverage rates and the corresponding estimated intervals start to approach the nominal levels, and then tend to overestimate the uncertainty of the parameters, as we incorporate more data. This overestimation happens when we have about 75 times more data than the original size.

The same trends apply to the parameters in the Fe XVII preliminary simulation study, as in Figure 6.10 and Figure 6.11. The error starts to dominate the bias if we include about 5 or more multiples of observed exposure for the Fe XVII preliminary simulation study.

In summary, the estimated biases are decreasing and the parameter estimates are better determined if we get more data, while the behaviour of coverage rate for the confidence intervals is conservative if the size of the data is too large. The results indicate we have to get about five to ten times more data to make reliable statements about atomic uncertainty.

6.6 Application to O VII observed counts

We have an observed dataset from 'amp' again, which comprises sets of observed counts from the source and the background exposures in 1200 channels. We also have a collection of $M = 10000$ emissivity realisations for seven

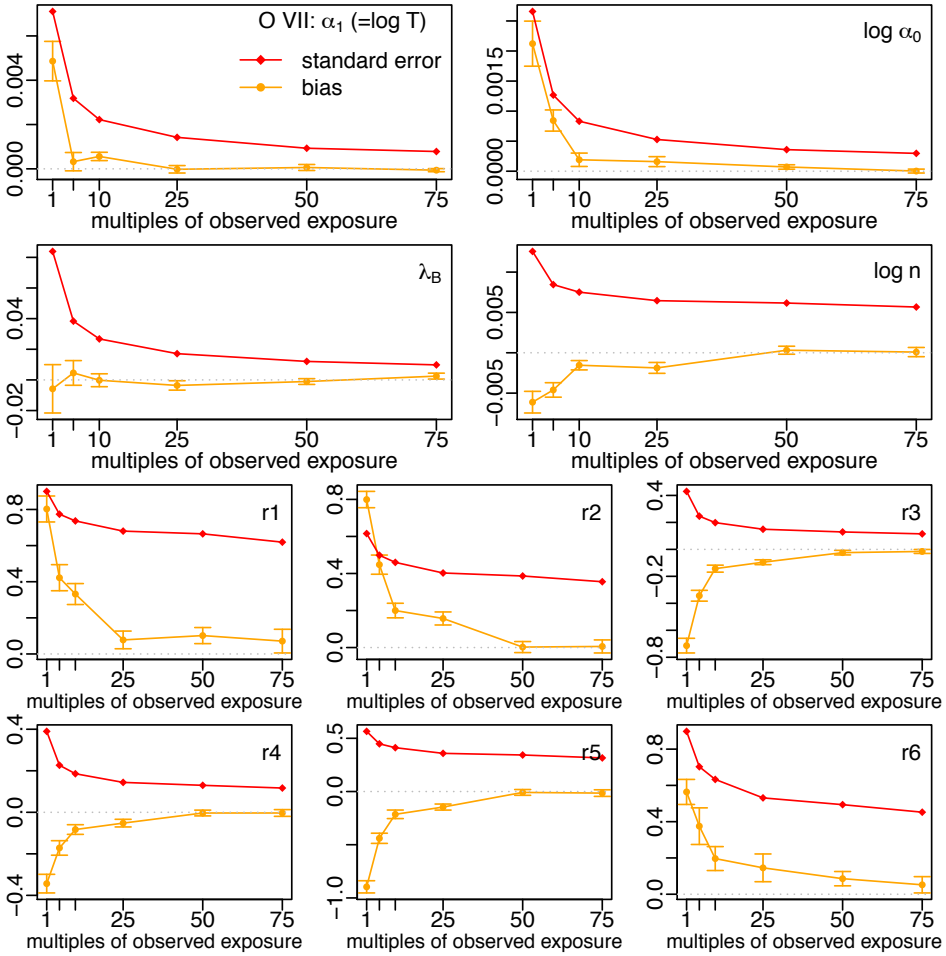

Figure 6.8: The bias of the average of the posterior means over all of the simulated replicates, in orange dots, and the corresponding average standard deviations in frequency, in red squares, versus the multiples of observed exposure, $(1, 5, 10, 25, 50, 75)$, for all the parameters in the O VII primary simulation study. The standard deviations for the bias estimates are marked as orange vertical intervals.

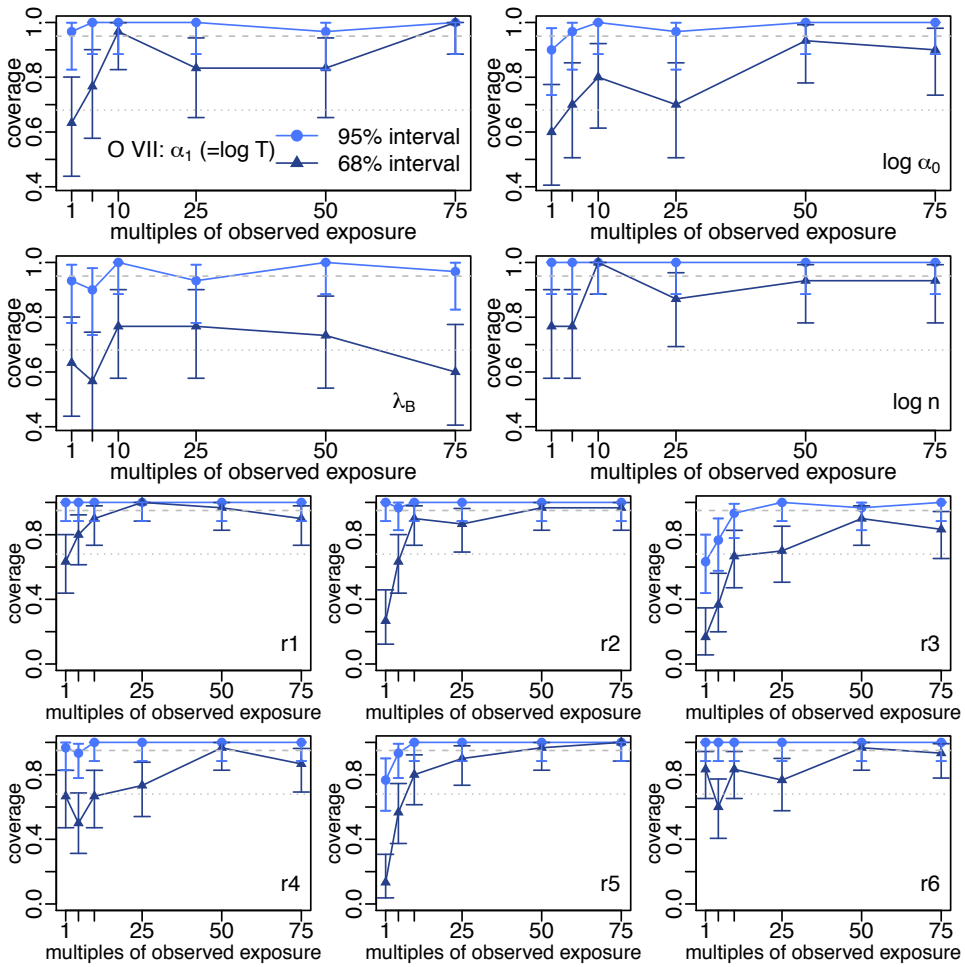

Figure 6.9: The empirical coverage rates of the 68% confidence intervals, in dark blue triangle, and the 95% confidence intervals, in light blue circle, of the posterior samples and the corresponding estimated intervals versus the multiples of observed exposure, $(1, 5, 10, 25, 50, 75)$, for all the parameters in the O VII primary simulation study.

