A Bayesian mixed effects model to capture the effect of interseasonal weather on phenology

Dan Cunha

Boston University

August 26, 2022

Table of Contents

1. Problem
2. Model
3. Inference and computation
4. Results

Harmonic regression for phenology modeling

Let h be the number of harmonics,

$$\epsilon_i \overset{i.i.d.}{\sim} N(0, \sigma^2), i = 1, ..., n$$

$$y_i | t_i, \beta \sim \beta_0 + \beta_1 t_i + \sum_{l=1}^{h} \beta_{2l} sin(\frac{2\pi t_i l}{365}) + \beta_{2l+1} cos(\frac{2\pi t_i l}{365}) + \epsilon_i$$

And define X as the corresponding design matrix.

Harmonic regression is too rigid for phenological modeling

1. Harmonic regression assumes the same phenology each year
2. But each aspect of phenology, such as the green-up or senescence, naturally change year to year
3. The phenology model should incorporate that flexibility
4. A repeated measurement model can do just that

Harmonic regression is too rigid for phenological modeling

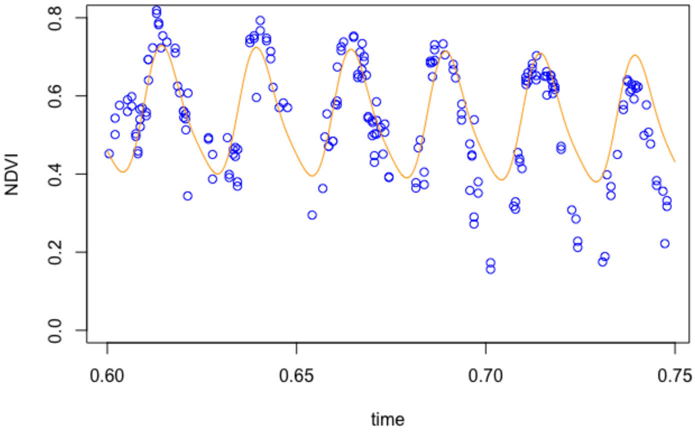


Table of Contents

1 Problem

2 Model

3 Inference and computation

4 Results

Towards a Bayesian repeated measurement model

1. Repeated measurement models can provide flexibility to learn natural changes in phenology each year
2. Common repeated measurement models include mixed effects models
3. For example, a mixed effects model could learn a random effect for the harmonics each year.
4. This would enable scientists to infer, for example, a separate green-up time each year.

Towards a Bayesian repeated measurement model con't

1. The frequentist treatment of mixed effects models considers the random effects as nuisance parameters
2. But in phenology modeling, we are interested in plotting the mean predictive function given the data, as well as the series of random effects through time.
3. We can use a Bayesian approach instead, capturing the posterior distribution of the yearly random effects, in order to understand how phenology naturally changes each year.

A Bayesian repeated measurement model

Let J be the number of years in the study, d the number of random effects, and n the number of observations,

$$\epsilon_i \overset{i.i.d.}{\sim} N(0, \sigma^2), i = 1, \ldots, n$$
$$y \sim X\beta + Z\gamma + \epsilon$$
$$\gamma \sim N(0_{Jd \times 1}, T)$$

Where $Z \in \mathbb{R}^{n \times Jd}$ is the design matrix for the random effects, and T is diagonal with parameters $\tau_{l=1,\ldots,d}$, each having J repeated entries,

$$T = diag(\tau_1, \ldots, \tau_1(totalJ), \ldots, \tau_d, \ldots, \tau_d)$$

Table of Contents

1 Problem

2 Model

3 Inference and computation

4 Results

Approach for inference and computation

1. We are tasked with modeling trillions of pixels, so computation needs to be fast
2. We also wish to take a Bayesian approach to make an inference about the expectation of the random effects given the data
3. Expectation Maximization is suitable for this setting

Posterior inference of the random effects

1. During EM, the posterior expectation of γ is under the distribution $p(\gamma|y, \beta^{(t)})$.
2. Since this distribution conditions on y and $\beta^{(t)}$, we can think of the likelihood as being a function of just the parameters γ, given the *working response* $(y - X\beta^{(t)})$. This likelihood is still normally distributed with respect to the working response, leading to an analytical posterior expectation calculation for γ.
3. Furthermore, we can use $(y - \mathbb{E}(Z\gamma))$ as the working response when performing the M-step for β.
4. The remaining M-step calculations for σ^2 and T are analytically tractable.

Table of Contents

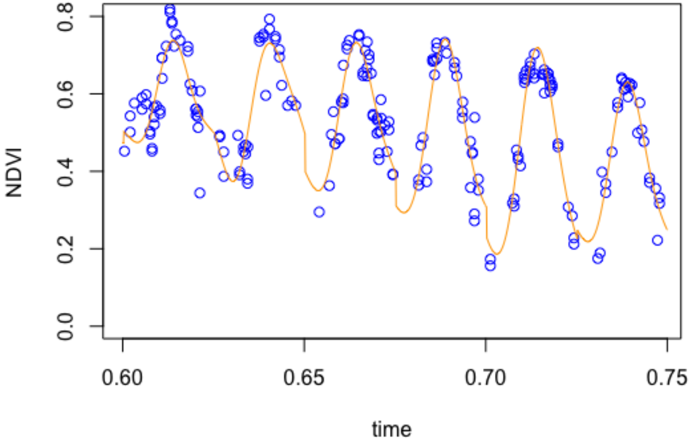
1 Problem

2 Model

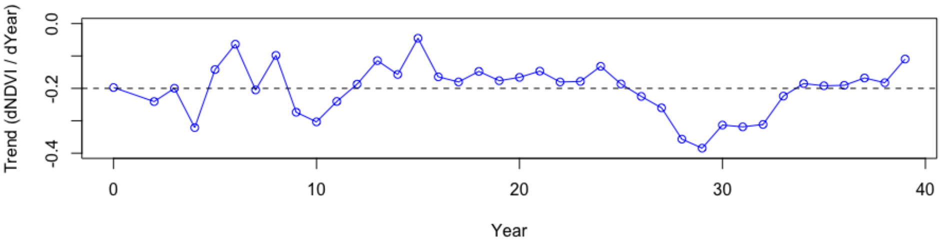
3 Inference and computation

4 Results

Random effects on intercept and first harmonics fits well



The random effects provide insight into growth and drought patterns



Next steps

1. Include uncertainty of posterior estimates
2. Use labeled change point time series to cross validate model flexibility