8-2019

# Spatio-Temporal Analysis of Tree Ring Chronology and Precipitation

Ruizhe Yin
*University of Arkansas, Fayetteville*

Spatio-Temporal Analysis of Tree Ring Chronology and Precipitation

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

by

Ruizhe Yin
Northwest University
Bachelor of Engineering in Resources Exploration, 2014
University of Arkansas
Master of Science in Geology, 2016

August 2019
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

John Tipton, Ph.D.
Thesis Director

Giovanni Petris, Ph.D.
Committee Member

Jyotishka Datta, Ph.D.
Committee Member

**Abstract**

Tree ring chronology data is known to reflect regional climate due to the strong impact of rainfall and temperature. Therefore, tree ring data can be used to reconstruct historical climate in order to understand how climate changed in the past and make prediction about the future behavior of the climate. For simplicity, this research only considers the influence of precipitation on tree ring growth within the New England area. A total of 94 measurement sites are used to record tree ring width over 881 years and corresponding precipitation data are given at some locations for 121 years. We developed a spatio-temporal model to describe the response of a tree growth to precipitation on an annual timescale and introduced the general hierarchical statistical framework from data, process and parameter models. To predict climate in the past, we considered an autoregressive process in time that accounts for temporal correlation of precipitation. Based on data collected at each observed location, geospatial Kriging allows us to predict reasonable precipitation at unobserved locations in a regional perspective.

**Acknowledgements**

I would like to express my deepest appreciation to various people for their help of my thesis and the whole graduate life in Department of Mathematical Science at University of Arkansas. Foremost, I would like to express my heartfelt thanks to Dr. John Tipton, for the continuous support of my master's study and research, for his patience, motivation, and immense knowledge. Im very pleased to work with him. His guidance helped me in all time of research and writing of this thesis.

I would like to express my sincere gratitude to my thesis committee members, Dr. Giovanni Petris and Dr. Jyotishka Datta, my committee members.for their useful advice, teachings, selfless support and excellent guidance throughout the duration of my study. I will keep all delighted memories and lasting friendships in my mind. Thanks to all faculties in Math Department for offering me such an enjoyable study place.

I would like to offer my special thanks to my friend, Xuan Gu and Tim He, for their kindness to make my study life filled with laughter. As I look ahead, I am strengthened by family and friendship. It is undeniable to express my love, respect, and thanks to my families. I cant have any opportunities to study in America without their raising and supports as they always encouraging me to not give up when I feel tired and confused. A special thanks to my husband, Yueyang Wang, who share many memorable moments with me and always stand behind me as my backer.

**Table of Contents**

**List of Figures**

## 1. Introduction

It is important to reconstruct the paleoclimate in order to comprehend how climate varied in the past and to forecast the future behavior of climate. Tree ring width data is typically considered to reflect the climate of a certain region due to the strong impact of rainfall and temperature to the tree growth (Guiot, 1982; Fritts, 1976; Schweingruber, 1996). The tree ring width is the amount of tree grows annually, which can reflect the climate condition of the year. Nowadays, two kinds of data are often compared to analyze climatology: one data source is the annual climate data which includes the total precipitation and average of temperatures, the other data source is the annual tree growth data which includes the tree ring width or densitometric parameters (Tessier, 1989). Climate data collected at each location only provides ”local information”, it is important and meaningful for us to understand and predict the ”regional patterns” based on the observed data. What's more, we also need to know whether the climate growth system is stable in time. To answer both questions together, we analyze our data using spatio-temporal statistical methods.

### 1.1. Tree ring data

The tree ring chronology data used for our research is measured from 12 different tree species at N=94 sites within the central New England region of USA. Each tree ring chronology is at least 160 years long for each species at most sites, some sites have records for the total of 881 years that date back to 1135 (Figure 1.1). The observation of annual tree ring widths observed from location $s$ at time $t$ are represented by $Y(\mathbf{s}, t)$.

As shown in the Figure 1.1, each line represents a chronology. There are plenty of tree ring chronology data from 1815 until 2015, but the datasets are less dense in the past. This leads to a more complicated situation to fit the statistical model because the model must be capable of dealing with missing data.

