

A Bayesian approach to comparing theoretic models to observational data: A case study from solar flare physics

S. Adamakis¹

C. L. Raftery³

R. W. Walsh²

P. T. Gallagher³

ABSTRACT

Solar flares are large-scale releases of energy in the solar atmosphere, which are characterised by rapid changes in the hydrodynamic properties of plasma from the photosphere to the corona. Solar physicists have typically attempted to understand these complex events using a combination of theoretical models and observational data. From a statistical perspective, there are many challenges associated with making accurate and statistically significant comparisons between theory and observations, due primarily to the large number of free parameters associated with physical models. This class of ill-posed statistical problem is ideally suited to Bayesian methods. In this paper, the solar flare studied by Raftery et al. (2009) is reanalysed using a Bayesian framework. This enables us to study the evolution of the flare’s temperature, emission measure and energy loss in a statistically self-consistent manner. The Bayesian-based model selection techniques imply that no decision can be made regarding which of the conductive or non-thermal beam heating play the most important role in heating the flare plasma during the impulsive phase of this event.

Subject headings: Sun: corona — Sun: flares — methods: statistical

1. Introduction

Solar flares are triggered by the instability of the magnetic field. This can result in the direct or indirect heating of chromospheric plasma leading to the process known as “chromospheric evaporation” (Milligan et al. 2006a,b). There are two ways to provoke evaporation (chromospheric

¹Decision Science, Lloyds Banking Group, 155 Bishopsgate, London, EC2M 3TQ, UK

²Jeremiah Horrocks Institute for Astrophysics and Supercomputing, University of Central Lancashire, Preston, PR1 2HE, UK

³Astrophysics Research Group, School of Physics, Trinity College, Dublin, Dublin 2, Ireland

plasma upflow): using thermal energy and/or using non-thermal energy according to the energy release mechanisms (Mariska et al. 1993). In the thermal energy model, heat is unleashed in the coronal portion of the loop (possibly because of magnetic reconnection). Then the thermal energy is carried downward to the upper chromosphere via conduction, where the deposited energy heats the plasma and stimulates it to move slowly upward. In the non-thermal energy model, the energy release (again in the corona) is in the form of a non-thermal electron beam. Electrons then collide with dense chromospheric plasma, which they heat, causing the plasma to expand both upward at very high speeds and downward at lower speeds.

This paper will focus on the statistical analysis of the observations outlined in Raftery et al. (2009) and will be compared with results from the Enthalpy Based Thermal Evolution of Loops (EBTEL) model (Klimchuk et al. 2008). Comparing to Raftery et al. (2009) this research will handle the statistics with care to derive which of the thermal or non-thermal heat flux is more dominant, during a single flare event, using different model comparison techniques (Kass 1993; Kass & Greenhouse 1989; Kass & Raftery 1995; Raftery 1996; Clyde et al. 2007; Aitkin et al. 2009). Furthermore, the method used here treats the uncertainties for the time constraints with equal respect as the uncertainties of the temperature and emission measure, which are difficult to incorporate into our analysis regarding Classical statistics. The thermal heat flux is proportional to the non-thermal heat flux plus a constant background heat flux. The parameter set is extended with the addition of the radius of the loop in order to make the model more realistic. Last but not least, the ratios between the radiative loss rate of the transition and the corona, the average coronal and apex temperature, and the coronal base and apex temperature are assumed to be free parameters so that their values will be decided by the data. For more information about these parameters refer to Section 3.

One advantage of the EBTEL model is that we can incorporate both a direct and a non-thermal heat input, which is not restricted by the data. Thus, in order to produce a temperature, density and/or emission measure profile, a specific form for the non-thermal heat flux function must be assumed. There appears to be a connection between the heating function form and the temperature profile, more so than with the emission measure profile. For example, a sudden increase in the heat input will lead to an abrupt uplift in the temperature, whereas a flat heating function will lead to a smoother temperature profile. Hence, it is important to understand and evaluate the impact any changes in the thermal and non-thermal heat flux have on the plasma evolution. Several forms of thermal heat fluxes have been tested in Adamakis (2009). Here we present only two of them: Half Gaussian profile and Full Gaussian profile. The Half Gaussian profile represents a sudden switching off of an electron beam, whereas the Full Gaussian profile imitates a gradual one. The rest of the paper has been structured as follows: Section 2 comments on the Raftery et al. (2009) observations and some assumptions in the EBTEL model. Section 3 discusses the details of the analysis regarding the data distribution and the parameter set. Section 4 addresses to the choice of the parameters in the prior distributions. Section 5 briefly presents the model selection methods applied here. Section 6 presents results of different models and parameter estimations. Finally, Section 7 summarise the

findings from this study and presents how this work can be further progressed.

2. Previous Work

2.1. Observations of a C-Class Solar Flare

The temporal evolution of temperature and emission measure in a C3.0 solar flare observed on March 26, 2002 has been analysed. During a typical solar flare, the temperature rises from less than one MK up to ~ 20 MK, and then cools down to the pre-flare temperatures. In the example under consideration here, Raftery et al. (2009) used different instruments to track this evolution: the Reuven Ramaty High Energy Solar Spectroscopic Imager (RHESSI; > 5 MK), GOES-12 (5 – 30 MK), the Transition Region and Coronal Explorer (TRACE 171 Å ; 1 MK), and the Coronal Diagnostic Spectrometer (CDS; $\sim 0.03 - 8$ MK). A notation should be made that TRACE data were not included in the emission measure analysis. This is as a result of the instrument being sensitive to multiple emission lines, making it complex to define the contribution function. The reader is referred to Raftery et al. (2009) which outlines the reasons why and at which particular time of the flare each particular instrument was employed. Figure 1 depicts a visual display of the observation results together with the underlying uncertainties.

The flare began with evidence of pre-flare heating at its onset. This was followed by explosive chromospheric evaporation during the impulsive phase and gentle chromospheric evaporation during the early decay phase. It is believed that the plasma reached a peak temperature of just more than 13 MK in approximately 10 min. Conduction losses dominated over radiative losses for the initial ~ 300 s of the post-flare decay, whereas for the next ~ 4000 s, radiative losses prevailed. Raftery et al. (2009) concluded that approximately equal direct and non-thermal heating mechanisms produced the data observed according only to an estimation of the ratio between the thermal and non-thermal heating rate, which is of course very critical. The best fit and parameter intervals were derived using an acceptable fit to the data by eye.

2.2. The EBTEL Model

The Enthalpy-Based Thermal Evolution of Loops (EBTEL) model (Klimchuk et al. 2008) is a zero-dimensional hydrodynamic (0D HD) model that takes into account the significance of introducing the enthalpy into the system. The difference with the 1D HD model is that the 1D HD model provides us with temperature, density, and velocity profiles along a given magnetic field line as time evolves (that is temperature, density, and velocity are a function of both space and time), whereas 0D HD models describe the average evolution of these values along a coronal strand as a function of time only. Subsequently, 0D HD models require less computing time than 1D models, but at the expense of losing the spatial resolution. In spite of this, the EBTEL code has been used

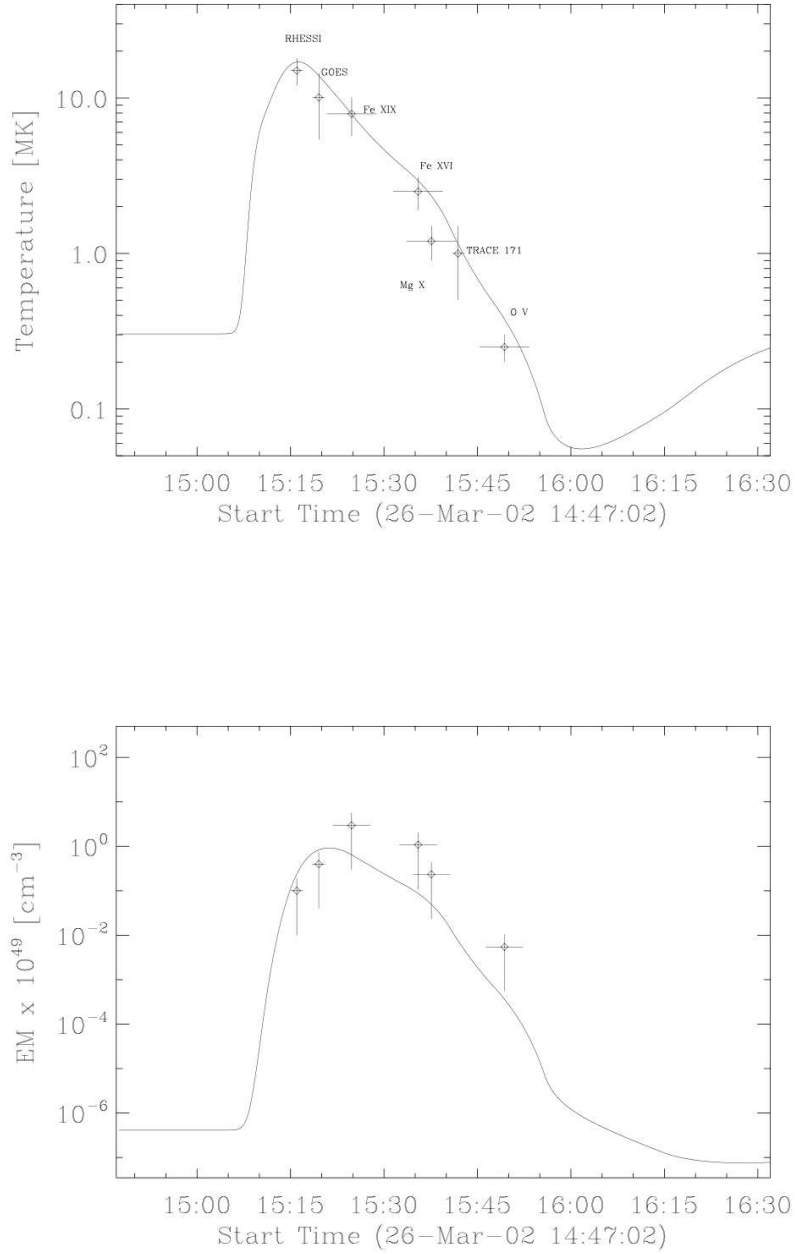


Fig. 1.—: Temperature and emission measure observations of a C3.0 solar flare reported in Raftery et al. (2009). *Top*: The EBTEL temperature evolution that best reproduced the observations, according to Raftery et al. (2009). Different instruments were employed to measure the temperature at different times. *Bottom*: The EBTEL model and observed emission measure evolution. TRACE observations were not included.

in a wide range of studies regarding different heat input forms (see Klimchuk et al. 2008, for more details). The importance of 0D models compared to the 1D models is that they can give similar results for the plasma response to a sudden heat input, despite the fact that they use up to four orders of magnitude less computing time.

