A Bayesian approach to estimate time of maximal “greenness” in Tropical Evergreen India

Manju M. Johny

December 5, 2016

1 Introduction

The behavior of natural vegetation is known to entrain to the cyclical rhythm of seasons. For example, cherry blossoms consistently bloom in the spring, and hoards of watermelon arrive at the supermarket in the summer followed by squashes and pumpkins in the fall. In this regard, a statistical model to estimate the optimal periods of growth for particular crops in a given region would be greatly beneficial to the agriculture industry.

The MERIS Terrestrial Chlorophyll Index (MTCI) is a satellite-measured index of chlorophyll content invaluable in studying the phenology of vegetation. The chlorophyll index of an area is a measure of the “greenness” of the region of interest, with large MTCI values corresponding to greater “greenness.” In this project, we explore MTCI of the Indian sub-continent measured over a span of 4 years. In particular, we limit our analysis to the phenology of one pixel from the available 50 x 50 matrix describing the area spanning the Indian sub-continent. Figure 1 shows a plot of the MTCI values over time for the given pixel of interest. The cyclical pattern in “greenness” is apparent in this plot – with 1 distinct peak of “greenness” in each of the 4 periods present. Accordingly, we utilize a Bayesian approach to determine the times at which maximum “greenness” is achieved per period.

2 Data

This project uses Level 3 Medium Resolution Imaging Spectrometer (MERIS) Terrestrial Chlorophyll Index (MTCI) data provided by NERC Earth Observation Data Center. MERIS operates at a spatial resolution of 4.6 km [2]. MTCI is calculated, using the standard MERIS band settings, as the ratio of differences in reflectance between spectral bands 10 ($\lambda = 753.75nm$) and 9 ($\lambda = 708.75nm$), and spectral bands 9 ($\lambda = 708.75nm$) and 8 ($\lambda = 681.25nm$) [1].

$$MTCI = \frac{R_{10} - R_9}{R_9 - R_8} \tag{1}$$

MTCI measurements were taken, in 8 day intervals between 2004 and 2007, of the area spanning the Indian sub-continent [2]. The MTCI measurements are arranged into 50 x 50 matrices where each cell gives the MTCI value of a particular pixel or area of land. The data set is composed of 4 different files for each year. Each file contains a list of matrices, each of dimension 50 x 50, arranged chronologically by time of measurement. All together, the 4 files contain 184 matrices of dimension 50 x 50. Additionally, we used information classifying the pixels in the matrix by the predominant vegetation type of the area. This information about vegetation type is summarized by another matrix, where each cell represents the predominant vegetation type of the corresponding pixel in our MTCI data set. There are 47 possible vegetation types, each of which are represented by numbers 0 through 46. For this project, we focus on just one pixel belonging to the vegetation type "Tropical Evergreen". Our pixel of interest is the 5^{th} element of the matrix (row = 1, col = 5). Figure 1 shows a plot of the MTCI values by time for the pixel of interest.

3 Methods

3.1 Model Rationale

Figure 1, the plot of MTCI values over time for our pixel of interest, exhibits obvious sinusoidal behavior. We employ Bayesian sinusoidal regression to model the cyclical behavior of the pixel's MTCI values over time. There are 4 distinct peaks in the total span of 4 years. Hence, the period appears to be the length of 1 year. For simplicity, we consider a fixed period in this project defined as the length of one tropical year per 8 days.

$$\omega = \frac{365.2421897}{8} = 45.6553 \tag{2}$$

Along with the sinusoidal terms, we will include an intercept term and a linear trend term in the model.

3.2 Model

For our pixel of interest, let y_t be the MTCI value at the t^{th} time point. Let t, ranging from 1 through 184, be the time point at which each MTCI measurement was taken. Recall that t is measured every 8 days. Additionally, we also assume that y_t are distributed independently.

