

# BAYESIAN SPECTRO-TEMPORAL PULSE DECOMPOSITION FOR GRB LIGHTCURVES

TBA

TBA

AND

## 1. INTRODUCTION

## 2. METHOD

## 2.1. Formulation of Statistical Problem

The data are binned photon counts  $Y_{ik}$ , for  $i = 1, 2, 3, 4$ , corresponding to the four energy channels used by the discriminator, and  $k = 1 \dots N$ , for the  $N$  time intervals. It is believed that the data have a Poisson distribution, since the discriminators have a very small amount of downtime and the photons arrive in a Poisson process.(cite?)

The underlying model for the amplitudes and onset time can be represented as a Gamma process with parameters  $a(dt, dE)$  and  $b$ . This results in a distribution over the *number* of non-trivial pulses in the domain as something that can be inferred from the data. Such a process is defined over a compact set - in this case, an interval of time with duration  $L_T$ .

The number of pulses with an amplitude larger than some threshold  $\epsilon$ ,  $J_\epsilon$ , is distributed Poisson:

$$J_\epsilon \sim \text{Po}(\alpha L_T E_1(\beta\epsilon)) \quad (1)$$

where  $E_1$  is the exponential integral function ?.

We will abbreviate  $J_\epsilon$  as  $J$ , since the threshold parameter  $\epsilon$  is typically chosen before any analysis is done; it is set such that any pulse of amplitude  $\epsilon$  is negligible in size. Each pulse represents a single component of the GRB, and so there is a separate set of parameters for each pulse. However, the subscripts indicating pulses have been omitted for this section.

The time at which the pulse arrives at the detector varies as a function of the energy of the photons. Typically higher-energy photons will arrive before lower-energy photons; we say the pulse arrives earlier at higher energies. We define the onset time,  $T$ , as the time the pulse arrives at  $E = 150$  keV. Since a pulse does not appear simultaneously at all energies, a linear lag parameter  $\gamma$  is introduced to shift the pulse onset at higher energies with units  $s/keV$ . At a given energy, as time increases past the onset of the burst, the flux falls exponentially with rate parameter  $\lambda$ . Finally, an amplitude parameter  $A$  is included to reflect the overall level of flux. So we can write the time component of the flux:

$$\Psi(t | E, T, \gamma, \lambda, A) = \begin{cases} \exp[-\lambda((t - T) + \gamma E)] & \text{if } (t - T) \geq 0 \\ 0 & \text{if } (t - T) < 0 \end{cases} \quad (2)$$

The energy spectrum of a pulse must also be modeled, as it varies between bursts. The Band model is a common functional form for the energy spectrum of a gamma-ray burst ?. The key feature of this model is that the photon rate is modeled by a power law (rate  $\alpha$ ) at energies lower than some cutoff ( $E_c = (\alpha - \beta)E_0$ ), and at a steeper power law (rate  $\beta$ ) at higher energies. That is, the energy spectrum is

$$\Phi(E | \alpha, \beta, E_0) \propto \begin{cases} \left(\frac{E}{100}\right)^{\alpha-\beta} \exp\{\beta - \alpha\} \left(\frac{E}{100}\right)^\beta & \text{if } E > E_c, \\ \left(\frac{E}{100}\right)^\alpha \exp\{-E/E_0\} & \text{if } E \leq E_c. \end{cases} \quad (3)$$

with all energies expressed in keV. Note that the parameters  $\alpha$  and  $\beta$  are negative. Thus  $\beta < \alpha < 0$ .

Finally, an amplitude parameter  $A$  is included to reflect the overall level of flux. So, each pulse is described by a seven-dimensional  $\theta = (T, A, \lambda, \gamma, \alpha, \beta, E_0)$ .

We can then write the total flux from set of pulses described by  $\theta_1 \dots \theta_J$  as a function of both time and energy:

$$F(t, E | \{\theta_j\}) = B + \sum_{j=1}^J \Psi(t | E, \theta_j) \times \Phi(E | \theta_j) \quad (4)$$

where  $B$  is some fixed baseline.

A photon that enters the detector may not be recorded at its true energy, or even at all, due to one of a number of possible reactions with other photons or the detecting material ?. For example, Compton scattering in the detector may result in the photon being recorded at a lower energy than that of the incident photon. In general, the error in energy estimation is biased; the energies estimated by the detector are systematically lower than the true energies. The “response function” of the detector is known due to experiments performed on the ground before the telescope was launched. There is a different response function for each energy channel and each GRB, since the functions also vary as a function of the incident angle of the photons (which is measured separately). We can then transform this incident flux using the response function for energy channel  $i$ , called  $R_i(E)$ . This gives the *observed* flux as a function of both energy and time. We can then integrate over the time bin and energy channel range to obtain the model estimate of the observed photon counts.

Then

$$Y_{ik} \sim \text{Po}(\mu_{ik})$$

where

$$\mu_{ik} = \int_{t_k}^{t_{k+1}} \int_0^\infty F(t, E) R_i(E) dE dt$$

## 2.2. Overdispersion

The model described above is necessarily an approximation to the true process generating the data. To compensate for model uncertainty, an overdispersion parameter,  $r$  is introduced. This parameter does not depend on the time interval or the energy channel. To allow for overdispersion of the data, instead of modeling  $Y_{ij}$  as Poisson, we can model it as negative binomial:

$$Y_{ij} \sim \text{NB} \left( n = r\mu_i, p = \frac{r}{1+r} \right)$$

Then  $\mathbb{E}(Y_{ij}) = \mu_{ij}$ , as before, but instead of constraining the variance of  $Y_i$  to  $\mu_i$ , as in the Poisson case,

$$\text{Var}(Y_{ij}) = \mu_{ij} \frac{1+r}{r}$$

Thus, when  $r$  is small, the variance of  $Y_i$  is inflated. This parameterization is not affected by merging two adjacent bins (and thus to any re-binning, by induction), because

$$\text{NB}(n_1, p) + \text{NB}(n_2, p) = \text{NB}(n_1 + n_2, p).$$

## 2.3. Priors

*A priori*, we model the location of the pulse in energy as uniform across the domain of the response function. The same modeling assumption is made for the location of the pulse in time, even though we know that a pulse is more likely to have occurred near  $t = 0$ , the trigger time, than significantly earlier or later.

The lag parameter  $\gamma$  is modeled as a  $\text{No}(0, 1)$  distribution, in units of  $\text{s}/\text{MeV}$ . The amplitude of the pulses are distributed as the jumps larger than some  $\epsilon$  in a gamma process in two dimensions with parameters  $\alpha$  and  $\beta$ .

The prior for  $E_0$  is a diffuse normal centered around 150 keV.

$$\alpha = -\text{Exp}(1), \quad \beta = -\text{Exp}(2)$$

and  $\alpha > \beta$ .

$\lambda$  was given a gamma prior centered at one, and  $\gamma$  was normally distributed *a priori*, with mean zero.  $\frac{r}{1+r}$ , the inverse of the inflation factor of the variance of the data, varies between 0 and 1, and is given a uniform prior on that range; the implied distribution for  $r$  is Generalized Pareto:  $\text{GP}(1, 1, 0)$ .

## 3. DETAILS OF IMPLEMENTATION

### 3.1. RJ-MCMC

The model is fit using Markov chain Monte Carlo (MCMC). Since the dimension of the parameter space depends on the unknown number of pulses that comprise the GRB, since the number of pulses comprising the gamma-ray burst, a technique known as reversible-jump MCMC is used. In addition to the typical Metropolis-Hastings step, where the proposed incremental change according to the Metropolis-Hastings rule, there are several types of transdimensional steps which adjust the dimension of the model. This results in posterior samples of varying dimension, from which the posterior distribution for the number of pulses,  $J$ , can be determined.

The basic reversible-jump algorithm involves three types of steps, “birth,” “death,” and “walk.” Each of these has an associated probability:  $p_B, p_D$  and  $p_W$  such that  $p_B + p_D + p_W = 1$ . For each iteration of the Markov chain, one type of step is taken.

#### 3.1.1. Birth Steps

In the birth step, a new pulse is proposed to be added to the  $J$  existing pulses. The parameters for this new pulse are drawn from the birth distribution  $\theta_* \sim b(\cdot)$ . We call the list of pulses before adding the new pulse  $\Theta = \langle \theta_1 \dots \theta_J \rangle$  and the list after adding the proposed pulse  $\Theta_* = \langle \theta_1 \dots \theta_J, \theta_* \rangle$ . This birth distribution must have the same support as the prior, but can otherwise be of any form.

We accept the proposed move from  $\Theta$  to  $\Theta_*$  using the Metropolis-Hastings ratio:

$$\frac{\mathcal{L}(Y \mid \Theta_*) \Pi(\Theta_*) q(\Theta_*, \Theta)}{\mathcal{L}(Y \mid \Theta) \Pi(\Theta) q(\Theta, \Theta_*)} \quad (5)$$

where  $\Pi(\cdot)$  is the prior for a set of pulses,  $\mathcal{L}(Y \mid \cdot)$  is the likelihood of the data given a set of pulses, and  $q(a, b)$  is the transition probability from state  $a$  to  $b$ .

Note that the prior for  $\Theta_*$  and  $\Theta$  are functions of  $J$ , as well as functions of the pulse parameters themselves. As in Eqn (2.1),

$$J_\epsilon \sim \text{Po}(\alpha L_T E_1(\beta \epsilon)) \quad (6)$$

For simplicity in notation, for a given pulse  $\theta$ , we say  $\theta \sim \pi(\cdot)$  *a priori*. Then, the prior distribution for  $\Theta$  can be written:

$$\Pi(\Theta \mid J, \theta_1 \dots \theta_J) = \prod_{j=1}^J \pi(\theta_j) \times \frac{(\alpha L_T E_1(\beta \epsilon))^J}{J!} \exp\{-\alpha L_T E_1(\beta \epsilon)\} \quad (7)$$

and we can write the ratio

$$\frac{\Pi(\Theta_* \mid J+1, \theta_1, \dots, \theta_J, \theta_{J+1})}{\Pi(\Theta \mid J, \theta_1 \dots \theta_J)} = \frac{\prod_{j=1}^J \pi(\theta_j) \pi(\theta_*) \times \frac{(\alpha L_T E_1(\beta \epsilon))^{J+1}}{(J+1)!} \exp\{-\alpha L_T E_1(\beta \epsilon)\}}{\prod_{j=1}^J \pi(\theta_j) \times \frac{(\alpha L_T E_1(\beta \epsilon))^J}{J!} \exp\{-\alpha L_T E_1(\beta \epsilon)\}} \quad (8)$$

$$= \pi(\theta_*) \frac{\alpha L_T E_1(\beta \epsilon)}{J+1} \quad (9)$$

The transition from  $\Theta$  to  $\Theta_*$ , written  $q(\Theta, \Theta_*)$  is  $p_B \times b(\theta_*)/(J+1)$ . The factor of  $J+1$  represents the arbitrary choice to place  $\theta_*$  at the *end* of the list of pulses to create  $\Theta_*$ .

