12-2018

Spatio-Temporal Reconstruction of Remote Sensing Observations

Kamrul Khan
University of Arkansas, Fayetteville

Spatio-Temporal Reconstruction of Remote Sensing Observations

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Statistics

by

Md Kamrul Hasan Khan

University of Dhaka
Bachelor of Science in Statistics, Biostatistics & Informatics, 2010
University of Dhaka
Master of Science in Statistics, Biostatistics & Informatics, 2011

December 2018
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

Avishek Chakraborty, PhD
Thesis Chair

Giovanni Petris, PhD
Thesis Chair

John Robert Tipton, PhD
Committee Member

Abstract

The USDA Forest Service aims to use satellite imagery for monitoring and predicting changes in forest conditions over time within the country. We specifically focus on a $230,400$ hectares region in north-central Wisconsin between 2003 - 2012. The auxiliary data collected from the satellite imagery of this region are relatively dense in space and time and can be used to efficiently predict how the forest condition changed over that decade. However, these records have a significant proportion of missing values due to weather conditions and system failures. To fill in these missing values, we build spatiotemporal models based on fixed effect periodic patterns, spatial random effects with conditional autoregressive prior and a first-order autoregressive temporal effect. Multiple validation and comparison diagnostics are run to identify the best performing model for each of the auxiliary variables as well as for basal area. Findings from our analysis are represented with a series of maps followed by a discussion of their agreement with known spatial patterns across the landscape.

Acknowledgements

First I would like to thank my thesis advisors Dr. Avishek Chakraborty and Dr. Giovanni Petris of the Department of Mathematical Sciences at University of Arkansas for their continuous guidance throughout my thesis work. Besides, I would like to thank my thesis committee member Dr. Tipton for his meaningful comments. I would also like to acknowledge Dr. Barry T. Wilson, USDA Forest Service, for providing the satellite data used in this thesis work.

I sincerely acknowledge continuous support from the USDA Forest Service Northern Research Station (15-JV-11242305-104) for conducting my thesis research. I utilized computational resources from the Arkansas High Performance Computing Center which is funded through multiple National Science Foundation grants and the Arkansas Economic Development Commission.

Finally, I would like to thank my family for their continuous support.

List of Abbreviations

$\mathcal{N}$ ..... Univariate Normal distribution

$\mathcal{N}_k$ ..... k-variate Normal distribution

IG ..... Inverse-Gamma distribution

I_n ..... $n \times n$ identity matrix

$\mathbf{0}$ ..... Column vector of all zeros with dimension as appropriate for the context

$\mathbf{1}$ ..... Column vector of all ones with dimension as appropriate for the context

$|A|$ ..... Number of elements inside the set A

Table of Contents

List of Figures

List of Tables

Chapter 1

Spatial and Spatio-Temporal Modeling

1.1 Introduction

Now-a-days the use of satellite in collecting data over a period of time is increasing rapidly because it is easy, quick and inexpensive. However, usually this kind of data contains lots of missing values due to bad weather or malfunction of sensor. But a full map is required before using it in a model. Hence, in this thesis a hierarchical model is developed for reconstructing complete spatiotemporal map.

1.2 Spatial Modeling

Spatial data are data that have geographic reference. It represents not only the location but also the size, shape and attributes of an object on planet like forest, lake, road, city etc. Due to the development of science and technology the use of spatial data in diverse areas such as geological science, climatology, environmental science, biological science, public health etc increases significantly from last few decades.

There are three types of spatial data. (i) Point-referenced data: Spatial data are called point-referenced data, also known as geostatistical data, if the response $y(s)$ at location $s \in D \subseteq \mathbb{R}^r$ is a random vector and s varies continuously over D. In the spatial context, usually $r > 1$, for instance, $r = 2$ means latitude and longitude, or $r = 3$ says latitude, longitude, and altitude above sea level. On the other hand, time series follows this approach with $r = 1$. (ii) Areal data: In this type of dataset the fixed subset, D, is divided into a finite collection of regular or irregular shaped areal units, say $\{W_1, W_2, \ldots, W_S\}$, that are separated with well-defined boundaries. The outcome y_s is a summary over the areal unit

W_s. Since this kind of data is used here, we are explaining it with an example. Figure 1.1 shows the proportion of a surveyed population whose household income falls below 200% of poverty level in several regions of Hennepin County, Minnesota (Banerjee et al., 2014). Here the areal units are separated by regional boundaries. (iii) Point pattern data: Here $y(s)$ represents the occurrence or not of an event and locations s are random that means given $y(s)$ we want to know which location it corresponds to.

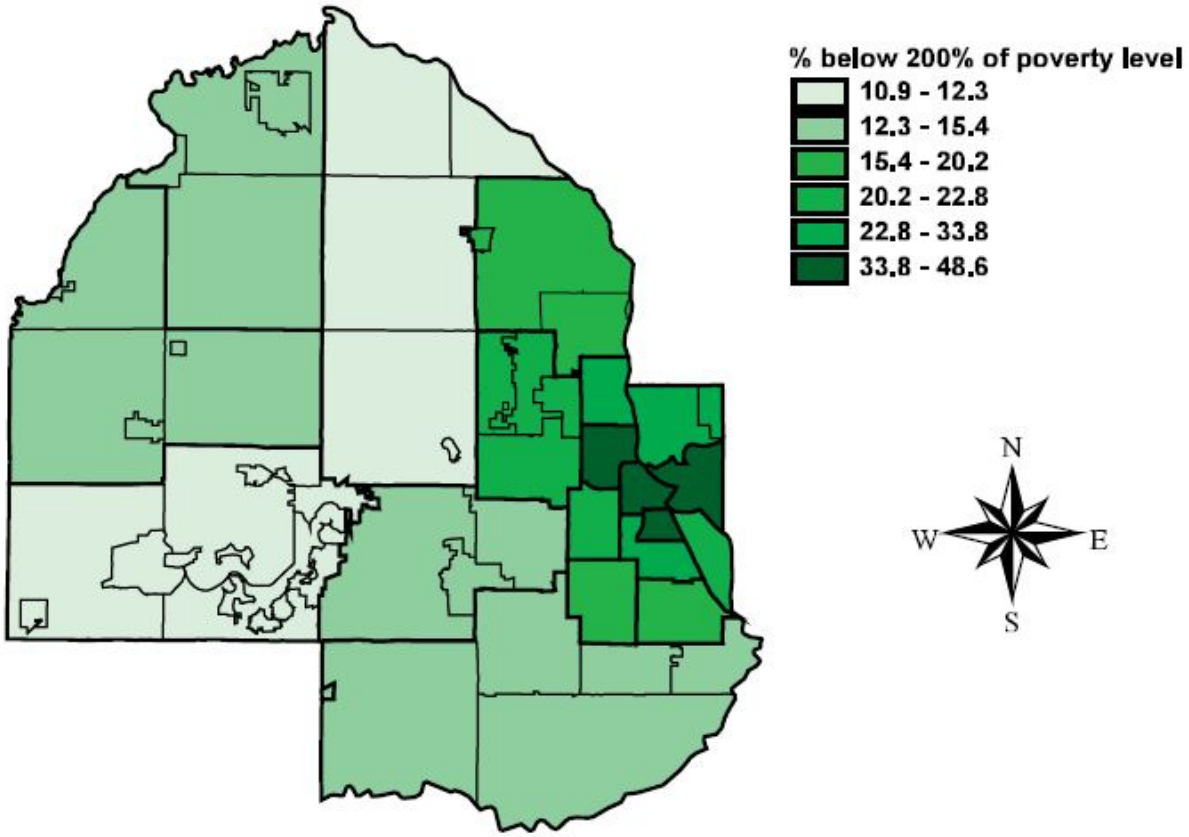


Figure 1.1: Map of percentage of surveyed population with household income below 200% of the federal poverty limit in several regions of Hennepin County, Minnesota (Figure 1.2 in Banerjee et al. (2014))

1.2.1 Modeling for Areal Data

This thesis work specifically focus on areal data. Let the given region, D, is partitioned into S areal units say $\{W_1, W_2, \ldots, W_S\}$ and the summary response at unit W_s is y_s. Now we

can write the regression model as:

$$y_s = X^T(s)\beta + e(s) \tag{1.1}$$

where $X(s)$ is the predictor variable at unit W_s, β is the regression coefficient and $e(s)$ is the residual at s-th areal unit. (1.1) is known as the multiple linear regression model. Now if we get a high value of adjusted R^2 we can use this model to do prediction. But if the adjusted R^2 is low, we first have to figure out the reason and based on the reason the model needs to be modified. Some possibilities are:

- The response and covariates are not strongly correlated.
- This model requires additional important covariates.
- Measurements from adjacent areal units may have some additional information that cannot be captured by the regressors.

Now if the third is the case, we need to modify the model by separating the residual into spatial, $\phi(s)$, and nonspatial, $\epsilon(s)$, random effects. Therefore, (1.1) becomes:

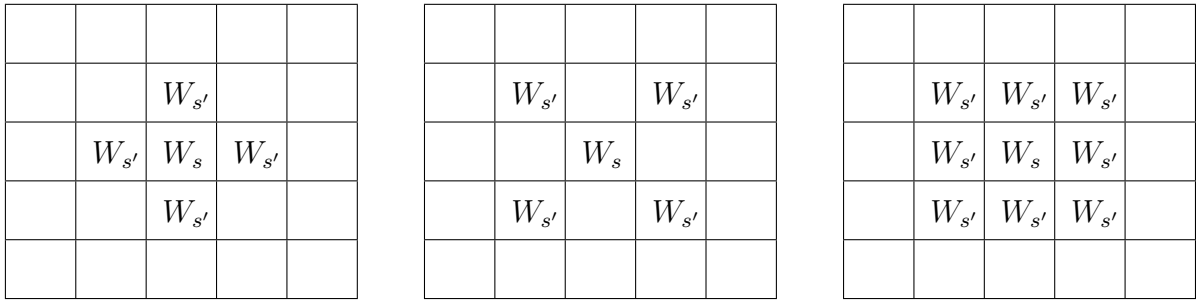
$$y_s = X^T(s)\beta + \phi(s) + \epsilon(s)$$

This model is known as areal-level spatial model.