To reflect the physical processes taking place within the standard flare model, the EBTEL model allows for both thermal and non-thermal heating of the plasma in the system. It has been noted that proton precipitation is not included (see Aschwanden 2005). Proton beams are expected to excite strong kinetic Alfvén waves. The turbulence caused by kinetic Alfvén waves contains enough energy to produce the non-thermal velocities observed in flares, and therefore could contribute to impulsive primary plasma heating between the reconnection regions and the flare footpoints.

The governing equations of the EBTEL model for the coronal part of the loop are (Klimchuk et al. 2008):

$$\frac{d\bar{P}}{dt} \approx \frac{2}{3} \left[\bar{Q} - \bar{n}^2 \Lambda(\bar{T})(1 + c_1) - \frac{\mathcal{F}}{L} \left(1 - \frac{3}{2} \frac{k\bar{T}}{\mathcal{E}} \right) \right], \quad (1)$$

$$\frac{d\bar{n}}{dt} = -\frac{c_2}{5c_3 k L \bar{T}} (F_0 + \mathcal{R}_{tr}) + \frac{\mathcal{F}}{\mathcal{E} L} \left(1 - \frac{c_2}{5c_3} \frac{\mathcal{E}}{k\bar{T}} \right), \quad (2)$$

$$\frac{d\bar{T}}{dt} \approx \bar{T} \left(\frac{1}{\bar{P}} \frac{d\bar{P}}{dt} - \frac{1}{\bar{n}} \frac{d\bar{n}}{dt} \right), \quad (3)$$

where P is the pressure, t the time, Q the direct (thermal) heating rate, n the electron number density, T the temperature, $\Lambda(\cdot)$ the optically thin radiative loss function, \mathcal{F} the non-thermal energy flux ($\mathcal{F} = \mathcal{E}\mathcal{J}$, with \mathcal{E} the mean energy of the accelerated (non-thermal) electrons impinging the chromosphere in keV and \mathcal{J} the nonthermal particle flux), k the Boltzmann's constant ($\approx 1.38 \times 10^{-16}$ erg K⁻¹), L the length of the loop from the coronal base to the apex, F_0 the thermal conduction (heat flux) at the beginning of the coronal region, R_{tr} the radiative cooling rate at the transition region, $c_1 = \frac{R_{tr}}{R_c}$ the ratio of the radiative loss function between the transition region and the corona, $c_2 = \frac{\bar{T}}{T_a}$ the ratio between the average temperature at the coronal part of the loop and the temperature at the apex of the loop, $c_3 = \frac{T_0}{T_a}$ the ratio between the temperature at the base of the corona and the temperature at the apex of the loop and the over bars indicate spatial averages along the coronal section of the loop. Furthermore, the transition region is treated separately.

Deviations of the observations from the model could be caused by two distinct reasons: First, the EBTEL model treats the effects of the non-thermal electron beam in a very simple way. It is assumed that all the beam's energy goes into evaporating plasma upwards into the loop, which is reasonable for gentle evaporation. However, for explosive evaporation, some of the beam energy will be used to drive chromospheric downflows, making the total observed non-thermal energy larger than that predicted by the EBTEL model. Second, the flare loop is almost certainly constructed of many strands that are heated at different times. Raftery et al. (2009) assume the flaring loop

consists of individual strands, all heated simultaneously. Therefore, they model the loop as one “fat strand”. However, some strands are likely to be heated at a different time compared to this solid structure.

3. Defining the Model Parameters

3.1. Data Distribution

The errors in temperature come from the width of the contribution function for the particular emission line (for the CDS points) and from the width of the instrument response function for GOES and TRACE. The RHESSI temperature error comes from the uncertainty in the fit to the Maxwell Boltzmann distribution.

As there is no information about the form of the errors, it is assumed that the errors between the observed temperatures and emission measures and those predicted from the model are normally distributed with mean zero and standard deviation derived from the error bars. In this particular case study we assume “ 3σ ” belief, *i.e.* the distance between the mean of the distribution and the upper or lower error bar is three standard deviations of the normal distribution (see Adamakis et al. 2010, for a link between error bars and statistical distributions). The reason for choosing a Normal distribution as the likelihood function is two-fold: (*i*) the error bars are symmetric and intuitively a symmetric likelihood is needed, and, (*ii*) it is the standard likelihood distribution that is applied in many parametric models in the absence of additional information.

This means that the combined likelihood which contains both temperature and emission measure information is of the form:

$$p(\mathbf{D}|\mathbf{P}) = \frac{1}{(2\pi)^{(n_1+n_2)/2} (\prod_{i=1}^{n_1} \sigma_{i1}) (\prod_{i=1}^{n_2} \sigma_{i2})} \exp \left[- \sum_{i=1}^{n_1} \frac{(T_i - \widehat{T}_i(\mathbf{P}))^2}{2\sigma_{i1}^2} - \sum_{i=1}^{n_2} \frac{(EM_i - \widehat{EM}_i(\mathbf{P}))^2}{2\sigma_{i2}^2} \right] \mathcal{I}(\max(T) < 3 \times 10^7),$$

where $\mathbf{D} = (\mathbf{T}, \mathbf{EM})$ with $\mathbf{T} = (T_1, \dots, T_{n_1})$ the temperature data-points and $\mathbf{EM} = (EM_1, \dots, EM_{n_2})$ the emission measure data-points, \widehat{T}_i and \widehat{EM}_i are the temperatures and emission measures respectively proposed by the EBTEL model, \mathbf{P} is the parameter set (see end of Section 3), n_1 is the number of observed temperature values, n_2 is the number of observed emission measure values, σ_{i1} are the standard deviations of the temperature errors, σ_{i2} are the standard deviations of the emission measure errors and $\mathcal{I}(\cdot)$ is the indicator function which is given by:

$$\mathcal{I}(\max(T) < 3 \times 10^7) = \begin{cases} 1, & \text{if } \max(T) < 3 \times 10^7 \text{ K} \\ 0, & \text{otherwise,} \end{cases}$$

where $\max(T)$ is the maximum temperature value of the cooling curve. The temperature profile is restricted to not exceed 30 MK because it is believed that temperatures above this threshold will create unphysical data. Temperature and electron density (T and n respectively) are calculated numerically from Equations (1) – (3). First order finite difference methods are employed to solve numerically the partial differential equations, after providing initial values for temperature, density and pressure. Emission measure (EM) can be computed from $2n^2\pi r^2L$, where r is the dimensionless radius of the loop and L is the dimensionless length of the loop.

3.2. Parameters for Non-Thermal Heat Flux

As discussed in Section 1, one of our aims is to compare which of the Half Gaussian and Full Gaussian functions for the non-thermal heat flux fits better to the data observed. In the case of the Half Gaussian the non-thermal heat flux function is given by the form:

$$\mathcal{F}(t) = \begin{cases} \frac{\mathcal{A}}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{(t-\mu)^2}{2\sigma_1^2}\right], & t \leq \mu \\ 0, & t > \mu, \end{cases}$$

where t is the time, μ is the time that maximise the non-thermal heat flux function (which can be adopted as a parameter), σ_1 is a value that defines the width of this distribution (which can be adopted as a parameter as well) and \mathcal{A} is the amplitude of the function (parameter). For the total non-thermal heat function \mathcal{F}_{tot} we have:

$$\mathcal{F}_{tot} = \int_0^\infty \mathcal{F}(t)dt = \int_0^\mu \mathcal{F}(t)dt = P(0 \leq t \leq \mu) \mathcal{A},$$

where $P(\cdot)$ defines the probability. We have $P(0 < t < \mu) = p \leq 0.5$ if and only if $\sigma_1 = \frac{\mu}{\Phi^{-1}\left[\frac{1+2p}{2}\right]}$, where $\Phi^{-1}(\cdot)$ is the cumulative distribution function of the Gaussian distribution with mean 0 and standard deviation 1. In the case that μ is big and σ_1 is small (*e.g.* $\sigma_1 \leq \mu/3$), then $P(0 \leq t \leq \mu) \approx \frac{1}{2}$. For these examples, we can assume $\mathcal{F}_{tot} = \frac{\mathcal{A}}{2}$. Moreover, if \mathcal{F}'_{tot} is the dimensionless total non-thermal heat flux, then $\mathcal{F}_{tot} = \mathcal{F}'_{tot} \times 10^9 \text{ ergs cm}^{-2}$.

In the case of the Full Gaussian the heat flux function is given by the form:

$$\mathcal{F}(t) = \frac{\mathcal{A}}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{(t-\mu)^2}{2\sigma_1^2}\right].$$

We have $P(t > 0) = p$ (with $0.5 < p \leq 1$) if and only if $\sigma_1 = \frac{\mu}{\Phi^{-1}(p)}$. If again, *e.g.*, $\sigma_1 \leq \mu/5$, then $P(t > 0) \approx 1$ and for these cases we can assume $\mathcal{F}_{tot} = \mathcal{A}$.