## 1.2. Climate data

The climate data used in this research are Parameter-Elevation Relationships on Independent Slopes Model (PRISM) gridded data for the total of 121 years from 1895 to 2015 (Group, 2011). In this analysis, we obtained both temperature and annual total precipitation (rain and melted snow) data from PRISM datasets collected from N=94 sites within the New England area of USA. The log transformation of precipitation data can be recognized as normally distributed within a given month. The average temperature is also normally distributed at a given month but follows the same pattern that the highest temperature occurs around June to August and the lowest temperatures can be recorded from December to February. We simplify the research to fit tree ring width data using log precipitation following results in Tipton et al. (2016) that suggest the tree ring chronologies are moisture sensitive. We model the log precipitation data with a Gaussian distribution as spatial random effect.

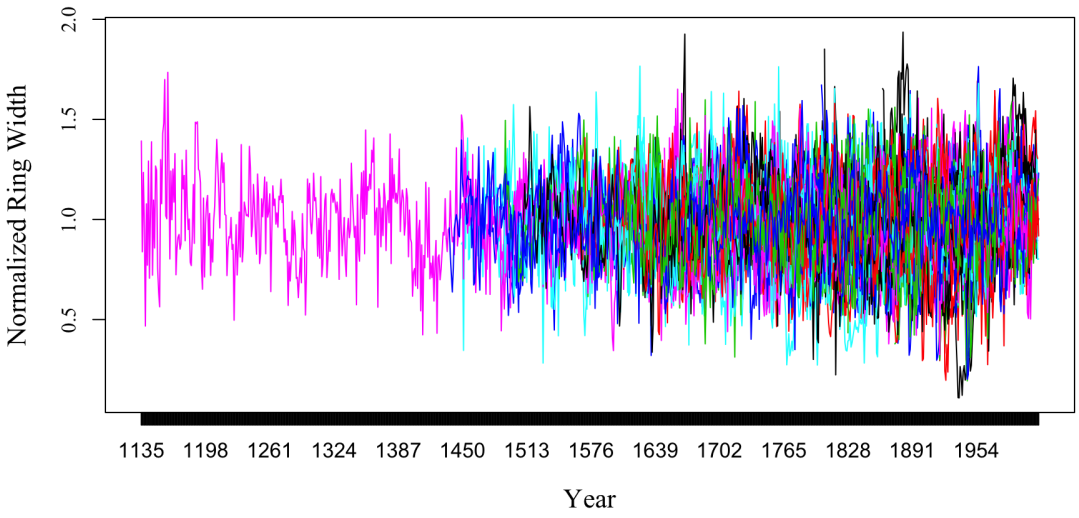



Figure 1.1: Tree-ring chronology data collected from 94 sites from 1135 to 2015. Each line is a time series data. Denser and more black lines can be noticed from 1815 to 2015, the color of lines becomes lighter as data density decreases as time goes back to 1135.