Data Model

For given time period, t = {1,2,..,184}:

$$y_t \overset{ind}{\sim} N\big(\mu_t, \tau^2\big) \tag{3}$$

$$\mu_t = \beta_0 + \beta_1 t + \beta_2 \text{cos}\bigg(\frac{2\pi t}{45.6553}\bigg) + \beta_3 \ \text{sin}\bigg(\frac{2\pi t}{45.6553}\bigg) \tag{4}$$

Priors

$$\begin{aligned} \tau &\sim Ca^{+}\big(0, 0.04\big) \\ \beta_0 &\sim U\big(0, 10\big) \\ \beta_1 &\sim U\big(-10, 10\big) \\ \beta_2 &\sim U\big(-10, 10\big) \\ \beta_3 &\sim U\big(-10, 10\big) \end{aligned} \tag{5}$$

Valid MTCI values range from 1 to 6 [2]. The uniform priors on the β parameters were chosen such that they allowed coverage of valid ranges of MTCI values. The intercept term, β_0's uniform prior has sufficient density over the possible values for the intercept (values between 1 and 6). Similarly, β_1's uniform prior has sufficient density over the possible values for the linear trend term, including both positive and negative trends. Consider the trigonometric equivalence

$$\beta_2 \text{cos}\bigg(\frac{2\pi t}{45.6553}\bigg) + \beta_3 \text{sin}\bigg(\frac{2\pi t}{45.6553}\bigg) = A\text{cos}\bigg(\frac{2\pi t}{45.6553} - \delta\bigg)$$
$$\text{where the amplitude, } A = \sqrt{\beta_2^2 + \beta_3^2}.$$

Plausible values for the amplitude are confined between 0 and 5 based on the valid ranges of MTCI values. Thus, the uniform priors on β_2 and β_3 were chosen such that they have sufficient density over possible values of the amplitude. The intervals for the β parameters' uniform priors are wider than necessary. We could have also chosen more narrow intervals for the priors on each of the β parameters and still maintained sufficient coverage over plausible values. Lastly, a half Cauchy prior was chosen as a diffuse prior for the standard deviation, τ. Since all of the priors used were proper, we have assurance of posterior propriety.

3.3 Model Fitting

JAGS was used with the rjags package to generate samples from the posterior distribution of the parameters. A 3 chain MCMC was run for 10,000 iterations with a burn-in period of 5,000 iterations using R's default starting values. Ultimately, this culminated in 30,000 MCMC samples generated for the β and τ^2 parameters. The potential scale reduction factors for the parameters, given in Appendix: Table A.1, were all 1. Additionally, the trace plots, shown in Appendix: Figure A.1, indicated well mixed chains. Based on the potential scale reduction factors and the trace plots, there was no indication of lack of convergence in the Markov Chains. Figure 3 shows a comparison of 19 replications from the posterior predictive distribution of y, and our observed data. A comparison of the replications to our observed data shows no indication of lack of fit in our model.

4 Results

We were able to obtain posterior distributions for the unknown parameters along with their estimates using MCMC. The posterior distribution for the β and τ^2 parameters are shown in Figure 4. Additionally, the posterior medians and 95% credible intervals for each of the parameters are reported in Table 1. For each iteration. the resulting posterior β parameters were used to calculate μ_t according to Eq (4). The 95% credible intervals and posterior median for the marginal posterior of μ_t is shown in Figure 2.

In Figure 2, there are 4 distinct peaks in μ_t, where each peak is confined to its respective period ($\omega = 45.6553$ time points). In order to answer the scientific question of interest, we create credible intervals for the time points at which each of these peaks appear. For each of the 30,000 posterior samples, the time point corresponding to maximum μ_t within a period was determined. So, each iteration gave 4 distinct time points per cycle at which maximum μ_t was reached – lets call these values peaktime$_1$, peaktime$_2$, peaktime$_3$ and, peaktime$_4$. From these peaktimes in each of the 30,000 samples, we were able to create 95% credible intervals for peaktime. These credible intervals are summarized in Table 2. Additionally, we may be interested in peaktime as measured from the end of the previous cycle. The posterior 95% credible intervals for this information is summarized in Table 3.

5 Discussion

The objective of this project was to identify the time points at which maximum "greenness" occurred per cycle. We estimated these values by creating posterior credible intervals for peaktimes. Peaktimes are defined as the time points at which maximum posterior μ_t is achieved per cycle. We obtained estimates for each of the four peaktimes as summarized in Table 2. The posterior median peaktimes, over the four year period, were at the 29^{th}, 75^{th}, 121^{st}, and 166^{th} time points. The 95% credible intervals around the posterior medians are tight, varying at most by 1 time point from the median. Additionally, we may further be interested in the peaktimes as measured from the end of the previous period. The posterior median peaktimes, at each of the four periods, occurred respectively at 29, 29, 30, and 29 time points after the end of the previous period. The 95% credible intervals for peaktimes, for all four periods, are between 29 and 30 time points after the end of the previous period. Converted to days, these time points correspond to 232 and 240 days respectively. The similarity of peaktimes over the four periods shows consistency among the times at which maximum "greenness" is achieved year to year from 2004 to 2007.