1.2.2 Construction of Neighborhood Matrix

In spatial data it is expected that measurements for areal units which are close to each other will tend to be more similar than those which are far way from each other. Therefore, we need to construct a neighborhood matrix, w, to define the relationship among areal units. Some common approaches to consider areal units as neighborhood are:

- Sharing boundary: Two areal units, say W_s and $W_{s'}$ are said to be neighborhood if they share common edges (left), vertex (middle) or both (right).



- Centroid distances: In this case either the number of neighbors, k, or the distance, d, beyond which there is no direct spatial influence between spatial units is fixed. For given k, all k closest units, based on centroid distance, of W_s are called neighbors of it. On the other hand, if distance, d, is an important criterion of spatial influence, a cell, $W_{s'}$, is called neighbor of W_s if distance between them is less than d.

Usually the entries of neighborhood matrix, w, are binary: $w(s, s') = 1$ if the cells W_s and $W_{s'}$ are neighbors and 0 otherwise.

1.2.3 A Measure of Areal Dependence

Before incorporating spatial random effects in the model, it is necessary to check whether there exists any spatial association among neighborhood areal units or not. Moran's I is a popular statistic that measure the strength of spatial association among areal units (Banerjee et al., 2014).

Moran's I is defined as

$$I = \frac{n \sum_s \sum_{s'} w(s, s')(y_s - \bar{y})(y_{s'} - \bar{y})}{(\sum_{s \neq s'}) \sum_s (y_s - \bar{y})},$$

where $I \in (-1, 1)$ but not strictly. Significantly different from zero values of I represents

strong spatial association and if I is close to zero, then there is no spatial association in the data.

1.2.4 Conditional Auto-Regressive Prior

A conditional autoregressive (CAR) prior, introduced by Besag (1974), is a common choice for spatial random effects due to the convenient computation in Gibbs sampling. The derivation of the Gaussian prior is discussed below:
We begin with conditionals distributions.

$$\phi_s|\{\phi_{s'} : s \neq s'\} \sim \mathcal{N}\Big(\sum_{s'} b_{ss'}\phi_{s'}, \tau_s^2\Big), \;\; i = 1, 2, \ldots, S \tag{1.2}$$

Now using Brook's lemma (Brook, 1964) we get the joint distribution as follows

$$\pi(\phi_1, \phi_2, \ldots, \phi_S) \propto \exp\Big\{ -\frac{1}{2}\phi^T D^{-1}(I - B)\phi\Big\} \tag{1.3}$$

where $B = \{b_{ss'} : s, s' \in D\}$, D is a diagonal matrix with $D_{ss} = \tau_s^2$ and $\phi = (\phi_1, \phi_2, \ldots, \phi_S)^T$. Hence, ϕ follows multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\Sigma_\phi = (I-B)^{-1}D$. Now Σ_ϕ^{-1} will be symmetric if $\frac{b_{ss'}}{\tau_s^2} = \frac{b_{s's}}{\tau_{s'}^2}$ for all s, s'. That means B is not required to be symmetric. Reset $b_{ss'} = w(s, s')/w_{s+}$ and $\tau_s^2 = \tau^2/w_{s+}$ where neighborhood matrix, w, is symmetric. Hence, (1.2) becomes $\pi(\phi_s|\{\phi_{s'}, s \neq s'\}) \sim \mathcal{N}\Big(\sum_{s'} w(s, s')\phi_{s'}/w_{s+}, \tau^2/w_{s+}\Big)$. And the joint distribution becomes:

$$\pi(\phi_1, \phi_2, \ldots, \phi_S) \propto \exp\Big\{ -\frac{1}{2\tau^2}\sum_{s\neq s'} w(s, s')(\phi_s - \phi_{s'})^2\Big\}$$

This density, being invariant to translation, is improper.

1.3 Temporal Modeling

When a sequence of data points are measured over different time periods, usually evenly spaced times, it is called time series. In autoregressive model, the variable of interest is modeled on the past values of the same variable and the order is the total number of old time points that are used to do the prediction. Let ψ_t be the value of outcome variable at time t. Hence, a p-th order autoregressive model can be written as

$$\psi_t = \beta_0 + \beta_1\psi_{t-1} + \beta_2\psi_{t-2} + \cdots + \beta_p\psi_{t-p} + \delta_t$$

where δ_t is white noise. When $p = 1$, it is called first-order autoregressive model and the form becomes

$$\psi_t = \beta_0 + \beta_1\psi_{t-1} + \delta_t$$

1.4 Spatio-Temporal Modeling for Areal Data

Spatio-temporal data are data sets that are observed both spatially and temporally. Hence, with an extension in the spatial model discussed in Chapter 1.2.1 a spatio-temporal model can be formulated. Let $\{W_1, W_2, \ldots, W_S\}$ be S areal units and y_{st} be the summary response at unit W_s at time t. The spatio-temporal model can be written as

$$y_{st} = \mu_{st} + e_{st} \tag{1.4}$$

where μ_{st} is the mean structure and e_{st} denotes the residual at unit s at time t. Now e_{st} can be rewritten as $e_{st} = w_{st} + \epsilon_{st}$ where w_{st} is the spatio-temporal random effect and ϵ_{st} is the white noise. Finally, the spatio-temporal model becomes

$$y_{st} = \mu_{st} + w_{st} + \epsilon_{st}$$

Chapter 2

Description of Remote Sensing Imagery

An active area of forestry research involves the use of auxiliary data that can be collected quickly and inexpensively, such as from satellite imagery. These data are then used to model the relationship between these auxiliary data and the field plot data in an effort to improve the precision of population estimates, particularly for smaller domains within the larger population. Since satellite-based sensors, such as Landsat 7's ETM+ instrument, detect reflectance from the Earth's surface, these data are expected to be closely correlated with land cover. Kauth and Thomas (1976) developed a linear transformation of the original Landsat Multispectral Scanner bands, named the tasseled cap (TC) transformation. Comparable transformations have since been developed for the Landsat Thematic Mapper, Enhanced Thematic Mapper Plus, and Operational Land Imager sensors. The TC features are related to growing vegetation, soil moisture, and overall surface brightness, and are correlated with the phases of vegetation development over time. With the advent of the data policy of 2008, granting unrestricted access to the entire USGS archive of Landsat, dense time series of TC features can now be used to model forest dynamics. Numerous studies have shown TC observations derived from Landsat imagery to be useful for mapping land cover (Yuan et al., 2005), as well as several forest characteristics such as growing stock volume (Zheng et al., 2014) and biomass (Karlson et al., 2015).

We are given the Landsat 7's ETM+ imagery, collected between the ten year period 2003 - 2012, from a $230,400$ hectares square-shaped region in north central Wisconsin. However, this stack of monthly TC imagery has many missing values both in space and time, due to failure of scan line correctors or presence of cloud cover. Hence, our goal is to fill in the missing parts of these images using a spatiotemporal regression. Sellers et al. (1994, 1996)

were among the first to use a mathematical model to describe time series of satellite imagery in an effort to correct for missing data, in their case primarily due to snow or residual cloud contamination. Their approach starts with adjustment of maximum value composite images by means of harmonic regression via Fourier series. Utility of harmonic regression for feature extraction from dense time series of Landsat data was also demonstrated in Wilson et al. (2018). In current work, we explore whether augmenting spatial and/or temporal random effects to these fixed-effect harmonic functions results in significant improvement in model performance and prediction. Multiple candidate specifications are explored and compared using diagnostic tools to select the best performing model and TC maps are reconstructed for subsequent use.

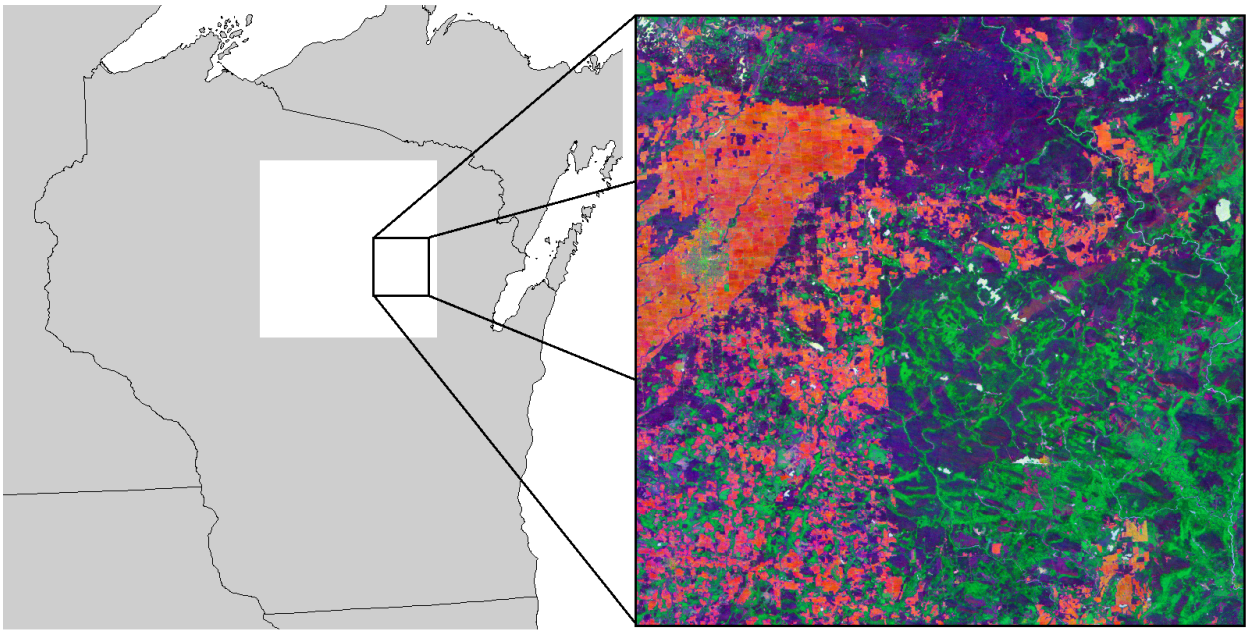


Figure 2.1: (left) In the map of Wisconsin, the study area (the smaller square) is shown in the southeastern part of the WELD tile H20V05 (the larger square). (right) Spatial variability of the land cover within the study area.