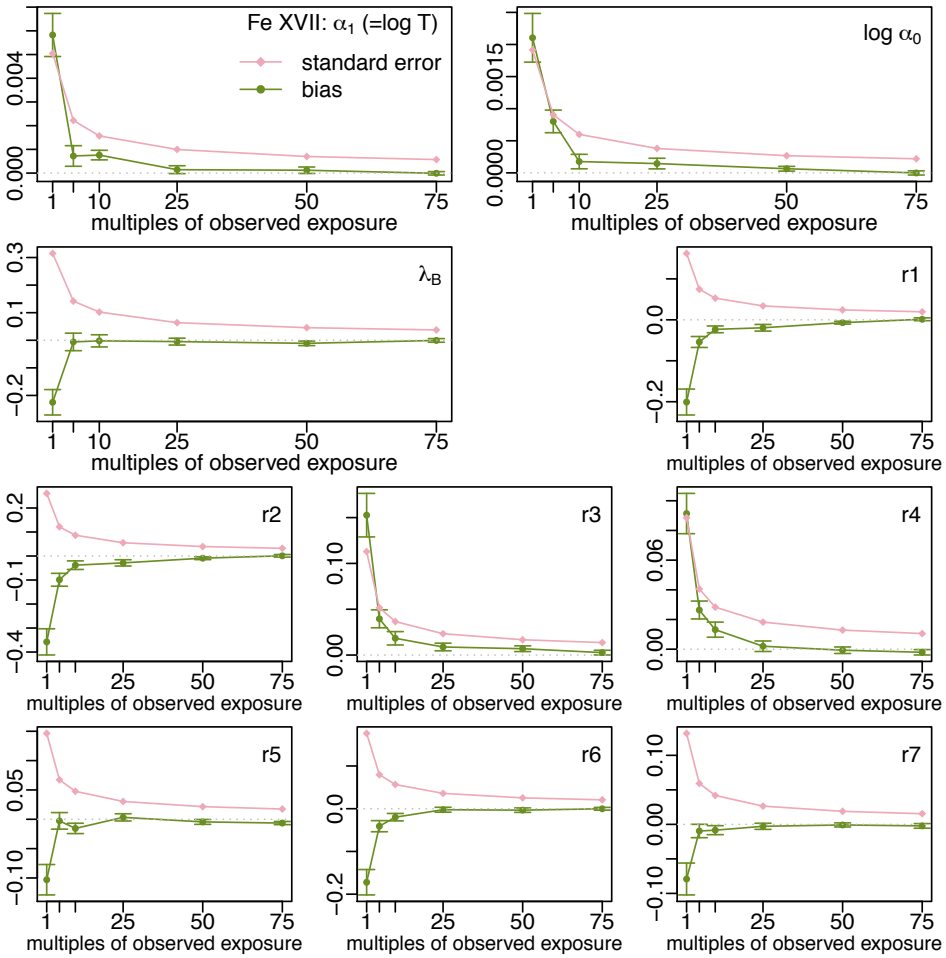

Figure 6.10: The bias of the average of the posterior means over all of the simulated replicates, in green dots, and the corresponding average standard deviations in frequency, in pink squares, versus the multiples of observed exposure, $(1, 5, 10, 25, 50, 75)$, for all the parameters in the Fe XVII preliminary simulation study. The standard deviations for the bias estimates are marked as green vertical intervals.

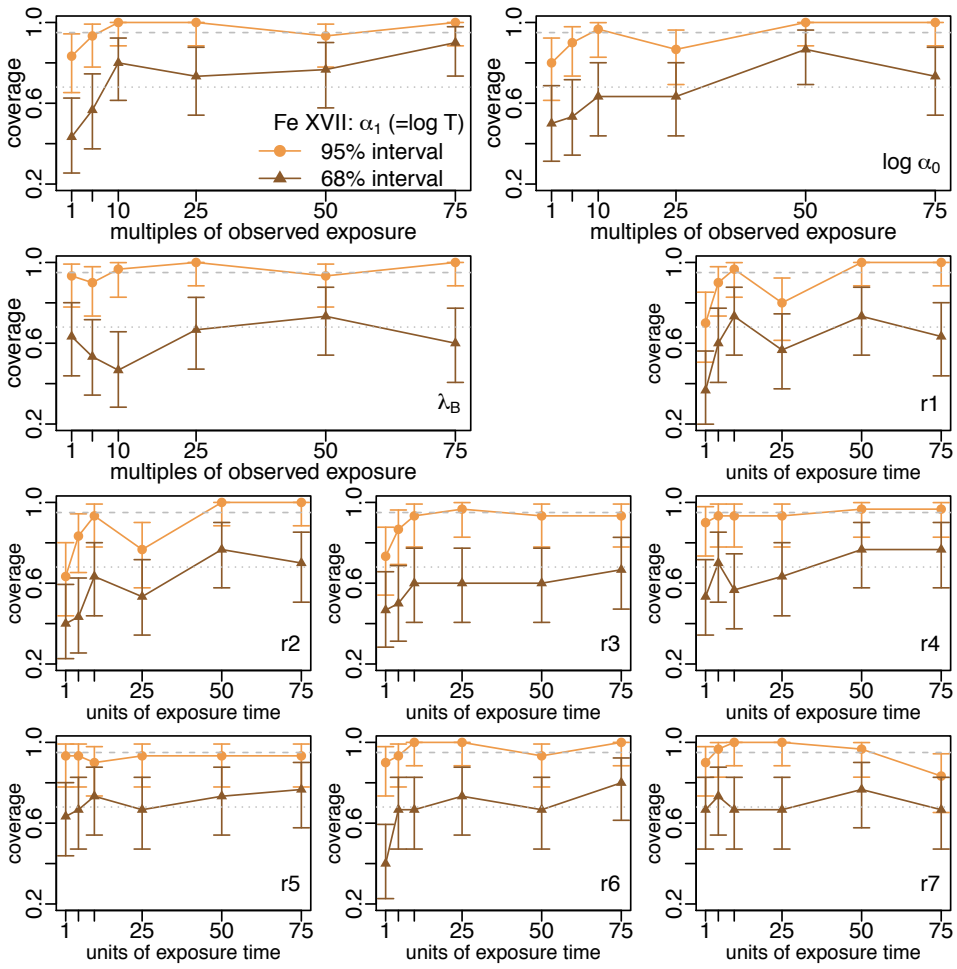

Figure 6.11: The empirical coverage rates of the 68% confidence intervals, in dark brown triangle, and the 95% confidence intervals, in light brown circle, of the posterior samples and the corresponding estimated intervals versus the multiples of observed exposure, $(1, 5, 10, 25, 50, 75)$, for all the parameters in the Fe XVII preliminary simulation study.

ω (Å)	$\hat{\sigma}$	$\hat{\omega}$ (Å)
17.396	0.0137	17.3969
17.768	0.0189	17.7736
18.627	0.0108	18.6319
21.602	0.0115	21.6030
21.805	0.0100	21.8045
22.101	0.0100	22.0995

Table 6.1: A summary of the best-fit width, $\hat{\sigma}$, and the adjusted centres, $\hat{\omega}$, for the O VII LRF.

spectral lines of O VII.