3.3. Parameters with the EBTEL Model

The thermal heating rate is assumed to be of the form:

$$Q(t) = \alpha\mathcal{F}_1(t) + B. \tag{4}$$

Here Q is the direct heating rate; $B = B' \times 10^{-5}$ ergs cm⁻³ s⁻¹ is a constant background heating rate that is occurring (B' is dimensionless and is included as a parameter); $\mathcal{F}_1(t)$ is the non-thermal heating rate with $\mathcal{F}_1(t) = \frac{\mathcal{F}(t)}{L}$ where $L = L' \times 10^9$ cm is the loop's length from the top of the chromosphere to the apex (parameter); and α is a factor that defines which of the two heating functions (thermal or non-thermal) is dominant assuming that the background heating rate is negligible comparing the other two heating sources. Under this assumption, if $\alpha = 1$ we have equal amounts of thermal and non-thermal heating, if $\alpha > 1$ then thermal heat flux is more dominant, whereas if $\alpha < 1$ then non-thermal heat flux is more dominant. In particular, we will be examining comparisons of the form: $H_0 : \alpha \neq 1$, $H_1 : \alpha = 1$, $H_2 : \alpha > 1$ and $H_3 : \alpha < 1$. Finally, we include the dimensionless radius of the loop r' (assuming the loop to be homogeneous), with $r = r' \times 10^9$ cm, as another parameter into our analysis. The importance of r' is that we calculate $2n^2\pi r^2 L$ in order to compare it with the observed emission measure values.

Initially, when this data-set was firstly assigned for analysis, it was suggested that c_i should be fixed to the values proposed in Klimchuk et al. (2008). However, it was interesting to leave them as parameters to check whether: *i*) we will get data profiles with values close to those proposed in Klimchuk et al. (2008), *ii*) if not, whether these changes affect other parameters and subsequently model comparisons. Therefore, in Section 6.1, these values are assumed to be fixed to given numbers, whereas in Section 6.2 they are assumed to be free parameters which can be determined by the data.

3.4. Dealing with time

The data values have uncertainties (error bars) not only on the y axis (temperature and emission measure), but on the x axis as well (time). These time errors were calculated as being the width of the spline interpolated light curves. There have been some attempts in the past to deal with problems of error bars in both axes (Winsor 1946; Berkson 1950; Jaynes 1991). Reviews of errors in variable models can be found in Casella & Berger (1990); Cheng & Van Ness (1999); Fuller (1987). For extragalactic astronomy applications of these methods the reader is referred to Akritas & Bershady (1996); Kelly (2007); Patriota (2009).

A Bayesian solution to the problem using the reversible jump MCMC algorithm can be found in Henderson et al. (2000). The latter application is utilised in the current paper and according to this, the *true* time¹ should be included as another parameter.

All in all, the parameter set would be: $\mathbf{P} = (\mathbf{P}_1, \mathbf{P}_2)$, with $\mathbf{P}_1 = (t_1, \dots, t_{n_1})$, where t_i is

¹The *true* value of a variable (*e.g.* T_{true}) is based on the *observed* value of that variable (T_{obs}) with the addition of an error (ΔT), *i.e.* $T_{true} = T_{obs} + \Delta T$.

the time for the i th observation, and $\mathbf{P}_2 = (L', \mathcal{F}'_{tot}, \mu, \sigma_1, \alpha, B', r')$ for Section 6.1 and $\mathbf{P}_2 = (L', \mathcal{F}'_{tot}, \mu, \sigma_1, \alpha, B', r', c_1, c_2, c_3)$ for Section 6.2.

4. Defining the Priors for the Parameters

A major difference between Classical statistics and Bayesian statistics is that the former does not consider any prior information for the parameters, whereas the latter can incorporate any available information about the parameters before we observe the data. Further discussion on choosing priors can be found in Kass & Wasserman (1996), while the importance of priors in model comparison can be found in Kass (1993) and Kass & Greenhouse (1989).

There are three main ways of choosing a prior:

1. Subjective: the prior expresses the experimenter’s personal probability that a parameter lies within a specified range.
2. Objective and informative: the experimenter may have information or historical data prior to the experiment being undertaken.
3. Non-informative: expresses ignorance about the value of a parameter and is usually dominated by the likelihood function.

For the current analysis, we incorporated subjective prior wherever it was possible and non-informative in any other case. It is worth noting that in case studies like this, prior distributions can prevent parameters from taking unphysical values or values that do not agree with what we observe. In this case, prior belief refers to any guess we have about a particular parameter that is related to the current observations under consideration, rather than prior belief from previous studies.

The belief we have *before* we undertake the analysis in Section 6 is summarised in Table 1. We conclude from the observations that L' should be around 3 and probably between 2 and 4. We have the option to give more weight to the value of 3 and less weight as we move away from that value. Hence, we assume that L' has a Gamma distribution, $L' \sim \mathcal{G}(37.64, 0.08)$, where $\mathcal{G}(\alpha, \beta)$ is a Gamma distribution with shape parameter α , scale parameter β and mean $\alpha\beta$, which will give a 95% probability between 2 and 4 with mode at 3. For the total non-thermal energy \mathcal{F}'_{tot} we do not have any information at all, therefore $\pi(\mathcal{F}'_{tot}) \propto 1$, where $\pi(\cdot)$ denotes the prior probability density function.

We assume that the flare observations are from the cooling phase of the temperature profile; from the highest temperature lines downwards, the temperature drops continually. Thus, it is very likely that the main flare energy release will occur at most up until the first observation. For this reason we choose the mean of the non-thermal heat flux (μ) to have an upper limit at

the lower error bar of the first observation (which is 884 s after the beginning of the observational period). Since we do not have any other belief about giving weights to any specific values we assume $\pi(\mu) \propto \mathcal{I}(\mu \leq 884)$, where $\mathcal{I}(\cdot)$ is the indicator function (see Section 3.1). Also, $\sigma_1 \sim \mathcal{G}(6.78, 17.28)$ will give 95% probability for the one standard deviation of the heating function to be between 20 and 200 with mode at 100. Since we do not have any knowledge about the background heating, we can assume an improper prior, *e.g.* $\pi(B') \propto 1$. In Section 6.2 we use $\mathcal{G}(1.70, 2.36)$ as a prior for c_1 and the uniform prior $\mathcal{U}(0, 1)$ for c_2, c_3 . The prior for c_1 will give $\sim 99.90\%$ probability for values below 20 and those for c_2 and c_3 come naturally as they will impose an upper limit of one, without favouring any values below unity. Note that the prior for c_2 can be conservative, as there might be evidence that this number is close to the value 0.87 (Klimchuk et al. 2008). Nevertheless, we decided not to include this information as this is not an output from the current observations and because we wanted to test whether our simulations will converge to values close to 0.87.

For the radius of the loop as observed in Figure 1 from Raftery et al. (2009), we assume that the ratio $R = \frac{r'}{L'}$ is probably 1/6. It is almost certainly no more than 1/2; so we give the ratio a 99.73% probability to be inside the interval $[0, 0.5]$. For this reason we have $R \sim \mathcal{G}(6.03, 0.03)$. Since $r' = R \times L'$ and R, L' have known prior distributions, we can simulate the prior distribution of r' to be $\mathcal{G}(5.23, 0.12)$, which will give 95% probability between $[0.20, 1.24]$ with mode at 0.49.

Since there is no prior information regarding the α parameter, it would be preferable to assign an uninformative prior just as with B' . However, due to the fact that it is the “important” parameter of the analysis (if we want to test the hypotheses $H_i, i = 0, \dots, 3$), an improper prior would lead to Bartlett’s paradox (Lindley 1957; Bartlett 1957)². Therefore, for sensitivity reasons, results from an informative prior with 50% probability for $\alpha > 1$ and 50% probability for $\alpha < 1$ will be presented as well. For the H_0 hypothesis ($\alpha \neq 1$), this probability density function will be of the form:

$$\pi(\alpha) = \begin{cases} 0.5, & 0 < \alpha < 1 \\ 0.5 \exp(1 - \alpha), & \alpha > 1; \end{cases}$$

for the H_2 hypothesis ($\alpha > 1$), $\pi(\alpha) = \exp(1 - \alpha)$; and for the H_3 hypothesis ($\alpha < 1$), $\alpha \sim \mathcal{U}(0, 1)$. As far as the true time is concerned we can assume that it is normally distributed about the observed time with standard deviation obtained from the time error bars.

Our last assumption involves the initial values of the temperature, density and emission measure. For this, we assume the pre-flare conditions to have an upper limit of 0.5 MK, $6 \times 10^7 \text{ cm}^{-3}$ and $4 \times 10^{43} \text{ cm}^{-3}$ respectively (Raftery et al. 2009). Since the non-thermal heating rate is almost zero in the beginning, then because of Equation (4), the dominant heat input will be the background heating rate. Hence, these upper limits can indirectly serve as an upper limit to the

²In brief, Bartlett’s paradox states that the less informative the prior of the “important” parameters is, the more the simpler models will be favoured. On the other hand, if the prior of the “important” parameters is very informative, then the more complex model will be favoured. This implies that the model selection method can be very sensitive to the prior information, which is somewhat bizarre.

background heating rate.