Our region of interest lies in the southeastern part of Web-enabled Landsat Data (WELD) tile H20V05 (Roy et al., 2010), located at the intersection of Langlade, Shawano, and Menominee counties in north-central Wisconsin, USA. Figure 2.1 shows its position within the tile and

the state in the left panel. The right panel shows a variety of land covers and uses, such as agriculture fields (orange pixels), deciduous (purple) and evergreen (green) forests, developed land (gray), as well as scattered water bodies (white) and wetlands (pink). This area falls within the Laurentian Mixed Forest Province of the United States Forest Service National Hierarchical Framework of Ecological Units (Cleland et al., 2007; McNab et al., 2007) with the the city of Wausau in northwest. The province experiences a continental climate, with some maritime influence near the Great Lakes. This leads to moderately long winters and warm summers, when most of the precipitation occurs. This landscape was shaped by past glaciation with a mix of agricultural fields, boreal and broadleaf deciduous forests. The eastern half of the study area is predominantly forested. On June 7, 2007, a tornado, with estimated winds of 225 to 255 kilometers per hour, touched down and followed a 65 kilometers long and 3/4 kilometer-wide northeasterly path through parts of Menominee, Langlade, and Oconto counties.

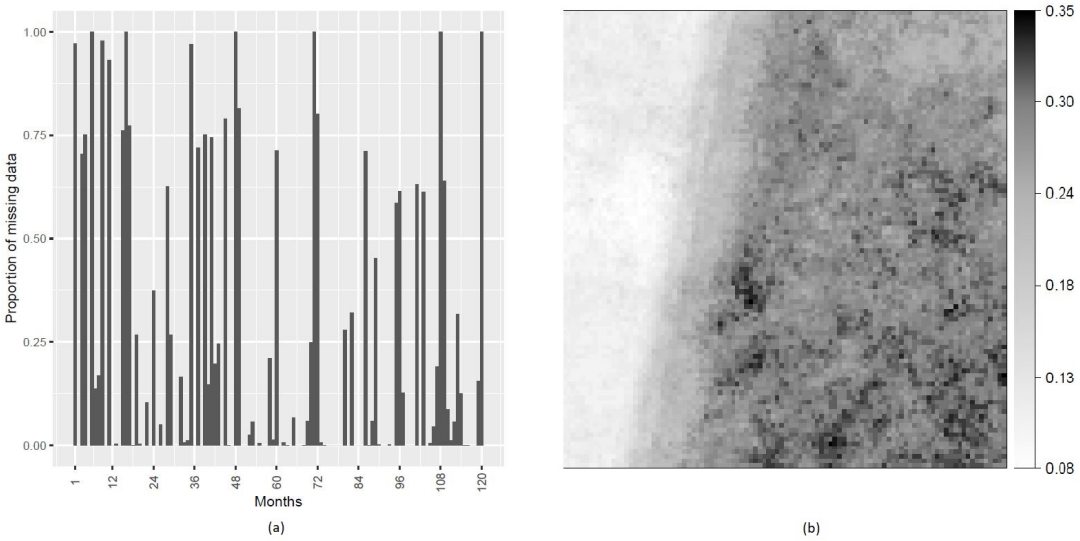


Figure 2.2: Proportion of missing TC observations (left) temporally, at each month and (b) spatially, at each cell, during 2003 - 2012

The auxiliary data used in the study were dense Landsat time series images from the WELD project. WELD imagery are composites of high fidelity data, determined on a pixel-by-

pixel basis, from all Landsat 7 ETM+ scenes collected over a compositing period. These composite images have been processed for the contiguous United States and Alaska over the decade of 2003 - 2012. The composite scenes have been ortho-rectified, transformed to top-of atmosphere reflectance and mosaicked into 5000 $\times$ 5000 pixel tiles at the native 30-meter pixel resolution using the Albers Equal Area projection with origin at 23°N and 96°W. Our study area, a part of tile H20V05, consists of 1600 $\times$ 1600 pixels. The WELD monthly composites for the entire decade of 2003 - 2012 from this area were used for the study. For each monthly composite, the reflectance values from ETM+ were transformed to the first three TC components: 1) brightness (TC1), 2) greenness (TC2), and 3) wetness (TC3) (Huang et al., 2002). The monthly TC features were then compiled into individual stacks. To control the dimension, we aggregated adjacent 16×16 pixels into a single gridcell of area 23.04 hectares by taking average on the pixels with available data. At this scale, our study area consists of $10,000$ gridcells, an area $230,400$ hectares in size, Easting between $532,330$ and $579,750$ meters, and Northing between $2,452,240$ and $2,499,760$ meters. However, each TC component has a high frequency of missing values ($\sim 23\%$) with 100% missing data for six months (April in 2004, June in 2003, November in 2011, and December in 2006, 2011 and 2012), due either to completely missing records in the WELD archive or to pixels that were masked out because of the presence of clouds, snow, or artifacts of sensor failure. Figure 2.2 (a) shows the temporal patterns of missingness and Figure 2.2 (b) shows the spatial pattern of missingness with a strong linear boundary, with larger values in the east and smaller values in the west, due to the flight track of the Landsat satellites. Areas to the west of the image are in the zone of overlap between neighboring Landsat scenes, while areas to the east have no overlap. Therefore, there are fewer pixel observations in the eastern portion of the image, meaning that it is more likely that data will be missing due to clouds, snow, or sensor error. Hence, in order to make the imagery complete, we perform a model-based reconstruction of missing TC observations in Chapter 3.

Chapter 3

Model for Reconstruction of Remote Sensing Measurement

Since the TC features, as outlined above, have large gaps across space and time, our aim is to make these imagery compete. Hence, we develop a hierarchical model for reconstructing the complete spatiotemporal map for each of these features.

3.1 Hierarchical model development

We begin by defining notation. The spatial domain of study, denoted by $\mathcal{W}$, is partitioned into disjoint areal units $W_1, W_2, ..., W_S$ such that the feature imagery are available at the resolution of these areal units. Let $\mathcal{T}$ be the temporal domain and $y_s(\tau)$ represents the value of a feature variable at any gridcell W_s at time $\tau \in \mathcal{T}$. Dataset for each of these features consists of composite values over disjoint time intervals within $\mathcal{T}$, so we partition $\mathcal{T} = \bigcup_{t=1}^{T} D_t$. For the current application, data were reported monthly, so $D_1, D_2, ..., D_T$ represent consecutive months spanning $\mathcal{T}$. Define $y_{st} = \int_{\tau \in D_t} y_s(\tau) d\tau$ as the summary feature over W_s during D_t. We assign a model for the X_{st} process below:

$$y_{st} = \mu_{st} + \epsilon_{st},\ s = 1, 2, ..., S,\ t = 1, 2, ..., T \tag{3.1}$$

where μ_{st} denotes the mean structure and $\epsilon_{st} \sim N(0, \sigma^2)$ represents the pure error uncorrelated in space and time. We propose several options for specifying μ_{st} below with gradual increase in the model complexity through increase in the number of parameters:

- **Model I :** We start with the customary fixed-effect linear regression model with a vector of covariates (including intercept) X_{st} and regression coefficients β. This model does not include any random effect. In practice, Xst can include any relevant co-

variate, spatial and/or temporal. Wilson et al. (2018) employs a harmonic regression via Fourier series that can be used to approximate a periodic function, such as the seasonality of a vegetation index. For example, the first harmonic captures the expected annual variation in vegetation response. The second harmonic accommodates some cropping patterns seen in the area, with spring and winter crops grown on the same fields in some cases. A closer approximation is given by using more harmonics in the series at the cost of potential overfitting. In general, if first q harmonics are used as covariates, X_{st} becomes a function of t only and can be written as $X_t = [1, \cos h_1 t, \sin h_1 t, \ldots, \cos h_q t, \sin h_q t]^T$ where $h_i = 2\pi i/12$ and the period of the i-th harmonic is $12/i$. It follows that

$$\mu_{st} = X_t^T \beta$$

- **Model II:** To capture the temporal pattern in residuals, we augment a time-dependent random effect in Model I as

$$\mu_{st} = X_t^T \beta + \psi_t.$$

Since ψ_t is free of s, it can only measure temporal variation in the spatially aggregated data. A first-order autoregressive (AR(1)) prior is assigned on ψ_t as:

$$\psi_t = \eta \psi_{t-1} + \delta_t,$$

where $\delta_t \overset{iid}{\sim} \mathcal{N}(0, \kappa^2)$ and $\psi(0) \sim \mathcal{N}(0, \kappa^2/(1-\eta^2))$ with $|\eta| < 1$. The joint prior distribution of $\Psi = (\psi(1), \psi(2), ..., \psi(T))^T$ can be written as $\mathcal{N}_T(\mathbf{0}, \Sigma_\psi)$ where $\Sigma_\psi[i,j] = \kappa^2 \eta^{|i-j|}/(1-\eta^2)$.