Similar preprocessed steps, including filtering wavelength, removing the contaminated lines, and tuning the LRF as in Section 5.6.1, need to be implemented on the O VII dataset before model fitting. A subset of channels based on the lines of interest are selected carefully, typically the adjacent 20 channels on both sides of a line and preventing any possible contaminations from other strong lines. There also could be gaps in $\mathcal{W}^O$. There is a strong line at 17.36 Å from the Fe XVIII mixed with the O VII 17.40 Å lines. We therefore need to ignore the channels in that area. There is another strong line from Ca XVIII at 18.69 Å which would be blending the O VII 18.67 Å resulting in the emissivity of line 18.67 Å is about a factor of 4 weaker than that of the other lines. Therefore, we decide to ignore this contaminated 18.67 Å line in the O VII analysis. As a result, there are in total only $H = 112$ sub-channels considered for $L = 6$ lines in this O VII analysis. We also need to fit width and centres of a t-distribution LRF to each individual lines, which are chosen from their corresponding fine grids by minimizing the RSS of the expected source photon counts, to get spectrum profiles line up with the data perfectly. Moreover, we set a boundary for the width in LRFs of at least 0.01 to match the fact that the spectra cannot be too narrow. A summary of the best-fit width and the adjusted centres is shown in Table 6.1 and is used throughout this section.

Now, we are ready to demonstrate the effects of the different types of models, the pragmatic and the fully Bayesian methods on emissivities with informa-

tion from the preliminary analysis and the alternative model with no information from the preliminary analysis, mentioned in Section 6.2, by applying them to this preprocessed datasets.

Four-step MC Sampler for the pragmatic Bayesian method on emissivities

We implement the four-step Gibbs Sampler for the pragmatic Bayesian method on emissivities. There are 1000000 MC iterations drawn from the joint posterior distribution in Eq (6.23) considering a t-distribution prior on $(\log \alpha_0, \alpha_1)$ where the first $L' = 1000$ are used to trigger the sampler and the following first half is discarded as burn-in. The first 6 principle components (out of 7038 features) are considered accounting for about 97% of the total variability of the given emissivities. Specifically, in Step 4, the adaptive Metropolis algorithm is used to sample $(\log \alpha_0, \alpha_1, \log \mathrm{n})$ from $p(\log \alpha_0, \alpha_1, \log \mathrm{n} \mid Y_S^O)$. The empirical estimate of the variance-covariance matrix is updated at every $\mathcal{L} = 50$ iterations to prevent unnecessary computing. A small $\beta = 0.05$ is used in the mixture proposal distribution. The prior variance matrix of the variables is $\Sigma_0 = \mathrm{diag}(\mathrm{var}(\log \alpha_0), \mathrm{var}(\alpha_1), \mathrm{var}(\log \mathrm{n}))$.

HMC for the fully Bayesian method on emissivities

Using the same 6 principle components, we implement the Stan software package to obtain sample of those model parameters via HMC by sampling directly from their joint posterior distribution in Eq (6.18) considering both a Gaussian prior, called the *Gaussian prior case*, and a t-distribution prior, called the *t-distribution prior case*, on $(\log \alpha_0, \alpha_1)$. For each case, there are 4 chains running, 4000 iterations sampled each, and the first 2000 iterations of each chain are discarded as burn-in. Different starting values are used for different chains making sure the starting values are located across the supports. Multiple chains and different starting values are applied to make sure that the convergence is achieved.

HMC for the alternative model with no information from the preliminary analysis

Using the same 6 principle components, we implement the Stan software package to obtain sample of those model parameters via HMC by sampling directly from their alternative joint posterior distribution in Eq (6.30) by considering two independent uniform distributions on $(\log \alpha_0, \alpha_1)$, called the *uniform prior case*. There are 4 chains running, 4000 iterations sampled each, and the first 2000 iterations of each chain are discarded as burn-in. Similarly, different starting values are used for different chains making sure the starting values are located across the supports.

Output data analysis on t-distribution prior case

We start with the t distribution prior case. The comparison of the posterior samples of those model parameters, $(\alpha_1, \log \alpha_0, \theta_B^{Fe}, \log \mathrm{n}, r^O)$, under the pragmatic and the fully Bayesian models on emissivities is shown in Figure 6.12. The fully Bayesian model on emissivities tends to shrink the error bars of $\log \mathrm{n}$ and $\log \alpha_0$ compared with that of under the pragmatic Bayesian model on emissivities. For the posterior samples of $\alpha_1, \log \alpha_0$, and $\log \mathrm{n}$, the fully Bayesian model is shifting the best estimates significantly since those parameters are sensitive to the choice of the emissivities. We are learning the emissivities from the data as well. Similar to the Fe XVII case study, as the posterior samples of some of the principal component variables under the fully Bayesian model, e.g., r_3^O and r_5^O in Figure 6.12, are up to three standard deviations away from the center of their prior distributions, it is possible to see the best-fit atomic emissivities for the six O VII lines, which are preferred by the data, could be quite far away from the default emissivity values from the CHIANTI, e.g., the line 21.602 Å in Figure 6.13.

Under the fully Bayesian model on emissivities with the t distribution prior case, we have obtained the parameter estimations with standard deviations: $\log \alpha_0 = 31.178 \pm 0.010$, $\alpha_1 = 6.701 \pm 0.010$, and $\log \mathrm{n} = 9.980 \pm 0.020$. The expected photon counts in the source exposure can be evaluated from