6 Improvements and Future Work

For simplicity, we assumed a fixed period defined as the length of 1 tropical year per 8 days. However, the true period need not be exactly that length. A better approach may have been to put a prior on the period, ω, and estimate it along with the other parameters. Additionally, the phenology of the Indian sub-continent is known to house diverse land types, some of which exhibit distinct phenological behaviors. For example, certain land types have more than one growing season – giving multiple peaks in MTCI values per year. It could be interesting to build a hierarchical model, grouping together pixels by vegetation type. Additionally, we could have also approached this problem from a dynamic linear modeling perspective. This project, while an illuminating first step, merely scratches the surface of greater statistical and phenological understanding that could be fostered from this data set.

7 Figures

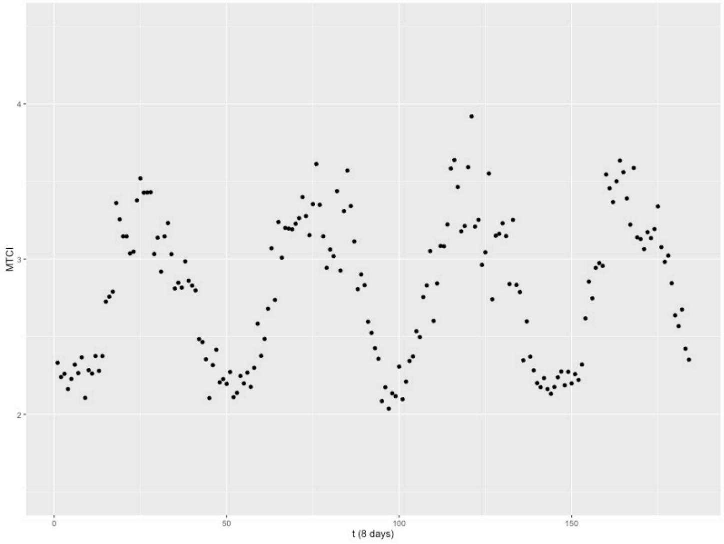

Figure 1: Plot of MTCI values over time for the pixel of interest.

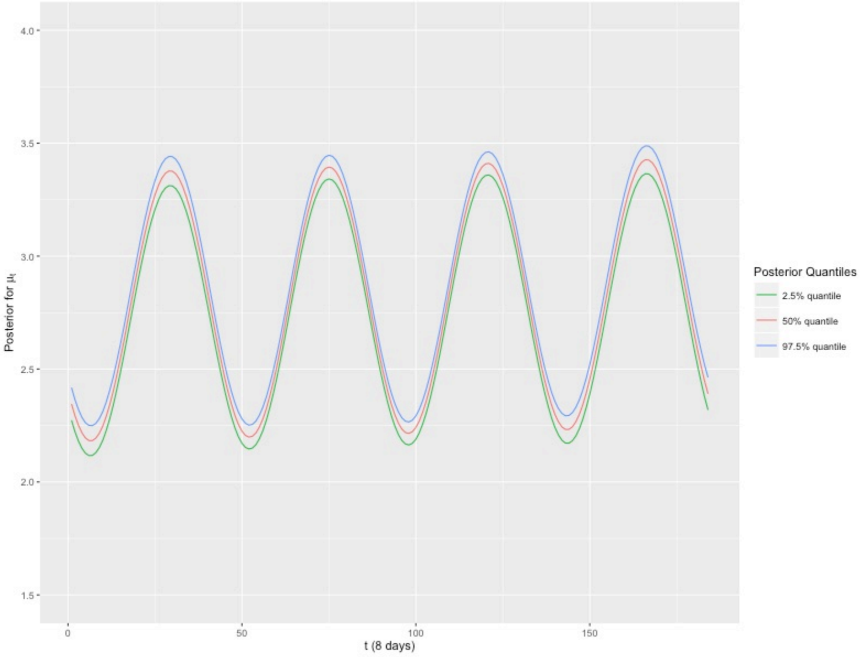

Figure 2: Quantiles for the posterior distribution of μ_t. The red line indicates the posterior median for μ_t. A 95% credible interval for μ_t is contained in the area between the green line (2.5^{th} quantile) and blue line (97.5^{th} quantile).