5. Model Comparison Methods

5.1. Direct prior inclusion

Model comparison methods using Bayesian statistics are reviewed in great detail in Clyde et al. (2007). An excellent description of how Bayes factor is applied in several problems can be found in Kass & Raftery (1995) and Raftery (1996). The key value for estimating the Bayes factor is to calculate the marginal density function:

$$p(\mathbf{D}|M_k) = \int_{\mathbf{P}} p(\mathbf{D}|\mathbf{P})\pi(\mathbf{P})d\mathbf{P},$$

where M_k is the hypothesis (model) under consideration, $p(\mathbf{D}|\mathbf{P})$ is the likelihood function and $\pi(\mathbf{P})$ is the prior distribution. Roughly speaking, the model with the largest marginal density will provide evidence of its favour among a given class of models. Essentially, the marginal distribution is nothing more than the belief that we have about the data after we have integrated (averaged) over all the parameters, for the specific model under consideration. Then, the Bayes factor for testing two hypotheses (M_k and M_l) is given by:

$$BF_{k,l} = \frac{p(\mathbf{D}|M_k)}{p(\mathbf{D}|M_l)}. \quad (5)$$

From Equation (5) it is clear that if $BF_{k,l} \gg 1$ there is more evidence in favour the M_k model, if $BF_{k,l} \ll 1$ there is more evidence in favour the M_l model, whereas if $BF_{k,l} \approx 1$ the data do not favour either model (but see Kass & Raftery 1995, for more information). Thus, the challenge here is to calculate the marginal densities. Since they are very difficult to calculate analytically, we will turn to numerical methods. In this paper a two-stage MCMC sampler (Mira 2001) is employed. Laplace method and importance sampling methods for marginal likelihood estimation as well as the MCMC sampler with transformation of the parameter set for better convergence are described in detail in Adamakis (2009).

5.2. Indirect prior inclusion

Although Bayes factor has an intuitive interpretation, it has two major drawbacks: (i) it is heavily dependent on the choice of the prior distribution for the important parameters, and, (ii) accurate and efficient computation of the marginal likelihood can be difficult. Attempts have been made to mitigate the first issue by using the intrinsic Bayes factor (Berger & Pericchi 1996, 1998) or the fractional Bayes factor (O’Hagan 1995). According to these methods, the reasearcher should have large enough data, so that the sample can be split into a training set to estimate the posterior

distribution of the parameters, which in turn will be used as the prior distribution in the remaining sample. However, the Bayes factor still depends to the choice of the training set. For this reason, one has to take into account each possible training set and calculate the new Bayes factor as the average of all the Bayes factors. The application of this was not feasible for the present study, where we have only 13 data-points: 7 temperature measures and 6 emission measures. Therefore, in this paper, results from information criteria and posterior deviances will be presented for comparison.

5.2.1. Information criteria

The Akaike Information Criterion (AIC), proposed by Akaike (1974), and the Bayesian Information Criterion (BIC), proposed by Schwarz (1978) are also employed in the current study. AIC propose to choose the model that minimises:

$$\text{AIC} = -2(\log \text{ maximised likelihood}) + 2\lambda,$$

whereas the latter chooses the model that minimises:

$$\text{BIC} = -2(\log \text{ maximised likelihood}) + \lambda \log n,$$

where λ is the number of the parameters and n is the number of the data-points. The difference between two BICs:

$$B_{k,l} = \exp \left[-\frac{1}{2} (BIC_k - BIC_l) \right]$$

can be viewed as an approximation to the Bayes factor without the researcher’s conformable choice of priors. However, one has to come to terms with the fact that, indirectly, a prior similar to the Jeffreys prior is incorporated (Kass & Wasserman 1995). The most important difference between AIC and BIC is that AIC was designed to find the model that produces estimates of the density which is close on average to the true density³, whereas BIC was designed to find the most probable model given the data (Wasserman 2000).

5.2.2. Posterior deviance

Another promising method for comparing different models can be to only use the posterior deviance distribution of each competing model (Aitkin et al. 2009; Liu & Aitkin 2008). According to this method, suppose that $p_k(\mathbf{P}^{[t]}|\mathbf{D})$ is the t th independent MCMC draw of the posterior likelihood function given the data \mathbf{D} and model k . Furthermore, assume that the t th independent MCMC draw of the posterior deviance function for model k is $D_k(\mathbf{P}^{[t]}|\mathbf{D}) = -2 \log p_k(\mathbf{P}^{[t]}|\mathbf{D})$, where $\log(\cdot)$ is the natural logarithm function. From this, the distribution of the difference between two

³Close is measured by the Kullback-Leibler distance.

model deviances can be calculated as $D_{k,l} = D_k - D_l$. Therefore, the posterior deviance difference distribution can be used to calculate the probability $P[D_{k,l} > \beta | \mathbf{D}]$. For large numbers of β , the higher this probability the stronger evidence there is in favour of model l against model k . As is always the case with model selection, subjectivity comes before objectivity. That is, in order to objectively choose a model, we need to subjectively specify a rule that will lead us to this choice. We present here three different ways that can help a researcher to use posterior deviances for model selection.

1. The choice of the value of β can vary for different applications. For example, for $\beta = 0$ our value of interest is translated to the probability of model k having just a higher deviance than model l . In order to make an association with Classical statistics, the difference between the deviances of two nested models is approximately a chi-square distribution with df degrees of freedom, where $df = \nu_l - \nu_k$, the difference between the number of parameters estimated. Then the difference between the two deviances is compared with $\chi_{1-\alpha;df}$, the value that leaves probability α to the right tail of a chi-square distribution with df degrees of freedom. For example, if $\alpha = 0.05$ and $df = 1$ then $\chi_{0.95;1} = 3.84$. In this case we can choose $\beta = \chi_{1-\alpha;df}$. Aitkin et al. (2009) use $\beta = 4.4$ as this gives a posterior probability of 0.9 for model l , under the assumption of equal prior probabilities on each model. Their suggestion is that if $P[D_{k,l} > 4.4 | \mathbf{D}] > 0.9$ then there is quite strong evidence in favour of model l against model k .
2. One can also try to find the value of β that gives:

$$P[D_{k,l} > \beta | \mathbf{D}] = 0.50, \quad \beta \geq 0. \quad (6)$$

If we adopt the table in Kass & Raftery (1995), then we will end up with Table 2 for the range of β . It is worth noting that these numbers are driven more from intuition, rather than a scientific justification.

3. Another way to quantify our beliefs is to calculate the value of γ that gives

$$P[D_{k,l} > 0 | \mathbf{D}] = \gamma, \quad \gamma \geq 0.50. \quad (7)$$

Comparison of the three methods There appears to be a connection between β, γ, df and ν_k . That is, for fixed values of β and df the more parameters in a model, the smaller the value of γ , in order to obtain the same information. For example, if we adopt the approximation $D_k(\mathbf{P} | \mathbf{D}) \approx D(\hat{\mathbf{P}}) + \chi_{\nu_k}$, where $D(\hat{\mathbf{P}})$ is the frequentist deviance and χ_{ν_k} has a chi-square distribution with ν_k degrees of freedom (Spiegelhalter et al. 2002), and also $\beta = 2, df = 2$ then for $\nu_k = 1$ we get $\gamma = 0.86$, whereas for $\nu_k = 10$, we get $\gamma = 0.63$. Figure 2 depicts how γ varies with ν_k for $df = \{1, 5, 10\}$ and $\beta = \{2, 6\}$. On the other hand, if we bin the values of γ , then β will vary according to γ, df and ν_k . This result implies that in order to quantify our belief one should bin the values of β or γ but not both.

In Section 6 results for Bayes factors, AIC, BIC and posterior deviances are presented. Regarding the posterior deviances in particular, although we show results from the above three methods, we favour the second method, *i.e.* comparing the value of β that satisfies Equation (6) with the values in Table 2.

6. Results

The aim of this analysis is to address the following questions:

1. Which of the Half Gaussian and Full Gaussian functions fits better to the data we have?
2. Which of the thermal and non-thermal heat fluxes is dominant?
3. Should the c_i values remain fixed or should they be free parameters?

The second question is the most important regarding the physics of the system. The other two questions are associated more with the statistical analysis. Nevertheless, they can indirectly affect which heat flux is most dominant because they contain relevant parameters. The estimations of the marginal densities of the different models that are shown in Figure 3 are presented in Table 3. For all the statistical models we have assumed that the mean energy of the accelerated (non-thermal) electrons impinging the chromosphere is $\mathcal{E} = 15$ keV. Different values of \mathcal{E} have a minor effect on the temporal profiles.

Our decision about α will be based on AIC, BIC and the posterior deviance. Bayes factors are presented as well for both proper and improper priors for α . However, since there is no real prior information for α , Bayes factors should be used for guidance and not for decisioning.

6.1. c_i Fixed

6.1.1. Model selection

The c_i parameters are fixed at $c_1 = 4, c_2 = 0.87, c_3 = 0.72$ due to an acceptable overall agreement with 1D HD simulations (Raftery et al. 2009; Klimchuk et al. 2008). From Table 3 and with a non-informative prior for α we can derive that the Bayes factor between the Full Gaussian and the Half Gaussian heating function is 42.10 using the Importance sampling estimator. This provides “strong” evidence in favour of the Full Gaussian, according to the table in Kass & Raftery (1995). Similar decision is reached when comparing the marginal densities for informative α prior. Comparing the solid and the dashed lines at the left panel of Figure 4 we can conclude that Full Gaussian gives a better fit to the data. Also, $P[D_{5,10} > 0|\mathbf{D}] = 0.86$, $P[D_{5,10} > 4.4|\mathbf{D}] = 0.64$ and $P[D_{5,10} > 6.3|\mathbf{D}] = 0.50$. According to Table 2 there is “strong” evidence in favour of the Full

Table 1:: Prior belief about the parameters.

	mode	95% probability	prior distribution
L'	3	[2,4]	$\mathcal{G}(37.64, 0.08)$
F'_{tot}	NA	NA	$\propto 1$
μ	NA	NA	$\propto \mathcal{I}(\mu \leq 884)$
σ_1	100	[20,200]	$\mathcal{G}(6.78, 17.28)$
α	NA	NA	$\propto 1$
B'	NA	NA	$\propto 1$
r'	0.49	[0.20,1.24]	$\mathcal{G}(5.23, 0.12)$
c_1	NA	NA	$\mathcal{G}(1.70, 2.36)$
c_2	NA	NA	$\mathcal{U}(0, 1)$
c_3	NA	NA	$\mathcal{U}(0, 1)$

Table 2:: Rule of thumb for quantifying our belief when comparing between two models. The values of β satisfy Equation (6).

