Chapter 6

Joint modelling of transit and stellar temperature using a Markov Chain Monte Carlo approach

When searching for small differences between the transit parameters of different light-curve de-trending approaches, we noted the presence of strong correlations between parameters. These were to some extent accounted for by using a residual permutation algorithm to estimate the uncertainties in the transit parameters. However, a more robust method to explore these correlations was desirable.

Additionally, as pointed out by many authors (e.g Seager & Mallén-Ornelas 2003), the detailed shape of the observed transit depends on the stellar density as well as the planet's parameters. This means that if we have any external (or prior) constraints on the stellar density, we should incorporate them in the transit fitting process, in a Bayesian manner. The spectroscopic measurement of the stellar temperature, discussed in Chapter 5, constitutes such a prior constraint. The stellar temperature is directly, although not uniquely, linked to the stellar density as stellar evolution models predict a particular temperature and density for stars of a given mass, age and composition.

The Markov Chain Monte Carlo (MCMC) technique uses a random walk approach to explore a parameter space of arbitrary dimensionality. The MCMC is well suited to the problem of modelling planetary transits, because a) it avoids getting trapped in local minima in the figure of merit space; b) it enables prior information to be incorporated in the figure of merit seamlessly and c) in the process of exploring the parameter space to find the best fit, it also samples the posterior distribution for each parameter, and thus yields robust uncertainty estimates.

In this chapter, I implement a Metropolis-Hastings MCMC algorithm for transit fitting, with specific Bayesian priors on stellar properties. The transits are modelled using the formalism of Mandel & Agol (2002), but care was taken to re-formulate the MCMC

such that the jump parameters (the parameters over which the MCMC random walk occurs) are parameters for which it is reasonable to assume a flat prior distribution. Our formalism, which is derived from that briefly described in Pont et al. (2009), can also optionally incorporate an external constraint on a particular stellar parameter, in the present case the effective stellar temperature, which is implemented as an additional term in the MCMC merit function. The conversion between stellar temperature and density is done using the theoretical stellar evolution models of Girardi et al. (2002).

The MCMC formalism is described in Section 6.1 and applied to the case of CoRoT-2b in Section 6.2. The results are discussed, along with possibilities for future work in this direction, in Section 6.2.3.

6.1 Markov Chain Monte Carlo

A Markov chain is a succession of states where the next state depends only on the current state. The term "Monte Carlo" means that the transition between states is randomly determined based on a probabilistic distribution. The strength of the Markov Chain Monte Carlo method (MCMC) is in the way it samples the grid of parameters, and in the fact that it can identify the global best fit model in the parameters space, rather than a locally best fit model. In the chain, when the current model is worse that the precedent one, instead of systematically disregarding it, the MCMC uses a probabilistic approach to invalidate or validate this step, with lower probability for poorer models. This property gives the MCMC the capability of exploring areas of the parameter space that would not have been explored otherwise. For instance, in a standard χ^2 minimisation technique, only the parameters improving the fit are considered, which means that the convergence algorithm can get stuck in a local minimum and never explore the rest of the parameter space to find the global minimum, especially if only one set of initial conditions are used. Another key advantage of the MCMC is that it enables the posterior probability distributions of the parameters to be estimated.

6.1.1 MCMC implementation

There are different types of Markov Chain Monte Carlo methods, which differ in their random walk nature. The one used in this chapter is a Metropolis-Hastings algorithm which generates a random walk from a proposal distribution of the next step (here, flat or Gaussian probability distributions are used as explained in Section 6.1.3) and proscribes a method to reject the proposed move if the new model is worse than the previous one (here, a function of how much worse the new model is compared to the previous one, is used).

The input to the MCMC are a) the data to be modelled (here, the IRF-filtered planet transit light curve), b) the model to be compared to the data (here, the transit formula-

tion of Mandel & Agol 2002), c) the initial values (which model the data sufficiently well to start the chain), d) the scale sizes (the typical step size in each parameter), and e) the constraints on the model (here, a Gaussian prior distribution on the stellar temperature). The output of the MCMC are the posterior probability distribution of the values of each adjusted parameter. These distributions are used to derive the 1σ uncertainty ranges on the parameters, as explained in Section 6.1.7.

The steps of the MCMC used in this chapter are as follow. The flow chart of this MCMC is presented in Figure 6.1.

1. Compute the model for the initial values set.
2. Calculate the merit function of this model. The merit function used here is the likelihood $L = e^{-\frac{\chi^2}{2}}$ where $\chi^2 = \frac{\sum(\mathrm{data}-\mathrm{model})^2}{\mathrm{DOF}}$ with the number of degree of freedom (DOF) being the difference between the number of data points and the number of adjusted parameters.
3. Select a new set of parameters by making a random step in each parameter. The size of the step in each parameter is drawn from a Gaussian distribution cantered at 0 with standard deviation equal to scale size inputed for this parameter (see section 6.1.4). The scale sizes of the steps need to be large enough to allow a wide exploration of the parameter space, and small enough to keep a reasonable number of accepted steps, steps which are used to build a statistically robust posterior distribution.
4. Compute the model associated to this new set of parameters and calculate its figure of merit (likelihood as explained in step 2).
5. Compute the ratio of the figure of merit of the new set of parameters and the current ones.
   (a) If the ratio is greater than 1, the likelihood of the model has improved and the step is validated by storing the new values of the parameters and the likelihood.
   (b) If the ratio is smaller than 1, the likelihood of the model has worsened and a random number is drawn from a uniform distribution between 0 and 1.
      i. If the value of the ratio is greater than the randomly pulled value, the step is validated by storing the new values of the parameters and the likelihood. The closer to 1 the value of the ratio is, the higher the probability of the new step to be kept, as there are more values between 0 and 1 smaller than the value of the ratio.
      ii. If the value of the ratio is smaller than this number, the new step and likelihood are disregarded and the current parameters are re-stored.

6. Go back to step 3, and iterate N times. The number of iterations used in this chapter is 500000.

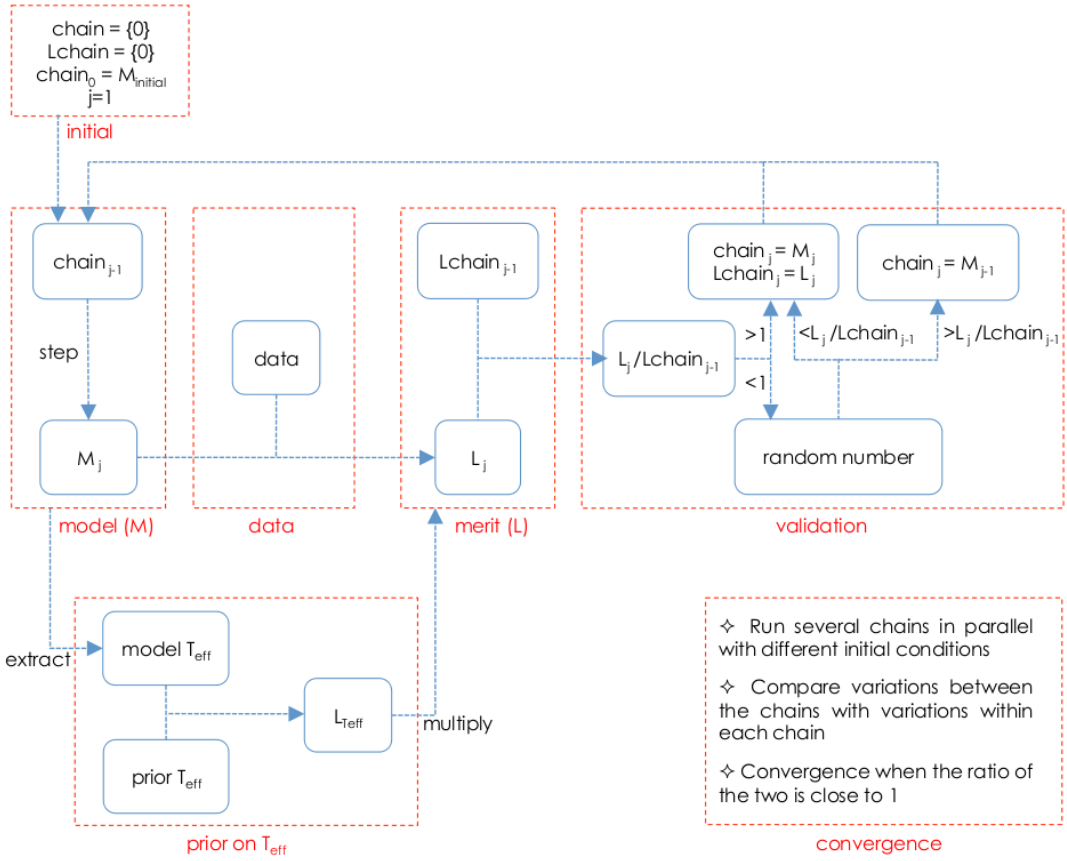

Figure 6.1: MCMC flow chart

6.1.2 Jump parameters

When using the MCMC chain to construct posterior probability distributions for the parameters of the model, a flat prior for the jump parameters are implicitly assumed. Another prior can be applied by multiplying the distributions with the relevant prior distribution, but it is easier and more robust to formulate the model in terms of parameters for which a flat prior is a reasonable assumption.