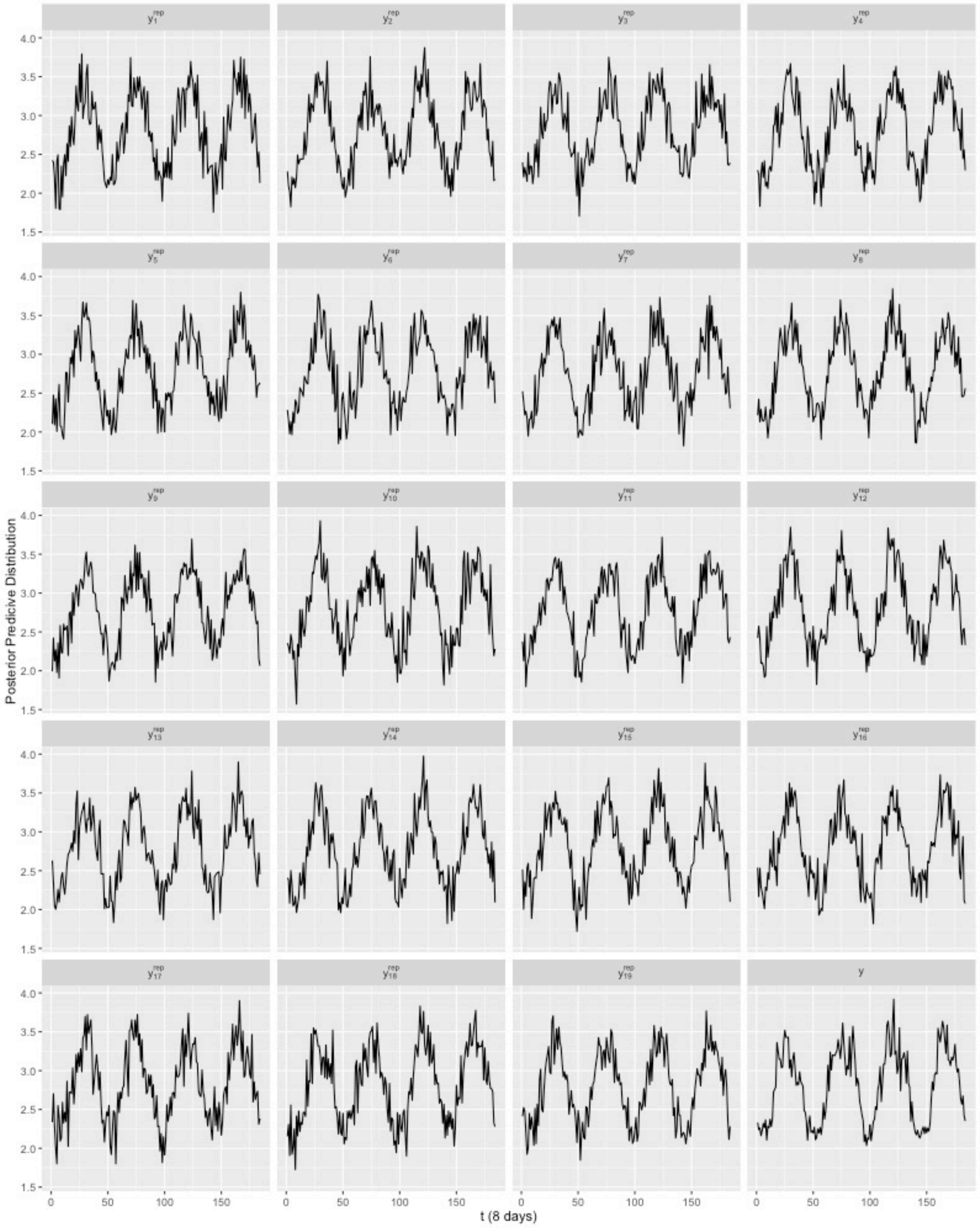

Figure 3: Comparison of 19 replications from the posterior predictive distribution of y (y^{rep}), and observed data (y).

