ST440/540 – Mid-term take-home exam – Due 4/9

THIS IS AN EXAM - DO NOT DISCUSS THE PROBLEM WITH ANYONE (INCLUDING OTHER STUDENTS OR THE TA)! If you have questions please visit office hours or email me.

The data for this analysis are simulated to represent satellite measurements of Normalized Difference Vegetation Index (NDVI), which is a common measure of "greenness" used in many fields ranging from ecology to agriculture. The true value of NDVI was simulated for 6 pixels over the course of 365 days. Let θ_{tj} be the true NDVI on day t in pixel j, and $\boldsymbol{\theta}_t = (\theta_{t1}, ..., \theta_{t6})^T$ be the collection of NDVI on day t. The true values were simulated as

$$\boldsymbol{\theta}_1 \sim \text{Normal}(\mu_1, \Sigma_1) \text{ and } \boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1} \sim \text{Normal}(\mu_2 + \rho\boldsymbol{\theta}_{t-1}, \Sigma_2),$$

where μ_1 and μ_1 are mean vectors, $\rho \in (0, 1)$ controls temporal dependence, and Σ_1 and Σ_2 are 6×6 covariance matrices.

The (Fake) data come from three (fake) satellites:

1. `Y1` is a 365×6 matrix of observations from the gold-standard satellite. The $[t, j]$ element of `Y1` is known to be an unbiased (but noisy) measurement of θ_{tj}, i.e., you can assume that the expected value of `Y1[t,j]` is θ_{tj}. Unfortunately `Y1` is missing 80% of the time (assume all missing data is missing completely at random).

2. `Y2` is a potentially biased and noisy measurement that is missing only 10% of the time. However, this satellite only provides a measurement of the NDVI in pixel 1, so that `Y2[t]` is measuring θ_{t1}.

3. `Y3` is a another potentially biased and noisy measurement that is missing only 10% of the time. However, this satellite only provides a measurement of the NDVI for the entire spatial domain, so that `Y3[t]` is measuring $\sum_{j=1}^{6} \theta_{tj}/6$.

These three variables are given in the `R` workspace `E2.RData` on the course webpage,

`https://www4.stat.ncsu.edu/~reich/ABA/assignments.html`

The data are also plotted on the final two pages on this exam. Your objective is to use Bayesian methods to estimate θ_{tj} for all $t = 1, ..., 365$ and $j = 1, ..., 6$. Denote $\hat{\theta}_{jt}$ as your estimate of the true NDVI, θ_{tj}.

You should summarize your work with both an analysis report and prediction table.

- **Analysis report**: The report should be no more than **4 pages** (11 font, 1-inch margins). You **must** have the following four sections, each on a separate page (although do not feel obliged to fill the entire page):