The input parameters of the Mandel & Agol (2002) formalism is the projected separation between the stellar and planetary disk centres z, in units of stellar radii, the planet-to-star radius ratio $R_{\rm p}/R_\star$, and the linear and quadratic limb-darkening coefficients u_a and u_b. Kepler's laws of planetary motion yield z as a function of time for a given period P, orbital inclination i relative to the plane of the sky, system scale $a/R_\star$, eccentricity e, longitude of periastron w, and time of passage at periastron T_0.

Random inclinations in three dimensions give rise to a flat distribution in $\cos i$ rather than in i. We therefore use the impact parameter $b \equiv \frac{a}{R_\star} \cos i$ as one of the jump parameters. At each step of the MCMC, b is converted to an inclination using:

$$i = \arccos\left(b \left[\frac{a}{R_\star}\right]^{-1} \right) \tag{6.1}$$

Whilst the radial velocity signal of a planet is exclusively sensitive to $e \sin w$, the light curve can (if a secondary eclipse occurs) constrain $e \cos w$. The jump parameters adopted in the most general implementation of the MCMC formalism used here are $e \cos w$ and $e \sin w$, rather than e and w, though in any case, we do not deal with radial velocity data or with eccentric orbits in the present thesis.

As the stellar density is directly constrained by the transit, it is a natural choice of jump parameters (instead of the system scale). However, in a random population of stars in the Galactic disk, there is no reason to expect a uniform density distribution (most stars would be found on the low-mass end of the main sequence). To take this into account, we use a grid of stellar evolutionary models (Girardi et al., 2002) interpolated by Aparicio & Gallart (2004) and resampled by Pont & Eyer (2004), with the density of models proportional to the number of stars expected at a given mass and evolutionary stage, assuming the initial mass function (IMF) derived for the Galaxy, and a uniform age distribution.

This resampled grid, kindly supplied by F. Pont (private communication) was sorted in density, and sorted in a file which we refer to as Padova2002 in the rest of this chapter. We use k, the index in the grid of models contained in the file Padova2002, as one of the jump parameters of the MCMC. To each k corresponds a particular star mass, luminosity and temperature. We use the luminosity and temperature to calculate a stellar radius, and obtain a system scale as follows:

$$\frac{a}{R_\star} = \left[\frac{P^2 G}{4\pi^2}\right]^{1/3} \frac{M_\star^{1/3}}{R_\star} \tag{6.2}$$

The stellar temperature in Padova2002 span from 2300 to 26800 K, the stellar masses span from 0.668 to 10.716 $M_\odot$, and the stellar radii from 0.628 to 455.361 $R_\odot$.

6.1.3 Incorporating external constraints

A gaussian prior is associated to the effective temperature of the star $T_{\rm eff}$ which can be derived from the spectroscopic analysis of its atmosphere. The transit model gives an estimate of the density of the star, from which using stellar evolution tracks, an associated stellar temperature can be derived. These two temperatures can be compared in the MCMC with a likelihood function which favours transit models with an associated temperature close to the effective temperature. This prior on the stellar temperature is included in the MCMC evaluation of best transit model by calculating the likelihood

$L = e^{-\frac{\chi^2}{2}}$ of the stellar temperature associated to each transit model compared to the spectroscopic temperature – where χ^2 measures the difference between the two temperatures –, and by multiplying the likelihood in step 2 of the MCMC by the likelihood on the temperature.

6.1.4 Step size

The size of the step in each parameter needs to be chosen so that in each parameter chain it takes the MCMC several steps to reach the extremes of the explored values. This gives confidence in the accuracy and coverage of the MCMC exploration, as small steps mean good sampling. However, if the steps are too small, it will take the MCMC a larger number of iterations to explore the same region of the parameter space. The method used to find the optimal scale size for each parameter is as follows:

1. Run a short chain (e.g. 1000 iterations) with initial step sizes equal to the uncertainty on the initial parameters
2. Calculate the standard deviation of the chain for each parameter and the number of accepted steps
3. Adjust the step size of each parameter so that it is smaller or of the same order as the standard deviation of the short chain for this parameter, and so that the number of accepted steps is close to 50%.

6.1.5 Chain length

The number of steps in the MCMC chain needs to be several times the correlation length of each parameter chain. There are two methods to test if the chain is long enough: 1) calculate the correlation length as explained below and compare to the length of the chain, or 2) calculate the Gelman & Rubin statistic of convergence as explained in Section 6.1.6.

A chain length several times the parameter correlation length means that the chain has explored the structure of the parameter space several times. There is thus less chance for the convergence to get stuck in a local minimum and the best model solution is thus more robust. The method used to check if the number of MCMC iterations used allows to derive statistically robust parameters is as follows:

1. Calculate the autocorrelation of the chain: $A_j(\theta) = \sum_i \theta_i \times \theta_{i+j}$ for each parameter θ. The autocorrelation length is given by the number of iterations needed to bring the autocorrelation from maximum to zero
2. The length of the chain should be several times (e.g. $>$10) the value of the autocorrelation length. The longer the chain the more robust the solution found for the best model.

When the MCMC has done a sufficient number of iterations, it will have sampled the posterior distributions of each parameter evenly, so that the distributions of the values of each parameter along the chain should be relatively smooth, and representative of the true underlying distributions. But running more iterations requires more computing time. For instance, running 100000 MCMC steps with four free parameters, on a 150-day-long and 516s-sampling light curve (CoRoT long run) takes close to 48h on an intel MAC pro (pre 2008 model) with 32Gb of RAM and 4 processor cores of 2.66GHz each. Here are two approaches used in this work to reduce the computing time:

1. The MCMC can be sped up by reducing the number of data points: 1) by binning the light curve, and 2) by using only the light curve around the transits. In the case of the transit light curve of CoRoT-2b studied here, the number of points was reduced from $\sim$19000 to $\sim$1200, by binning each 8 points in the phase folded light curve and by keeping only the points within 0.2 phase units of the transits. This speeds up the MCMC by a factor 10.

2. To increase the number of iterations with limited computing time, several independent chains can be run simultaneously. Each chain has a different starting point, and needs to be longer than the burn-in phase. The burn-in phase of a chain is the number of steps the chain takes before converging on the area of the true parameter and starts exploring the uncertainty area of this parameter. In this chapter, after visual check of the chains, the burn-in phase is taken as 7000 steps. A longer burn-in phase made little difference in the posterior distributions while shortening the valid length of the chain. The chains can then be combined together to obtain distributions with a larger number of iterations, i.e more statistically robust distributions with shapes that will not evolve with more iterations.

6.1.6 Gelman & Rubin diagnostic for convergence

To get a long chain, another option is to combine independent chains which have different starting points but same statistics, i.e. which have converge to their common distribution.

The Gelman & Rubin statistic test of convergence (Gelman & Rubin, 1992; Brooks & Gelman, 1997) checks if independent chains of the same length have converged to the same probability distribution, i.e. have similar statistics. It does so by comparing the variance between several chains B (Equation 6.5) with separate starting points and the variance within each chain W (Equation 6.6). If all the chains are sampling the same distribution and doing it in a complete way, these variances should be the same, i.e S=1 for S defined in Equation 6.3.

When the chains have converged, the statistic S (equation 6.3) is equal to 1, as the variations between the chains (V, equation 6.4) should be of the same order as the variations within the chains (W, equation 6.6). The independent chains can then

be combined together to form a longer chain with a more statistically robust posterior distribution. In this chapter, all chains are started within one scale size of the initial value.