The transition  $q(\Theta_*, \Theta)$  represents the probability of moving from  $\Theta_*$  to  $\Theta$ . This is a death step, which will be described in more detail shortly. The probability of choosing a death step is  $p_D$ , and a pulse is chosen uniformly from  $\theta_1, \dots, \theta_J, \theta_*$  to be removed. In order to move from  $\Theta_*$  to  $\Theta$ ,  $\theta_*$  must be selected to be removed from the model. The probability that  $\theta_*$  is chosen is  $1/(J+1)$  and thus  $q(\Theta_*, \Theta) = \frac{p_D}{J+1}$ .

The ratio  $q(\Theta_*, \Theta)/q(\Theta, \Theta_*)$  is then  $\frac{p_D}{p_B \times b(\theta_*)}$

Thus the Metropolis-Hastings ratio can be written

$$H = \frac{\mathcal{L}(Y \mid \Theta_*)}{\mathcal{L}(Y \mid \Theta)} \times \pi(\theta_*) \frac{\alpha L_T E_1(\beta \epsilon)}{J+1} \times \frac{p_D}{p_B \times b(\theta_*)}$$

As in the traditional Metropolis-Hastings procedure, the proposed move is accepted with probability  $\max(1, H)$ .

### 3.1.2. Death Steps

When performing a death step, as alluded to above, we remove a pulse chosen uniformly-at-random  $\theta_k$  from the list of pulses  $\Theta$  to create  $\Theta_*$ . Here again  $J$  is the number of pulses in  $\Theta$ .

The ratio of priors  $\Pi(\Theta_*)/\Pi(\Theta)$  is found in the same way as in the birth step to find:

$$\frac{\Pi(\Theta_*)}{\Pi(\Theta)} = \frac{\prod_{1 \leq j \leq J, j \neq k} \pi(\theta_j) \times \frac{(\alpha L_T E_1(\beta \epsilon))^{J-1}}{(J-1)!}}{\prod_{1 \leq j \leq J} \pi(\theta_j) \times \frac{(\alpha L_T E_1(\beta \epsilon))^J}{J!}} \quad (10)$$

$$= \frac{J}{\pi(\theta_k) \alpha L_T E_1(\beta \epsilon)} \quad (11)$$

The transition kernel for the death move  $q(\Theta_*, \Theta)$  is  $p_D/J$ . The transition kernel for the reverse move, the birth of  $\theta_k$ , is found in the same way as described in the previous section, with the caveat that  $\Theta$  has  $J-1$  pulses instead of  $J$ :

$$q(\Theta, \Theta_*) = \frac{p_B b(\theta_k)}{J}$$

So the Metropolis-Hastings ratio for a death step is

$$H = \frac{\mathcal{L}(Y \mid \Theta_*)}{\mathcal{L}(Y \mid \Theta)} \times \frac{J}{\pi(\theta_k) \alpha L_T E_1(\beta \epsilon)} \times \frac{p_D}{p_B b(\theta_k)} \quad (12)$$

and the move is accepted with probability  $\max(1, H)$ .

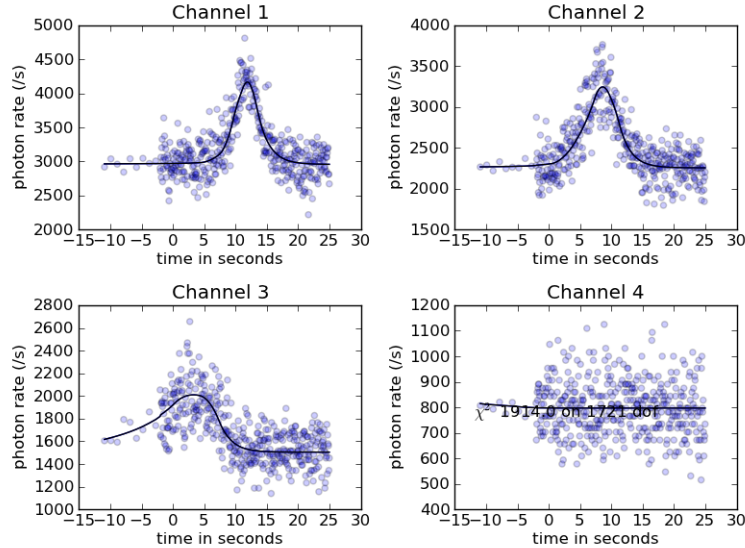
### 3.1.3. Walk Steps

Walk steps are performed at each iteration with probability  $p_W$ . This step is no different from a standard Metropolis-Hastings update. It is not a problem from a theoretical perspective to attempt to innovate on more than one pulse at once, but since the other steps described also affect only a single pulse, we use the same approach for the walk steps. Typically, a single pulse,  $\theta_k$ , is chosen uniformly at random from the pulses in  $\Theta$ , and its parameters are changed with some innovation kernel  $\theta_* \sim p(\cdot \mid \theta_k)$ . A common choice for  $p$  is a Normal distribution centered around  $\theta_k$ . In this case, the Metropolis-Hastings ratio is very simple:

$$H = \frac{\mathcal{L}(Y \mid \Theta_*)}{\mathcal{L}(Y \mid \Theta)} \times \frac{\pi(\theta_*)}{\pi(\theta_k)} \times \frac{p(\theta_k \mid \theta_*)}{p(\theta_* \mid \theta_k)} \quad (13)$$

Interpretation	Parameter	Value
Start time	$T$	5
Linear time lag	$\gamma$	0.5
Time decay	$\lambda$	0.1
Power law decay (low energy)	$\alpha$	-0.7
Power law decay (high energy)	$\beta$	-2
Function of cutoff between two power laws	$E_0$	125

**Table 1**  
Parameters for Simple Simulations



**Figure 1.** Sample draw from the posterior distribution for the simulation in Section (4.1)

#### 3.1.4. Other Steps

For this problem we use in addition “merge” and “split” steps, which will merge two pulses or split a single pulse into two. In this case the iterations are split with probabilities  $p_B, p_D, p_W, p_S$ , and  $p_M$  which all sum to one. This improves mixing in cases where a feature in a GRB may be resolved as either one or two pulses, but it is hard for the Markov chain to explore both cases with the traditional birth, death and walk steps.

#### 3.2. Parallel Thinning

These transdimensional Markov chains typically have lower acceptance rates than those with fixed dimension, an impediment to efficient computation of posterior distributions. Additionally, the posterior distribution for the pulse structure in a GRB may be highly multimodal. In order to efficiently mix between modes, a tempering algorithm known as parallel thinning is used (cite). This algorithm creates auxiliary chains which put more emphasis on the prior instead of the data. This results in faster mixing in these auxiliary chains, and it is possible to construct a Metropolis Hastings step to “swap” two adjacent chains while maintaining the correct limiting distribution for the posterior chain. Details can be found in (cite).

### 4. EXAMPLE: SIMPLE SIMULATION

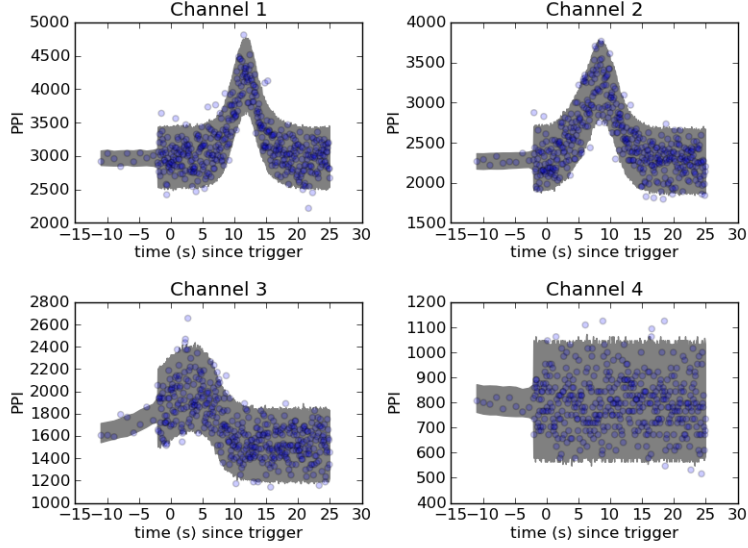
In this section we look at data simulated from our model such that the simulated data appears similar to a GRB. The data is taken to be on a time interval of -12 to 25 seconds, with 0 representing the trigger time. As in many of the BATSE GRBs, the bin lengths are 1.024s at times before -2s, and 0.064 after -2s. A single pulse with the parameters described in Table(1) is transformed by the discriminators for GRB 1406. The data is then drawn from Poisson distributions with the means described by this transformed pulse, along with a fixed background rate.

#### 4.1. Trivial

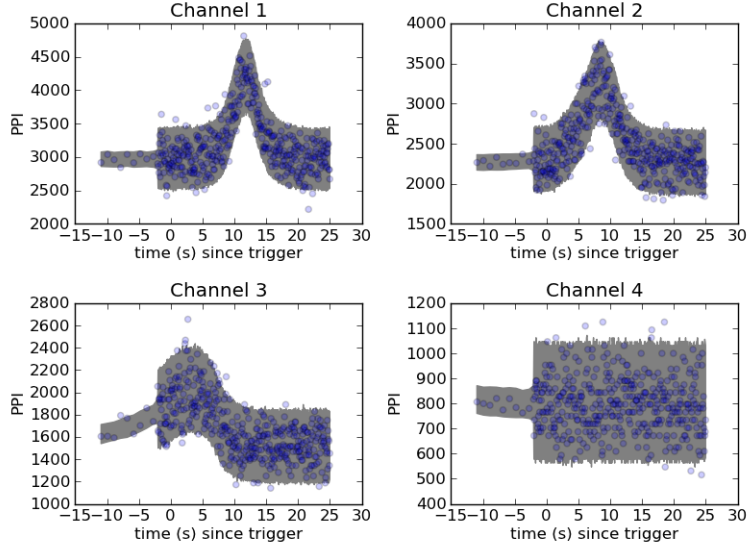
In this simulation, the number of pulses ( $J = 1$ ) is treated as known. We see that the fit of the model under this assumption is very good in one example of a posterior draw in Figure(1). Here we can see the  $\chi^2$  statistic for this particular iteration, which is 1862 on 1717 degrees of freedom.