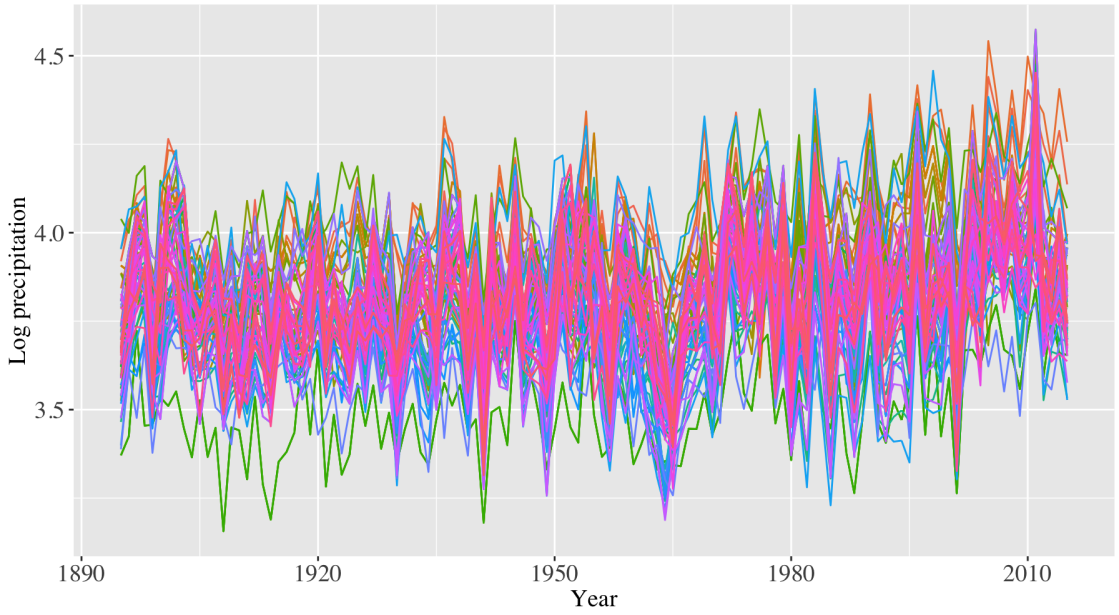



Figure 1.2: Log total annual precipitation data collected from 94 sites within the same area as tree ring width data. Each line represents the log annual precipitation at each site.

Figure 1.2 is the time series plot of log annual total precipitation at each site represented by different colored lines from 1895 to 2015. From this figure, we notice the log total precipitation recorded at some sites is lower than other sites, but the average is approximately constant in time with, perhaps, a slight increase from year 1980 to 2015.

We have tree ring width data dating back to 1135 however, the precipitation data were only collected from 1895 to 2015, meaning there are unobserved climate data over 760 years. When fitting the model, it is necessary to reconstruct unknown precipitation in the past using a relationship between tree ring width and precipitation learned from the overlapping data sources. Using this learned relationship, we then predict the annual log precipitation over the spatial domain of interest.

## 2. Statistical Method

Spatio-temporal data analysis is a growing area of statistical research because of the development of computational techniques for large datasets. Virtually all data occur at some locations in space and time and data commonly found in the real world involve interactions across various spatial and temporal scales. A primary aim of spatio-temporal data analysis is to predict the process in both space and time by describing the characteristics of the process with a statistical model.

A general hierarchical statistical model (HM) is often an approach to derive the posterior distribution for both spatial and spatio-temporal processes. HM often uses a shinkage estimator, pulling the posterior distribution towards the prior mean of individual parameters and reducing the mean squared errors (Zhao et al., 2010). Three commonly used methods to estimate Hierarchical Models are maximum likelihood estimation (MLE), Bayesian hierarchical model (BHM), and empirical Bayes model (EBM). To find the MLE, one need to estimate parameters by finding their values that can maximize the likelihood function. A fully BHM arises from modeling parameter uncertainty at a parameter model level, with the advantage that BHM gives us in-depth thought and collaboration between climatologists and statisticians (Hoeting, 2009). We usually find the joint posterior of all parameters and make inference with Markov Chain Monte Carlo (MCMC). Finding the MLE of hyper-parameters, then plugging the estimates into a BHM gives rise to an empirical Bayes model (EBM). BHM produces a shrinkage estimator for all parameters towards their population means, but can be difficult to compute. EBM can be viewed as an approximation to a BHM. EBM neglects the uncertainty of hyper-parameters by assigning them with most likely values instead of integrating over a distribution. Therefore, EBM normally produces narrower credible intervals and more intensive shrinkage estimators than BHM. To solve a complex problem, it is straightforward to break the problem into three stages: data models (likelihood), process models (prior distribution), and parameter models (hyper-prior distribution). We begin by describing the general spatial model, and introduce the general spatio-temporal model by adding the influence of temporal behavior into the model, then derive a spatio-temporal model

based on our real problem in what follows.

### 2.1. Spatial Model

Consider a process occurring at spatial locations $\mathbf{s}$, where the elements of $\mathbf{s}$ are latitude and longitude, and $\mathbf{s} \in \mathcal{D}$ is a fixed finite area. The process is observed at N locations $\{\mathbf{s}_1, ..., \mathbf{s}_N\}$ where $\mathbf{Y}$ is the observed geostatistical random field:

$$\mathbf{Y} = (Y(\mathbf{s}_1), ..., Y(\mathbf{s}_N))'. \tag{2.1}$$

The geostatistical model is an application of regionalized variables theory to prediction of the values in spatial or spatio-temporal phenomena (Matheron, 1971). The geostatistical model assumes that spatially discrete data are sampled from an unobserved continuous process over a finite area. By learning the spatial relationships of the continuous process, one can predict at new locations (Diggle et al., 2010).