For each parameter, the statistic S comparing several chains (m chains or n iterations each after cutting away the burn-in phase) can be calculated with:

$$S = \sqrt{\frac{V}{W}} \tag{6.3}$$

where V is calculated as:

$$V = \left(1 - \frac{1}{n}\right) W + \frac{1}{n} B \tag{6.4}$$

with B, the variance between the chains, calculates as:

$$B = \frac{n}{m-1} \sum_{j=1}^{m} (\bar{C}_j - \bar{C})^2 \tag{6.5}$$

where $\bar{C}_j$ is the mean of the j^{th} chain, and $\bar{C}$ is the mean of all the chains.
W, the sum of the variance of each chain, is calculated as:

$$W = \frac{1}{m(1-n)} \sum_{j=1}^{m} \sum_{l=1}^{n} (C_{j,l} - \bar{C}_j)^2 \tag{6.6}$$

where $C_{j,l}$ is the parameter value of the l^{th} step of the j^{th} chain.

The MCMC runs are presented by a group of three chains, to evaluate the convergence of the individual chains, before combining them into a longer one. Usually, groups of five chains are run and compared to each other to check their convergence. Here, three chains were used to start with. Future work will include running the MCMC on more individual chains.

6.1.7 Posterior distributions

Once the chain has converged, the shape of the posterior distribution is statistically robust and will not change significantly with more iterations. This distribution is a good representation of the true posterior distributions of the parameter. The best model parameters and associated uncertainties can be derived from it as follows. The best model is taken as the one with the largest likelihood, i.e. the minimum χ^2 to the data. The uncertainty on the best model value is defined by the 1σ spread of the posterior distribution, which is calculated as the range in value enclosing the 68% of the chain around the median, i.e. 34% each side of the median value of the chain.

The set of parameters with the largest occurrence in the posterior distributions can be slightly different from the parameter set with the largest likelihood, but should be

consistent within the 1σ spread of the distribution. If this is not the case, it is an indication that the length of the chain is not long enough.

The choice of the "best" value depends on what the value is going to be used for. The "best-fit" set of parameters maximises the figure of merit. It is an appropriate choice for the "best" value when creating a transit model to compare to data of the same nature, or when comparing between different data analysis techniques such as in Table 6.3. The most probable value of the chain maximises the posterior probability distribution. This value is a better estimate of the true physical value of that parameter, which is true despite the fact that the model generated from the median or most probable values of each parameter may not have a high merit function. The median or most probable values should be used, for example, when making comparisons with theory. Usually, the median is used rather than the most probable value because the median is a better single-number representation of a distribution (e.g. the case of a distribution with two peaks).

6.1.8 Limitations

The main limitation of the MCMC method is that it requires a lot of computing time to derive chains which have convergence, especially if the number of data point is large. However, it is still faster than a systematic exploration of the same parameter space at similar resolution using a regular grid.

The computing time of the MCMC can be reduced by decreasing the number of data points. One way of reducing the number of data points in a light curve is by binning the points. However, too much binning reduces the resolution of the light curve in the ingress and egress of the transit, which translates into a lower precision on the derived planet parameters.

In this chapter, the chain is left to start not far from the parameter values in the discovery paper. The advantage of this choice is that the burn-in phase of the chain is shorter (fewer iterations) and the valid part of the chain longer. The drawback is that if the region around the initial value is a local minimum and if the MCMC steps are too small, the MCMC will spend a lot of time in this area before exploring other regions of the parameter space. If the step size is too large, more iterations will be rejected on the basis of being worse than the previous steps, and the chain will need to be longer to count sufficient accepted steps and build statistically robust posterior distributions.

An MCMC is also only as good as the priors assumed in it. It is a caveat that should be kept in mind when comparing the method to other fitting techniques.

Another limit to the current MCMC is the sampling of the stellar evolution model that it uses. The accuracy and precision of the derived parameters can only be as good as the finest mesh of the grid of models.

6.2 Application to CoRoT-2b

6.2.1 Method

The IRF-filtered light curve of CoRoT-2 is corrected for the 5.6% contaminant flux due to a star falling into the CoRoT mask of CoRoT-2. This is done by subtracting 0.056 to the normalised IRF-filtered light curve and re-normalising the resulting light curve.

The MCMC is applied to fit the transit of CoRoT-2b in the IRF-filtered CoRoT light curve of CoRoT-2. The IRF-filtered light curve used is filtered with a time scale for stellar variability of 0.25 days, as this version has less residual stellar variability than the 0.5 day version while the transit shape is still well recovered (Chapter 3 Section 3.2.3). To speed up the MCMC and allow us to run a larger number of iterations, the number of data points are reduced by binning the phase folded filtered light curve with bin size of 0.0006 phase units (0.06% of the planet's orbit), and keeping only the section within $\pm$0.2 phase units of the transit centre. The uncertainty associated to each bin is taken as the standard deviation of the points binned together in this bin. In the MCMC, the uncertainty associated to the data set is taken as the maximum of these uncertainties. The data points were re-normalised by fitting a 2^{nd} order polynomial function about the transit and the data points were divided by this fit.

The MCMC is run varying the time of transit T_0, the impact parameter b, the planet-to-star radius ratio $R_{\rm p}/R_\star$, and k the line index of stellar density in Padova2002. P, e and w are fixed to the value in Alonso et al. (2008). The limb darkening coefficients are fixed to the quadratic limb darkening coefficients corresponding to the CoRoT bandpass, derived from Sing (2010). The adjusted parameters are then translated into T_0 ($T_{0p}+cst$) the epoch of the transit centre, $a/R_\star$ and i.

When the MCMC is run with the prior on the $T_{\rm eff}$ the stellar atmosphere parameters of CoRoT-2 are $T_{\rm eff}$=5516$\pm$33 K, $\log g$=4.3$\pm$0.2, (M/H)=0.0$\pm$0.1. The value of $T_{\rm eff}$, used here, is the one derived using the equivalent width ratios, the other parameters have not been re-calculated and are kept the same as in the discovery paper. The associated quadratic limb darkening coefficients in the CoRoT bandpass are u_a=0.478$\pm$0.010 and u_b= 0.205$\pm$0.007. These are the coefficient used in all the fitting procedures below , including the LMA, to allow a direct comparison of the derived parameters.

First, no prior on the stellar temperature is applied, then a gaussian prior on the $T_{\rm eff}$ is added later. The stellar temperature used as prior is the temperature derived for CoRoT-2 in Chapter 5 using the temperature calibrated equivalent width line ratios: $T_{\rm eff}$=5516$\pm$33 K.

The scale stepe sizes chosen when no prior on the $T_{\rm eff}$ was applied, is $2\ 10^{-5}$ for T_0, $4\ 10^{-3}$ for b, $2\ 10^{-4}$ for $R_{\rm p}/R_\star$, and 1000 for k. The scales were obtained by trial and errors starting from scale sizes equal to the uncertainty on the initial values and reducing the scale size until the number of accepted steps was close to 50% when no prior on the $T_{\rm eff}$ was used, and close to 30% when a prior on the $T_{\rm eff}$ was applied. When the prior

on the $T_{\rm eff}$ is used, the number of accepted steps decreases due to the additional constraint on the $T_{\rm eff}$. If less steps are accepted, a longer chain is needed to keep the statistic of the posterior distributions robust. To avoid having to run longer chains, two adjustments are made to increase the number of accepted steps when a prior on the $T_{\rm eff}$ is applied. First, prior to running the chain, the stellar evolution model file was re-arranged to keep only the models with a $T_{\rm eff}$ at the prior temperature plus or minus 4$\times$ the uncertainty on this prior, so that when stepping in k the $T_{\rm eff}$ of more models will be compatible with the prior on the $T_{\rm eff}$. Secondly, the scale size in k was reduced to 100.