The pointwise posterior predictive interval is also available in Figure(2). Here we can see that coverage is quite good. Note that the discontinuity in the posterior predictive intervals is due to the non-constant widths of the bins.

Since we know the true values of the parameters used in the simulation, we can examine their traceplots to see if the true values are recovered. This will help us identify identifiability issues in the model. Figure(4), although it is



**Figure 2.** Pointwise posterior predictive interval (95%) for the model described in Section (4.1)



**Figure 3.** Credible interval (95%) and posterior median for the model described in Section (4.1)

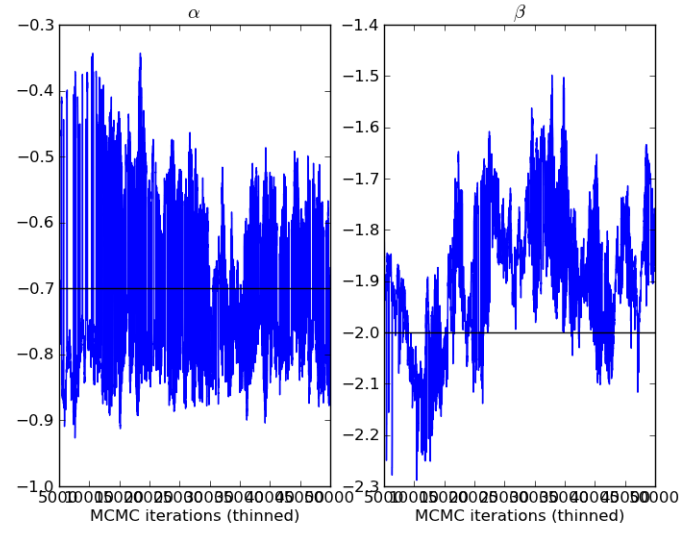
very clear that mixing is poor and the MCMC has not converged, shows that the parameters that define the decay in the energy spectrum,  $\alpha$  and  $\beta$ , are approaching their true values.

In Figure (5) we can see that  $\gamma$  approaches its true value of 0.1. Figure (6) and Figure (7) shows that the time decay parameter, as well as the onset time of the pulse  $T$ , are also well-recovered.

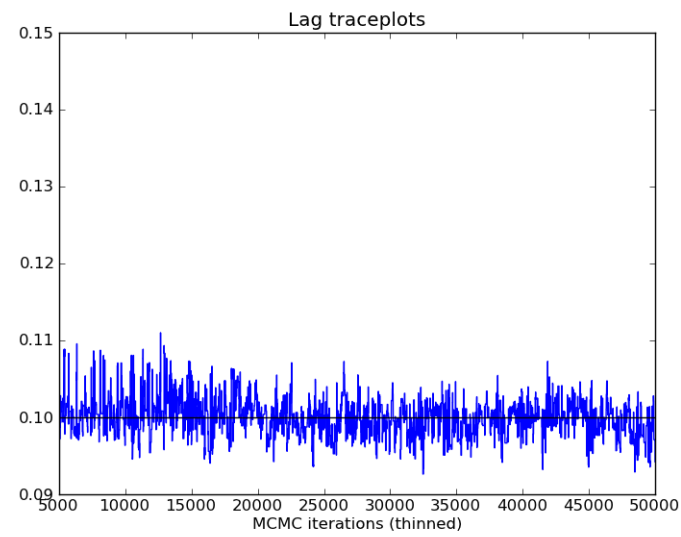
However the  $E_0$  parameter seemed to oscillate between two modes (Figure(8)). It is clear that the posterior chain and the least-thinned chain ( $p = 0.9$ ) had two different modes for  $E_0$ , and the two chains were swapped by parallel thinning.

Figure(9) shows the factor by which the variance is inflated relative to a Poisson model. Since the data in this simulation are distributed Poisson, the true value of  $r$  is  $\infty$ , and we should expect this histogram to be near to 1. We see a mode around 1.05 which indicates a variance of 105% of the true variance.

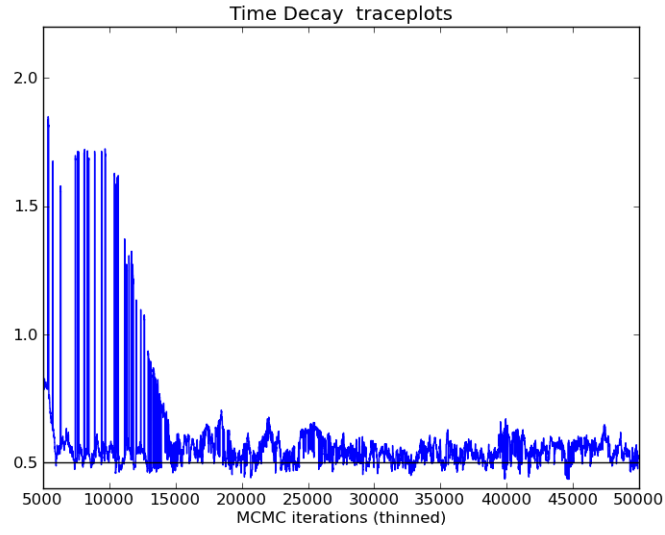
#### 4.1.1. Pairs plots!!



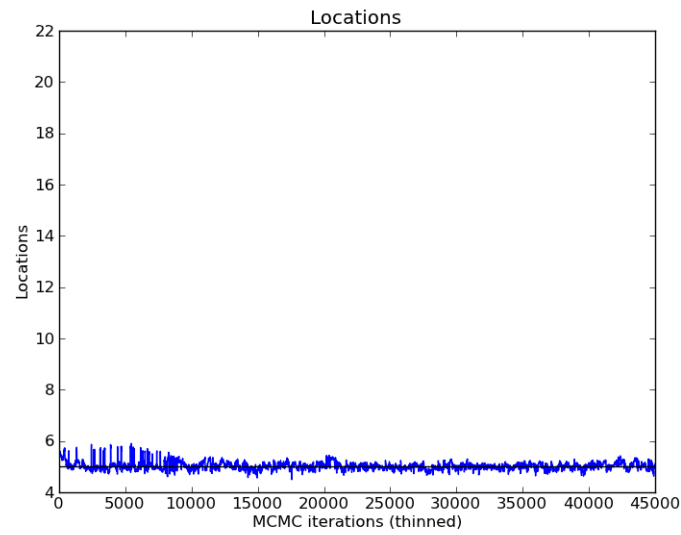
**Figure 4.** The energy spectrum rates  $\alpha$  and  $\beta$