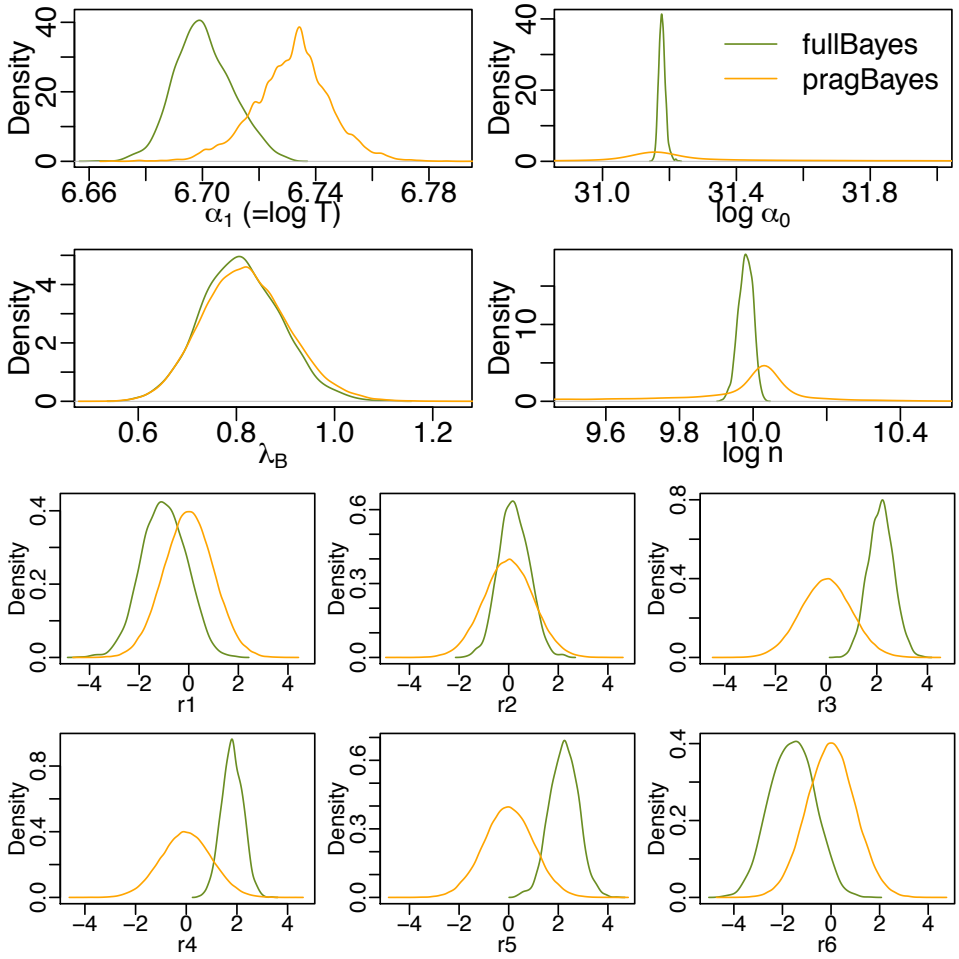

Figure 6.12: Comparison of the posterior samples of those model parameters in O VII case study, $(\alpha_1, \log \alpha_0, \theta_B^O, \log \mathrm{n}, r^O)$, under the pragmatic (amber) and the fully (green) Bayesian models on emissivities when considering a t distribution prior on $(\alpha_1, \log \alpha_0)$.

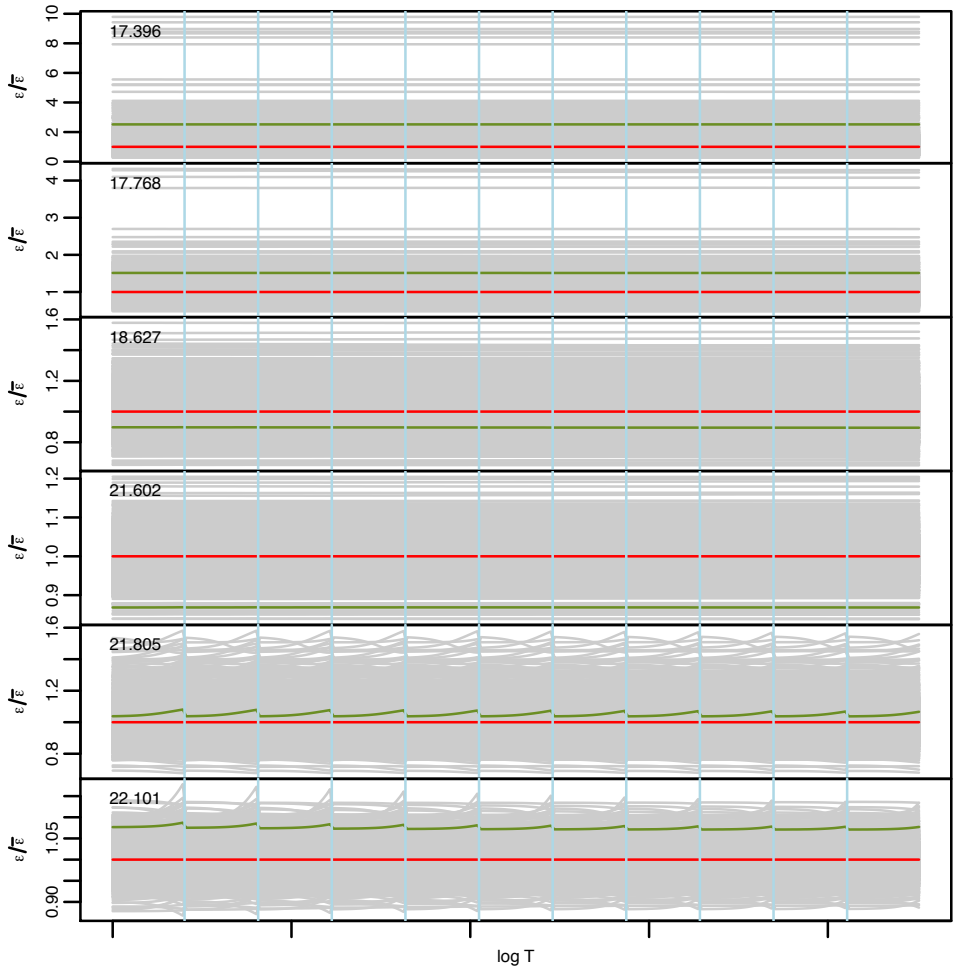

Figure 6.13: Ratios of emissivities for the six O VII lines considered in this section to the default emissivity. The grey lines represent the ratios from $10,000$ realizations of the O VII CHIANTI atomic data. The red curve is for the default value from the CHIANTI. The green curve is for the best emissivities preferred by the data under the fully Bayesian model. In each panel, the horizontally-arranged sub-panels correspond to each value of the equally-spaced temperature grid, $\log \mathrm{T} \in (6.5, 7.5)$, and the temperature is increasing from left to right. Within each sub-panel, for a given temperature, the density, $\log \mathrm{n} \in (9.0, 11.0)$, is increasing.

the posterior means of those model parameters sampled from HMC under the fully Bayesian model on emissivities. There is a good mismatch between the expected photon counts and the observed photon counts in the source exposure along the wavelength, as seen in the upper panel of Figure 6.14. The middle panel is a zoom in version, which is built up in the same way while the gaps in the wavelength are removed, to look at the detailed features of the spectrum profiles for those O VII lines. The contribution of the continuum counts in the spectrum is about 6 counts per channel on average. There are about 82% of the standardised residuals of the expected photon counts fall in 95% confidence interval, $(-2, 2)$, without any dependency, in the lower panel of Figure 6.14, indicating the current model can fit the data very well.

Comparison of the models with different prior distributions

Besides the t distribution prior case, we also test the Gaussian prior case and the uniform prior case to check the sensitivity of the results to the shape of the prior distributions. Figure 6.15 compares the posterior and the prior sample of those model parameters from all three cases.

The posterior distributions for $\log \alpha_0$ and α_1 shift a little bit away from their prior distributions in the Gaussian prior case indicating the preliminary analysis on the Fe XVII has provided more information and resulting in narrower error bars while adding the primary O VII analysis has not brought much effect on it. This is due to the fact that the Gaussian prior obtained in the preliminary stage is too narrow to incorporate the information from the likelihood in the primary stage and the sampling does not go across the whole space.