The planetary parameters (T_0, $R_{\rm p}/R_\star$, b) from Alonso et al. (2008) were first used as initial values to run the MCMC. The MCMC translates k into $a/R_\star$ using stellar evolution models, and b and $a/R_\star$ into i. As the sampling of the grid in the stellar evolution models is finite, the derived $a/R_\star$ and i do not correspond exactly to the values in the discovery paper, and the transit model derived from these value is too different from the data (large χ^2) to allow the MCMC to start. The initial parameters used are T_0 from the discovery paper, b=0.05, $R_{\rm p}/R_\star$=0.16, and k=509965 (no prior on $T_{\rm eff}$) or k=81492 (prior on $T_{\rm eff}$). k=509965 correspond to $T_{\rm eff}$=5521 K, $M_\star$=0.888 $M_\odot$, $R_\star$=0.881 $R_\odot$ in the full stellar evolution model file used. k = 81492 corresponds to the same model in the file re-arranged to increase the number of accepted steps when using a prior on the $T_{\rm eff}$. $T_{\rm eff}$=5521 K is the closest to $T_{\rm eff}$=5516$\pm$33 K in our list of stellar evolution models. $M_\star$=0.888 $M_\odot$ and $R_\star$=0.881 $R_\odot$ are the closest to $R_\star$=(0.88$\pm$0.03)$R_\odot$ and $M_\star$=(0.89$\pm$0.09)$M_\odot$ corresponding to $T_{\rm eff}$=5516$\pm$33 K. The corresponding $R_\star$ is calculated using $L_\star = 4\pi\sigma R_\star^2 T_{\rm eff}^4$. $L_\star$ calculated using the values in Alonso et al. (2008): $R_\star$=(0.90$\pm$0.02)$R_\odot$ and $T_{\rm eff}$=5625$\pm$120 K . The corresponding $M_\star$ is derived from the range of models in Padova2002 corresponding to the range in $T_{\rm eff}$ and $R_\star$.

Three chains of 500000 iterations each are run simultaneously with different starting points, all starting points being within a scale size from the initial value. The first 7000 steps of each chain are cut away, as considered part of the burn-in phase. A visual check of the chains, showed that it takes that many steps for a chain to fall within the 1σ range of the total chain. These chains are then compared to each other using Gelman and Rubin's statistic (described in section 6.1.6) to check their convergence. If the convergence criteria is reached, the three chains are combined together (put one after the other) to form a longer chain, statistically more robust. The distribution of the values stored in the MCMC chain for each parameter is plotted, and used to derive the 1σ uncertainty range of each parameter as described in Section 6.1.7.

6.2.2 Results

Without a prior on the stellar temperature

Three 500000-step MCMC chains are run on the binned and truncated IRF-filtered transit light curve of CoRoT-2b. The resulting chains (Figure 6.3) have 56 to 57% of the steps accepted which shows that the scale sizes allowed the MCMC to explore the surrounding parameter space while keeping a large number of steps around the best model. The transition between value extrema in each parameter chain is checked to be done in several steps, which confirms that the scale size in each parameter is not too large.

The 7000 steps of the burn-in phase are cut away in each chain. The resulting three chains have similar position of maximum and 1σ range in their posterior distributions, and the best fit model of each chain are consistent with each other within the uncertainty range of their posterior distributions. The correlation length of each parameter chain is several times smaller than the length of the chain, i.e. the correlated features of the parameter space were explored several times. The Gelman & Rubin's convergence statistic run on the three 500000-step chains returns S_{T_0} = 1.02, $S_{R_p/R_\star}$ = 1.40, S_k = 1.47, and S_b = 1.87. These numbers show that the 500000-step chain has not fully converged in $R_\mathrm{p}/R_\star$, k and b, although the correlation length of the chain was several times smaller than the total length of the chain. Gelman & Rubin's convergence statistic is a more robust test for chain convergence than the correlation length of a chain. In future work, longer chains will be run to ensure that the individual chains have fully converged in all the parameters. An alternative to lengthening the chains will be to increase the scale sizes. This allows the chain to explore more of the parameter space, but at the detriment of the number of accepted steps. Chains with slightly larger scale sizes than the ones giving the optimal number of accepted steps, have shown to have S values closer to 1, i.e. to converge faster.

The three chains are combined together to increase the statistical robustness and the precision of the posterior distributions. The posterior distributions of the combined chain are plotted in Figure 6.2. The planetary parameters derived from the model with the highest likelihood, and the associated 1σ range derived from the posterior distributions are shown in Table 6.1.

Using equation 6.2 and the posterior distribution of k translated into $M_\star$ and $R_\star$ the posterior distribution of $a/R_\star$ can also be derived. In addition, using equation 6.1, the posterior distribution of b and the posterior distribution of $a/R_\star$, the posterior distribution of i can also be derived. The posterior distribution of $a/R_\star$ and i are shown in Figure 6.4, along with the posterior distribution in T_eff, $M_\star$ and $R_\star$ derived from k. The best model value and the 1σ uncertainty range are shown in Table 6.1.

k, b and $R_\mathrm{p}/R_\star$ show a strong correlation with each other, which is expected as they are all dependent on the stellar radius $R_\star$.

Table 6.1: Table of parameters for CoRoT-2b derived from the posterior distribution of the MCMC with no prior on the stellar temperature. The table presents the model with the highest likelihood (best-fit), and the median value (median) and 1σ uncertainty range (1σ range) of each distribution of parameters.

	Best-fit	Median	1σ range
P (d)	<—— 1.7429964 $\pm$ 0.0000017 (fix) ——->		
T_0-2454237 (d)	0.53518	0.53518	(0.53511 - 0.53524)
$R_{\rm p}/R_\star$	0.1618	0.1615	(0.1608 - 0.1623)
b	0.25	0.22	(0.18 - 0.29)
$a/R_\star$	6.56	6.60	(6.48 - 6.72)
i (°)	87.9	88.1	(87.5 - 89.0)
u_a	<———— 0.478 $\pm$ 0.010 (fix) ————>		
u_b	<———— 0.205 $\pm$ 0.007 (fix) ————>		
e	<———————— 0 (fix) ————————->		

With a prior on the stellar temperature

Three chains of 500000 steps each are run, with the prior on the temperature equal to 5516$\pm$33 K, and a smaller scale size in k as mentioned previously. The chains are presented in Figure 6.5. The number of accepted steps was between 29 and 31%. This is smaller than in the case with no prior on $T_{\rm eff}$ despite the reduced scale sizes. With the new constraint on the temperature, the MCMC disregards the models which have a $T_{\rm eff}$ incompatible with the prior on the $T_{\rm eff}$, while these models were kept in the chains without a prior on the $T_{\rm eff}$.

The first 7000 steps of each chains are part of the burn-in phase and cut away. The Gelman & Rubin's convergence statistic is applied and returned S_{T0} = 1.02, S_b = 1.87, $S_{R_p/R_\star}$ = 1.40, and S_k = 1.47. These numbers show that the 500000-step chains have not fully converged in b, $R_{\rm p}/R_\star$, and k, and that longer chains still need to be run to ensure the statistical robustness of the posterior distributions.

The three chains are combined together. The planet parameters are derived from the model with the highest likelihood and the 1σ uncertainty range from the posterior distributions (Figure 6.6); the values are shown in Table 6.2. The model with the highest likelihood is for a k corresponding to $T_{\rm eff}$=5521 K, $M_\star$=0.947 $M_\odot$, $R_\star$=0.912 $R_\odot$. The median of the k corresponds to $T_{\rm eff}$=5495 K, $M_\star$=0.952 $M_\odot$, $R_\star$=0.910 $R_\odot$.

The posterior distributions of $T_{\rm eff}$, $M_\star$, $R_\star$, $a/R_\star$ and i are shown in Figure 6.7. The distribution in $a/R_\star$ is calculated from $M_\star$ and $R_\star$ extracted from the distribution in k. The posterior distribution of the inclination is derived from the posterior distributions of b and $a/R_\star$; the values are shown in Table 6.2.

The MCMC coverage in the 2-D parameter spaces is plotted in Figure 6.6. The correlation between k, b and $R_{\rm p}/R_\star$ is present and a range in $R_\star$ values is covered under the uncertainty range of the prior on $T_{\rm eff}$. The posterior distribution of $T_{\rm eff}$ reproduces the prior distribution of $T_{\rm eff}$. This is as expected given that the prior distribution on the $T_{\rm eff}$ is applied as a constraint to select the models with compatible $T_{\rm eff}$. The current distribution in k has two peaks (at 75000 and 85000), corresponding $T_{\rm eff}$ of 5407 and 5470 K, $R_\star$

of 0.897 and 0.884 $R_\odot$, and $M_\star$ of 0.895 and 0.922 $M_\odot$respectively. The stellar density associated to these two peaks is different ($M_\star/R_\star^3 \sim$ 1.240 and 1.336 $M_\odot/R_\odot^3$), which translated into a double peak distribution in $a/R_\star$ peaking at $\sim$ 6.55 and 6.75. The first peak ($a/R_\star$=6.55) is the same as the peak of the probability distribution of $a/R_\star$ in the MCMC run with no prior on the $T_{\rm eff}$ (Figure 6.4). The second peak is therefore added by the prior in $T_{\rm eff}$. This indicates that the current prior on $T_{\rm eff}$ is different from the $T_{\rm eff}$ that would naturally be derived from models of the transit with no prior constraint on the stellar temperature.