**Figure 5.** Time lag parameter  $\gamma$ .

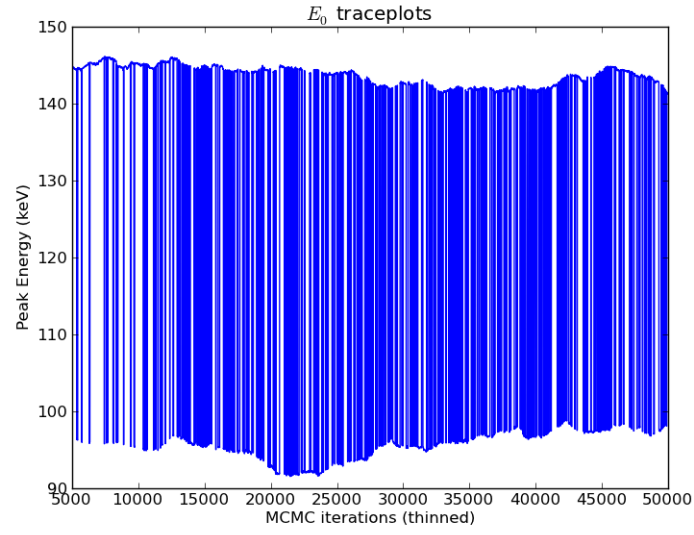


**Figure 6.** Time decay parameter  $\lambda$  (trivial simulation)

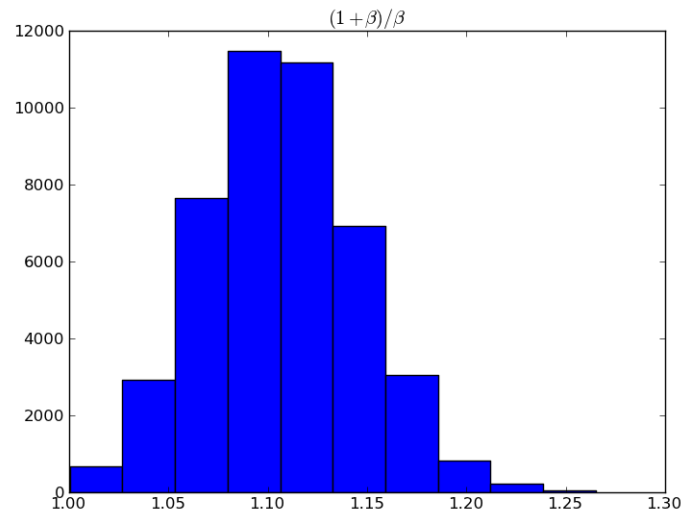


**Figure 7.** Time onset parameter  $T$  (trivial simulation).

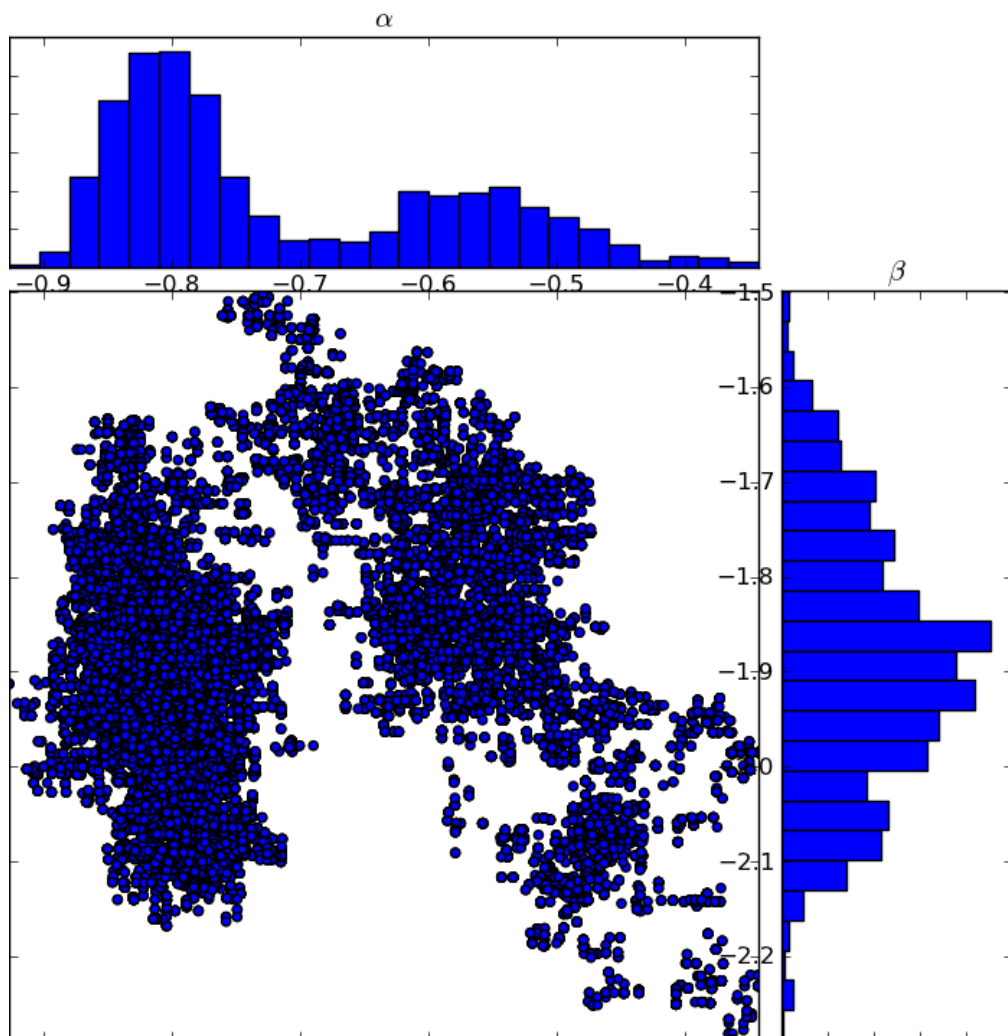




**Figure 8.** Energy spectrum parameter  $E$  (trivial simulation)



**Figure 9.** Factor by which the variance is inflated, trivial simulation



**Figure 10.**  $\alpha$  and  $\beta$ , the Band model power law rates

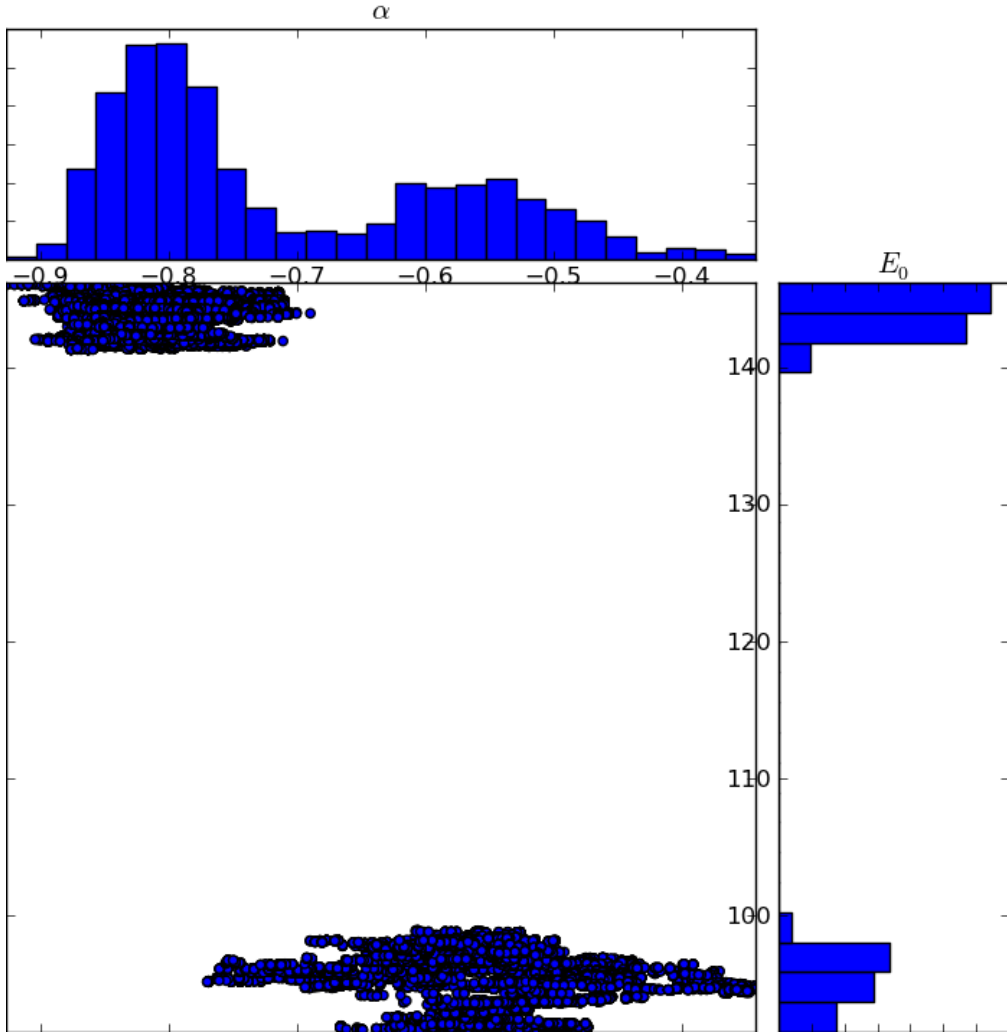


Figure 11.  $\alpha$  and  $E_0$