β	Evidence
0-2	Not worth more than a bare mention
2-6	Positive
6-10	Strong
> 10	Very strong

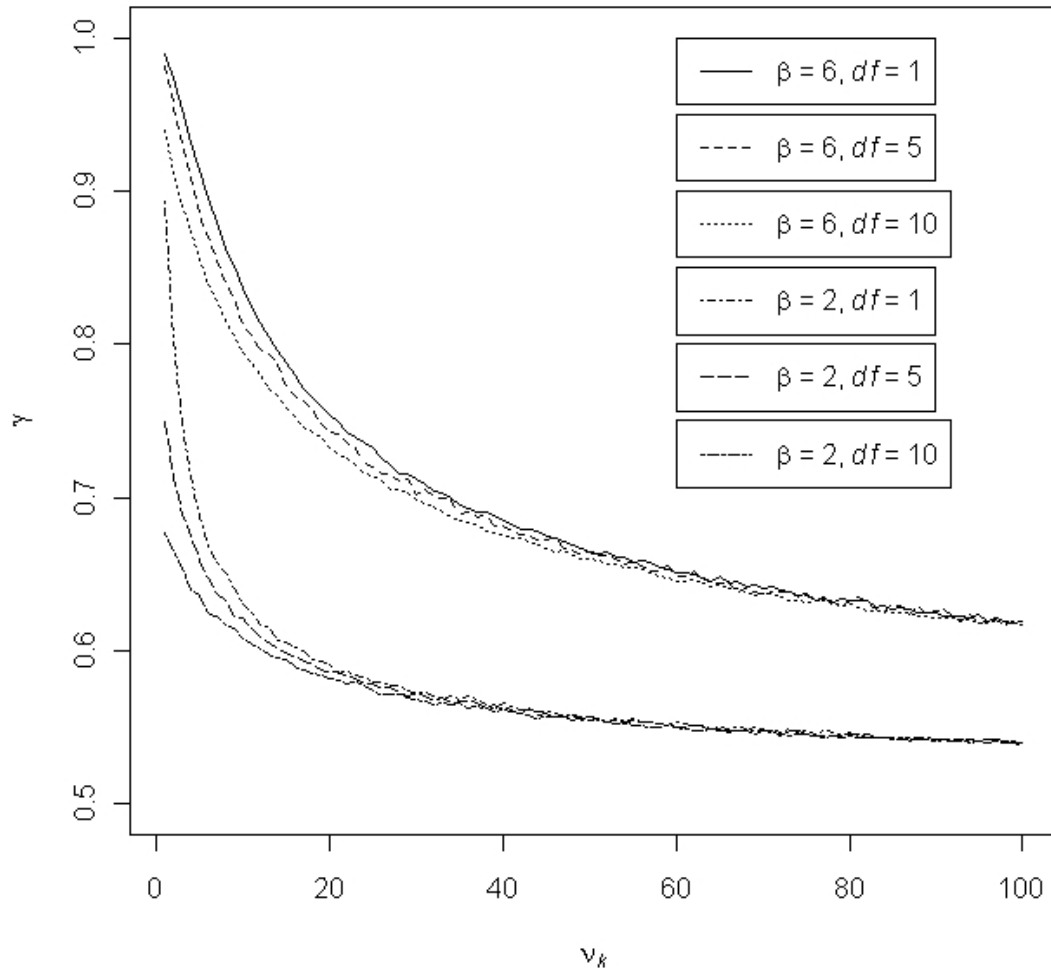


Fig. 2.—: γ as a function of β, df and ν_k . If we assume that the values of β that satisfy Equation (6) can be binned as in Table 2 in order to quantify our beliefs, then for a given value of β the value of γ that satisfies Equation (7) varies according to df and ν_k .

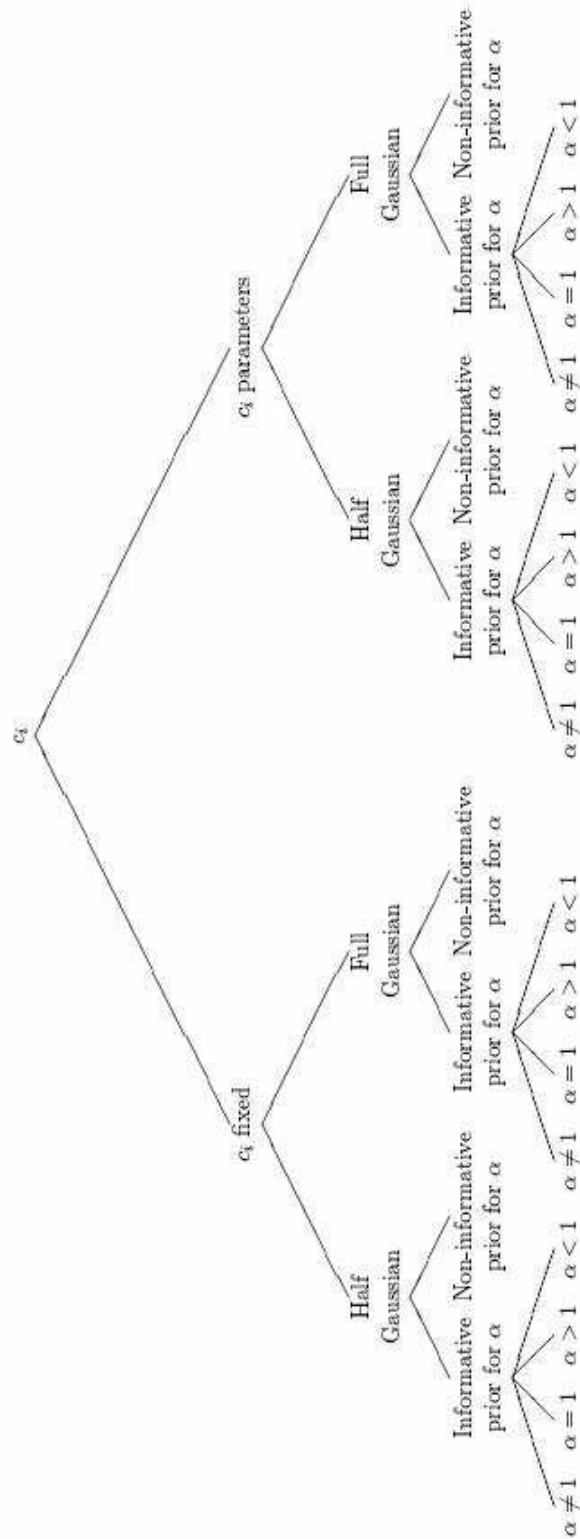


Fig. 3.—: Tree diagram that depicts all the different hypotheses we have used. The purpose of using these hypotheses is to compare them and select the one that best describes the data-set we have.

Table 3:: Model selection criteria according to marginal densities, AIC and BIC produced under several hypotheses using the EBTEL model to compare with temperature and emission measure profiles (Raftery et al. 2009). This table produces results for both c_i fixed and free parameters for comparison tests. The logarithmic marginal densities are estimated using: 1: Laplace method with posterior covariance matrix, 2: Laplace method with robust posterior covariance matrix, 3: Importance sampling estimation with the probability density from the first stage of the two-stage sampler as the additional probability density. See Adamakis et al. (2010) for more information about these estimations.

				Log-marginal densities			AIC	BIC
c_i	Heat	α prior	Models	1	2	3		
fixed	HG	Informative	$M_1 : \alpha \neq 1$	-796.43	-797.78	-797.66	1589.83	1597.74
			$M_2 : \alpha = 1$	-797.38	-798.19	-797.37	1587.22	1594.57
			$M_3 : \alpha > 1$	-797.20	-798.29	-797.53	1589.83	1597.74
			$M_4 : \alpha < 1$	-797.34	-798.54	-797.99	1590.17	1598.08
		Non-informative	$M_5 : \alpha \neq 1$	-792.88	-793.68	-795.95	1589.83	1597.74
	FG	Informative	$M_6 : \alpha \neq 1$	-793.82	-795.28	-794.84	1582.32	1590.23
			$M_7 : \alpha = 1$	-794.45	-795.39	-794.40	1581.13	1588.48
			$M_8 : \alpha > 1$	-793.87	-794.93	-794.42	1582.32	1590.23
			$M_9 : \alpha < 1$	-794.17	-795.31	-794.73	1582.73	1590.64
		Non-informative	$M_{10} : \alpha \neq 1$	-789.07	-790.48	-792.21	1582.32	1590.23
free par.	HG	Informative	$M_{11} : \alpha \neq 1$	-793.26	-795.98	-798.47	1573.52	1583.12
			$M_{12} : \alpha = 1$	-793.36	-796.26	-798.06	1571.88	1580.92
			$M_{13} : \alpha > 1$	-794.07	-795.91	-797.96	1573.52	1583.12
			$M_{14} : \alpha < 1$	-793.23	-796.60	-797.62	1573.98	1583.58
		Non-informative	$M_{15} : \alpha \neq 1$	-789.65	-792.36	-796.98	1573.52	1583.12
	FG	Informative	$M_{16} : \alpha \neq 1$	-791.06	-793.25	-794.28	1568.66	1578.26
			$M_{17} : \alpha = 1$	-790.26	-792.38	-794.39	1567.62	1576.66
			$M_{18} : \alpha > 1$	-789.99	-792.37	-794.32	1568.66	1578.26
			$M_{19} : \alpha < 1$	-791.48	-793.32	-794.71	1569.05	1578.65
		Non-informative	$M_{20} : \alpha \neq 1$	-786.85	-788.49	-792.02	1568.66	1578.26

Gaussian. Furthermore, both information criteria prefer the Full Gaussian model. Finally, the first column of Figure 5 shows the best fit of the parameters regarding the posterior distributions. Even by eye we can distinguish between the Full Gaussian and the Half Gaussian function.

From all the above, we can conclude that there is enough evidence to support that the Full Gaussian function for the heating profiles is much more adequate than the Half Gaussian, at least for the particular data-set under analysis.