A longer chain should be run to check the statistical robustness of this double peak probability distribution in k, or if one of the peaks will dominate the distribution. Additionally, another chain should be run with a different prior on the $T_{\rm eff}$ to investigate the relative position of the peaks with different priors on the stellar temperature.

Table 6.2: Table of parameters for CoRoT-2b derived from the posterior distribution of the MCMC with a prior on $T_{\rm eff}$ of 5516$\pm$33 K. The table presents the model with the highest likelihood (best-fit), the median value and the 1σ uncertainty range of each distribution of parameters.).

	Best-fit	Median	1σ range
P (d)	<—— 1.7429964 $\pm$ 0.0000017 (fix) ——->		
T_0-2454237 (d)	0.53518	0.53518	(0.53511 - 0.53524)
$R_{\rm p}/R_\star$	0.1618	0.1616	(0.1607 - 0.1624)
b	0.24	0.23	(0.10 - 0.29)
$a/R_\star$	6.57	6.59	(6.48 - 6.74)
i (°)	87.9	88.0	(87.4 - 89.1)
u_a	<———— 0.478 $\pm$ 0.010 (fix) ————>		
u_b	<———— 0.205 $\pm$ 0.007 (fix) ————>		
e	<———————— 0 (fix) ————————->		

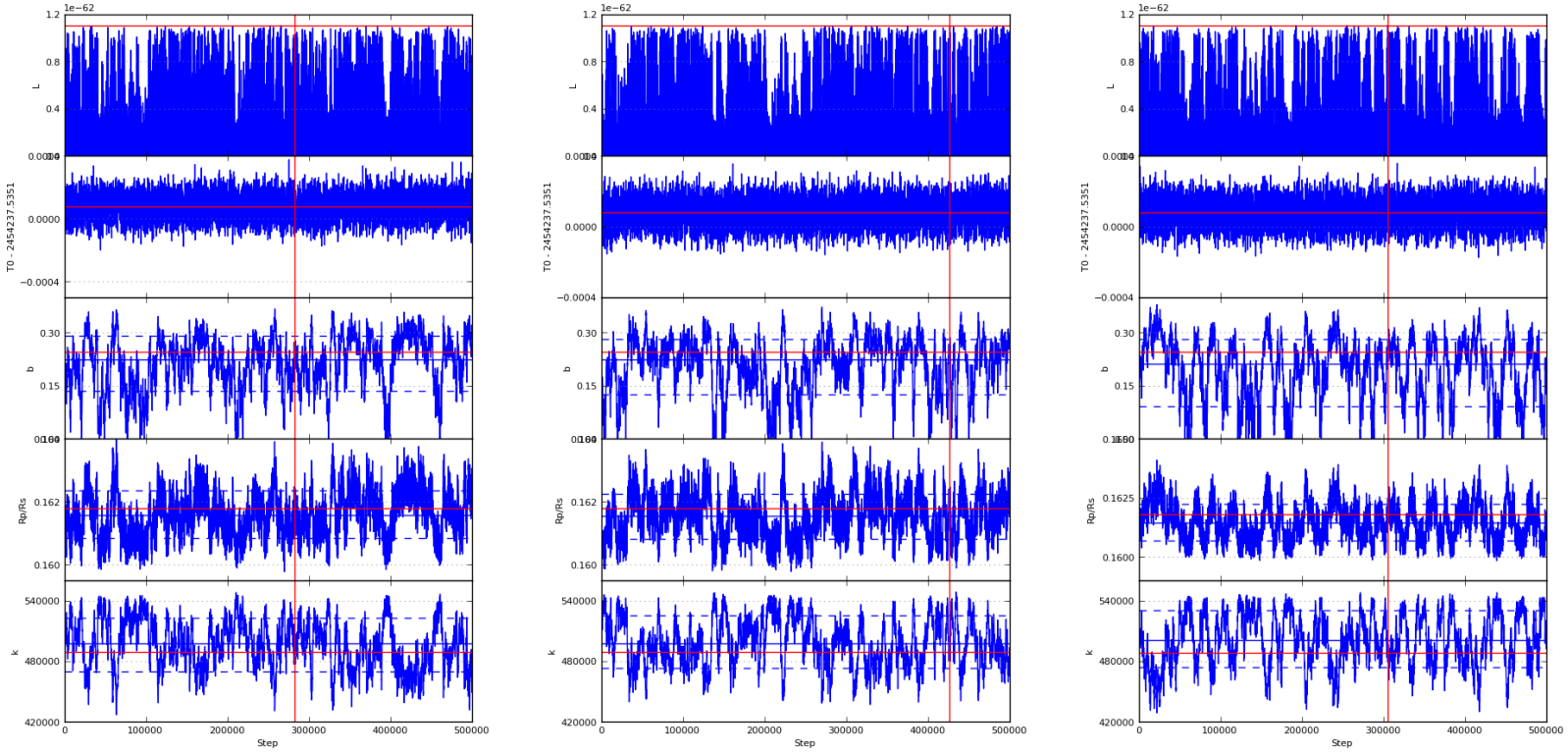

Figure 6.2: The three independent 500000-step MCMC runs on the transit of CoRoT-2b, without a prior on the $T_{\rm eff}$. The adjusted parameters are the transit epoch T_0, the impact parameter b, the planet to star radius ratio $R_{\rm p}/R_\star$, and k the stellar density line index in the re-arranged Padova2002 stellar evolution models. The likelihood L of the model stored at each step is plotted in the top panel. The model with the highest likelihood is marked by the red lines. The 1σ uncertainty range on this model is marked by the blue dashed lines. The median value of each distribution is marked by a vertical blue line. These three MCMC runs are combined together in Figure 6.3.