- **Model III:** Incorporating a spatiotemporal random effect inside Model I, we get

$$\mu_{st} = X_t^T \beta + \phi_{st} \tag{3.2}$$

We explore two different choices for specifying the distribution of ϕ_{st}, also suggested in Gelfand et al. (2004).

- **Model III-A:** Consider the additive form of temporal and spatial effects which estimate spatially averaged temporal effect and temporally averaged spatial effect respectively for ϕ_{st} in (3.2). Its form resembles a two-way ANOVA model without interaction term:

$$\phi_{st} = \psi_t + \phi_s$$

where ϕ_s is the spatial random effect at gridcell W_s. A conditional autoregressive (CAR) prior (Besag and Kooperberg, 1995) is assigned to ϕ. This requires us to define a neighborhood matrix w such that, for $1 \leq s, s' \leq S$, $w(s, s') = 1$ if the pair of gridcells W_s and $W_{s'}$ are neighbors and 0 otherwise. The prior joint distribution for $\{\phi_s\}$ is specified by the series of univariate conditional distributions:

$$\phi_s | \{\phi_{s'} : s' \neq s\} \sim \mathcal{N}\left(\frac{\sum_{s'} w(s, s')\phi_{s'}}{w_{s+}}, \frac{\tau_\phi^2}{w_{s+}} \right),$$

where τ_ϕ is the scale parameter of the conditional distribution and $w_{s+} = \sum_{s'} w(s, s')$ denotes the number of neighbors of gridcell W_s.

- **Model III-B:** Model ϕ_{st} in (3.2) as a temporally independent sequence of spatial processes. Define $\phi_{s*} = (\phi_{s1}, \phi_{s2}, ..., \phi_{sT})^T$ for $1 \leq s \leq S$. Given neighborhood matrix w as in Model III-A, the prior conditional distribution for ϕ_{s*} is defined by the series of multivariate conditional distributions:

$$\phi_{s*}|\{\phi_{s'*} : s' \neq s\} \sim \mathcal{N}_T(M_s^\phi, \Sigma_s^\phi),$$

where M_s^ϕ is a vector of length T with its t-th entry $= \sum_{s'} w(s,s')\phi_{s't}/w_{s+}$ and Σ_s^ϕ is a $T \times T$ diagonal matrix with its t-th diagonal $= \tau_{\phi,t}^2/w_{s+}$.

- **Model IV:** Model III-B lacks any parameter to capture potential temporal association, so we augment it with a temporal random effect as,

$$\mu_{st}^{\mathrm{x}} = X_t^T\beta + \phi_{st} + \psi_t,$$

with ϕ_{st} and ψ_t defined as in Model III-B and Model II respectively.

We assign the usual Gaussian prior $\mathcal{N}_{2q+1}(\alpha_0, \nu_\beta^2 I_{2q+1})$ for β and Inverse-Gamma prior $\mathrm{IG}(a_0, b_0)$ for all variance parameters, i.e., σ^2, τ_ϕ^2, $\tau_{\phi,t}^2$, κ^2. A truncated Gaussian prior $\mathcal{N}(0, \nu_\eta^2)I_{(-1,1)}$ is assigned to η. The hyperparameters that appear inside these prior distributions are treated as constants with their values chosen in order to diffuse the prior while maintaining propriety.

3.2 Posterior estimation and assessment

We employ an MCMC scheme to draw posterior samples for model parameters and use them to simulate from the posterior predictive distributions of missing TC features. Since missing observations lead to unbalanced structure of the available measurements in space and time, derivation of posterior distributions of ϕ_{st} and ψ_t, for the models where one or both of them appear, requires attention and is described below.

To start with update for ϕ_{st}, we introduce some notations. For $1 \leq s \leq S$, let $\mathcal{A}_{s*} = \{t : y_{st} \text{ is available}\}$ and $n_{s+} = |\mathcal{A}_{s*}|$. Define two vectors, each of length n_{s+}, as $\widetilde{y}_s = \{y_{st} : t \in \mathcal{A}_{s*}\}$, $\widetilde{\psi}_s = \{\psi(t) : t \in \mathcal{A}_{s*}\}$ and an $n_{s+} \times (2q+1)$ matrix $\widetilde{X}_s$ with rows from

$\{X_t^T : t \in \mathcal{A}_{s*}\}$. The likelihood of the data involving ϕ_{s*} follows from $\widetilde{y}_s^{\phi} \sim \mathcal{N}_{n_{s+}}(\tilde{I}_s^{\phi}\phi_{s*}, D_s^{\phi})$ where $\widetilde{y}_s^{\phi} = (\widetilde{y}_s - \widetilde{X}_s\beta)$ and $(\widetilde{y}_s - \widetilde{X}_s\beta - \widetilde{\psi}_s)$ for Model III-B and Model IV respectively, $\tilde{I}_s^{\phi}$ is a $n_{s+} \times T$ submatrix of I_T obtained by retaining only the rows with indices in $\mathcal{A}_{s*}$ and $D_s^{\phi} = \sigma^2 I_{n_{s+}}$. Therefore, the posterior distribution of ϕ_{s*} becomes $\mathcal{N}_T(V_s^{\phi}\mu_s^{\phi}, V_s^{\phi})$ where $\mu_s^{\phi} = (\tilde{I}_s^{\phi}(s))^T(D_s^{\phi})^{-1}\widetilde{y}_s^{\phi} + (\Sigma_s^{\phi})^{-1}M_s^{\phi}$ and $(V_s^{\phi})^{-1} = (\tilde{I}_s^{\phi})^T(D_s^{\phi})^{-1}\tilde{I}_s^{\phi} + (\Sigma_s^{\phi})^{-1}$.

Turning our attention to updating Ψ, we define $\mathcal{E} = \{t : \exists s \text{ such that } y_{st} \text{ is available}\}$ and $T_{\mathcal{E}} = |\mathcal{E}|$. For $t \in \mathcal{E}$, define $\mathcal{A}_{*t} = \{s : y_{st} \text{ is available}\}$ and $n_{+t} = |\mathcal{A}_{*t}|$. Now, the likelihood of the data involving Ψ follows from $\widetilde{y}^{\psi} \sim \mathcal{N}_{T_{\mathcal{E}}}(\tilde{I}_{\psi}\Psi, D_{\psi})$ where $\widetilde{y}^{\psi}$ is a vector of length $T_{\mathcal{E}}$ with its t-th entry being $\sum_{s \in \mathcal{A}_{*t}} (y_{st} - {X_t}^T\beta)/n_{+t}$ for Model II, $\sum_{s \in \mathcal{A}_{*t}} (y_{st} - {X_t}^T\beta - \phi_s)/n_{+t}$ for Model III-A and $\sum_{s \in \mathcal{A}_{*t}} (y_{st} - {X_t}^T\beta - \phi_{st})/n_{+t}$ for Model IV, $\tilde{I}_{\psi}$ is a $T_{\mathcal{E}} \times T$ submatrix of I_T constructed by retaining the rows with indices in $\mathcal{E}$ and D_{ψ} is a $T_{\mathcal{E}} \times T_{\mathcal{E}}$ diagonal matrix with t-th diagonal entry σ^2/n_{+t}. Hence, the posterior distribution of Ψ becomes $\mathcal{N}_T(V_{\psi}\mu_{\psi}, V_{\psi})$ where $\mu_{\psi} = \tilde{I}_{\psi}^T D_{\psi}^{-1}\widetilde{y}^{\psi}$ and $V_{\psi}^{-1} = \tilde{I}_{\psi}^T D_{\psi}^{-1}\tilde{I}_{\psi} + \Sigma_{\psi}^{-1}$.

All other parameters except η, have posterior distributions in standard forms since we use conjugate priors. The posterior distribution of η is non-standard and is drawn via slice sampling (Neal, 2003). All posterior distributions, used in MCMC, are summarized below. It is noted that in this application there is no cell with 100% missing time points. However, data with all missing values at any grid cell requires modification in the posterior distributions accordingly.

Model I:

$$\beta|- \sim \mathcal{N}_{2q+1}(\Sigma_{\beta}\,\mu_{\beta}, \Sigma_{\beta}); \sigma^2|- \sim \text{IG}(a_1, b_1)$$

where

$$\mu_{\beta} = \sum_{s=1}^{S} \widetilde{X}_s^T\widetilde{y}_s/\sigma^2 + \alpha_0/\nu_{\beta}^2 \text{ and } \Sigma_{\beta}^{-1} = \sum_{s=1}^{S} \widetilde{X}_s^T\widetilde{X}_s/\sigma^2 + I_{2q+1}/\nu_{\beta}^2$$

$$a_1 = a_0 + \sum_{s=1}^{S} n_{s+}/2 \text{ and } b_1 = b_0 + \sum_{s=1}^{S} (\widetilde{y}_s - \widetilde{X}_s\beta)^T(\widetilde{y}_s - \widetilde{X}_s\beta)/2$$