In order to construct a spatially correlated regression model, we need to define a general hierarchical model for observations. Here, we introduce the general hierarchical statistical model in sections first describing the data model, then the process model, and finally we describe the parameter model (Arab, 2015).

#### 2.1.1. Data Model

The observed random field $Y(.)$ is a noisy measurement of value of the geostatistical process $Z(.) = \{Z(\mathbf{s}) : \mathbf{s} \in \mathcal{D}_s\}$ at locations $\mathbf{s} \in \mathcal{D}$. Then, we assume the observation model:

$$Y_i(\mathbf{s}) = Z(\mathbf{s}) + \varepsilon_i(\mathbf{s}), \tag{2.2}$$

where $\varepsilon_i(\mathbf{s})$ is a white noise process with mean zero and variance $\sigma^2_\varepsilon \geq 0$. In this equation, $i$ is the number of repeated measures at site $\mathbf{s}$. In most cases, $i = 1$, so for simplicity, we drop the repeated

measures notation in what follows. However, in the application, we have some sites with repeated measures.

Equivalently, we can write:

$$Y(\mathbf{s})|Z(\mathbf{s}), \sigma_\varepsilon^2 \overset{iid}{\sim} \mathrm{N}(Z(\mathbf{s}), \sigma_\varepsilon^2). \tag{2.3}$$

We assume that $Z(.)$ and $\varepsilon(.)$ are independent Gaussian processes and characterize the probability distribution using the first two moments.The assumption of second order stationarity of the random process $Z(.)$, also called weak stationarity or covariance stationarity, meaning the mean is constant and the covariances $cov(Z(\mathbf{s}), Z(\mathbf{s}+\mathbf{h}))$ depend only on lag $\mathbf{h}$. Therefore, we can define the covariance operator

$$C_Z(\mathbf{h}) = cov(Z(\mathbf{s}), Z(\mathbf{s}+\mathbf{h})), \quad \text{for all } \mathbf{s}, \mathbf{s}+\mathbf{h} \in \mathcal{D}_\mathbf{s}. \tag{2.4}$$

When $\mathbf{h} = 0$, define $C_Z(\mathbf{0}) \equiv \sigma_Z^2$ and

$$\lim_{\mathbf{h}\to 0}(C_Z(\mathbf{0}) - C_Z(\mathbf{h})) = C_Z(\mathbf{0}+) = \sigma_0^2 \geq 0, \tag{2.5}$$

where $\sigma_0^2 \leq \sigma_Z^2$ and $\sigma_0^2$ is called the micro scale variance of the random process $Z(.)$. In early development of spatial and spatio-temporal statistics, it was assumed that there is no measurement error of any observation taken on the process, which is an unreasonable assumption in general. Then, the covariance of the data is,

$$cov(Y(\mathbf{s}), Y(\mathbf{s}+\mathbf{h})) \equiv C_Y(\mathbf{h}) = \begin{cases} \sigma_Z^2 + \sigma_\varepsilon^2, \ \mathbf{h} = 0 \\ C_Z(\mathbf{h}), \ \mathbf{h} \neq 0, \end{cases} \tag{2.6}$$

when $\mathbf{h} = 0$, we have $\sigma_Y^2 = \sigma_Z^2 + \sigma_\varepsilon^2$ which is referred to as the sill. With $var(\varepsilon(\mathbf{s})) = \sigma_\varepsilon^2$, we know that

$$\lim_{\mathbf{h}\to 0}(C_Y(\mathbf{0}) - C_Y(\mathbf{h})) = C_Y(\mathbf{0}+) = \sigma_0^2 + \sigma_\varepsilon^2 = c_0 \geq 0. \tag{2.7}$$

In this equation, the quantity $c_0$ containing two non-negative parts is called the nugget effect in the geostatistical area, where $\sigma_0^2$ represents the micro scale component of $Z(.)$ and $\sigma_\varepsilon^2$ stands for the measurement error of the whole process $Y(.)$ (Cressie and Wikle, 2015, p.123).