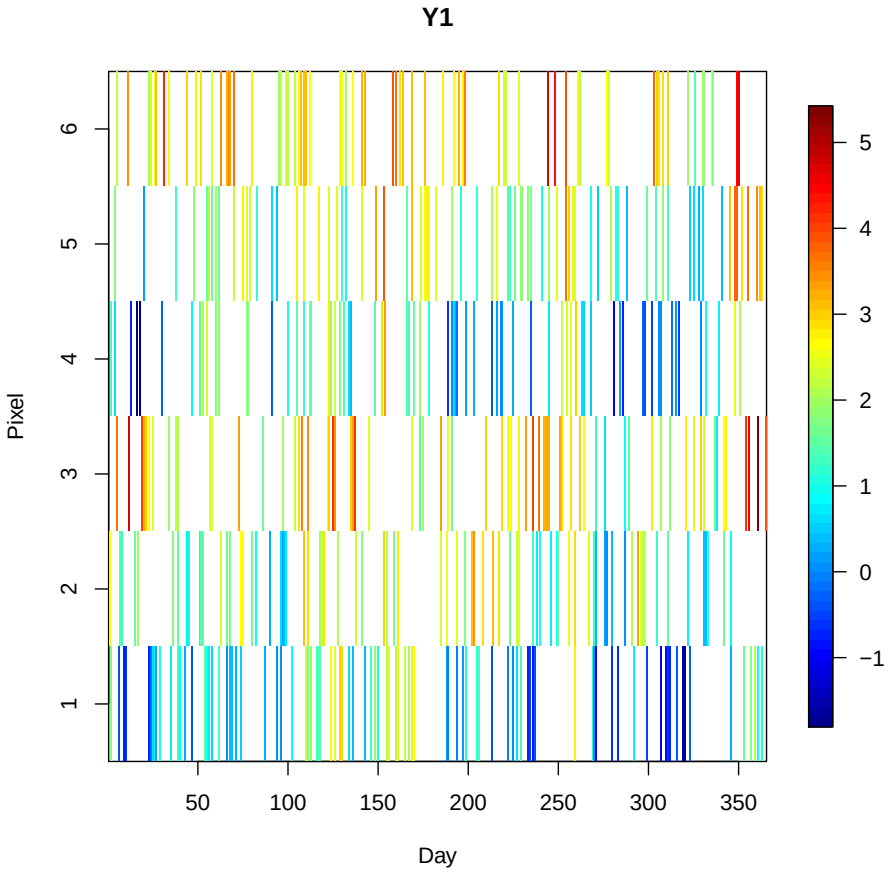
    1. **Statistical model**: This should be written in paragraph form with both a mathematical description of your model and complete sentences arguing that your model is reasonable. Be sure to include sufficient detail that a classmate could reproduce your work.
    2. **JAGS code**: Include only JAGS model statement (i.e., the string that defines the model). When possible use the same variable names in JAGS as you did the description of the model on Page 1, and add some comments to identify the key lines in the code. If you use a function other than JAGS, then give the code you used as well as a detailed description of each step in the code.
    3. **Convergence diagnostics**: Examine convergence using figures and/or tables and summarize the results in a paragraph with complete sentences.
    4. **Final results**: Justify that your model fits well by comparing the posterior means $\hat{\theta}_{tj}$ with the observed data. Summarize the plots and/or tables in a paragraph with complete sentences.

    Staple the four (or two if you print on both sides) pages together.

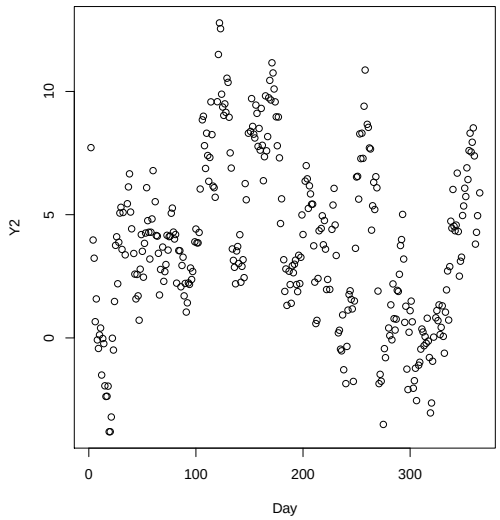
- **Prediction table**: Email a .csv file (use the write.csv function) to your TA that includes the 365×6 matrix of $\hat{\theta}_{jt}$. The matrix should have 6 columns (no row labels!) and 365 rows (no column labels!). **Your file must follow this exact format** or we will not be able to score your predictions. Save the file as FirstnameLastname.csv (i.e., mine would be saved as BrianReich.csv) and email it to your TA. Your prediction mean squared error, i.e., $MSE = \sum_{j=1}^{6} \sum_{t=1}^{365} (\theta_{tj} - \hat{\theta}_{tj})^2$, will be part of your exam grade. More importantly, the top scores will present their results in class and win a *huge* prize.

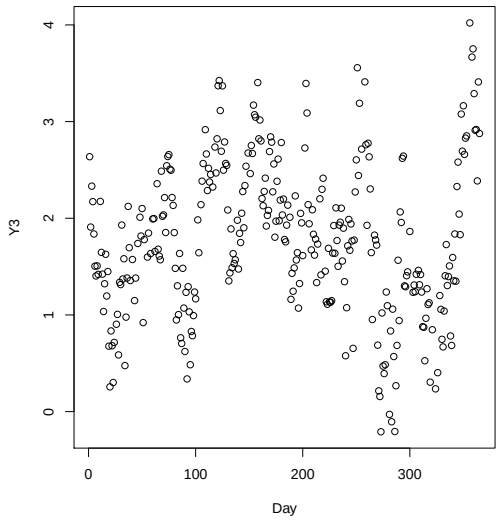
Plots of the satellite data

(a) Data from satellite 1 (missing is white)



(b) Data from satellites 2 and 3





(c) Data from all three satellites