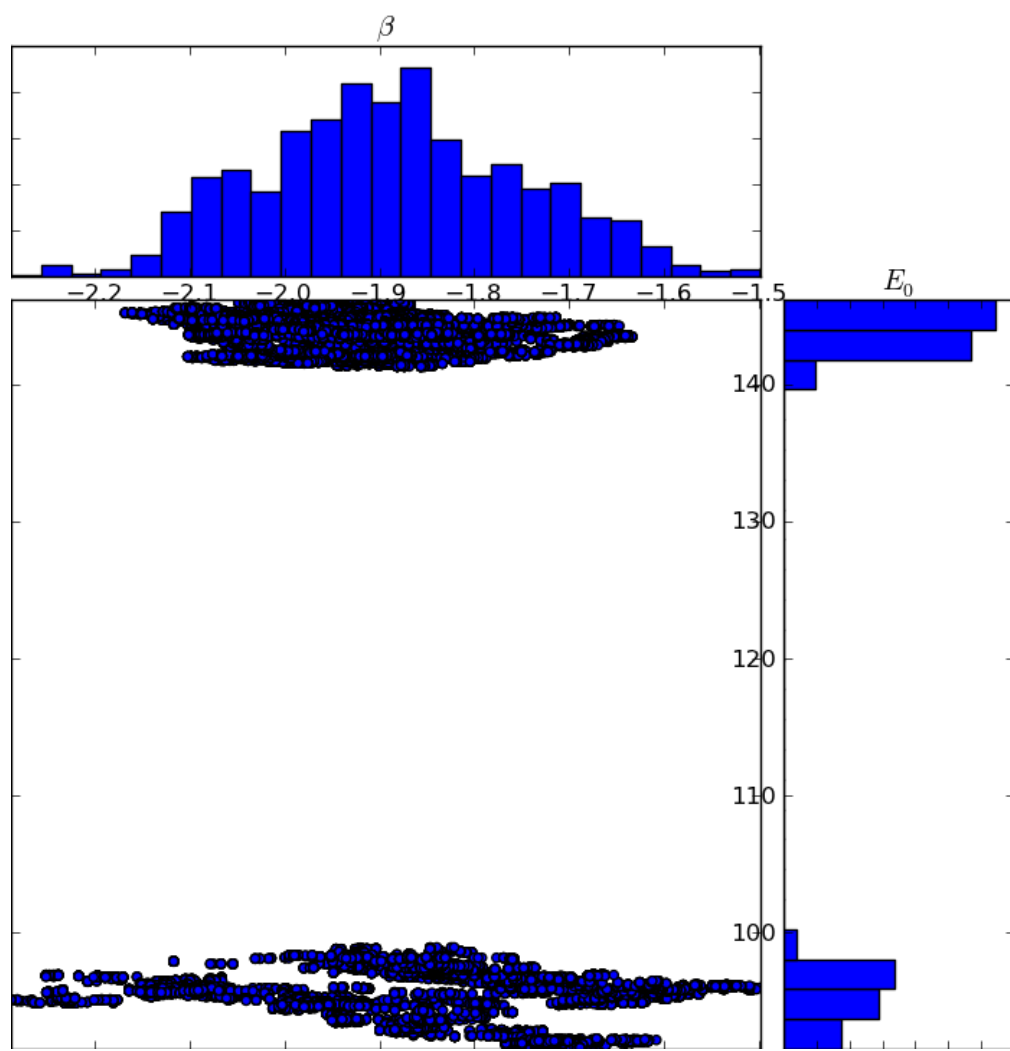


Figure 12.  $\beta$  and  $E_0$

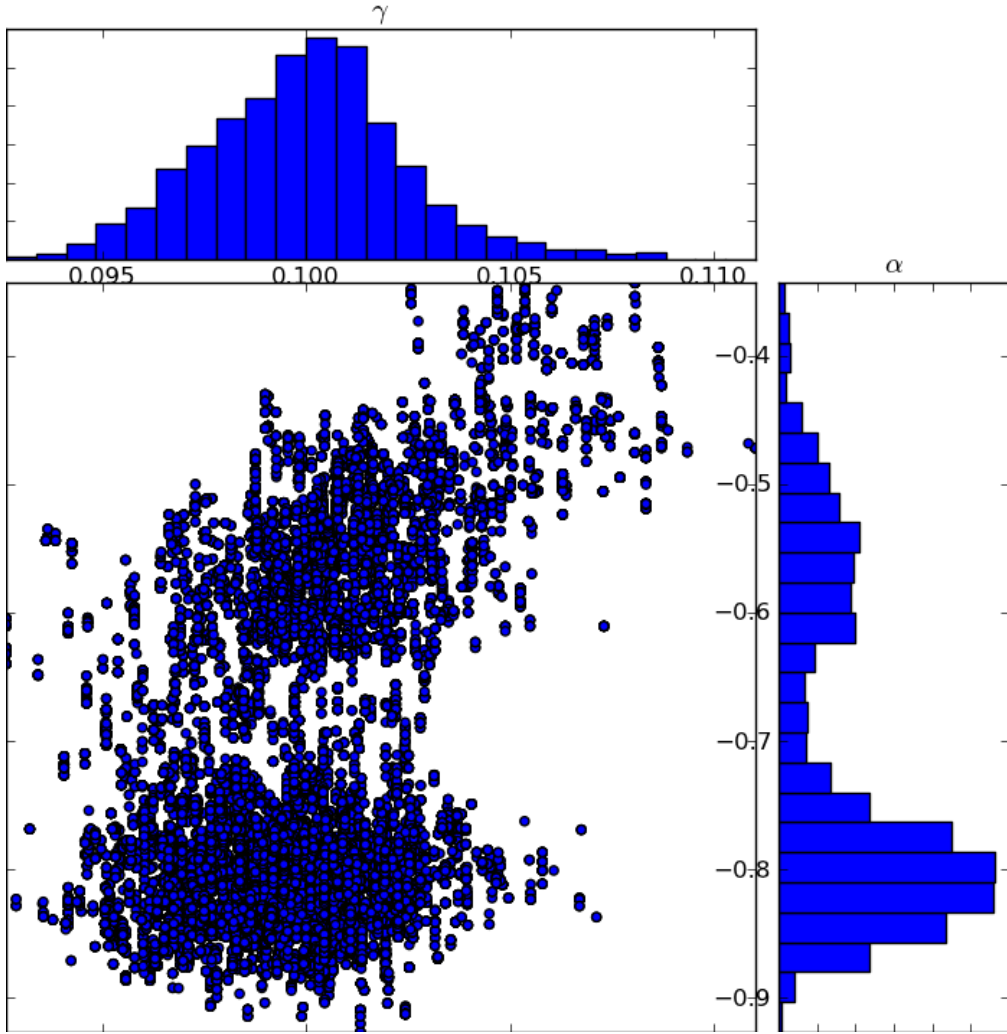


Figure 13.  $\gamma$  and  $\alpha$

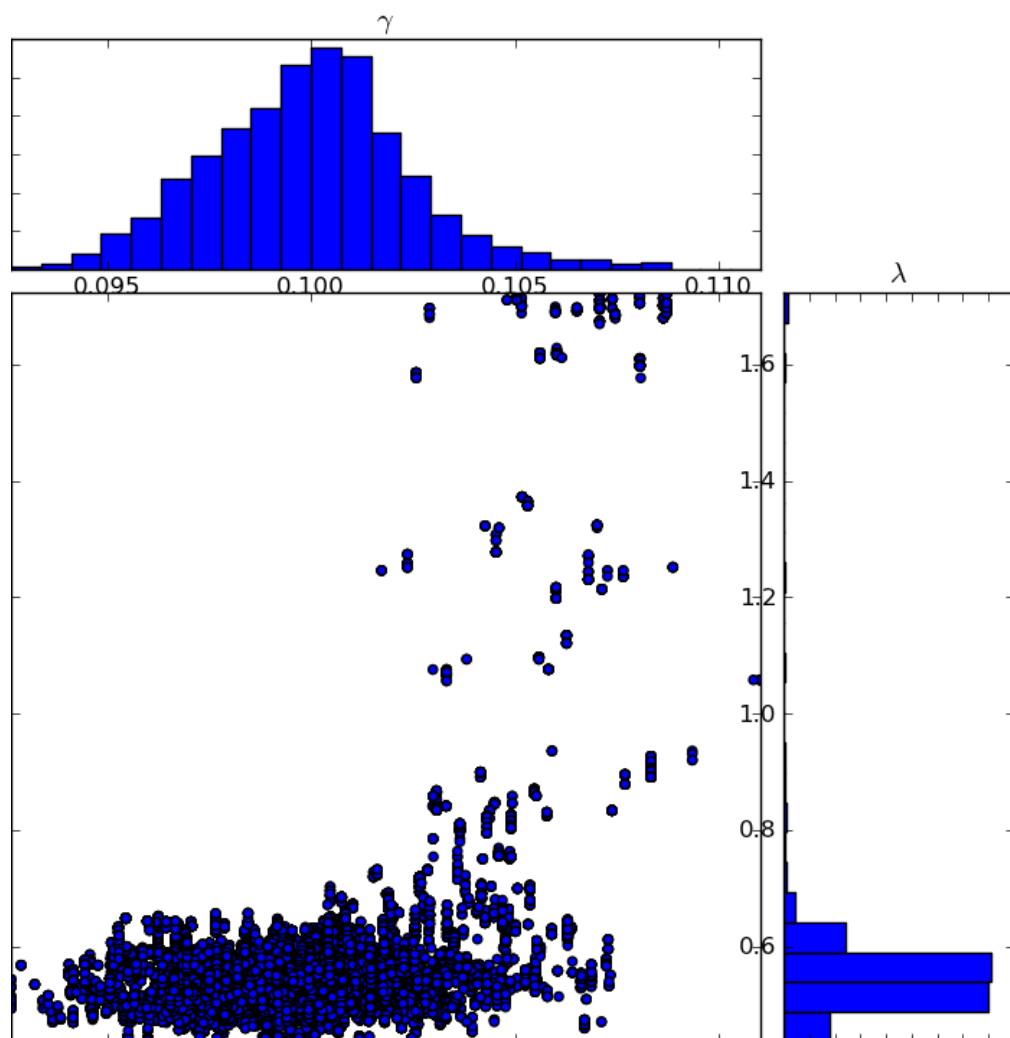


Figure 14.  $\gamma$  and  $\lambda$

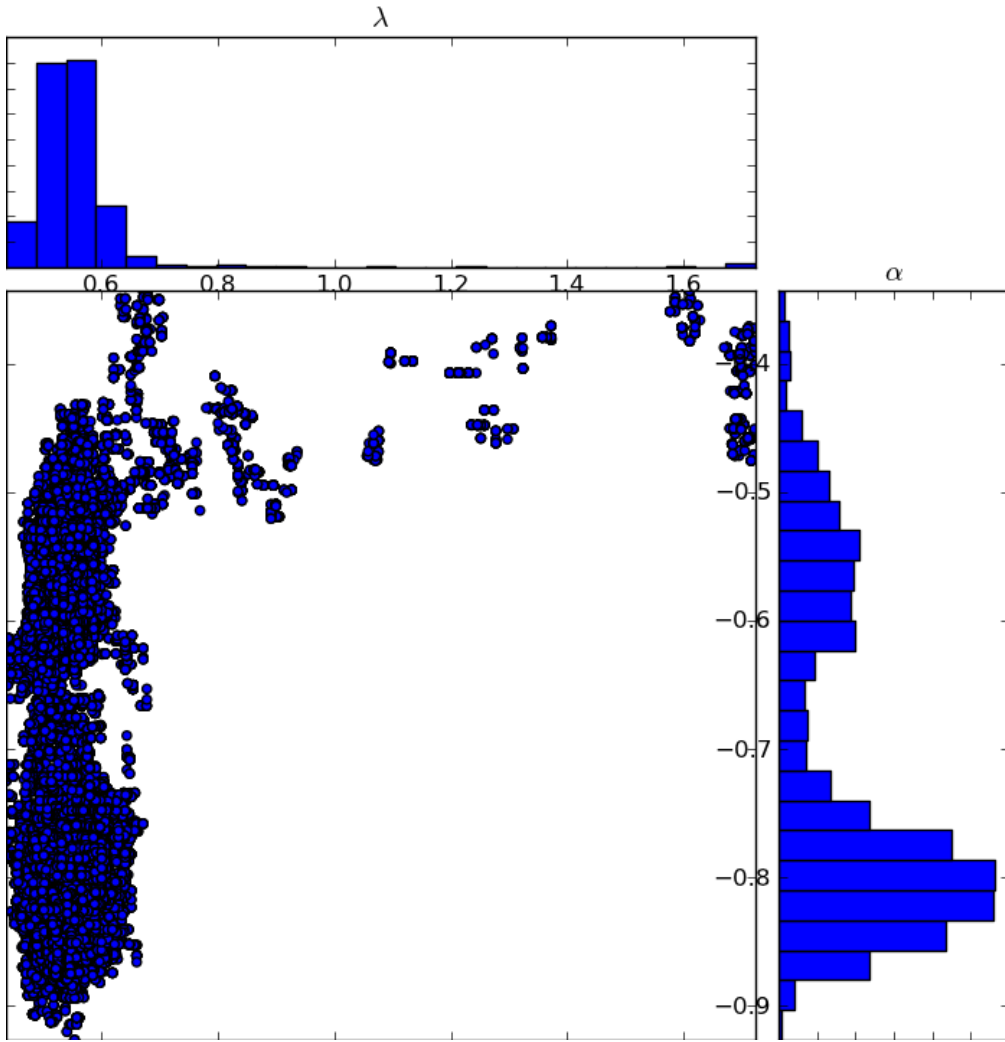


Figure 15.  $\lambda$  and  $\alpha$

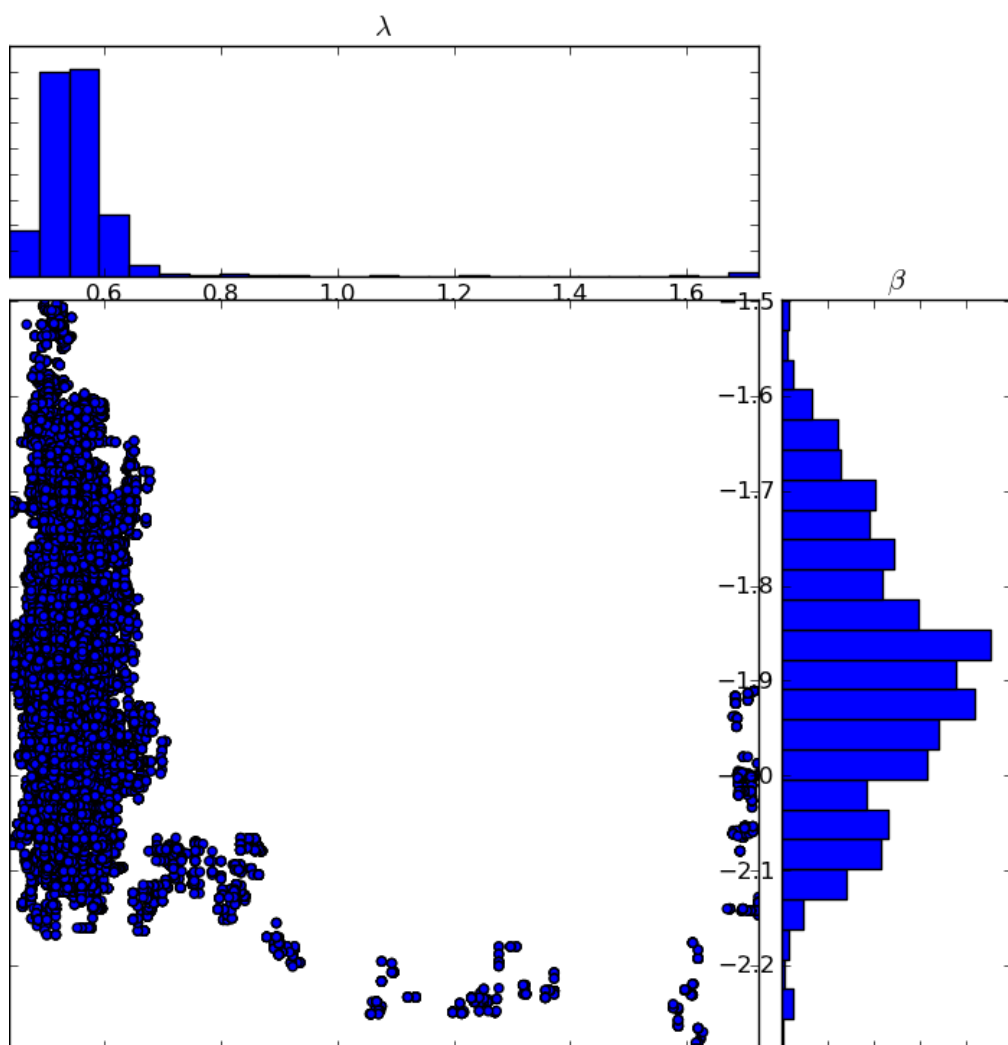


Figure 16.  $\lambda$  and  $\beta$



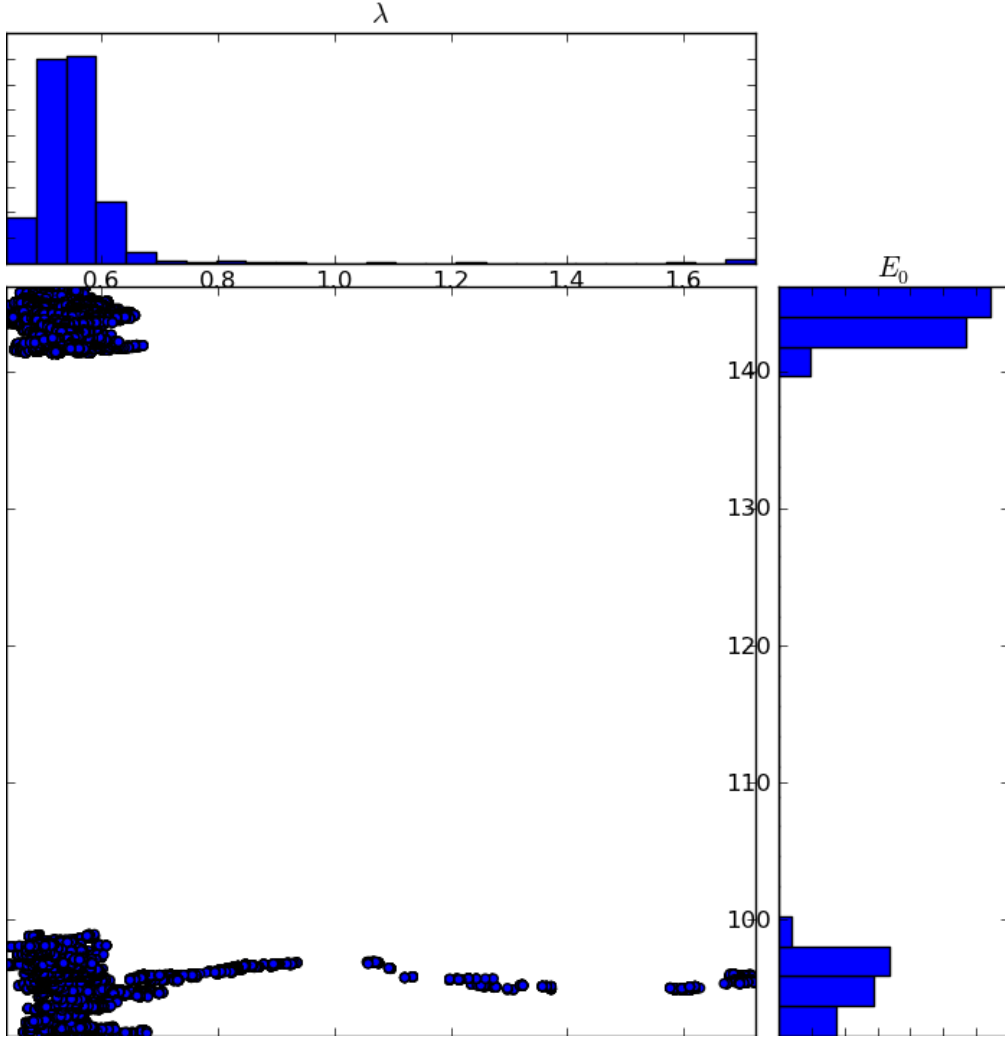


Figure 17.  $\lambda$  and  $E_0$

#### 4.2. Less trivial

We then attempt the same analysis, while allowing the number of pulses comprising the GRB to vary. The mixing for  $J$ , the number of pulses, is very poor. It is embarrassing to include this traceplot, but I have to refer you to Figure (23).

However, a sample iteration shown in Figure (24) indicates that the MCMC does approach some reasonable values for the parameters. The posterior predictive intervals in Figure (25) also indicate good coverage.

Figure (27) shows the factor by which the variance is inflated relative to a Poisson model. Since the data in this simulation are distributed Poisson, the true value of  $r$  is  $\infty$ , and we should expect this histogram to be near to 1.

#### 5. EXAMPLE: GRBS

Coming Soon!