Model II:

$$\beta|- \sim \mathcal{N}_{2q+1}(\Sigma_\beta\, \mu_\beta, \Sigma_\beta), \sigma^2|- \sim \text{IG}(a_1, b_1),$$
$$\psi_0|- \sim \mathcal{N}(\eta\psi_1, \kappa^2), \Psi|- \sim \mathcal{N}_T(V_\psi\, \mu_\psi, V_\psi), \kappa^2|- \sim \text{IG}(a_2, b_2),$$
$$u|- \sim \text{ Unif}(0, \sqrt{1-\eta^2}\,), \eta|- \propto \mathcal{N}(\mathrm{v}_\eta^2\mu_\eta, \mathrm{v}_\eta^2) I_{\{-\sqrt{1-u^2}, \sqrt{1-u^2}\,\}}$$

where

$$\mu_\beta = \sum_{s=1}^{S} \widetilde{X}_s^T(\widetilde{y}_s - \widetilde{\psi}_s)/\sigma^2 + \alpha_0/\nu_\beta^2 \text{ and } \Sigma_\beta^{-1} = \sum_{s=1}^{S} \widetilde{X}_s^T\widetilde{X}_s/\sigma^2 + I_{2q+1}/\nu_\beta^2$$
$$a_1 = a_0 + \sum_{s=1}^{S} n_{s+}/2 \text{ and } b_1 = b_0 + \sum_{s=1}^{S} (\widetilde{y}_s - \widetilde{X}_s\beta - \widetilde{\psi}_s)^T(\widetilde{y}_s - \widetilde{U}_s\beta - \widetilde{\psi}_s)/2$$
$$\mu_\psi = \tilde{I}_\psi^T D_\psi^{-1}\widetilde{y}^\psi \text{ and } V_\psi^{-1} = \tilde{I}_\psi^T D_\psi^{-1}\tilde{I}_\psi + \Sigma_\psi^{-1}$$
$$a_2 = a_0 + (T+1)/2 \text{ and } b_2 = b_0 + \sum_{t=1}^{T} (\psi_t - \eta\psi_{t-1})^2/2 + (1-\eta^2)\psi_0^2/2$$
$$\mu_\eta = \sum_{t=1}^{T} \psi_t\psi_{t-1}/\kappa^2 \text{ and } \mathrm{v}_\eta^{-2} = \sum_{t=2}^{T} \psi_{t-1}^2/\kappa^2 + 1/\nu_\eta^2$$

Model III-A:

$$\beta|- \sim \mathcal{N}_{2q+1}(\Sigma_\beta\, \mu_\beta, \Sigma_\beta),\ \sigma^2|- \sim \text{IG}(a_1, b_1),$$
$$\psi_0|- \sim \mathcal{N}(\eta\psi_1, \kappa^2), \Psi|- \sim \mathcal{N}_T(V_\psi\, \mu_\psi, V_\psi), \kappa^2|- \sim \text{IG}(a_2, b_2),$$
$$u|- \sim \text{ Unif}(0, \sqrt{1-\eta^2}\,), \eta|- \propto \mathcal{N}(\mathrm{v}_\eta^2\mu_\eta, \mathrm{v}_\eta^2) I_{\{-\sqrt{1-u^2}, \sqrt{1-u^2}\,\}},\ \tau_\phi^2|- \sim \text{IG}(a_3, b_3)$$

$$\phi(s)|- = \begin{cases} \mathcal{N}(\widetilde{\mathrm{v}}_\phi^2\widetilde{\mu}_\phi, \widetilde{\mathrm{v}}_\phi^2), & \text{if } n_{s+} > 0 \\ \mathcal{N}(\mathrm{v}_\phi^2\mu_\phi, \mathrm{v}_\phi^2), & \text{if } n_{s+} = 0 \end{cases}$$

where

$$\mu_\beta = \sum_{s=1}^{S} \widetilde{X}_s^T(\widetilde{y}_s - \widetilde{\psi}_s - \phi_s)/\sigma^2 + \alpha_0/\nu_\beta^2 \text{ and } \Sigma_\beta^{-1} = \sum_{s=1}^{S} \widetilde{X}_s^T \widetilde{X}_s/\sigma^2 + I_{2q+1}/\nu_\beta^2$$

$$a_1 = a_0 + \sum_{s=1}^{S} n_{s+}/2 \text{ and } b_1 = b_0 + \sum_{s=1}^{S} (\widetilde{y}_s - \widetilde{X}_s\beta - \widetilde{\psi}_s - \phi_s)^T(\widetilde{y}_s - \widetilde{X}_s\beta - \widetilde{\psi}_s - \phi_s)/2$$

$$\mu_\psi = \tilde{I}_\psi^T D_\psi^{-1} \widetilde{y}^\psi \text{ and } V_\psi^{-1} = \tilde{I}_\psi^T D_\psi^{-1} \tilde{I}_\psi + \Sigma_\psi^{-1}$$

$$a_2 = a_0 + (T+1)/2 \text{ and } b_2 = b_0 + \sum_{t=1}^{T} (\psi_t - \eta\psi_{t-1})^2/2 + (1-\eta^2)\psi_0^2/2$$

$$\mu_\eta = \sum_{t=1}^{T} \psi_t\psi_{t-1}/\kappa^2 \text{ and } \mathrm{v}_\eta^{-2} = \sum_{t=2}^{T} \psi_{t-1}^2/\kappa^2 + 1/\nu_\eta^2$$

$$a_3 = a_0 + (S-1)/2 \text{ and } b_3 = b_0 + \sum_{s=1}^{S} \sum_{s' \neq s} w(s,s')(\phi(s) - \phi(s'))^2/4$$

$$\widetilde{\mu}_\phi = \sum_{t \in A_{s*}} (\widetilde{y}_s - \widetilde{X}_s\beta - \widetilde{\psi}_s)/\sigma^2 + \sum_{s'} w(s,s')\phi(s')/\tau_\phi^2 \text{ and } \widetilde{\mathrm{v}}_\phi^{-2} = n_{s+}/\sigma^2 + w_{s+}/\tau_\phi^2$$

$$\mu_\phi = \sum_{s'} w(s,s')\phi(s')/w_{s+} \text{ and } \mathrm{v}_\phi^{-2} = \tau_\phi^2/w_{s+}$$

Model III-B: We introduce an additional notation $\widetilde{\phi}_s = \{\phi_{st} : t \in \mathcal{A}_{s*}\}$.

$$\beta|- \sim \mathcal{N}_{2q+1}(\Sigma_\beta\, \mu_\beta, \Sigma_\beta),\ \sigma^2|- \sim \mathrm{IG}(a_1, b_1),\ \tau_\phi^2|- \sim \mathrm{IG}(a_2, b_2), \phi_{s*}|- \sim \mathcal{N}_T(V_s^\phi\, \mu_s^\phi, V_s^\phi)$$

where

$$\mu_\beta = \sum_{s=1}^{S} \widetilde{X}_s^T(\widetilde{y}_s - \widetilde{\phi}_s)/\sigma^2 + \alpha_0/\nu_\beta^2 \text{ and } \Sigma_\beta^{-1} = \sum_{s=1}^{S} \widetilde{X}_s^T \widetilde{X}_s/\sigma^2 + I_{2q+1}/\nu_\beta^2$$

$$a_1 = a_0 + \sum_{s=1}^{S} n_{s+}/2 \text{ and } b_1 = b_0 + \sum_{s=1}^{S} (\widetilde{y}_s - \widetilde{X}_s\beta - \widetilde{\phi}_s)^T(\widetilde{y}_s - \widetilde{X}_s\beta - \widetilde{\phi}_s)/2$$

$$a_2 = a_0 + (S-1)T/2 \text{ and } b_2 = b_0 + \sum_{t=1}^{T} \sum_{s=1}^{S} \sum_{s \neq s'} w(s,s')(\phi_{st} - \phi_{s't})^2/4$$

$$\mu_s^\phi = (\tilde{I}_s^\phi(s))^T (D_s^\phi)^{-1} \widetilde{y}_s^\phi + (\Sigma_s^\phi)^{-1} M_s^\phi \text{ and } (V_s^\phi)^{-1} = (\tilde{I}_s^\phi)^T (D_s^\phi)^{-1} \tilde{I}_s^\phi + (\Sigma_s^\phi)^{-1}$$

Model IV:

$$\beta|- \sim \mathcal{N}_{2q+1}(\Sigma_\beta\, \mu_\beta, \Sigma_\beta),\ \sigma^2|- \sim \text{IG}(a_1, b_1),$$

$$\tau_\phi^2|- \sim \text{IG}(a_2, b_2), \phi_{s*}|- \sim \mathcal{N}_T(V_s^\phi\, \mu_s^\phi, V_s^\phi),$$

$$\psi_0|- \sim \mathcal{N}(\eta\psi_1, \kappa^2), \Psi|- \sim \mathcal{N}_T(V_\psi\, \mu_\psi, V_\psi), \kappa^2|- \sim \text{IG}(a_3, b_3),$$

$$u|- \sim\ \text{Unif}(0, \sqrt{1-\eta^2}\,), \eta|- \propto \mathcal{N}(\text{v}_\eta^2\mu_\eta, \text{v}_\eta^2) I_{\{-\sqrt{1-u^2}, \sqrt{1-u^2}\,\}}$$

where

$$\mu_\beta = \sum_{s=1}^{S} \widetilde{X}_s^T(\widetilde{y}_s - \widetilde{\phi}_s - \widetilde{\psi}_s)/\sigma^2 + \alpha_0/\nu_\beta^2 \text{ and } \Sigma_\beta^{-1} = \sum_{s=1}^{S} \widetilde{X}_s^T \widetilde{X}_s/\sigma^2 + I_{2q+1}/\nu_\beta^2$$

$$a_1 = a_0 + \sum_{s=1}^{S} n_{s+}/2 \text{ and } b_1 = b_0 + \sum_{s=1}^{S} (\widetilde{y}_s - \widetilde{X}_s\beta - \widetilde{\phi}_s - \widetilde{\psi}_s)^T(\widetilde{y}_s - \widetilde{X}_s\beta - \widetilde{\phi}_s - \widetilde{\psi}_s)/2$$

$$a_2 = a_0 + (S-1)T/2 \text{ and } b_2 = b_0 + \sum_{t=1}^{T}\sum_{s=1}^{S}\sum_{s\neq s'} w(s,s')(\phi_t(s) - \phi_t(s'))^2/4$$

$$\mu_s^\phi = (\tilde{I}_s^\phi(s))^T(D_s^\phi)^{-1}\widetilde{y}_s^\phi + (\Sigma_s^\phi)^{-1} M_s^\phi \text{ and } (V_s^\phi)^{-1} = (\tilde{I}_s^\phi)^T(D_s^\phi)^{-1}\tilde{I}_s^\phi + (\Sigma_s^\phi)^{-1}$$

$$\mu_\psi = \tilde{I}_\psi^T D_\psi^{-1}\widetilde{y}^\psi \text{ and } V_\psi^{-1} = \tilde{I}_\psi^T D_\psi^{-1}\tilde{I}_\psi + \Sigma_\psi^{-1}$$

$$a_3 = a_0 + (T+1)/2 \text{ and } b_3 = b_0 + \sum_{t=1}^{T}(\psi_t - \eta\psi_{t-1})^2/2 + (1-\eta^2)\psi_0^2/2$$

$$\mu_\eta = \sum_{t=1}^{T} \psi_t\psi_{t-1}/\kappa^2 \text{ and } v_\eta^{-2} = \sum_{t=2}^{T} \psi_{t-1}^2/\kappa^2 + 1/\nu_\eta^2$$