### 2.1.2. Process Model

The process model describes the latent, unobserved process of interest. For the geostatistical spatial model, defined on $\mathcal{D} \in \mathbb{R}^d$, Gaussian geostatistical models (GGMs) are commonly used to model continuous spatial Gaussian processes using a spatial covariance, which is a function of spatial distance and direction. There are two common assumptions of GGMs, second order stationarity and isotropy (Song et al., 2008). Second order stationarity means that the process has a constant mean and the covariance function is determined by the distance between two measurement locations over space. The term isotropic implies that the covariance function of the process only depends on Euclidean distance, meaning direction is not considered. One can model the covariance of a second order stationary and isotropic spatial process using a parametric function of Euclidean distance (Song et al., 2008).

In the data model provided above, the independent Gaussian process $Z(.)$ is the geostatistical process. We use the geostatistical process as a process model in the HM context and define the process model as

$$Z(\mathbf{s}) = X(\mathbf{s})'\boldsymbol{\beta} + \eta(\mathbf{s}),\ \mathbf{s} \in \mathcal{D}, \tag{2.8}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_{p-1})'$ is a $p$-dimensional fixed effect, $X(\mathbf{s}) = (1, X_1(\mathbf{s}), ..., X_{p-1}(\mathbf{s}))$ is a known $N \times p$ matrix (p < N) of covariates, containing the constant 1 and $p-1$ columns explanatory covariates potentially observed at location $\mathbf{s}$, and $\eta(\mathbf{s})$ is the random spatial effect at location $\mathbf{s}$ which is assumed to from isotropic Gaussian process $\boldsymbol{\eta}$ with mean zero and stationary covariance function $C_\eta$. We use $\mu(\mathbf{s}) \equiv X(\mathbf{s})'\beta$ to represent the mean of the spatial process.

Consequently, the simplified geostatistical data model can be written as:

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \eta(\mathbf{s}) + \varepsilon_Y(\mathbf{s}) \tag{2.9}$$

where $\boldsymbol{\eta} = (\eta(\mathbf{s}_1), ..., \eta(\mathbf{s}_N))' \sim N(\mathbf{0}, C_\eta)$ and the covariance of the $ij$th element $C_{\eta_{ij}} = cov(\eta(\mathbf{s}_i), \eta(\mathbf{s}_j)) = \tau^2 \mathbf{R}_{ij}$ where $\tau^2$ is the variance of the spatially correlated process and the matrix $\mathbf{R}$ is an isotropic correlation matrix that has ijth element $\mathbf{R}_{ij}$ as a function of the distance $|\mathbf{s}_i - \mathbf{s}_j|$ between location $\mathbf{s}_i$ and $\mathbf{s}_j$.

Common choices for the correlation functions are the Matérn family, of which the Gaussian and exponential covariance functions are special cases. In our model, we choose the Gaussian covariance function, also known as the double exponential function as our correlation function. The Gaussian correlation function matrix has $ij$th element $\mathbf{R}_{ij}(\phi) = \exp(\frac{-d_{ij}^2}{\phi^2})$ where $d_{ij}$ is the function of distance between $\mathbf{s}_i$ and $\mathbf{s}_j = \|\mathbf{s}_i - \mathbf{s}_j\|^{\frac{1}{2}}$, and $\phi$ is a range parameter that controls the rate of decrease of the correlation between observations as distance increases. The two hyper-parameters $\tau$ and $\phi$ are not identifiable, meaning that no matter what methods are used, neither parameter can be estimated consistantly even with larger datasets. However, the product of $\tau$ and $\phi$ can be estimated consistantly which is the quantity that is important to the spatial interpolation (Zhang, 2004).

### 2.1.3. Parameter Model

Parameter models for Bayesian Hierarchical models, are prior distributions for unknown parameters. We should define appropriate probability densities to unknown parameters according to our knowledge about their underlying support. If a priori knowledge is unavailable, it is often better to use non-informative distributions.