In order to distinguish between the thermal and non-thermal heat fluxes, we have calculated $P[D_{7,8} > 0.6|\mathbf{D}] = 0.50$, $P[D_{9,7} > 0.6|\mathbf{D}] = 0.50$ and $P[D_{9,8} > 1.2|\mathbf{D}] = 0.50$. Otherwise, $P[D_{7,8} > 0|\mathbf{D}] = 0.54$, $P[D_{9,7} > 0|\mathbf{D}] = 0.54$ and $P[D_{9,8} > 0|\mathbf{D}] = 0.58$. According to Table 2 none of the models is “worth more than a bare mention”. On the other hand, both information criteria choose the equal amounts of thermal and non-thermal heating model (M_7), although its AIC and BIC values are very close to the ones of M_8 and M_9 . Furthermore, the marginal densities of M_7, M_8, M_9 are very similar to each other, something that supports the idea that the difference between the models is “not worth more than a bare mention”. Finally, according to the quantiles of Table 4, the 95% credible interval with informative prior for α contains 1, whereas the 95% credible interval with non-informative prior for α does not contain 1. In other words, $P(\alpha > 1)$ is close to 0.50 with informative prior for α , whereas it is close to 1 with non-informative prior for α .

In summary, it is very difficult to distinguish which of the two heating forms is more dominant. Better data are required to address this issue. Since the Full Gaussian is more preferable than the Half Gaussian we will choose to adopt the Full Gaussian results for estimating the parameters.

6.1.2. Parameter estimation

The results of this analysis with non-informative prior for α can be viewed in Table 5. We estimate the mean of L to be 29.20 Mm, which is close to what was used by Raftery et al. (2009) (*i.e.* 30 Mm). The mean total non-thermal heat flux is calculated to be around 79.83×10^9 ergs cm^{-2} , while the mean time for the peak of this function is 853.5 s after the beginning of the flare. For the α parameter a 95% credible interval gives [1.30, 101.30], whereas the probability $P(\alpha > 1) = 55.86\%$ with informative prior for α . We expected the posterior probability of $\alpha > 1$ to be close to 1/2, as the Bayes factor could not provide enough evidence in favour of the hypotheses $\alpha > 1$ or $\alpha < 1$ (Table 3). An interesting point to note here is that if we use non-informative prior for α , we can conclude that since $\hat{\alpha} = 20.90$ then the $\alpha > 1$ hypothesis is preferable. This is something that is misused in astrophysics and can lead to false conclusions. More discussion about this is presented in Section 7.

Table 4: Posterior distribution information for the α parameter. Results are split according to c_i (fixed/parameters), heating flux (Half Gaussian/Full Gaussian) and α prior (informative/non-informative).

			Quantiles			$P(\alpha > 1)$
c_i	Heat	α prior	2.5%	50%	97.5%	
fixed	HG	Informative (M_1)	0.16	0.91	3.11	0.44
		Non-informative (M_5)	0.34	8.10	48.47	0.90
	FG	Informative (M_6)	0.22	1.10	3.70	0.56
		Non-informative (M_{10})	1.30	5.74	101.30	~ 1
free par.	HG	Informative (M_{11})	0.07	0.90	2.93	0.45
		Non-informative (M_{15})	0.42	11.63	32.31	0.91
	FG	Informative (M_{16})	0.13	1.13	4.08	0.57
		Non-informative (M_{20})	1.13	23.48	47.09	0.98

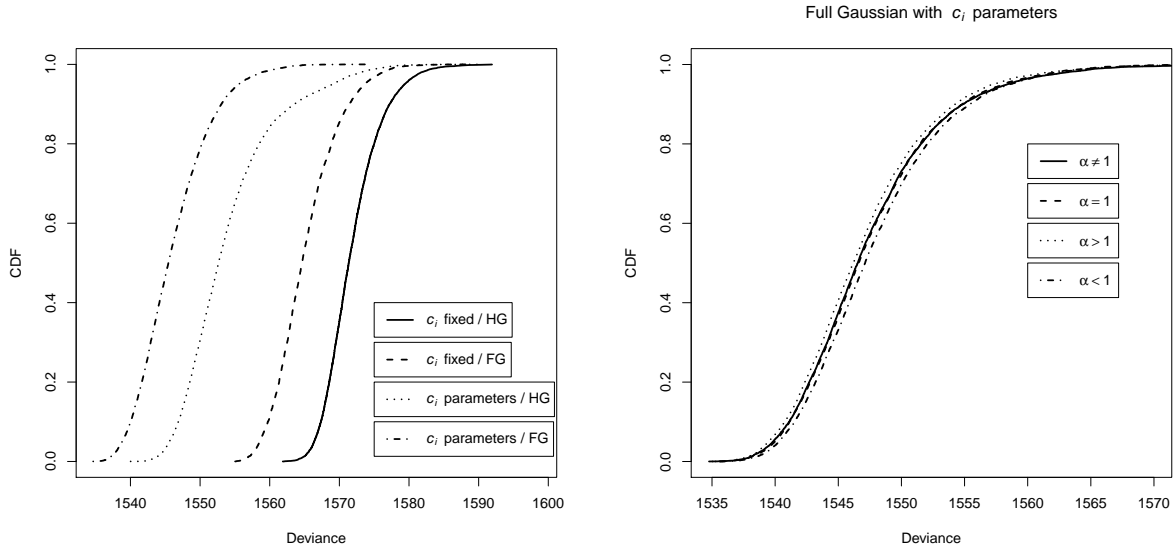


Fig. 4.— *Left*: Cumulative distribution functions for posterior deviance for models with c_i fixed and Half Gaussian (solid), c_i fixed and Full Gaussian (dashed), c_i parameters and Half Gaussian (dotted), c_i parameters and Full Gaussian (dot-dashed). All curves are with non-informative prior for α . The further to the left, the better the model. *Right*: Cumulative distribution functions for posterior deviance for c_i parameters and Full Gaussian are presented for a closer comparison. The dashed line ($\alpha = 1$) over plots the solid line ($\alpha \neq 1$). The dotted line ($\alpha > 1$) provides a slight better fit than the dashed line and the dot-dashed line ($\alpha < 1$) provides a slight worst fit than the dashed line. However, all the hypotheses provide very close lines, an indication that distinction between these models will be very difficult.

Table 5:: Summary of the posterior inference for both c_i fixed and free parameters with **non-informative** prior for α . Results steam from a Full Gaussian non-thermal heat flux profile.

		mean	mode	s.d.	2.5%	50%	97.5%
c_i fixed	L'	2.92	2.86	0.12	2.69	2.91	3.19
	\mathcal{F}'_{tot}	79.83	44.72	62.96	3.77	75.73	179.88
	μ	853.5	881.1	31.6	763.0	863.8	883.1
	σ_1	101.8	94.2	26.9	54.4	99.6	154.6
	α	20.90	8.46	28.23	1.30	5.74	101.30
	B'	0.19	0.25	0.11	0.01	0.20	0.36
	r'	0.47	0.51	0.12	0.25	0.45	0.73
c_i free parameters	L'	3.16	2.80	0.43	2.43	3.13	4.12
	\mathcal{F}'_{tot}	31.99	8.35	40.11	4.84	13.57	153.98
	μ	864.6	880.6	18.7	813.2	870.3	883.5
	σ_1	92.6	82.0	24.2	51.8	90.5	146.2
	α	23.51	38.09	15.94	1.13	23.48	47.09
	B'	0.12	0.21	0.07	0.01	0.17	0.24
	r'	0.67	0.57	0.17	0.39	0.66	1.05
	c_1	2.07	1.39	0.77	0.81	1.99	3.79
	c_2	0.86	0.88	0.07	0.73	0.87	0.99
	c_3	0.75	0.69	0.13	0.48	0.75	0.98