Instead of taking into account the information in the Fe XVII photon counts, there is only the information from the O VII photon counts considered in the uniform prior case. The posterior distributions for $\log \alpha_0$ and $\log \mathrm{n}$ tend to be as flat as their corresponding prior distributions indicating the O VII datasets does not provide any information on the estimation of $\log \alpha_0$ and $\log \mathrm{n}$. The result is consistent with the fact that, without the prior temperature information from the Fe XVII analysis, the O VII photon counts cannot

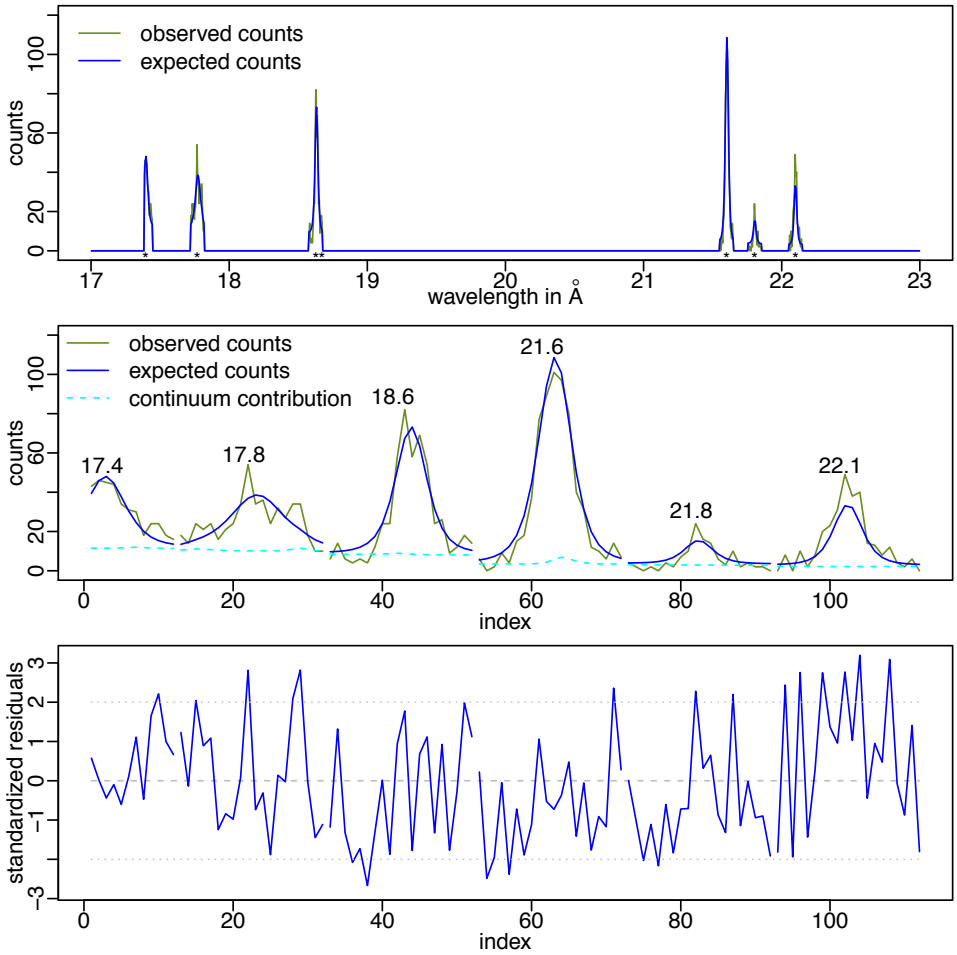

Figure 6.14: Comparison of the expected spectrum profiles under the fully Bayesian model with the observed O VII spectrum profiles. The upper panel compares the expected photon counts, marked as blue curve, with the observed photon counts, marked as green curve, in the source exposure along the wavelength (Å). The lines of interest are marked as '*'. The middle panel is a zoom in version focusing on the areas covered by the lines of interest. It is constructed in the same way while removing the gaps in the wavelength to compare the spectrum profiles closely and x-axis is the indices of model wavelength. The six wavelengths of interest are marked on top of each line. The light blue dashed curve represents the contribution of the continuum counts in the spectrum. The standardized residuals, which is the ratio of the difference between the observed counts and the expected counts to the square root of the expected counts since the counts follow Poisson distributions, are plotted along the indices of model wavelength in the lower panel.

provide enough information to estimate log n since the O VII line systems are sensitive to both the temperature and the density.

In contrast to the narrow Gaussian prior, the t-distribution prior has fatter tails and cover a wider range. In the t-distribution prior case, the O VII dataset does provide information on the estimation of model parameters. The posterior distributions for $\log \alpha_0$ and $\log$ n are significantly wider than that of in the Gaussian prior case and their corresponding prior distributions. As expected by the astronomers, once the information from the O VII photon counts is added, it introduces more uncertainties compared to the Gaussian prior case where the primary O VII analysis has no effect. The posterior distribution for $\log \alpha_0$ and $\log$ n are significantly narrower than that of in the uniform prior case. It is consistent with the fact that the O VII line systems have information about the density which still depends on the temperature. Incorporating the temperature information from the preliminary Fe XVII analysis in an appropriate way does reduce the uncertainties of density estimation significantly.

In summary, the prior information from the preliminary Fe XVII analysis plays a major role in the Gaussian prior case while the likelihood in the primary O VII analysis contribute almost nothing. The uniform prior case depends only on the O VII dataset. The t-distribution prior case combines the information from both the Fe XVII and the O VII datasets. These three cases are equivalent to a weighting strategy where different weights are assigned to the information provided by the Fe XVII and the O VII photon counts respectively. Depending on the weights we put on each of those analyses, different inferences about the temperature are obtained. The Gaussian prior case puts most of the weight on the Fe XVII analysis. The uniform prior case puts all of the weight on the O VII analysis. The t-distribution prior case is mixing over the two analyses properly by carrying forward the temperature information from the preliminary Fe XVII analysis into the primary analysis on O VII.