Predictive performance of five candidate models for each TC is examined using three different criteria. First, we calculate log likelihood and Bayesian predictive information criterion (BPIC, Ando, 2007; Li et al., 2017). Model with the lowest BPIC is considered to be the best. Additionally, cross-validation is performed using holdout method by repeatedly splitting the data into training sets and test sets, and evaluating the predictive accuracy of the model on the latter. We use three measures for assessment: (i) absolute bias: computed as the

difference between the hidden test observations and their corresponding posterior medians, (ii) uncertainty: measured as the width of 90% highest posterior density (HPD) interval and (iii) empirical coverage: calculated as the proportion of test observations that are within the corresponding 90% HPD interval. The model that attains the desired coverage level with minimum bias and uncertainty is deemed to be the best. Finally, to determine whether a candidate model adequately captures the spatial dependence, we assess its residuals using Moran's I statistic (Banerjee et al., 2014). A near zero or negative value of this statistic suggests lack of empirical evidence of spatial similarity between the residuals.

Chapter 4

Data Analysis

We begin by implementing candidate models from Chapter 3.1 on the Landsat imagery and assessing their performance using the tools described in Chapter 3.2. Then reconstruct the imagery from the best candidate model.

In our Gaussian linear model of (3.1), y represents TC features in logarithmic scale. In the original dataset, TC1 has strictly positively values whereas TC2 and TC3 can be both positive and negative. Prior to applying the logarithmic transformation, we first translate and rescale the measurements of TC2 and TC3 to the same range as that of TC1. We choose first $q = 2$ harmonics as fixed-effect covariates following the findings of Wilson et al. (2018) that 3rd or higher order harmonics fit extraneous noise rather than meaningful signal in TC components. Now, we estimate the five candidate models discussed in Chapter 3.1, separately for each TC feature, using MCMC algorithms from Appendix A.1. In both of Model III-B and IV, we set the sequence of hyperparameters $\tau_\phi(t) = \tau_\phi$, free of t. This is reasonable as each TC component contains significant percentage of missing data at many of the 120 months (including 6 months with no data) resulting in insufficient information to separately identify these hyperparameters for those months. For constructing the binary adjacency matrix in the CAR prior, we define any two cells as neighbors only if they share boundaries, resulting in exactly 4 neighbors for any interior cell. The chains are run for $60,000$ iterations with a burn-in length of $10,000$ and a thinning interval of 50. Now, we compare the model performance using the measures described in Chapter 3.2 starting with likelihood and BPIC. Then, using holdout cross-validation for each TC, we randomly remove $25,000$ observations into a test set, fit the candidate models on the remaining data points and post-MCMC, simulate these test observations from their respective posterior predictive distributions. We replicate this

holdout method 36 times with mutually exclusive test sets to approximately cover the entire set of observations and compute the three performance measures from Chapter 3.2, averaged across replications. Table 4.1 presents a detailed comparative analysis along with model-specific computation times, based on a single-processor implementation.

Table 4.1: Comparison of five candidate models for each TC feature

Measure	Feature	Model I	Model II	Model III-A	Model III-B	Model IV
Likelihood-based Model Comparison						
LL† ($\times 10^5$)	TC1	6.906	8.007	10.193	11.331	29.207
	TC2	13.700	14.561	15.869	18.865	31.928
	TC3	16.219	18.244	18.452	18.998	33.911
BPIC ($\times 10^5$)	TC1	-13.811	-16.012	-20.198	-18.477	-40.640
	TC2	-27.401	-29.120	-31.560	-32.644	-46.921
	TC3	-32.438	-36.485	-36.755	-35.663	-51.385
Cross Validation						
Absolute Bias	TC1	966.545	853.638	669.487	647.509	406.122
	TC2	391.557	340.310	324.370	275.120	173.625
	TC3	503.845	396.970	386.249	382.587	192.803
Uncertainty	TC1	3756.835	3367.608	2705.282	2717.386	1844.386
	TC2	1974.386	1801.597	1574.553	1333.331	972.392
	TC3	2144.116	1725.701	1694.420	1702.369	970.789
Empirical Coverage	TC1	0.894	0.892	0.893	0.906	0.910
	TC2	0.915	0.908	0.908	0.919	0.918
	TC3	0.894	0.901	0.901	0.896	0.913
Computational Cost						
Minutes per 1000 iterations	–	< 1	~ 2	~ 5	~ 10	~ 11

† LL refers to posterior median of log-likelihood

As expected, the likelihood increases with increase in model complexity from left to right of Table 4.1. The jump is most evident in case of Model IV which also registers the steepest decline in BPIC values relative to its competitors. Models III-A and III-B perform comparably in most of the cases. In terms of out-of-sample cross-validation measures, although all candidates attain the desired coverages, they exhibit significant difference in bias and uncertainty values. Model IV again achieves the shortest bias and narrowest credible intervals

among the five candidates with improvement ranging between 33-50% relative to the 2nd best performing model. Models III-A and III-B perform comparably with respect to these measures as well. As expected, the computational cost rises sharply from left to right of the table, owing to the presence of spatial random effects in three rightmost models.

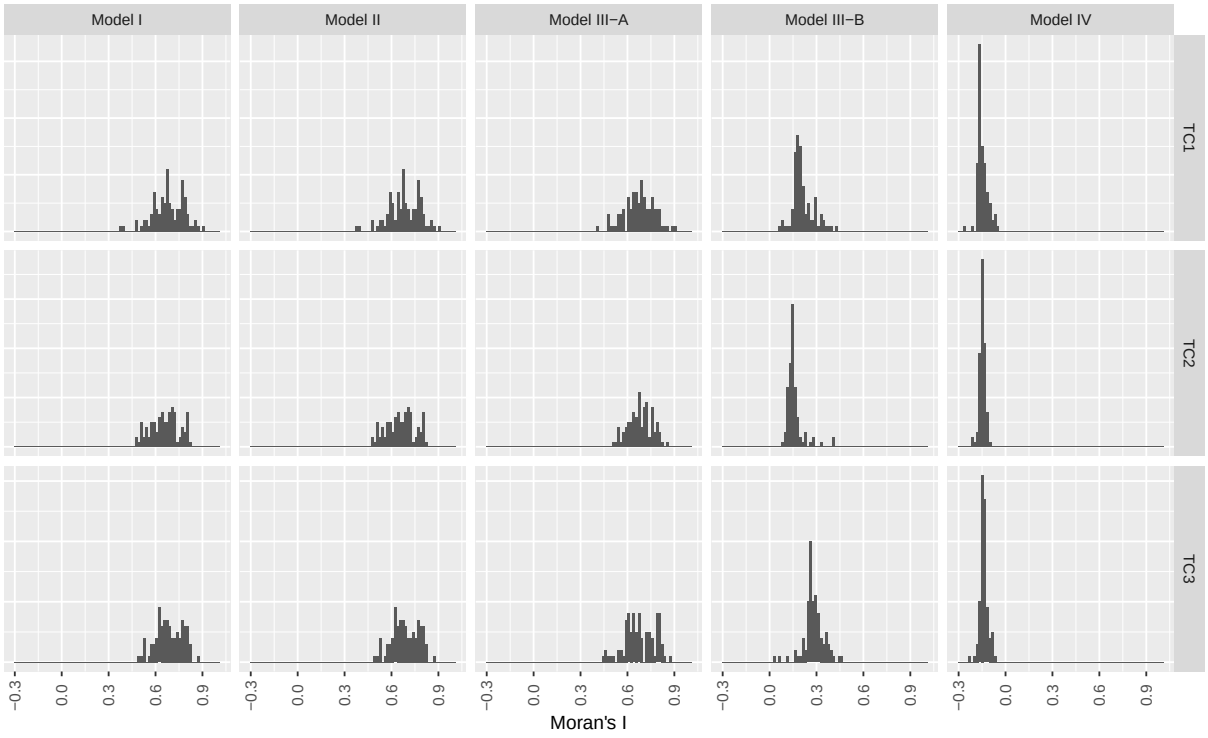


Figure 4.1: Histograms of 92 months' Moran I statistic, when more than 50% observations are available, calculated for five candidate models for each TC component

We perform one more assessment for presence of any leftover spatial association in the model residuals. As the given data contains a significant proportion of missing observations at many of the 120 months, we compute Moran I statistic, to measure the strength of spatial association among residuals, only for months when data are available from more than 50% of the cells. We use the R package ape (Paradis and Schliep, 2018) to compute the statistic based posterior median residuals, separately for each of 92 such months. Figure 4.1 represents the histogram of these 92 Moran's I values for all candidate models.