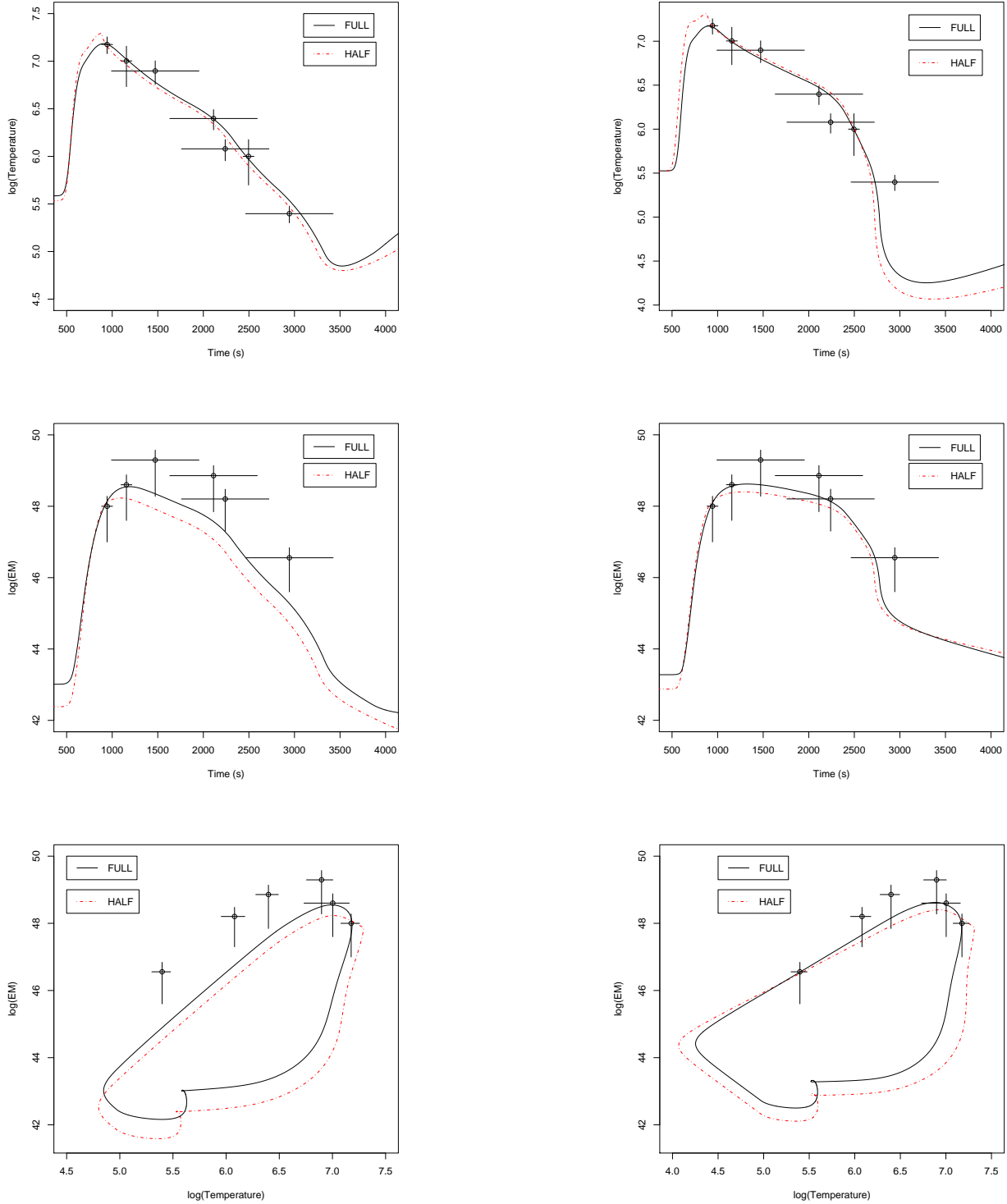


Fig. 5.—: Temperature evolution (*first row*), emission measure evolution (*second row*) and temperature against emission measure (*third row*) using the EBTEL model that best reproduced the observations (maximised the posterior distributions). Solid lines represent the fit from the Full Gaussian function for the thermal and non-thermal heat inputs, whereas dashed lines represent the fit from the Half Gaussian function. For all the curves a **non-informative** prior for α was employed. The *first column* depicts solutions using the c_i fixed, whereas the *second column* depicts

6.2. c_i Parameters

6.2.1. Model selection

We follow the same procedure as in Section 6.1.1 in order to derive the models that best describe the data. Regarding which of the two non-thermal heat fluxes is more dominant, there is no question that the Full Gaussian is more preferable as can be seen from Table 3 (the Bayes factor in favour of the Full Gaussian is 142.54 with a non-informative prior for α , according to the Importance sampling estimator). This can be characterised as “very strong” evidence in favour of the Full Gaussian. The same preference for the Full Gaussian function can be derived by comparing the log-marginal densities for $\alpha \neq 1$ with an informative prior for α . Looking at the dotted and dot-dashed lines at the left panel of Figure 4 we can conclude that Full Gaussian gives a better fit to the data. Also, $P[D_{15,20} > 0|\mathbf{D}] = 0.84$, $P[D_{15,20} > 4.4|\mathbf{D}] = 0.66$ and $P[D_{15,20} > 7.3|\mathbf{D}] = 0.50$. Again, this is “strong” evidence in favour of the Full Gaussian. Furthermore, both information criteria favour the Full Gaussian model. Finally, even by eye, the Full Gaussian model produces better results in the second column of Figure 5. Therefore, the fact that we included the c_i as free parameters, did not alter the outcome compared to the analysis in Section 6.1.1 where the c_i are fixed.

Regarding the thermal and non-thermal heat fluxes, from the right panel of Figure 4 it is apparent that there is a slight better fit to the data with the $\alpha > 1$ hypothesis. Nevertheless, this improvement in fit does not seem to be substantial. We have also calculated $P[D_{17,18} > 0.5|\mathbf{D}] = 0.50$, $P[D_{19,17} > 0.5|\mathbf{D}] = 0.50$ and $P[D_{19,18} > 1.0|\mathbf{D}] = 0.50$. Otherwise, $P[D_{17,18} > 0|\mathbf{D}] = 0.52$, $P[D_{19,17} > 0|\mathbf{D}] = 0.52$ and $P[D_{19,18} > 0|\mathbf{D}] = 0.55$. According to Table 2 none of them is “worth more than a bare mention”. On the other hand, both information criteria favour M_{17} , just as when c_i are fixed. Moreover, the marginal densities of M_{17}, M_{18}, M_{19} are very similar to each other, something that supports the idea that the difference between the models is “not worth more than a bare mention”. Finally, according to the quantiles of Table 4, the 95% credible interval with informative prior for α contains 1, whereas the 95% credible interval with non-informative prior for α does not contain 1. In other words, $P(\alpha > 1)$ is close to 0.50 with informative prior for α , whereas it is close to 1 with non-informative prior for α . Therefore, we reach the same conclusion as in Section 6.1.1: although it is clear that the Full Gaussian function is preferable, the data are not sufficient in order to distinguish between the different heating mechanisms.

6.2.2. Parameter estimation

All the estimations from the posterior distributions of the parameters can be viewed in Table 5, using the Full Gaussian function. We estimate the mean of L to be 31.6 Mm, the mean of the total non-thermal heat flux is 31.99×10^9 ergs cm^{-2} , the mean time for the peak of this function is 864.6 s after the beginning of the flare, the mean of the standard deviation of the Full Gaussian

is 92.6 s, the mean of the α parameter is 23.51, the mean background heating rate is 0.12×10^{-5} ergs $\text{cm}^{-3} \text{s}^{-1}$ and the radius of the loop is 6.7 Mm.

Regarding the c_i ratios, the mean of the ratio between the radiative loss rate of the transition and the corona (c_1) is 2.07, the mean of the ratio between the average coronal temperature and the apex temperature (c_2) is 0.86 and the mean of the ratio between the coronal base temperature and the apex temperature (c_3) is 0.75. Once again, we should bear in mind that although the estimation for α is greater than unity (23.51) with non-informative prior, a more detailed analysis shown in Section 6.2.1 suggests that we cannot distinguish which of the thermal or non-thermal heat fluxes is more dominant. Furthermore, assuming informative prior for α , α is greater than unity with probability 57.13%.

6.3. Comparison Between the Hypotheses: c_i Fixed and c_i Free Parameters.

Considering the non-thermal heat flux, both hypotheses (c_i fixed and c_i free parameters) propose the Full Gaussian statistical model. However, none of them can distinguish which of the two heating mechanisms is dominant (if any). Regarding Question 3, from Table 3 we reach different conclusions depending on the estimation method for the marginal distribution: Importance sampling does not seem to favour any hypothesis, but the other two methods favour the hypothesis c_i free parameters. This can be concluded when a comparison between models M_{10} and M_{20} (or M_6 and M_{16}) is made. On the other hand, the posterior deviances of Figure 4 depict preference to the models where c_i are set as free parameters and the form of the non-thermal heat flux is Full Gaussian. We have also calculated $P[D_{10,20} > 19.2|\mathbf{D}] = 0.50$, $P[D_{10,20} > 0|\mathbf{D}] = 0.995$ and $P[D_{10,20} > 4.4|\mathbf{D}] = 0.98$ which indicates a “very strong” evidence in favour of model M_{20} against model M_{10} . Similar conclusions can be driven with the information criteria, as both of them show clear preference to the c_i free parameter models. Last but not least, the difference between the two hypotheses in Figure 6 does not seem to be very important for the thermal evolution (first row), at least by eye. However, for the data-set under consideration, the difference is more profound for the emission measure evolution (second row). This indicates that the c_i parameters affect the emission measure values more than they affect the temperature values. The fact that there are better values for c_i than 4, 0.87 and 0.72 is even more clear in the third row where temperature is plotted against emission measure.

All the above indicate that the values introduced for c_i when they are fixed are not very good, in terms of maximising the likelihood function. The information added when c_i are free parameters is greater than the price we have to pay for introducing three additional parameters.

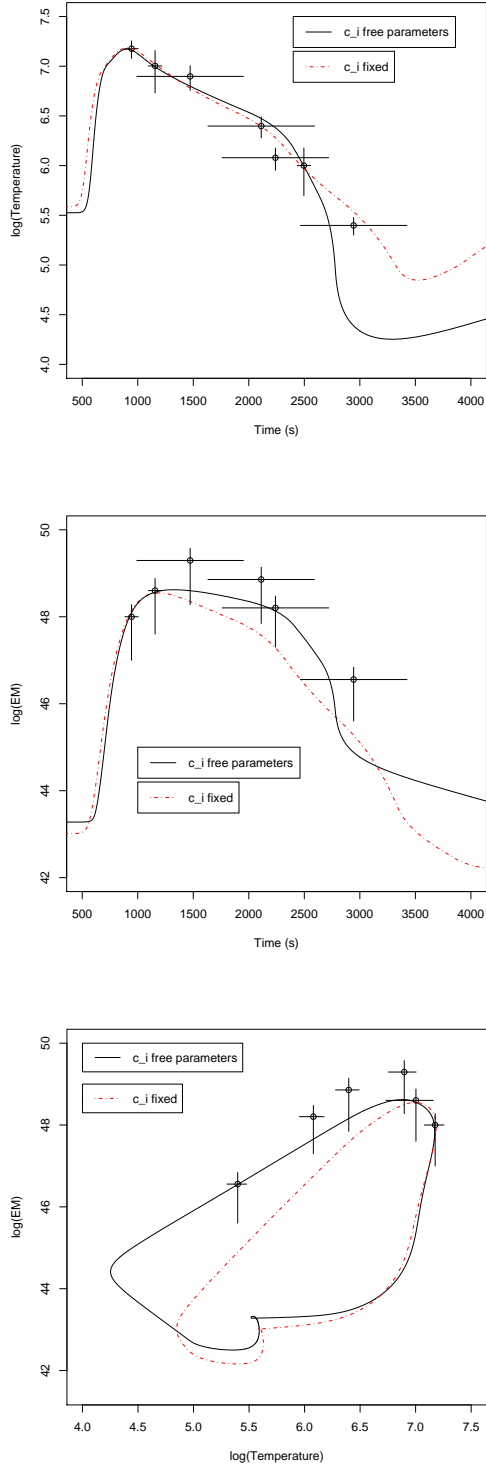


Fig. 6.—: Temperature evolution (*first row*), emission measure evolution (*second row*) and temperature against emission measure (*third row*) using the EBTEL model that best reproduced the observations. Solid lines represent the fit when c_i are floating as free parameters, whereas dot-dashed lines assume c_i to be fixed numbers. Each curve depicts the best fit that maximised the posterior distribution of the Full Gaussian statistical models. A **non-informative** prior for α was employed. Temperature is measured in K, while emission measure is in cm^{-3} .