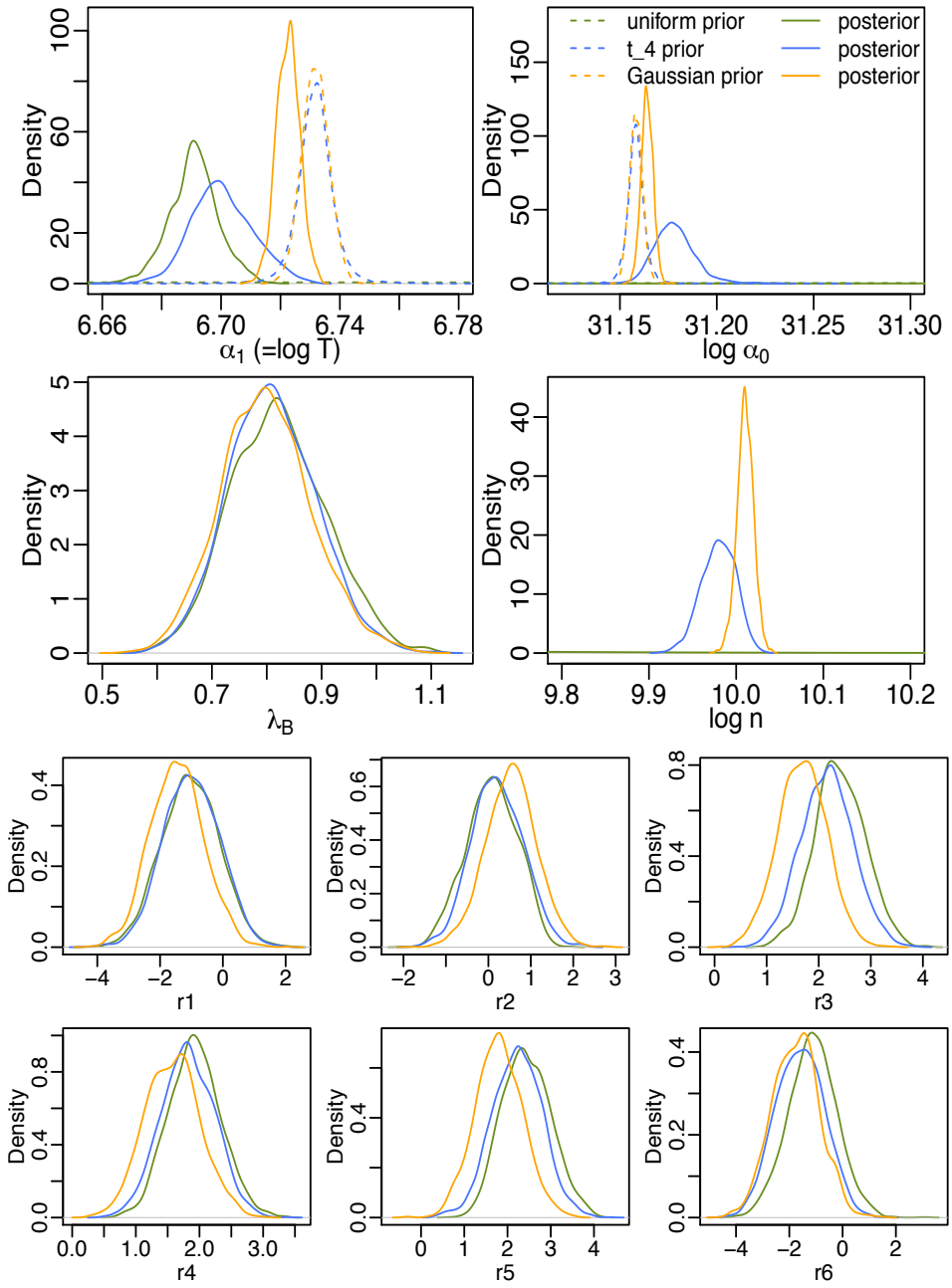

Figure 6.15: Comparison of the posterior sample of those model parameters in O VII case study under the fully Bayesian models with different prior distributions for $(\log \alpha_0, \alpha_1)$. There are three different prior distributions considered, with blue representing t distribution prior, amber a Gaussian prior distribution, and green a uniform prior distribution. The dashed lines represent the prior distributions and the solid lines represent the corresponding posterior distributions.

6.7 Conclusions and discussion

We have incorporated three different sources of uncertainties, uncertainties of the O VII atomic data, uncertainties for the common parameters in both of the O VII and the Fe XVII analyses, and uncertainties for all other plasma parameters in the O VII analysis, via a three-stage analysis. Reasonable parameter inferences are obtained when carrying forward the results from the two preliminary analyses into the primary analysis. We have considered a Bayesian framework to interpret the observed O VII photon counts in terms of a given ensemble of atomic data. A fully Bayesian model, where the observed photon counts are informative to the uncertainty in the atomic emissivities, reduces the uncertainties in the plasma parameters, compared to that of from a pragmatic Bayesian model, where the observed photon counts do not affect the uncertainties in the emissivities. The former model achieves an accurate and precise performance. We also check the sensitivity of the fitted results to the shape of different prior distributions. The t distribution prior case is able to assign appropriate weights to the information provided by the datasets in multiple stages. As a result, incorporating the temperature information from the preliminary Fe XVII analysis via a t distribution prior, together with the atomic uncertainties, into the primary O VII analysis does reduce the uncertainties of density estimation significantly.

We have been working on a single value temperature parameter on the whole atmosphere to explain all the issues in respect of astronomy throughout Chapter 5 and Chapter 6. However, different estimates for temperature have been obtained from different models with different weighting scheme on the sources of information after careful preprocessing and model fitting works. It indicates that a single temperature value on the whole atmosphere is probably not enough. Generalizing this single value temperature to a distribution of temperature by including more parameters, e.g., assuming the square root of temperature follows a normal distribution, might be a following step. Finding out a distribution of temperature could be a future aim of this spectral analysis.

7

Conclusions and discussion

7.1 A general three-stage analysis

We summarize a general framework for a three-stage analysis where all kinds of uncertainties could be involved. By generalizing the mathematical framework from Section 3.2, a general three-stage analysis can be summarized as

$$\text{Stage 1: } X_1 \mid \psi_1, \psi_1^{'}. \tag{7.1}$$

$$\text{Stage 2: } X_2 \mid \psi_1, \psi_2, \psi_2^{'}. \tag{7.2}$$

$$\text{Stage 3: } X_3 \mid \psi_1, \psi_2, \psi_3. \tag{7.3}$$

In Stage 1, the dataset, X_1, is modelled with respect to the unknown parameters, ψ_1 and $\psi_1^{'}$. In Stage 2, the dataset, X_2, is modelled with respect to the unknown parameters, ψ_1, ψ_2, and $\psi_2^{'}$. In Stage 3, i.e., the primary stage, the dataset, X_3, is modelled with respect to the unknown parameters, ψ_1, ψ_2, and ψ_3. Note that ψ_1 represents an unknown parameter or a set of unknown parameters that is common among all three models in the Stage 1, Stage 2, and Stage 3 analyses. In the meanwhile, ψ_2, an unknown parameter

or a set of unknown parameters, is common between the two models in the Stage 2 and Stage 3 analyses. The output information, typically estimates or posterior distributions, for ψ_1 and ψ_2 from the two individual preliminary analyses is required for and will be carried forward into the primary stage. Practically those parameters can represent systematic data like atomic data, parameters of interest, or any information from a preliminary analysis in the astrophysical study or other areas of interest.

Looking back on the special three-stage analysis in Chapter 6 and comparing it with the general form above, ψ_1 represents the O VII atomic parameter that is common in Stage 1 and Stage 3 only, $\psi_2 = (\log \alpha_0, \alpha_1)$ the common parameters in the Stage 2 Fe XVII and the Stage 3 O VII analyses only, and ψ_3 the parameters only exist in the Stage 3 O VII analysis.

We have been working on a pragmatic Bayesian method with respect to the atomic emissivities assuming the Fe XVII and the O VII spectra do not provide any information on the choice of atomic emissivities. Generally, this is a pragmatic Bayesian method with respect to the parameter in Stage 1 assuming the datasets in Stage 2 and Stage 3 do not provide any information on the choice of Stage 1 parameter. Mathematically, it is expressed as,

$$p(\psi_1 \mid X_1, X_2, X_3) = p(\psi_1 \mid X_1), \tag{7.4}$$

i.e., ψ_1 and X_2, X_3 given X_1 are independent. From a Bayesian statistical point of view, given X_1, the prior distribution for ψ_1 is the same as the posterior distribution for ψ_1 given X_2 and X_3.

Similarly, we can work on another pragmatic Bayesian method with respect to the common parameters in both of the Fe XVII and the O VII analyses assuming the O VII spectrum and the model for generating O VII atomic data do not provide any information on the estimations of those common parameters. In other words, only the Fe XVII spectrum provides information on the estimation of the common parameter. Generally, this is a pragmatic Bayesian method with respect to the Stage 2 parameter assuming the datasets in Stage 1 and Stage 3 do not provide any information on the choice

of Stage 2 parameter. Mathematically, it is expressed as,

$$p(\psi_2 \mid X_1, X_2, X_3) = p(\psi_2 \mid X_2), \tag{7.5}$$

i.e., ψ_2 and X_1, X_3 given X_2 are independent.