Prior distributions we choose for the geostatistical model mentioned above are $\phi \sim \text{Unif}(\phi_L, \phi_U)$, $\tau^2 \sim \text{InverseGamma}(\alpha_\tau^2, \beta_\tau^2)$, $\sigma_y^2 \sim \text{InverseGamma}(\alpha_\sigma^2, \beta_\sigma^2)$.

### 2.1.4. Spatial Kriging

The term Kriging was proposed by Georges Matheron (1963) in honor of D.G. Krige, a South African mining engineer. In geostatistics, Kriging can use a set of scattered data points to produce an estimated surface based on their spatial autocorrelation. In general, Kriging is an interpolation method for spatial prediction of geostatistical data based on the theory of stochastic processes.

Kriging is a prediction method based on statistical autocorrelation among measurement points. Because the assumption of a spatial correlation is reflected by the distance or direction between sample points that can be used to explain variation in the surface, we use Kriging to generate a prediction surface with associated uncertainty. The estimated quantity of Kriging is interpreted as a random variable of a process at an unknown location, given observed quantities in neighbor as well as the estimated variances.

One can estimate random effects in linear mixed models using Best Linear Unbiased Prediction (BLUP) (Robinson, 1991). Although it is similar to Best Linear Unbiased Estimate (BLUE), there are difference between these two methods. The BLUP is used to predict random effects while BLUE is used to estimate fixed effects. In general, one uses maximum likelihood or least square method to estimate fixed effects, while random effects are estimated using restricted maximum likelihood (REML)(Robinson, 1991). Furthermore, random effects are assumed to be sampled randomly from the population we are interested in, while fixed effects are assumed to be chosen consciously because they are parameters of interests (Searle et al., 2009). When fitting a mixed model, we first estimate the mean and covariance of the random effects, then the random effect can be calculated from the estimated mean and covariance as well as the known data. As an analogy, in the Bayesian framework the estimated mean and covariance specify a normal prior and the known data can be treated as likelihood. Then, the BLUP can be considered as the posterior mean as it takes account of both known data and estimated prior. In practice, the parameters,

including variances and residuals associated with the random field, are normally unknown. By simply plugging in the estimated values of these parameters into the predictors, the Empirical Best Linear Unbiased Predictor (EBLUP) can be developed. EBLUP can cause overly optimistic prediction as parameter variability will not be accounted for.

Kriging is the interpolation method derived from BLUP theory. Given appropriate assumptions of priors (for example, estimated mean and covariance), Kriging is similar to BLUP of a random field at unobserved locations. Assuming a Gaussian process, one can predict unobserved values at any new spatial locations, based on the prior and likelihood function associate with observed values at some spatial locations. In other word, Kriging can be explained as Bayesian inference (Williams, 1998), meaning that Kriging starts with a Gaussian prior distribution and the covariance function evaluated between any two sample locations. The posterior is also Gaussian distributed with a mean and covariances that can be computed from observed values. An analysis using Kriging consists of multiple steps, including exploratory statistical analysis of the data, variogram modeling, creating the surface, and optionally exploring a variance surface (Sen et al., 2008; Al-Hasnawi et al., 2017).

Briefly speaking, Kriging fits a mathematical model to all sample points within a specified area, weights the surrounding observed values, then derive a prediction for an unobserved location which is a linear combination of observations. The general formula is:

$$\hat{\eta}(\mathbf{s}_0) = \sum_{i=1}^{N} \lambda_i \eta(\mathbf{s}_i), \tag{2.10}$$

where $\eta(\mathbf{s}_i)$ is the observed value of the spatial random effect at the $i$th location, $\lambda_i$ is an unknown weight for the random effect at $i$th location, which is determined by the covariance between the observed and unobserved points, $\mathbf{s}_0$ is the predicted location and $N$ is the number of observed points. In our log precipitation example, we predict log precipitation at unobserved locations and $\mathbf{s}_0$ according to data measured at observed locations $\mathbf{s}_i$.