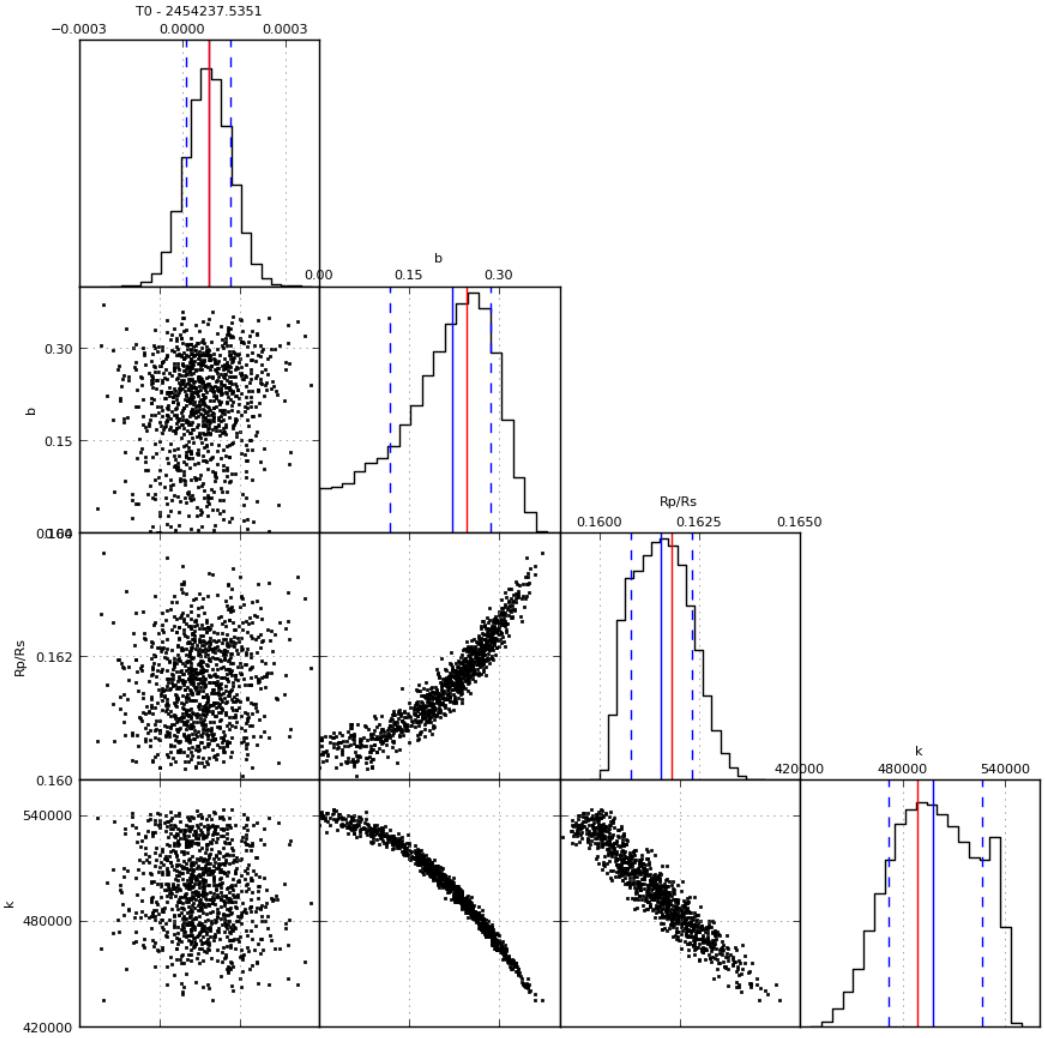

Figure 6.3: The 2D MCMC chains with no prior on the $T_{\rm eff}$, and their posterior distributions. The chains are 1479000 steps long. Same colour line legend as Figure 6.2.

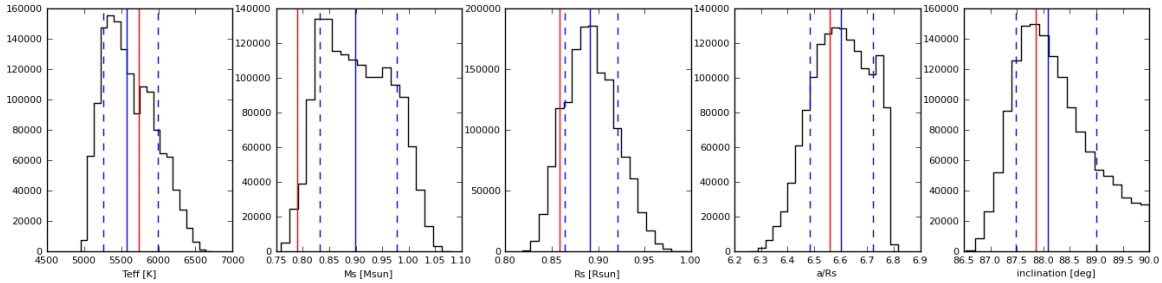

Figure 6.4: The posterior distributions of $a/R_\star$ and i derived from the posterior distributions of k and b. The posterior distributions of $T_{\rm eff}$, $M_\star$ and $R_\star$ derived from k, are shown for comparison with Figure 6.7. Same colour line legend as Figure 6.2.

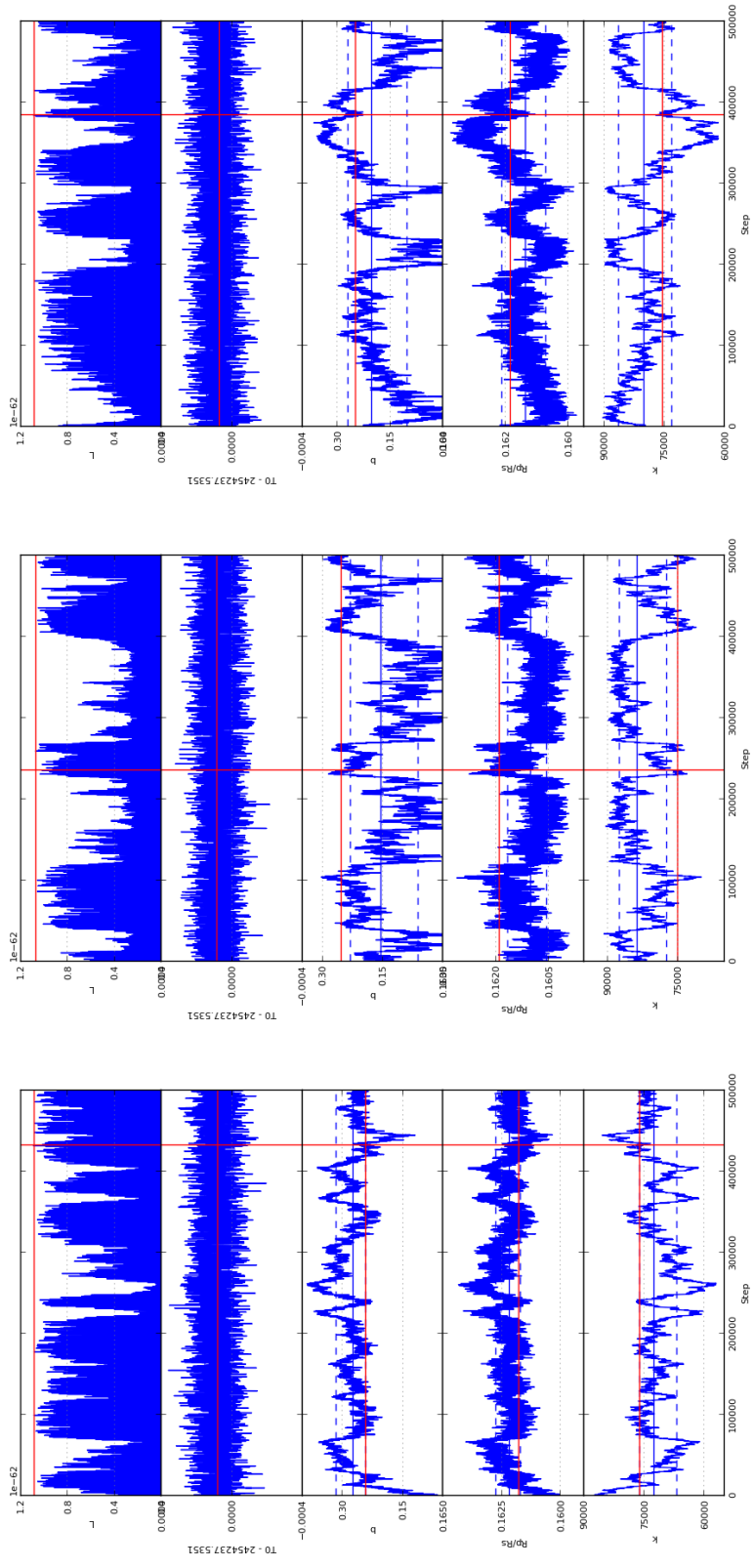

Figure 6.5: Same legend as Figure 6.2, but for the MCMC run with a prior on the $T_{\rm eff}$ equal to 5516±33 K, temperature found for CoRoT-2 in the previous chapter using the equivalent width line ratios. These three MCMC runs are combined together in Figure 6.6.