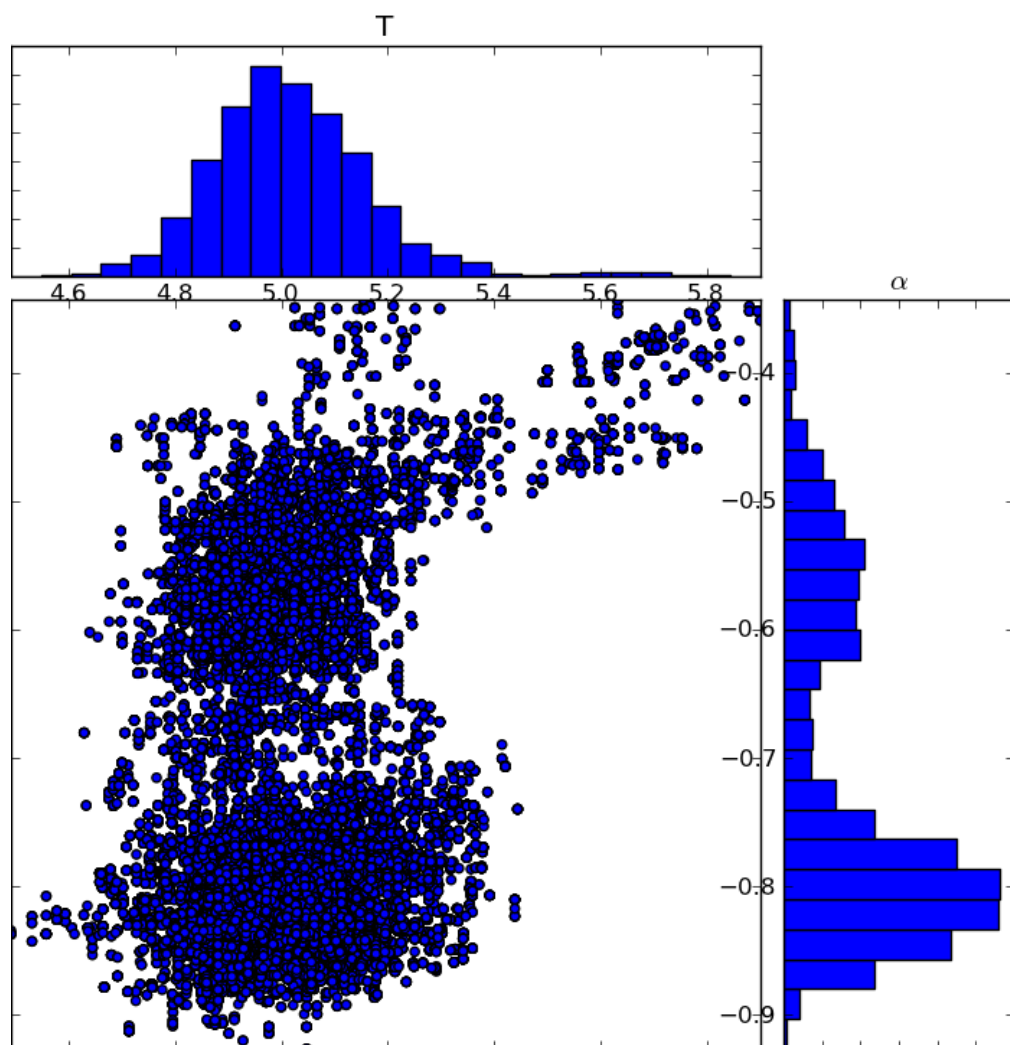


Figure 18.  $T$  and  $\alpha$

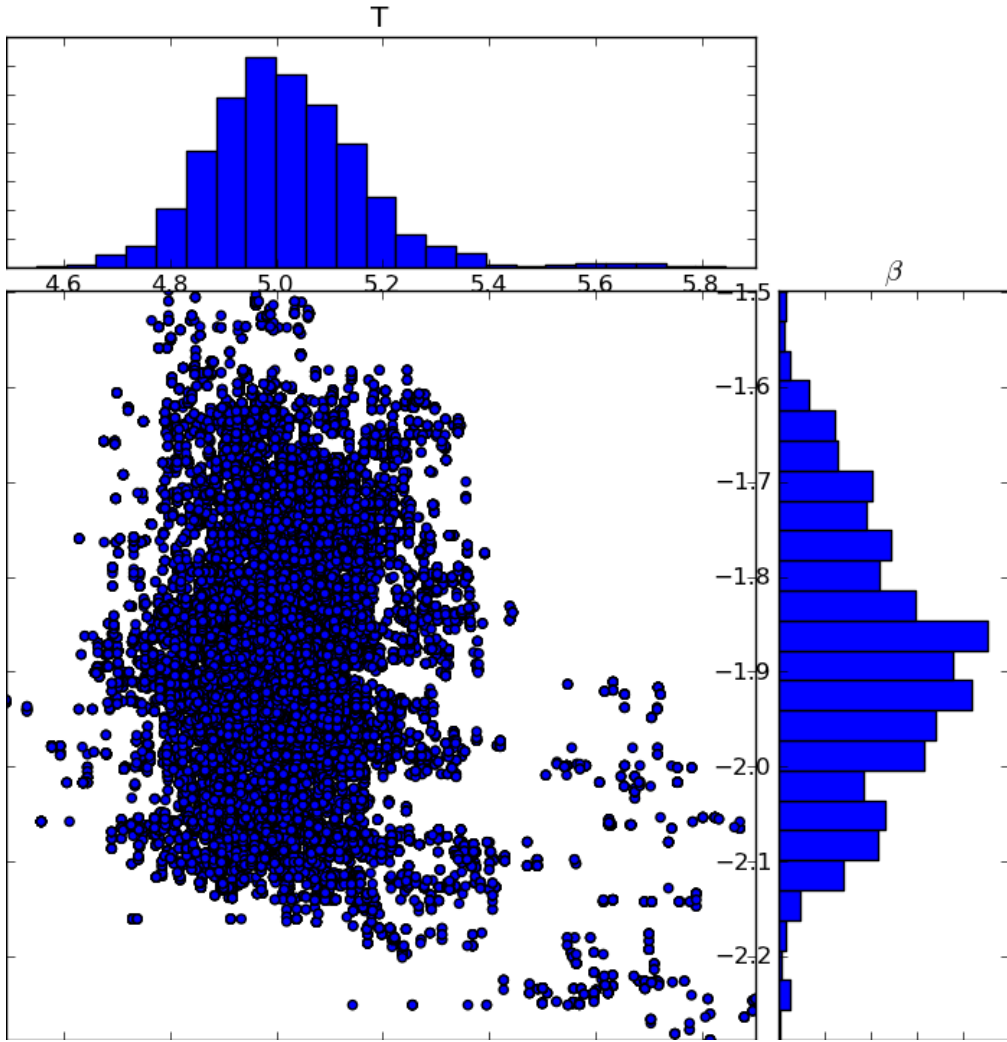


Figure 19.  $T$  and  $\beta$

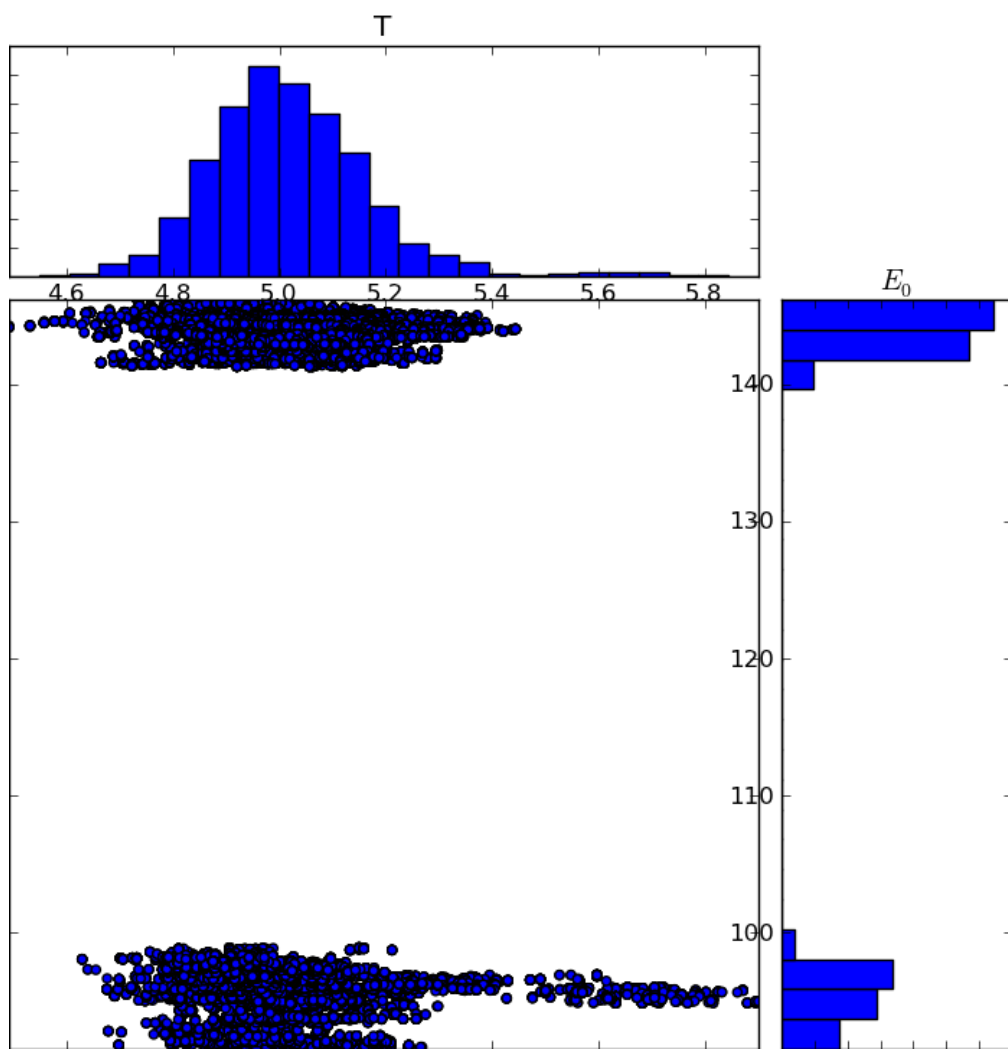


Figure 20.  $T$  and  $E_0$

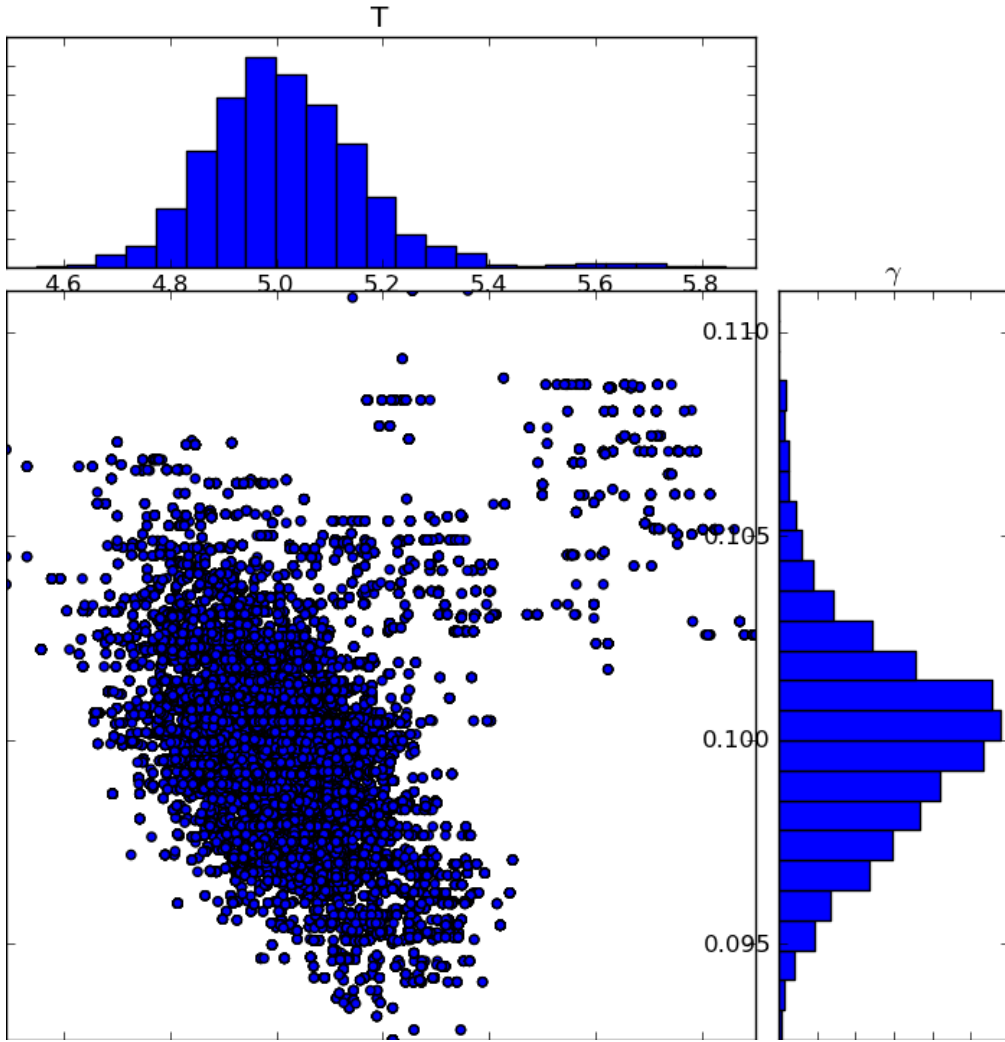


Figure 21.  $T$  and  $\gamma$

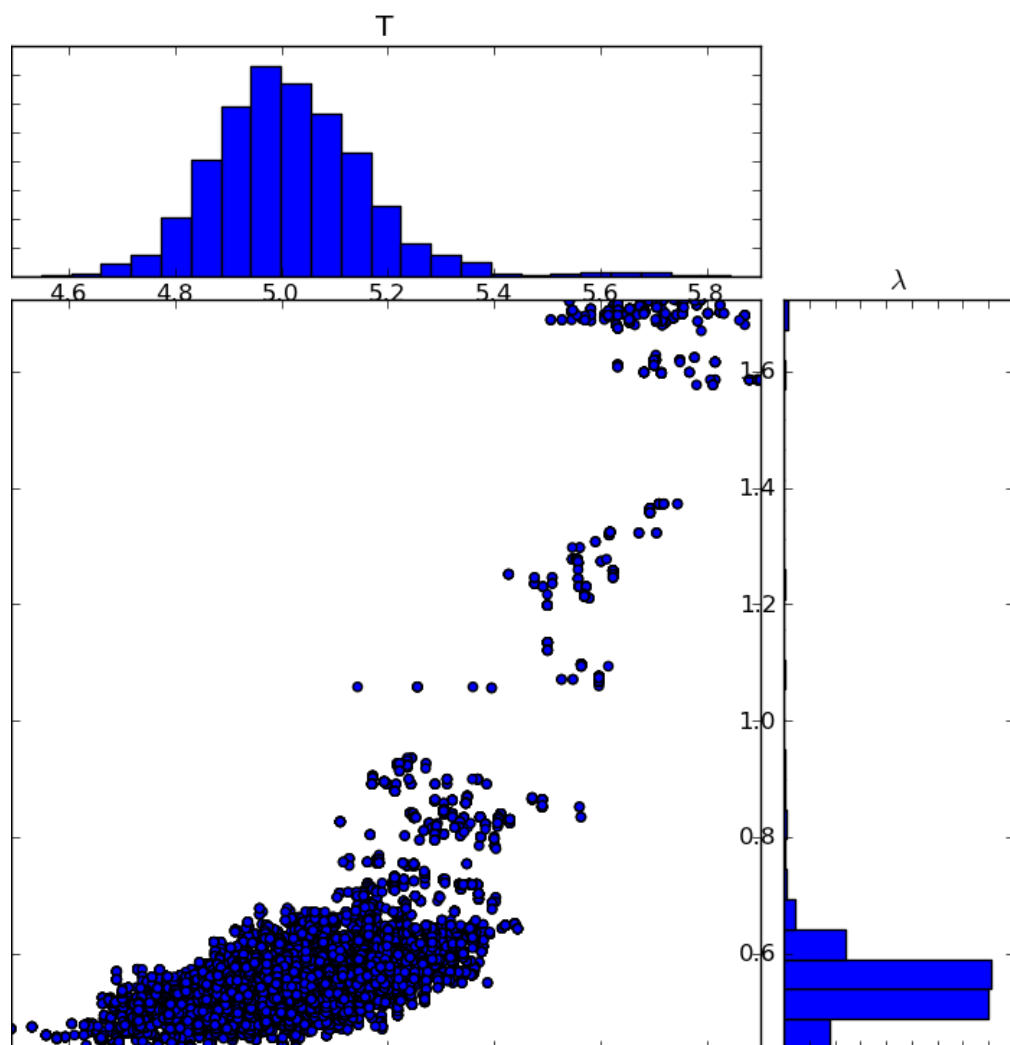
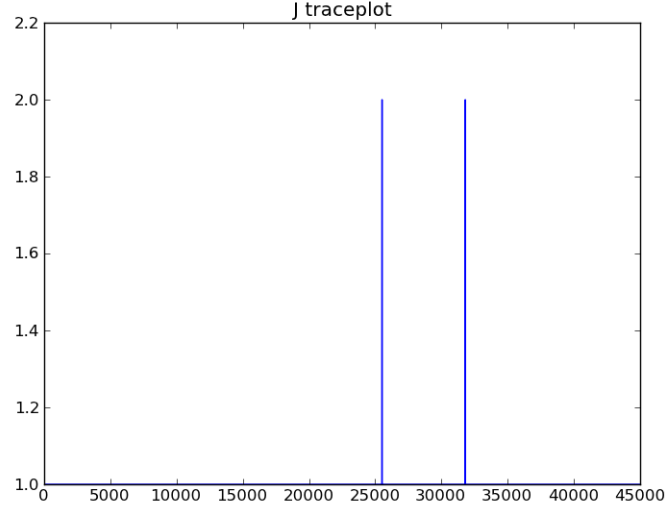
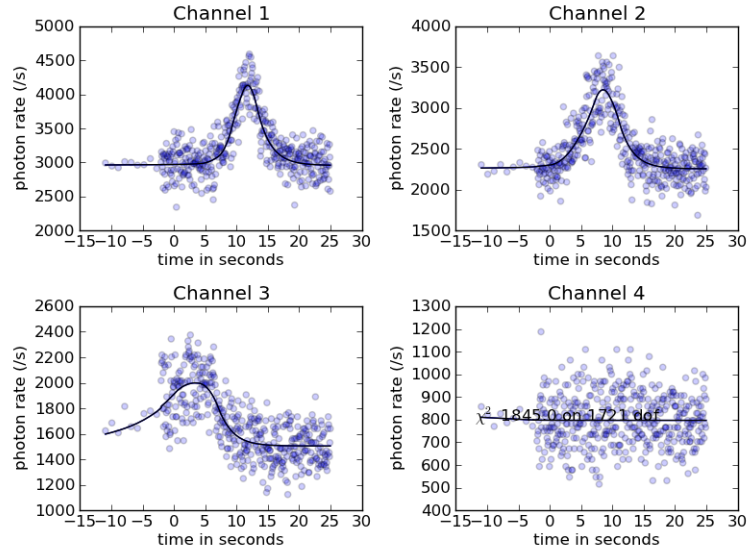


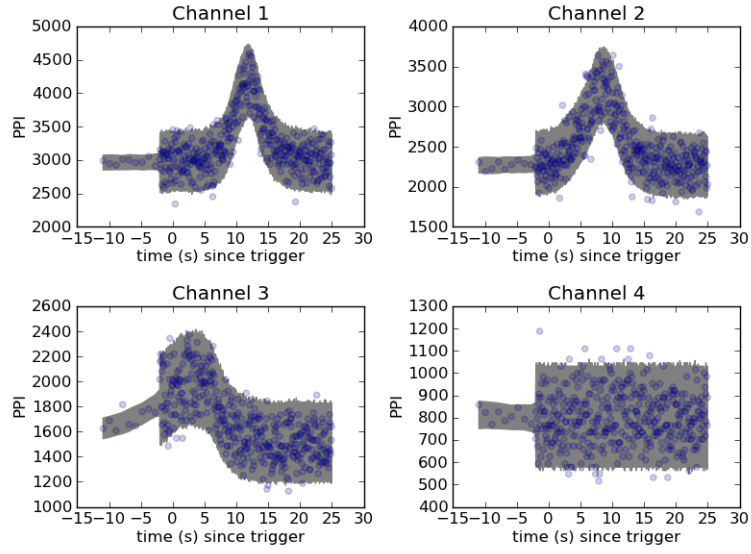
Figure 22.  $T$  and  $\lambda$