A pragmatic Bayesian method means we would rely completely on the prior distribution of a subset of parameters instead of learning about these parameters from any of the following analyses. Therefore, the prior and the posterior distributions of this subset of parameters will be exactly the same under the pragmatic Bayesian assumption. Compared with the corresponding fully Bayesian method where the assumption is eliminated, we will learn how large the effect of the information of the data from the following analyses on the uncertainties of the subset of parameters, i.e., quantities of interest. However, for the O VII analysis in Chapter 6, we would not work on the second pragmatic Bayesian method above since there is no meaning regarding the assumption in physics.

7.2 Summary

From a statistical point of view, we have designed the multistage analysis to improve the accuracy of model parameter inference when different sources of uncertainties coming from different stages need to be incorporated into subsequent analyses. It is a robust principled method that can be applied to any problems under the setting where the output from one or several preliminary analyses is required for a following primary analysis. A simple two-stage analysis is studied by accounting for uncertainties in atomic data from a preliminary analysis and carrying forward it into a primary analysis with observed spectral data in Chapter 4 and Chapter 5. Following this work, a more complicated three-stage analysis is studied where both uncertainties in atomic data from a preliminary analysis and uncertainties for a subset of parameters from another preliminary analysis are incorporated into a primary analysis with observed spectral data in Chapter 6. This methodology can also be generalized to incorporate more sources of uncertainties coming from

multiple preliminary analyses into a subsequent primary analysis.

We have used a Bayesian framework to interpret the observed spectral data in the context of the atomic data. A pragmatic Bayesian method, where each realization of emissivities is considered as fully specified and uncorrectable, yields larger uncertainties in the estimations of parameters than the uncertainty coming from the observed data alone. It is considered as an imprecise but accurate approach to parameter estimations. A fully Bayesian method, where we allow for the data to update the atomic data uncertainties, reduces the uncertainties in the parameters and shifts the estimates towards more accurate values. The advantages are evident and it is considered as a precise and accurate approach to parameter estimations.

Besides a brute force discrete approach where the atomic realizations are considered individually, a continuous analysis where principal component analysis is used to fully summarize the atomic uncertainties via a multivariate Gaussian distribution, to incorporate uncertainties in atomic data into our highly structured statistical model. It provides a concise statistical compression when there are complex correlations in the given emissivity ensemble.

From an astrophysical point of view, the methodology that we have developed in this thesis represents a breakthrough in how atomic data uncertainties are brought into a spectral analysis. We have been able to use the observed spectral data to narrow the uncertainty in the atomic data. In the meanwhile, we have estimated the electron density and path length of solar corona using the Fe XIII line spectrum observed from the EUV Imagining spectrometer on the *Hinode* satellite, and have estimated the coronal density and temperature of Capella using the Fe XVII and the O VII line systems in the soft X-ray regime over the course of the *Chandra* mission. Future improvements to the methodology and the structure of atomic databases will no doubt improve the process and make it more accessible. In addition to the uncertainty in the atomic data, commonly known as the dominant source of error in the analysis of solar and stellar spectra, our methodology is able to account for other sources of systematic uncertainties.

References

Anderson, T. W. (2003), *An Introduction to Multivariate Statistical Analysis (Wiley Series in Probability and Statistics)*, 3 edn, Wiley.

Arnaud, K., Smith, R. & Siemiginowska, A. (2011), *Handbook of X-ray Astronomy*, Vol. 7, Cambridge University Press.

Berger, T., De Pontieu, B., Fletcher, L., Schrijver, C., Tarbell, T. et al. (1999), 'What is moss?', *Solar Physics* **190**(1), 409–418.

Bishop, C. M. (2006), *Pattern Recognition and Machine Learning*, Vol. 1, Springer.

Blocker, A. W., Meng, X. L. et al. (2013), 'The potential and perils of preprocessing: Building new foundations', *Bernoulli* **19**(4), 1176–1211.

Brickhouse, N., Dupree, A., Edgar, R., Liedahl, D., Drake, S., White, N. & Singh, K. (2000), 'Coronal structure and abundances of Capella from simultaneous EUVE and ASCA spectroscopy', *The Astrophysical Journal* **530**(1), 387–402.

Brickhouse, N., Raymond, J. & Smith, B. (1995), 'New model of iron spectra in the Extreme Ultraviolet and application to SERTS and EUV observations: A solar active region and Capella', *The Astrophysical Journal Supplement Series* **97**, 551–570.

Canizares, C. R., Davis, J. E., Dewey, D., Flanagan, K. A., Galton, E. B., Huenemoerder, D. P., Ishibashi, K., Markert, T. H., Marshall, H. L., McGuirk, M. et al. (2005), 'The Chandra high-energy transmission grating: Design, fabrication, ground calibration, and 5 years in flight', *Publications of the Astronomical Society of the Pacific* **117**(836), 1144.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P. & Riddell, A. (2016), 'Stan: A probabilistic programming language', *Journal of Statistical Software* **20**, 1–37.

Chandra, X. (2019), 'The chandra proposers' observatory guide', *Chandra Project* .

Chen, Y., Meng, X. L., Wang, X., van Dyk, D. A., Marshall, H. L. & Kashyap, V. L. (2019), 'Calibration concordance for astronomical instruments via multiplicative shrinkage', *Journal of the American Statistical Association* **114**(527), 1018–1037.

Clopper, C. J. & Pearson, E. S. (1934), 'The use of confidence or fiducial limits illustrated in the case of the binomial', *Biometrika* **26**(4), 404–413.

Culhane, J., Harra, L., James, A., Al-Janabi, K., Bradley, L., Chaudry, R., Rees, K., Tandy, J., Thomas, P., Whillock, M. et al. (2007), 'The EUV imaging spectrometer for Hinode', *Solar Physics* **243**(1), 19–61.

Del Zanna, G. (2011), 'Benchmarking atomic data for astrophysics: Fe XVII X-ray lines', *Astronomy & Astrophysics* **536**, A59.

Del Zanna, G. (2013), 'The multi-thermal emission in solar active regions', *Astronomy & Astrophysics* **558**, A73.

Del Zanna, G., Berrington, K. & Mason, H. (2004), 'Benchmarking atomic data for astrophysics: Fe X', *Astronomy & Astrophysics* **422**(2), 731–749.

Del Zanna, G., Dere, K. P., Young, P. R., Landi, E. & Mason, H. E. (2015), 'CHIANTI - An atomic database for emission lines. Version 8', *Astronomy & Astrophysics* **582**, A56.

Del Zanna, G. & Mason, H. (2014), 'Elemental abundances and temperatures of quiescent solar active region cores from X-ray observations', *Astronomy & Astrophysics* **565**, A14.

Del Zanna, G. & Storey, P. J. (2012), 'Atomic data for astrophysics: Fe XIII soft X-ray lines', *Astronomy & Astrophysics* **543**, A144.

Dere, K., Landi, E., Mason, H., Fossi, B. M. & Young, P. (1997), 'CHIANTI-an atomic database for emission lines-I. Wavelengths greater than 50 Å', *Astronomy and Astrophysics Supplement Series* **125**(1), 149–173.