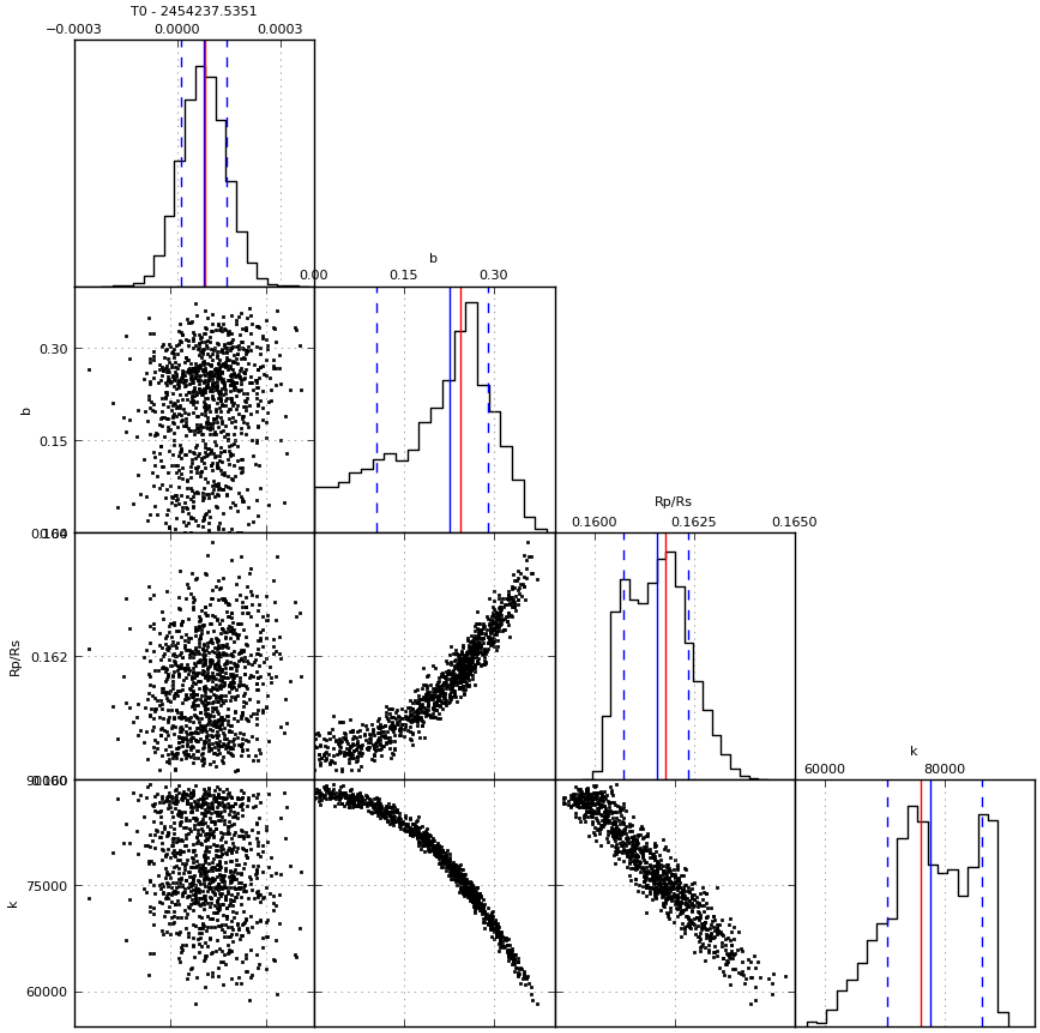

Figure 6.6: 2D MCMC chains, run with a prior of $T_{\rm eff}$=5516±33 K, and their posterior distributions. Same legend as Figure 6.3.

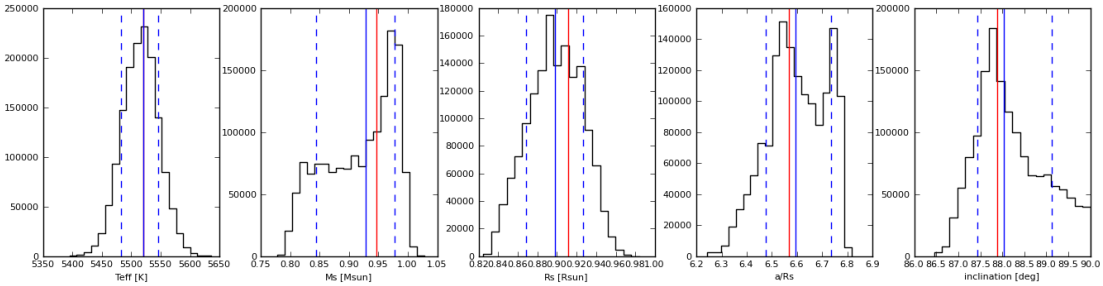

Figure 6.7: The posterior distributions of $T_{\rm eff}$, $a/R_\star$ and i derived from the posterior distributions of k and b of Figure 6.6. The posterior distribution of $M_\star$ and $R_\star$, derived from k, are also shown as used to derive $a/R_\star$. Same legend as Figure 6.4.

6.2.3 Discussion

The MCMC best-fit models of CoRoT-2b's IRF-filtered transit light curve are presented in Figure 6.8, and the parameters are summarised in Table 6.3. These models and values are compared to the parameters from Alonso et al. (2008) and the parameters derived with the Levenberg-Marquardt algorithm (fitting method used in Chapter 3).

Table 6.3: Comparison table of the parameters of CoRoT-2b presented in the discovery paper, derived using the Levenberg-Marquardt algorithm (LMA), and derived using the Markov Chain Monte Carlo (MCMC) without and with a prior on the $T_{\rm eff}$.

	Alonso et al. (2008)	LMA	MCMC	
			no $T_{\rm eff}$	$T_{\rm eff}$=5516±33K
P (d)	<———	1.7429964 ± 0.0000017	———	->
T_0-2454237 (d)	0.53562 ± 0.00014	0.53534 ± 0.00002	$0.53518^{+0.00006}_{-0.00006}$	$0.53518^{+0.00006}_{-0.00007}$
$R_{\rm p}/R_\star$	0.1667 ± 0.0006	0.1621 ± 0.0003	$0.1618^{+0.0005}_{-0.0010}$	$0.1618^{+0.0006}_{-0.0011}$
b	0.26 ± 0.01	0.25 ± 0.02	$0.25^{+0.04}_{-0.13}$	$0.24^{+0.05}_{-0.14}$
$a/R_\star$	6.70 ± 0.03	6.56 ± 0.04	$6.56^{+0.16}_{-0.08}$	$6.57^{+0.17}_{-0.09}$
i (°)	87.8 ± 0.1	87.8 ± 0.2	$87.9^{+1.1}_{-0.4}$	$87.9^{+1.2}_{-0.4}$
u_a	0.41 ± 0.03	<———	0.478 ± 0.010 (fix)	———>
u_b	0.06 ± 0.03	<———	0.205 ± 0.007 (fix)	———>
e	<———	0 (fix)	———	->

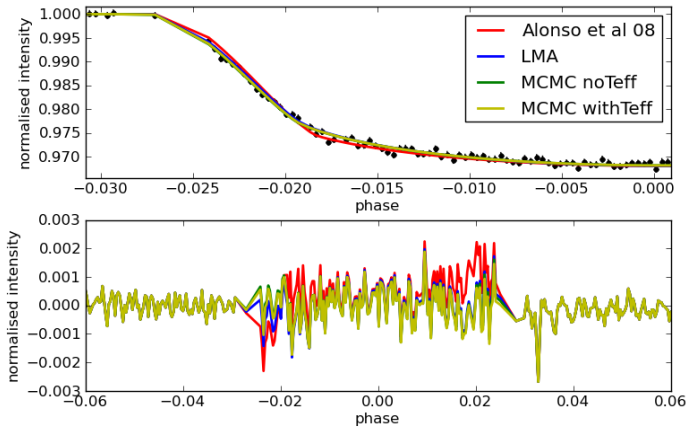

Figure 6.8: Top panel: the binned phase-folded transit of CoRoT-2b, zoomed over the first half of the IRF-filtered transit. Over-plotted are the best transit models derived with the LMA and with the different MCMC runs. Bottom panel: the residual to the data of the LMA and MCMC transit models.

Table 6.3 and Figure 6.8 show that both the LMA and the MCMC derive consistent planet parameters and transit model to the data. However, the uncertainties on the parameters derived from the MCMC are more conservative and better representative of the shape of the parameter space around the best model. The comparison show that the LMA has also found the global best minimum for this light curve. It also shows that adding a constraint on the $T_{\rm eff}$ does not change the planet parameters derived

from this light curve. However, it has changed the shape of the posterior distribution in k, adding a second peak to the distribution. This also creates a second peak in the distribution of $a/R_\star$ at $\sim$6.75. A longer chain needs to be run to check the robustness of the shape of this distribution and see if the peak due to the prior on the $T_{\rm eff}$ shifts the value of $a/R_\star$.

The prescribed MCMC walk in stellar density allows us to derive a posterior distribution for $T_{\rm eff}$. This provides an additional method to derive the stellar temperature from the photometric transit of an orbiting planet, and to map the likelihood space in $T_{\rm eff}$ around the best value. In the case of CoRoT-2, the temperature derived from the posterior probability distribution of a chain stepping in stellar densities, is 5741 K for the best-fit value, 5571 K for the median value and $\sim$5400 K for the most probable value (Figure 6.4). These values are smaller, but consistent within the 1σ uncertainty range of the distribution in $T_{\rm eff}$ (5260 - 5998), with the stellar temperature derived from the equivalent width ratios in Chapter 5 (5516$\pm$33 K), and with the stellar temperature published in Alonso et al. (2008) (5625$\pm$120 K).