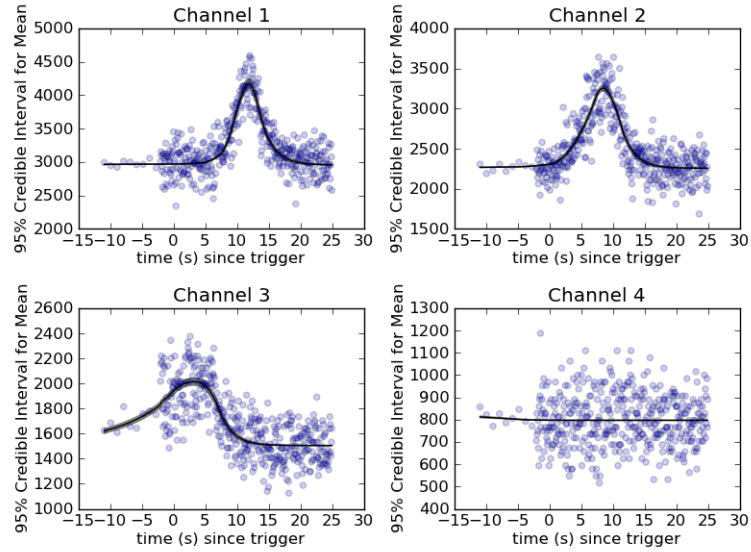
**Figure 23.** Number of pulses  $J$  (less-trivial simulation)



**Figure 24.** Sample posterior draw, (less-trivial simulation)

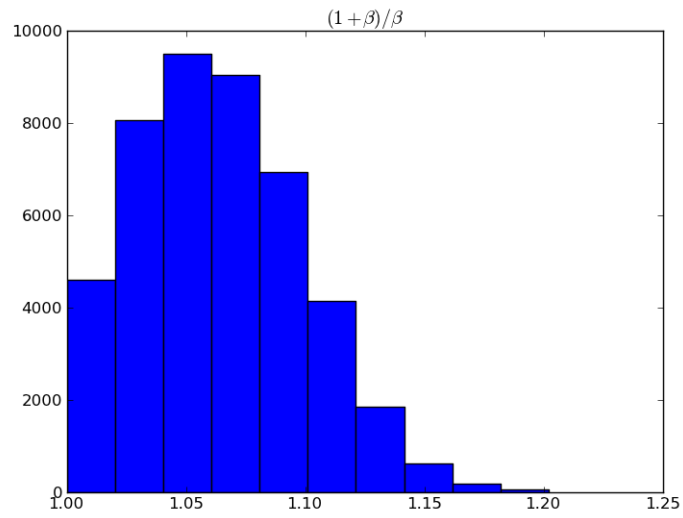


**Figure 25.** Posterior predictive interval (95%) for less-trivial simulation



**Figure 26.** Credible interval (95%) and posterior median for less-trivial simulation





**Figure 27.** Factor by which the variance is inflated, less-trivial simulation