Now we recall that both data and process models of the HM are assumed to be Gaussian processes with spatial random process. Because all joint probability and conditional probabilities are Gaussian distributions, the simple kriging predictor is the optimal spatial predictor $E(\boldsymbol{\eta}(\mathbf{s}_0)|\mathbf{Y}, \boldsymbol{\mu}, C_\eta, \sigma^2_\varepsilon)$ which is linear in the data (Cressie and Wikle, 2015). In this case, we have to generate predictions for the unobserved locations given observed data. Now we define the random geostatistical spatial process, mean of the spatial process and data at observed locations as $\boldsymbol{\eta}_{obs}$, $\boldsymbol{\mu}_{obs}$, $\mathbf{Y}_{obs}$ and use *unobs* to represent values at unobserved locations which need to be predicted, our model can be written as:

$$\begin{pmatrix} \mathbf{Y}_{obs} \\ \mathbf{Y}_{unobs} \end{pmatrix} \sim N\left( \begin{pmatrix} \boldsymbol{\mu}_{obs} \\ \boldsymbol{\mu}_{unobs} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\eta}_{obs} \\ \boldsymbol{\eta}_{unobs} \end{pmatrix}, \sigma^2_\varepsilon \mathbf{I} \right) \tag{2.11}$$

$$\begin{pmatrix} \boldsymbol{\eta}_{obs} \\ \boldsymbol{\eta}_{unobs} \end{pmatrix} \sim N\left( \begin{pmatrix} \mathbf{0}_{obs} \\ \mathbf{0}_{unobs} \end{pmatrix}, \begin{pmatrix} C_{obs,obs} & C_{obs,unobs} \\ C_{unobs,obs} & C_{unobs,unobs} \end{pmatrix} \right), \tag{2.12}$$

where $C_{obs,obs} = [cov(\eta(\mathbf{s}_i), \eta(\mathbf{s}_j))]^{n}_{i,j=1}$, $C_{unobs,obs} = [cov(\eta(\mathbf{s}_{oi}), \eta(\mathbf{s}_j))]^{n,n_{unobs}}_{i=1,j=1}$, $C_{unobs,unobs} = [cov(\eta(\mathbf{s}_{oi}), \eta(\mathbf{s}_{oj}))]^{n_{unobs}}_{i,j=1}$. To estimate the latent $\boldsymbol{\eta}$, we use composition sampling. Using the conditional multivariate normal distribution for sampling $\boldsymbol{\eta}_{unobs}$, we have the model (Rencher, 2002):

$$\boldsymbol{\eta}_{unobs}|\boldsymbol{\eta}_{obs} \sim N(\bar{\mu}, \bar{\Sigma}), \tag{2.13}$$

where

$$\bar{\mu} = \boldsymbol{\mu}_{unobs} + C_{unobs,obs} C^{-1}_{obs,obs} (\boldsymbol{\eta}_{obs} - \boldsymbol{\mu}_{obs}) \tag{2.14}$$

$$\bar{\Sigma} = C_{unobs,unobs} - C_{unobs,obs} C^{-1}_{obs,obs} C_{obs,unobs}. \tag{2.15}$$

## 2.2. Spatio-Temporal Model

When data vary spatially and temporally, we model the data using a spatio-temporal process. Combined with data model introduced in section 2.1.1, the general form of a spatio-temporal model can be generalized as:

$$Y(\mathbf{s},t) = \mu(\mathbf{s},t) + \eta(\mathbf{s},t) + \varepsilon_Y(\mathbf{s},t) \tag{2.16}$$

for $\mathbf{s} \in D$ and $t \in [1,T]$.