The larger residuals to the model from Alonso et al. (2008) in Figure 6.8 show that the phase-folded IRF-filtered transit light curve is slightly shifted in T_0 compared to the processing done by Alonso et al. (2008). Additionally, the stellar limb darkening coefficients of Alonso et al. (2008) do not reproduce the shape of the IRF-filtered transit as well as the ones of Sing (2010) used in this chapter.

The uncertainty on $R_{\rm p}/R_\star$ derived from the phase-folded IRF-filtered transit light curve of CoRoT-2b is smaller than the value published in Alonso et al. (2008), although derived more robustly. This shows an improvement in the light curve processing when using the IRF.

The planetary parameters derived from the LMA and the MCMC applied to the IRF-filtered transit light curve of CoRoT-2b makes the planet appear smaller and closer to its star than published in Alonso et al. (2008). CoRoT-2b is classified as an inflated planet, i.e. its radius is larger than what can be explained with the current planet composition and evolution models as discussed in Alonso et al. (2008). This new set of parameters makes CoRoT-2b appear less inflated and thus less challenging for the models.

One difference between the two analyses is the choice of different limb darkening coefficients. Using different limb darkening coefficients changes the shape of the transit model forcing the other parameters to adjust to reproduce the data points. In the discovery paper of CoRoT-2b, the limb darkening coefficients were fitted at the same time as the planet parameters, thus the degeneracy of their values with the value of $R_{\rm p}/R_\star$ depend strongly on the broadness and the finesse of the exploration of the parameter space performed by the authors, as well as on the light curve processing. In this thesis, the limb darkening coefficients were kept fixed to aid the comparison of the derived planet parameters obtained with the different methods used.

Stellar isochrones can be plotted in stellar luminosity versus temperature diagrams.

The constraint in stellar density, from the transit shape, allows us to derive probability distributions for the stellar ages, along with the other stellar parameters, using the resampled Padova stellar evolution models. In the MCMC, each k is related to a combination of stellar density, temperature and age. The MCMC posterior distribution in k is dependent on the constraints from the stellar density, from the transit shape, and from the stellar effective temperature from the equivalent width ratios. The distribution in stellar age can be directly produced from the distribution in k, as each k is related to a stellar age in the resampled Padova stellar evolution models. However, one should keep in mind that the validity of the derived age distribution will be both stellar-evolution-model and contaminant-flux-free-light-curve dependent. When a prior on the temperature is applied, the derived age distribution will also be stellar-temperature dependent.
Quantifying the improvement in the determination of the stellar age from the probabilistic distribution of the stellar density, and the new stellar temperature, will be an interesting study to perform as future work. It will be interesting to see if the new age estimate makes the host star of CoRoT-2b younger than currently thought. A younger star would also infer that the planet is younger, which better explains the inflated aspect of CoRoT-2b. Younger planets are intrinsically hotter and larger, as a planet cools and contracts after formation.

6.3 Conclusion and future work

The work presented in this chapter is in progress. Improvements on the code and the science included in the MCMC are ongoing.

6.3.1 Conclusions

The greatest strength of the MCMC is its capability to map the parameter space around the best model, showing structures in the likelihood of the surrounding models and correlations between parameters. If the MCMC is run for long enough, it derives a robust uncertainty range on the value of the parameter and a finesse of exploration sufficient to derive accurate values of the best model.

The other major advantage of the MCMC is that it allows the inclusion of a-priori knowledge to the search for best model. This makes the solution a better representation of the true reality. However, the probability distributions it returns are only as good as the priors assumed.

The greatest drawback of the MCMC approach is that it takes time to run. A lot of iterations are needed to derive statistically robust probability distributions. The Levenberg-Marquardt algorithm is much faster but more sensitive to the initial conditions, as it converges towards the first minimum in χ^2 it finds in the parameter space

which might not be the global minimum.

The set-up stage of an MCMC can also be lengthy, as opposed to the setting up of a grid search for instance. Factors to take into account in the setting-up stage on a MCMC include: a) the format of the data (number of points) as it can lengthen the calculation time, b) the choice of priors and their distributions, c) the initial values, d) the typical step sizes, e) the length of the chain to ensure convergence. However, once the MCMC is set-up, for the same resolution and statistical robustness of the solution, running an MCMC can be faster than a grid search.

The transit modelling using the MCMC approach described in this chapter makes use of a new photometric method to derive the stellar temperature. This method uses stellar evolution models to translate the stellar densities (adjusted in the MCMC) into stellar temperature. The precision of the derived $T_{\rm eff}$ will depend on the finesse and intrinsic accuracy of the grid of stellar evolution models.

Applied to the phase-folded IRF-filtered transit light curve of CoRoT-2b, the MCMC approach derived is a more robust method to determine error bars on the planet parameters. It also provides an independent measurement of the stellar temperature from the transit shape. The addition of a prior in the stellar temperature was not found to change the final values of the planet parameters, but did change the shape of the distribution in stellar density.

6.3.2 Future work

Longer chains still need to be run to ensure the convergence for each parameter. This will secure a robust solution for the parameters, for the given data and priors.

The uncertainties on the contaminant flux needs to be taken into account in the uncertainties derived for the parameters. The suggested approach is to add to the transit model, at each MCMC iteration, a constant flux drawn from a Gaussian distribution (with a zero mean and a standard deviation equal to the uncertainty on the contaminant flux), then re-normalise the model and use it in the calculation of the likelihood. The posterior distribution of the parameter will thus include the uncertainty of the contaminant flux.

The red noise in the light curve needs to be taken into account as a good merit function needs to use a true estimate of the noise. The red noise can be taken into account by replacing the σ_{white} in the merit function by $\sigma_{pink} = \sigma_{white} + N * \sigma_{red}$ (Pont et al., 2006). σ_{red} can be evaluated by binning the unfolded light curves with different bin sizes and evaluating each time the standard deviation of the resulting signal. The σ_{white} will go as $\sigma_{no\ bin}/N$, N being the number of points binned together, while σ_{red} should be constant, so the difference between the σ_{bin} and $\sigma_{no\ bin}/N$, e.g. for bin size of 1 or 2h, is σ_{red}.

The IRF filtered light curve was binned in order to reduce the number of data points to speed up the MCMC. It will be interesting to see how this binning affects the ac-

curacy and precision of the planet parameters and their 1σ uncertainty range. The binning of the light curve by half, for instance.

It will be also be interesting to see how the posterior probability distributions vary with different values of the prior in $T_{\rm eff}$, especially for the distribution in k and thus $a/R_\star$. The MCMC can be run with the prior on the $T_{\rm eff}$ set to the value published in Alonso et al. (2008) (5625$\pm$120 K).

The grid of stellar evolution models used in the MCMC can also be refined to include a finer sampling in the stellar parameters. This should improve the accuracy of the posterior distribution in k, and thus in $a/R_\star$ and i.

The MCMC performs steps in k, and at each k is associated a $T_{\rm eff}$. The limb darkening of a star depends on the stellar temperature (as well as on the stellar surface gravity and metallicity, and on the observational bandpass). Currently the limb darkening coefficients are not adjusted when the MCMC steps into another value of $T_{\rm eff}$. To be more consistent, the MCMC should be adjusted to allow the limb darkening coefficients to vary according to the $T_{\rm eff}$ associated to the k of each step. For a chosen limb darkening law and filter, the limb darkening coefficients can be calculated given the stellar atmosphere parameters ($T_{\rm eff}$, $\log g$, (M/H)). Claret (2000) and Claret (2004) give tables of limb darkening coefficients for different standard filters and Sing (2010) for CoRoT and Kepler bandpasses.

Finally, it will be interesting to homogeneously derive the planet parameters (1σ uncertainty range and best model) of the other CoRoT planets using the MCMC approach presented in this chapter, on the IRF-filtered transit light curve of the planet, including the prior knowledge on the stellar temperature. The value of the prior in $T_{\rm eff}$ can be set to the value from the discovery paper, or to the values derived from the equivalent width ratios. The derived planetary parameters can then be compared to each other, e.g. in a mass-radius diagram. This approach will reduce current biases on the type of light curve processing and fitting used to analyses the individual planets. It also provides the advantage of systematically taking into account more information on the planets and their host star.