7. Discussion

In order to reveal the mysteries of the Sun, we can break down our investigations into three major stages: (i) constructing theoretical models, (ii) gathering observations, and (iii) applying a statistical analysis in order to compare different statistical models and/or to restrict the parameters of the models. All of them are of equal importance and give us confidence in our results.

Making inferences solely from the mean or mode of Table 5 would warrant utmost criticism. For model comparison purposes the 95% credible interval is more robust. However, even with the 95% credible interval we can not compare with the $\alpha = 1$ hypothesis. Also, we do not directly include the likelihood function in our calculations. Model comparison techniques using Bayesian statistics (*e.g.* Bayes factor) take into account both the spread of the parameter posterior distributions and the information of the likelihood function. However, Bayes factor is sensitive to the choice of prior distributions and great attention should be paid when applying this method. On the other hand, techniques that are based only on posterior likelihood distributions do not depend on the prior distributions of the parameters — at least not directly. We strongly encourage researchers to employ and compare various statistical techniques when embarking on the subtle topic of model selection.

In this paper, temperature and emission measure profiles produced by the EBTEL model were compared with solar flare observations. The data distribution, the parameter set and the priors employed in this analysis were described in detail. The form of thermal and non-thermal heat input is much better described using Full Gaussian energy profile than Half Gaussian, which is what Raftery et al. (2009) also used.

Apart from choosing which energy profile function was more appropriate for the data-set we analysed, we were also interested in determining which of the thermal or non-thermal heat fluxes was more dominant. The data obtained were not able to provide an answer with great confidence. More data-points may be required in order to address this question. It has also been suggested that it might be possible to have mostly non-thermal heating in the impulsive phase and mostly thermal heating thereafter. To test this, Equation (4) can be replaced by:

$$Q(t) = \begin{cases} \alpha_1 \mathcal{F}_1(t) + B, & t \leq t_1 \\ \alpha_2 \mathcal{F}_1(t) + B, & t > t_1, \end{cases}$$

where α_1, α_2 and t_1 parameters under consideration. If the above statement were true then one would expect that t_1 should be close to where the temperature peak of Figure 5 is observed, as well as $\alpha_1 < 1$ and $\alpha_2 > 1$.

The c_i parameters were given fixed values in the Klimchuk et al. (2008) paper so that the 0D HD model will approximate the 1D HD model. If we assume that the range of these parameters provides sufficient approximations, then we should include them as free parameters. In any case, we disagree with fixing some parameters to certain values, in order to reduce the parameter set, as this might affect the results — unless we have high confidence about these fixed numbers.

The fact that the Bayes factor is not so decisive in choosing between the hypotheses c_i fixed or free parameters is partially an outcome of the conservative prior distributions we have chosen for the c_i parameters (see Section 4). If we had added more information in the prior distribution of c_2 , say $\mathcal{U}(0.5, 1)$ or even better the Beta distribution $\mathcal{B}(38.49, 5.75)$, this would have been in favour of the c_i parameters hypothesis. In comparison, posterior likelihood techniques and information criteria clearly show that c_i should be set as free parameters. More improved estimations for these parameters can be seen from the mean or mode of Table 5 (for the particular data-set we analyse).

Finally, an obstacle presented in this analysis was that of constraining the profiles produced by the EBTEL model. For example, we might not want to restrict the initial values of temperature, density and/or emission measure profiles. This can produce model profiles that are closer to the data profiles, but the initial values might take exceptionally high numbers. For instance, we had undertaken the same analysis without fixing the initial values of temperature, density and emission measure. This resulted in model profiles with initial temperatures⁴ of ~ 3 MK, initial electron density of $\sim 500 \times 10^7 \text{ cm}^{-3}$ and initial emission measures of $\sim 3 \times 10^{47} \text{ cm}^{-3}$. This was because the background heating rate was three orders of magnitude higher than that in Table 5. Apart from unrealistic estimations of the parameters of interest, this could have also led to unreliable Bayes factor estimations with false conclusions regarding the three posed questions in the beginning of Section 6. Naturally, the quality of the output of the analysis is dependent of the quality of the input.

Apart from improved statistical techniques, of equal importance is that improvement upon the observations should be made. This means that future missions with new instrumentation should provide data-sets with a large enough number of observations in order to distinguish between different heating mechanisms. The data-set under consideration provided information only upon the decay phase of the temporal evolution. However, a better time resolution for the rise phase of the temperature will be needed in order to provide a better estimate for the form of thermal/non-thermal heat flux. And of course, a large sample of solar flares will be required.

An assumption made in this paper is that the thermal and non-thermal fluxes have the same form, based on a lack of information on the thermal distribution. However, it would be interesting to test fluxes of different forms that do not depend on each other. Additionally, a further improvement in the EBTEL model is required regarding the non-thermal heat flux, as it is efficient for gentle chromospheric evaporation but suffers from inadequately representing explosive chromospheric evaporation. Last but not least, several other forms of heating input, like proton beams, could be included in an attempt to make the model more realistic.

SA has been supported by a STFC grant. CLR is supported by an ESA/Prodex grant, administered by Enterprise Ireland. The authors would like to express their gratitude to the referee

⁴Initial temperature is assumed to be at the base of the transition region.

for the useful comments.

REFERENCES

- Adamakis, S. 2009, PhD thesis, University of Central Lancashire
- Adamakis, S., Walsh, R., & Morton-Jones, T. 2010, *Solar Physics*, 262, 117
- Aitkin, M., Liu, C., & Chadwick, T. 2009, *Annals of Applied Statistics*, 3, 199
- Akaike, H. 1974, *IEEE Transactions on Automatic Control*, 19, 716
- Akritas, M. & Bershadsky, M. 1996, *The Astrophysical Journal*, 470, 706
- Aschwanden, M. 2005, *Physics of the Solar Corona: An Introduction with Problems and Solutions* (Springer Praxis Books/Astronomy and Planetary Sciences)
- Bartlett, M. S. 1957, *Biometrika*, 44, 533
- Berger, J. & Pericchi, L. 1996, *Journal of the American Statistical Association*, 91, 109
- . 1998, *Sankhya*, 60, 1
- Berkson, J. 1950, *Journal of the American Statistical Association*, 45, 164
- Casella, G. & Berger, R. 1990, *Statistical Inference* (Wadsworth & Brooks, Pacific Grove, CA)
- Cheng, C.-L. & Van Ness, J. 1999, *Statistical Regression with Measurement Error* (Kendall's Library of Statistics 6, Arnold, London)
- Clyde, M., Berger, J., Bullard, F., et al. 2007, in *Statistical Challenges in Modern Astronomy IV*, ed. G. J. Babu & E. D. Feigelson (ASP Conference Series, Vol. 371), 224–240
- Fuller, W. 1987, *Measurement Error Models* (Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York)
- Henderson, R., Morton-Jones, A., & McKnespiey, P. 2000, *Applied Statistics*, 49, 563
- Jaynes, E. 1991, *Straight Line Fitting – A Bayesian Solution* [Internet] (Updated 21 April 1999), Technical Report. Available at <http://bayes.wustl.edu/etj/node2.html> [Accessed 22 March 2008]
- Kass, R. E. 1993, *Statistician*, 42, 551
- Kass, R. E. & Greenhouse, J. 1989, *Statistical Science*, 4, 310
- Kass, R. E. & Raftery, A. 1995, *Journal of the American Statistical Association*, 90, 773

- Kass, R. E. & Wasserman, L. 1995, *Journal of the American Statistical Association*, 90, 928
- . 1996, *Journal of the American Statistical Association*, 91, 1343
- Kelly, B. 2007, *The Astrophysical Journal*, 665, 1489
- Klimchuk, J., Patsourakos, S., & Cargill, P. 2008, *The Astrophysical Journal*, 682, 1351
- Lindley, D. V. 1957, *Biometrika*, 44, 187
- Liu, C. & Aitkin, M. 2008, *Journal of Mathematical Psychology*, 52, 362
- Mariska, J., Doschek, G., & Bentley, R. 1993, *The Astrophysical Journal*, 419, 418
- Milligan, R., Gallagher, P., Mathioudakis, M., et al. 2006a, *The Astrophysical Journal*, 638, L117
- Milligan, R., Gallagher, P., Mathioudakis, M., & Keenan, F. 2006b, *The Astrophysical Journal*, 642, L169
- Mira, A. 2001, *Metron*, 59, 231
- O’Hagan, A. 1995, *Journal of the Royal Statistical Society Series B*, 57, 99
- Patriota, A. 2009, *Statistical Methodology*, 6, 408
- Raftery, A. E. 1996, in *Markov Chain Monte Carlo in Practice*, ed. W. Gilks, S. Richardson, & D. Spiegelhalter (Chapman and Hall), 163–188
- Raftery, C., Gallagher, P., Milligan, R., & Klimchuk, J. 2009, *Astronomy & Astrophysics*, 494, 1127
- Schwarz, G. 1978, *Annals of Statistics*, 6, 461
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. 2002, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 538
- Wasserman, L. 2000, *Journal of Mathematical Psychology*, 44, 92
- Winsor, C. 1946, *Biometrics Bulletin*, 2, 101