The histograms show that there exists strong spatial association among residuals in models

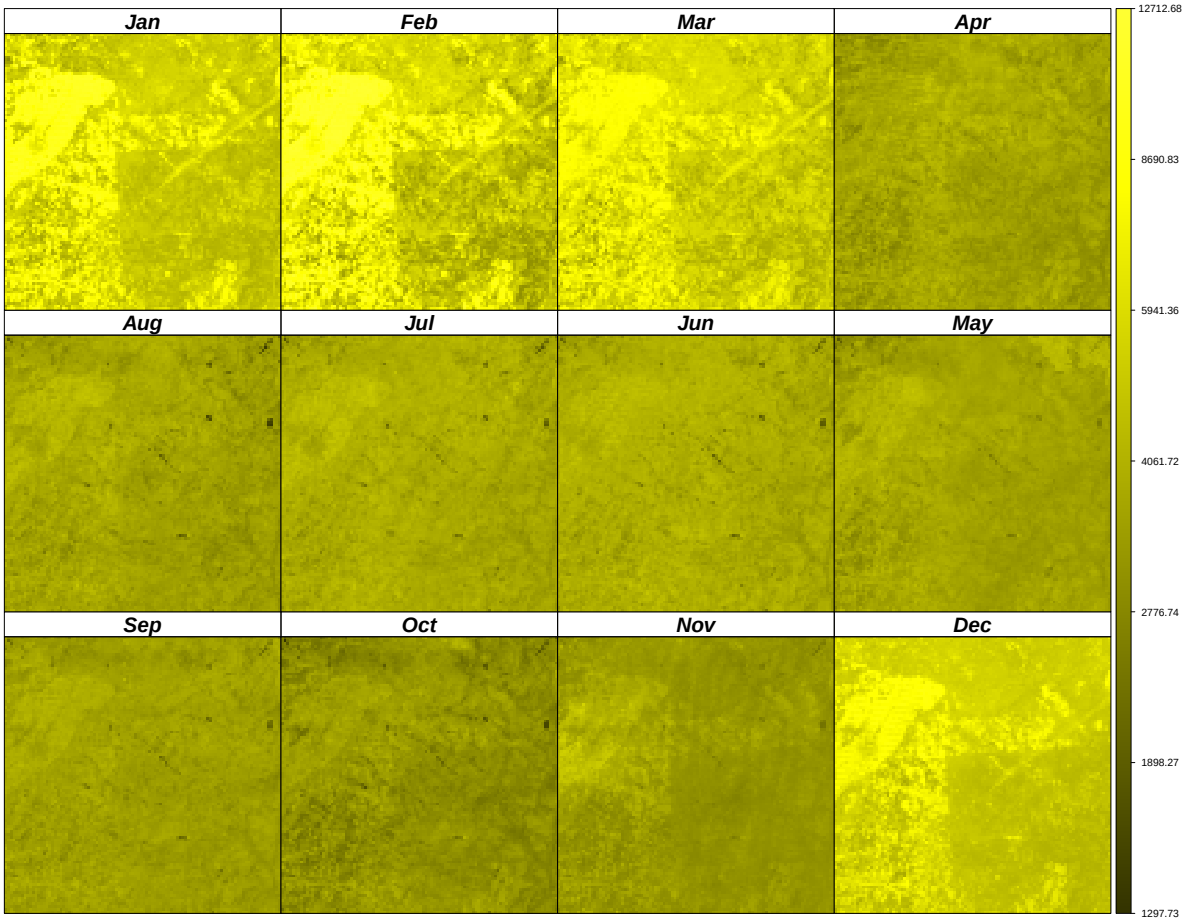


Figure 4.2: Monthly Map of 10-year average TC1 (Brightness) reconstructed using Model IV

without any spatial random effect (Model I and Model II) since in most of the months the statistic lies between 0.5 to 0.8 for all TC components. After introducing a single spatial random effect for all months at each cell (Model III-A), the strength of the association remains pretty similar to what has been observed for the first two models. Introduction of time-specific spatial random effects in Model III-B significantly reduces the value of the statistic mostly to a range of 0.1 to 0.4 but the p-value remains 0 in one tailed (greater) test at all months for all TCs. Model IV, that adds a temporal random effect to Model III-B, once again performs optimally with the Moran I values falling in the negative range and p-values close to or equal to 1 at all of these time points indicating no strong spatial

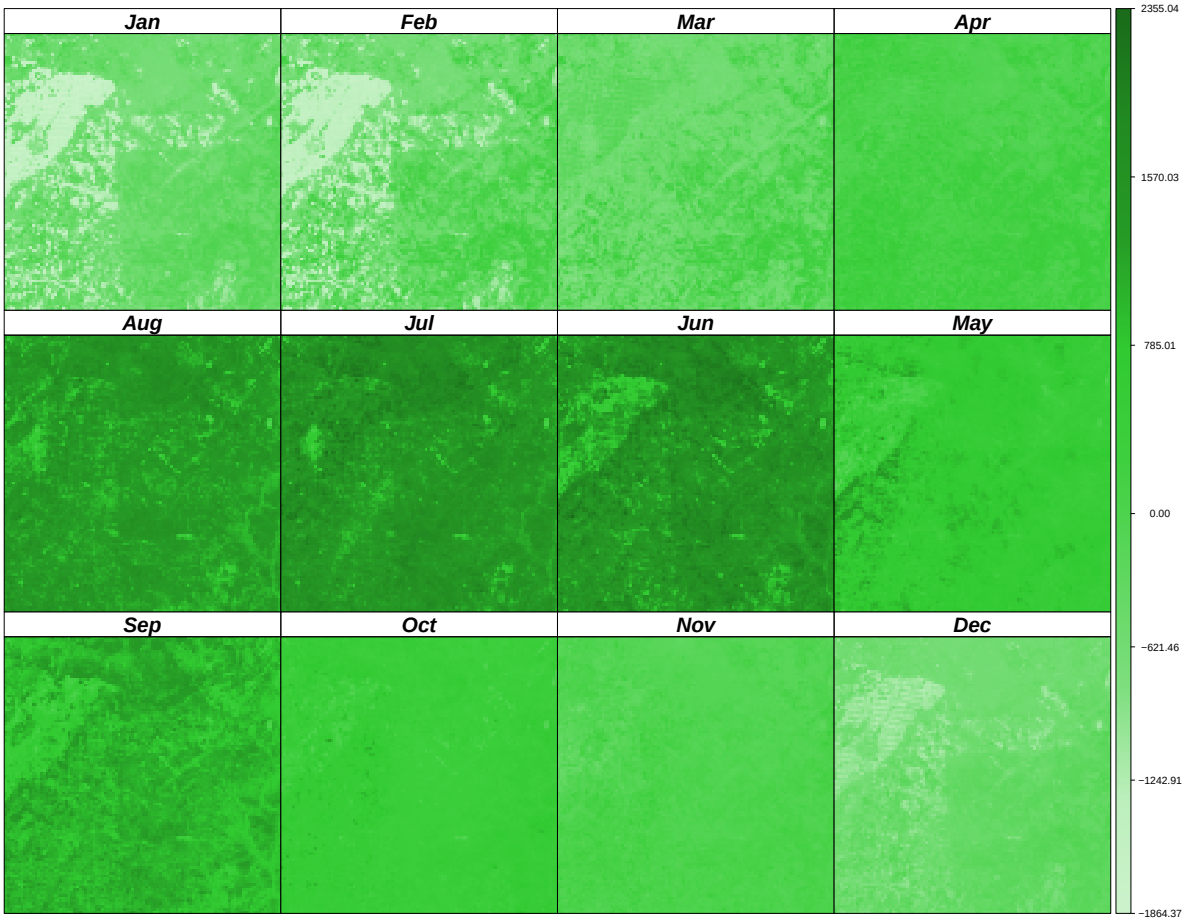


Figure 4.3: Monthly Map of 10-year average TC2 (Greenness) reconstructed using Model IV

association left among residuals.

The findings from Table 4.1 and Figure 4.1 suggest that despite the highest computational cost, performance of Model IV is far superior compared to the remaining models, for all three TC components. Hence, we choose Model IV to impute each of missing TC observations using median of the draws from its posterior predictive distribution. Figures 4.2 - 4.4 depict 10-year averaged monthly maps for reconstructed TC components in the original scale of these measurements.

Both the spatial and monthly variation seen in Figures 4.2 - 4.4 can be explained by differences in land cover across the study area, as well as seasonality in vegetation and weather. TC brightness is related to overall surface albedo. Snow-covered fields have larger albedo