Drake, J. J., Ratzlaff, P., Kashyap, V., Edgar, R., Izem, R., Jerius, D., Siemiginowska, A. & Vikhlinin, A. (2006), 'Monte Carlo processes for including Chandra instrument response uncertainties in parameter estimation studies', **6270**, 471–482.

Duane, S., Kennedy, A. D., Pendleton, B. J. & Roweth, D. (1987), 'Hybrid Monte Carlo', *Physics Letters B* **195**(2), 216–222.

Fletcher, L. & De Pontieu, B. (1999), 'Plasma diagnostics of transition region 'moss' using SOHO/CDS and TRACE', *The Astrophysical Journal Letters* **520**(2), L135.

Foster, A., Smith, R., Brickhouse, N., Kallman, T. & Witthoeft, M. (2010), 'The challenges of plasma modeling: Current status and future plans', *Space Science Reviews* **157**(1-4), 135–154.

Freeland, S. & Handy, B. (1998), 'Data analysis with the SolarSoft system', *Solar Physics* **182**(2), 497–500.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2014), *Bayesian Data Analysis*, Vol. 2, CRC press Boca Raton, FL.

Gelman, A., Roberts, G. O., Gilks, W. R. et al. (1996), 'Efficient Metropolis jumping rules', *Bayesian Statistics* **5**, 599–608.

Geman, S. & Geman, D. (1984), 'Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-6**(6), 721–741.

Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. (1996), 'Introducing Markov chain Monte Carlo', *Markov chain Monte Carlo in Practice* **1**, 19.

Haario, H., Saksman, E., Tamminen, J. et al. (2001), 'An adaptive Metropolis algorithm', *Bernoulli* **7**(2), 223–242.

Harel, O. & Zhou, X.-H. (2005), 'Multiple imputation: review of theory, implementation and software', *Statistics in Medicine* **26**(16), 3057–3077.

Hastings, W. K. (1970), 'Monte Carlo sampling methods using Markov chains and their applications', *Biometrika* **57**(1), 97–109.

Hoffman, M. D. & Gelman, A. (2014), 'The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo.', *Journal of Machine Learning Research* **15**(1), 1593–1623.

Jolliffe, I. T. (2002), 'Principal components in regression analysis', *Principal Component Analysis* pp. 167–198.

Jones, D. E., Kashyap, V. L. & Van Dyk, D. A. (2015), 'Disentangling overlapping astronomical sources using spatial and spectral information', *The Astrophysical Journal* **808**(2), 137.

Kallman, T. & Palmeri, P. (2007), 'Atomic data for X-ray astrophysics', *Reviews of Modern Physics* **79**(1), 79.

Kashyap, V. & Drake, J. J. (1998), 'Markov-Chain Monte Carlo reconstruction of emission measure distributions: application to solar extreme-ultraviolet spectra', *The Astrophysical Journal* **503**(1), 450.

Le Cam, L. (1986), 'The central limit theorem around 1935', *Statistical Science* pp. 78–91.

Lee, H., Kashyap, V. L., Van Dyk, D. A., Connors, A., Drake, J. J., Izem, R., Meng, X.-L., Min, S., Park, T., Ratzlaff, P. et al. (2011), 'Accounting for calibration uncertainties in X-ray analysis: effective areas in spectral fitting', *The Astrophysical Journal* **731**(2), 126.

Li, K.-H., Raghunathan, T. E. & Rubin, D. B. (1991), 'Large-sample significance levels from multiply imputed data using moment-based statistics and an F reference distribution', *Journal of the American Statistical Association* **86**(416), 1065–1073.

Mariska, J. T. (1992), *The Solar Transition Region*, Vol. 23, Cambridge University Press.

Mernier, F., Werner, N., Lakhchaura, K., de Plaa, J., Gu, L., Kaastra, J. S., Mao, J., Simionescu, A. & Urdampilleta, I. (2020), 'How do atomic code uncertainties affect abundance measurements in the intracluster medium?', *Astronomische Nachrichten* **341**(2), 203–209.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953), 'Equation of state calculations by fast computing machines', *The Journal of Chemical Physics* **21**(6), 1087–1092.

Mewe, R., Raassen, A., Drake, J., Kaastra, J., Van Der Meer, R. & Porquet, D. (2001), 'CHANDRA-LETGS X-ray observations of Capella-Temperature, density and abundance diagnostics', *Astronomy & Astrophysics* **368**(3), 888–900.

Neal, R. M. et al. (2011), 'MCMC using Hamiltonian dynamics', *Handbook of Markov Chain Monte Carlo* **2**(11), 2.

Roberts, G. O., Gelman, A., Gilks, W. R. et al. (1997), 'Weak convergence and optimal scaling of random walk Metropolis algorithms', *The Annals of Applied Probability* **7**(1), 110–120.

Roberts, G. O. & Rosenthal, J. S. (2001), 'Optimal scaling for various Metropolis-Hastings algorithms', *Statistical Science* **16**(4), 351–367.

Roberts, G. O. & Rosenthal, J. S. (2007), 'Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms', *Journal of Applied Probability* **44**(2), 458–475.

Roberts, G. O. & Rosenthal, J. S. (2009), 'Examples of adaptive MCMC', *Journal of Computational and Graphical Statistics* **18**(2), 349–367.

Roberts, H. V. (1965), 'Probabilistic prediction', *Journal of the American Statistical Association* **60**(309), 50–62.

Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*, Vol. 81, John Wiley & Sons.

Schwartz, D. A. (2014), 'Invited review article: The Chandra X-ray Observatory', *Review of Scientific Instruments* **85**(6), 061101.

van Dyk, D. A. (2003), Hierarchical models, data augmentation, and Markov chain Monte Carlo, *in* 'Statistical Challenges in Astronomy', Springer, pp. 41–55.

Van Dyk, D. A., Connors, A., Kashyap, V. L. & Siemiginowska, A. (2001), 'Analysis of energy spectra with low photon counts via Bayesian posterior simulation', *The Astrophysical Journal* **548**(1), 224.

Warren, H. P., Winebarger, A. R. & Brooks, D. H. (2012), 'A systematic survey of high-temperature emission in solar active regions', *The Astrophysical Journal* **759**(2), 141.

Watanabe, T., Hara, H., Yamamoto, N., Kato, D., Sakaue, H., Murakami, I., Kato, T., Nakamura, N. & Young, P. (2009), 'Fe XIII density diagnostics in the EIS observing wavelengths', *The Astrophysical Journal* **692**(2), 1294.

Xu, J., Van Dyk, D. A., Kashyap, V. L., Siemiginowska, A., Connors, A., Drake, J., Meng, X.-L., Ratzlaff, P. & Yu, Y. (2014), 'A fully Bayesian method for jointly fitting instrumental calibration and X-ray spectral models', *The Astrophysical Journal* **794**(2), 97.

Young, P., Watanabe, T., Hara, H. & Mariska, J. (2009), 'High-precision density measurements in the solar corona - I. Analysis methods and results for Fe XII and Fe XIII', *Astronomy & Astrophysics* **495**(2), 587–606.

Yu, X., Del Zanna, G., Stenning, D. C., Cisewski-Kehe, J., Kashyap, V. L., Stein, N., van Dyk, D. A., Warren, H. P. & Weber, M. A. (2018), 'Incorporating uncertainties in atomic data into the analysis of solar and stellar observations: A case study in Fe XIII', *The Astrophysical Journal* **866**(2), 146.