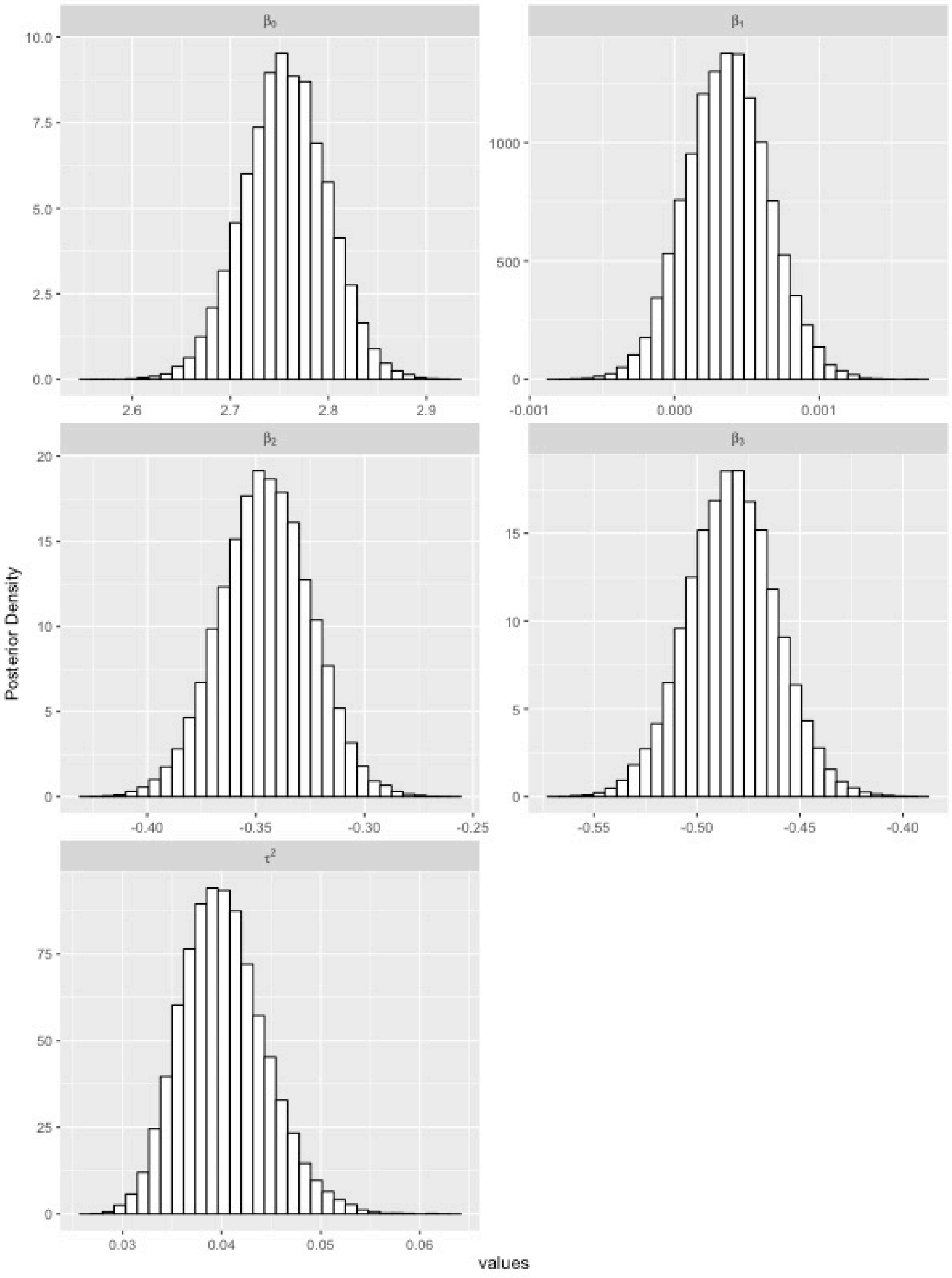

Figure 4: Histograms of the posterior distribution of $\beta_0, \beta_1, \beta_2, \beta_3, \tau^2$ based on 30,000 posterior samples generated using MCMC.

8 Tables

Table 1: Summaries of the posterior distributions of β and τ^2 parameters. The posterior median, and central 95% posterior credible intervals for each parameter is shown below.

	Posterior median	95% posterior credible interval
β_0	2.76	[2.67, 2.84]
β_1	0.0004	[-0.0002, 0.0009]
β_2	-0.35	[-0.39, -0.30]
β_3	-0.48	[-0.53, -0.44]
τ^2	0.040	[0.033, 0.049]

Table 2: Summaries of the posterior distributions of peaktimes. "peaktime$_p$" represents the time point at which μ_t is maximized over the p^{th} period. The posterior median, and central 95% posterior credible interval for peaktimes at each of the 4 periods is shown below.