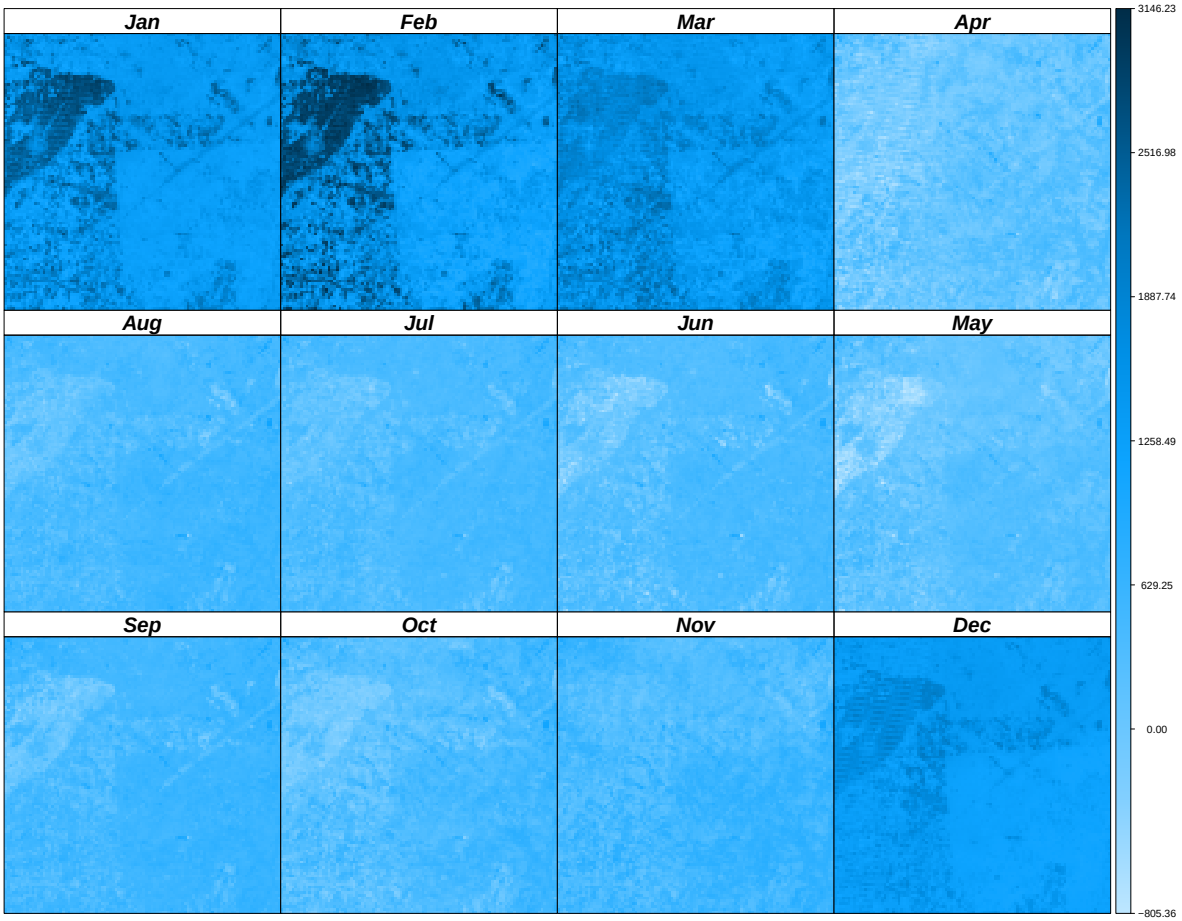


Figure 4.4: Monthly Map of 10-year average TC3 (Wetness) reconstructed using Model IV

values than forests. This pattern is most obvious in Figure 4.2 during the winter months, with the western portion of the image being brighter than the eastern portion. The contrast in brightness between forests and agricultural fields is much smaller during the summer months, in the absence of snow. TC greenness is related the amount of photosynthetically active vegetation on the land surface. During the winter months, agricultural fields have been previously harvested and are dormant. Similarly, deciduous trees have shed their leaves, while evergreen trees retain their needles. During the spring, fields are planted and crops begin to grow and deciduous trees leaf-out. Conversely in autumn, crops are harvested and deciduous trees begin to drop their leaves. These patterns are visible in Figure 4.3, with an overall increase in greenness during the spring and summer months. Also, the larger

greenness values of evergreen trees stand out in the southeast potion of the image during the winter months. Finally, TC wetness is related to soil moisture, as well as the presence of water and snow. As with TC brightness, wetness values are larger during the winter months. However, Figure 4.4 shows less contrast between forests and agricultural fields during the growing season of approximately April through October.

Chapter 5

Discussion

We have developed a hierarchical approach to fill in missing tasseled cap imagery. Certain aspects of this analysis are worth mentioning. The parametric structure developed in Chapter 3 can be generalized to any spatial or temporal scale, even if different from ours. For example, if one wants to summarize the TC features seasonally, that can easily be accommodated in the present setting. Additionally, that we established the feasibility of proposed approach to fill in tasseled cap features, its natural extension would be to scale up the analysis to the entire H20V05 WELD tile. This will necessitate focus on the computational aspects of the model. More specifically, the model for filling up the missing values in TC imagery involves spatiotemporal random effects and use of low-rank approximations should be explored to control the computational cost. In our analysis, we ignore any georeferencing errors in the satellite imagery.

The national forest inventory (NFI) of the United States, conducted under the Forest Inventory and Analysis (FIA) program of the USDA Forest Service, is designed to provide consistent and unbiased strategic-level information about the status and trends of the Nation's forest resources (Bechtold and Patterson, 2005). An important forest characteristics of interest is live tree basal area. We want to build a novel and flexible hierarchical model leveraging the relationship between the TC imagery, for which the current thesis reconstructs the complete set of measurements over each and every population unit at regular time intervals, and the basal area data, for which we have relatively few observed values only from the FIA sample units at certain years, in order to make annual predictions of the latter for every population unit. The majority of studies use remote sensing imagery either at a single point in time or as a composite of images over the study duration. However, we want to use

the entire time series of monthly images over the decade of the study period. The temporal patterns of these features, being correlated with land cover, are shown to be informative in distinguishing deciduous from evergreen forests, as well as different tree species from one another (Wilson et al., 2012). Hence, in the regression equation for forest characteristic, we are planning to use the TC imagery as functional predictors (Morris, 2015; Ramsay and Dalzell, 1991).

Bibliography

Ando, T. (2007), "Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models," *Biometrika*, 94, 443–458.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014), *Hierarchical modeling and analysis for spatial data*, CRC press.

Bechtold, W. A. and Patterson, P. L. (2005), "The enhanced forest inventory and analysis program-national sampling design and estimation procedures," Tech. Rep. SRS-80, US Department of Agriculture, Forest Service, Southern Research Station, Ashville, NC.

Besag, J. (1974), "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 192–236.

Besag, J. and Kooperberg, C. (1995), "On conditional and intrinsic autoregressions," *Biometrika*, 82, 733–746.

Brook, D. (1964), "On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems," *Biometrika*, 51, 481–483.

Cleland, D., Freeouf, J., Keys, J., Nowacki, G., Carpenter, C., and McNab, W. (2007), "Ecological subregions: sections and subsections for the conterminous United States," *General Technical Report WO-76D*, 76.

Gelfand, A. E., Ecker, M. D., Knight, J. R., and Sirmans, C. (2004), "The dynamics of location in home price," *The journal of real estate finance and economics*, 29, 149–166.

Huang, C., Wylie, B., Yang, L., Homer, C., and Zylstra, G. (2002), "Derivation of a tasselled cap transformation based on Landsat 7 at-satellite reflectance," *International Journal of Remote Sensing*, 23, 1741–1748.

Karlson, M., Ostwald, M., Reese, H., Sanou, J., Tankoano, B., and Mattsson, E. (2015), "Mapping tree canopy cover and aboveground biomass in Sudano-Sahelian woodlands using Landsat 8 and random forest," *Remote Sensing*, 7, 10017–10041.

Kauth, R. J. and Thomas, G. (1976), "The tasselled cap–a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat," in *LARS Symposia*, p. 159.

Li, Y., Jun, Y., and Zeng, T. (2017), "Deviance information criterion for Bayesian model selection: Justification and variation," *Working Paper, available at: https://ink.library.smu.edu.sg/soe_research/1927/.*

McNab, W., Cleland, D., Freeouf, J., Keys Jr, J., Nowacki, G., and Carpenter, C. (2007), "comps. 2007. Description of ecological subregions: sections of the conterminous United States [CD-ROM]," *Gen. Tech. Report WO-76B. Washington, DC: US Department of Agriculture, Forest Service.*

Morris, J. S. (2015), "Functional regression," *Annual Review of Statistics and Its Application*, 2, 321–359.

Neal, R. M. (2003), "Slice sampling," *Annals of statistics*, pp. 705–741.

Paradis, E. and Schliep, K. (2018), "ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R," *Bioinformatics*, bty633.

Ramsay, J. O. and Dalzell, C. (1991), "Some tools for functional data analysis," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 539–572.

Roy, D. P., Ju, J., Kline, K., Scaramuzza, P. L., Kovalskyy, V., Hansen, M., Loveland, T. R., Vermote, E., and Zhang, C. (2010), "Web-enabled Landsat Data (WELD): Landsat ETM+ composited mosaics of the conterminous United States," *Remote Sensing of Environment*, 114, 35–49.

Sellers, P. J., Tucker, C. J., Collatz, G. J., Los, S., Justice, C., Dazlich, D., and Randall, D. (1994), "A global 1 by 1 NDVI data set for climate studies. Part 2: The generation of global fields of terrestrial biophysical parameters from the NDVI," *International Journal of remote sensing*, 15, 3519–3545.

Sellers, P. J., Tucker, C. J., Collatz, G. J., Los, S. O., Justice, C. O., Dazlich, D. A., and Randall, D. A. (1996), "A revised land surface parameterization (SiB2) for atmospheric GCMs. Part II: The generation of global fields of terrestrial biophysical parameters from satellite data," *Journal of climate*, 9, 706–737.

Wilson, B. T., Lister, A. J., and Riemann, R. I. (2012), "A nearest-neighbor imputation approach to mapping tree species over large areas using forest inventory plots and moderate resolution raster data," *Forest Ecology and Management*, 271, 182–198.

Wilson, B. T., Knight, J. F., and McRoberts, R. E. (2018), "Harmonic regression of Landsat time series for modeling attributes from national forest inventory data," *ISPRS Journal of Photogrammetry and Remote Sensing. 137: 29-46.*, 137, 29–46.

Yuan, F., Sawaya, K. E., Loeffelholz, B. C., and Bauer, M. E. (2005), "Land cover classification and change analysis of the Twin Cities (Minnesota) Metropolitan Area by multitemporal Landsat remote sensing," *Remote Sensing of Environment*, 98, 317–328.

Zheng, S., Cao, C., Dang, Y., Xiang, H., Zhao, J., Zhang, Y., Wang, X., and Guo, H. (2014), "Retrieval of forest growing stock volume by two different methods using Landsat TM images," *International journal of remote sensing*, 35, 29–43.