	Posterior median	95% posterior credible interval
peaktime$_1$	29	[29, 30]
peaktime$_2$	75	[75, 76]
peaktime$_3$	121	[120, 121]
peaktime$_4$	166	[166, 167]

Table 3: Summaries of the posterior distributions of peaktimes as measured from the end of the previous period. "peaktime$_p$" represents the time point (starting from the end of the previous period) at which μ_t is maximized over the p^{th} period. The posterior median, and central 95% posterior credible interval for peaktimes at each of the 4 periods is shown below.

	Posterior median	95% posterior credible interval
peaktime$_1$	29	[29, 30]
peaktime$_2$	29	[29, 30]
peaktime$_3$	30	[29, 30]
peaktime$_4$	29	[29, 30]

9 Appendix

Table A.1: Potential scale reduction factors for estimated parameters, β_0, β_1, β_2, β_3, τ^2, as calculated using the coda package. Point estimates and upper confidence limits of the potential scale reduction factors are reported.

	Point estimate	Upper confidence limit
β_0	1	1
β_1	1	1
β_2	1	1
β_3	1	1
τ^2	1	1

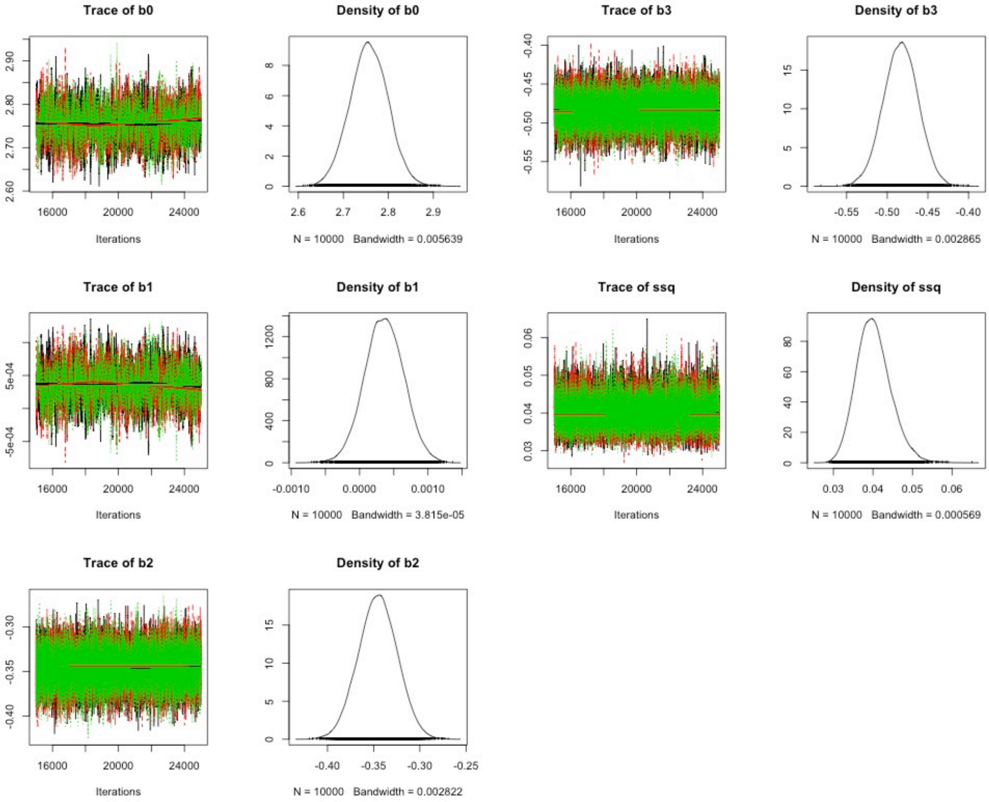

Figure A.1: Trace plots and densities of the estimated parameters. β_0, β_1, β_2, β_3, and τ^2 are represented as "b0", "b1", "b2", "b3", and "ssq" respectively in the plots.

References

[1] P.M. Atkinson C. Jeganathan, J. Dash. Mapping the phenology of natural vegetation in india using a remote sensing-derived chlorophyll index. *International Journal of Remote Sensing*, 31(22):5777–5796, 2010.

[2] P.M. Atkinson J. Dash*, C. Jeganathan. The use of meris terrestrial chlorophyll index to study spatio-temporal variation in vegetation phenology over india. *Remote Sensing of Environment*, 114:1388–1402, 2010.