Define the spatio-temporal random effect to be a single stationary Gaussian process, $\boldsymbol{\eta} = [\eta(\mathbf{s}_1,1),\cdots,\eta(\mathbf{s}_N,1),\eta(\mathbf{s}_1,2),\cdots,\eta(\mathbf{s}_N,T)]' \sim \mathrm{N}(0,\boldsymbol{\Sigma})$, where the covariance function $cov(\eta(\mathbf{s}_i,t_i),\eta(\mathbf{s}_j,t_j)) = c(\|\mathbf{s}_i-\mathbf{s}_j\|,\|t_i-t_j\|)$ depends on both $\|\mathbf{s}_i-\mathbf{s}_j\|$ and $\|t_i-t_j\|$ with the dimension of $NT \times NT$ (Genton, 2007). The covariance is a separable structure which means that there is no conditional dependence of observations between space and time. Thus, the covariance of the process is the product of a spatial and a temporal covariance function:

$$cov(\eta(\mathbf{s}_i,t_i),\eta(\mathbf{s}_j,t_j)) = \tau^2 C_1(\|\mathbf{s}_i-\mathbf{s}_j\|;\phi)C_2(\|t_i-t_j\|;\theta). \tag{2.17}$$

Then the covariance can be written as the Kronecker product between spatial and temporal covariance matrices:

$$\boldsymbol{\Sigma} = \Sigma_\eta(\tau^2,\phi,\theta) = \tau^2 \mathbf{R}_s(\phi) \otimes \mathbf{R}_t(\theta), \tag{2.18}$$

where $\mathbf{R}_s(\phi)$ and $\mathbf{R}_t(\theta)$ are spatial and temporal correlation matrices with sizes $N \times N$ and $T \times T$ respectively.

Dynamic spatio-temporal models are usually used to describe a process that is discrete in space and time (Wikle, 2015). For our annual precipitation example, the data we have is discrete in time, therefore, considering the spatial random effect $\eta(\mathbf{s},t)$ to be a discrete time series process is reasonable. By applying the same idea of Hierarchical Models to dynamic spatio-temporal

models, one generally makes Markov assumptions for the process, with the requirement of the following form (Wikle, 2015):

$$\eta_t(\cdot) = \mathcal{M}(\eta_{t-1}(\cdot); \boldsymbol{\theta_n}; \delta_t(\cdot)), \tag{2.19}$$

where $\mathcal{M}$ is the function of parameters $\boldsymbol{\theta_n}$ that can control the evolution of the process, in our model, we use $\rho$ instead, the most recent past observation $\boldsymbol{\eta}_{t-1}$ and the spatial error process $\boldsymbol{\delta}_t$. In general, the model can have a Gaussian or non-Gaussian distribution. We consider an AR(1) time series process for the spatial random effect $\eta(\mathbf{s}, t)$. Defining $\boldsymbol{\eta}_t = (\eta(\mathbf{s}_1, t), \eta(\mathbf{s}_2, t), ..., \eta(\mathbf{s}_N, t))$,we have the dynamic linear model representation of the space-time process:

$$\boldsymbol{\eta}_t = \mathbf{M}_t \boldsymbol{\eta}_{t-1} + \boldsymbol{\delta}_t, \tag{2.20}$$

where the $N \times N$ transition matrix $\mathbf{M}_t = \begin{pmatrix} \rho & & 0 \\ & \ddots & \\ 0 & & \rho \end{pmatrix}$ and $\boldsymbol{\delta}_t \sim N(0, \tau^2 \mathbf{R}_\mathbf{s}(\phi))$ that accounts for spatial correlation. For $t = 1, \cdots, T$, define $\mathbf{Y}_t = (Y(\mathbf{s}_1, t), Y(\mathbf{s}_2, t), ..., Y(\mathbf{s}_n, t))'$. Then the model can be written as:

$$\mathbf{Y}_t \sim N(\boldsymbol{\mu}_t + \boldsymbol{\eta}_t, \sigma_\varepsilon^2 I) \tag{2.21}$$

$$\boldsymbol{\eta}_t \sim N(\mathbf{M}_t \boldsymbol{\eta}_{t-1}, \Sigma), \tag{2.22}$$

where $\mathbf{M}_t$ and Eq.2.22 models the dynamics of the spatial process in time and $\Sigma = \tau^2 \mathbf{R}_\mathbf{s}(\phi)$ models the spatial correlation.

### 2.2.1. Spatio-Temporal Kriging

Extending the concept of spatial Kriging in Chapter 2.1.4, we generalize the the same idea in a spatio-temporal context in what follows.

Cressie and Wikle (2015, p.321) illustrated that the assumption of stationarity is for conveniently estimating parameters and computation, however, Kriging doesn't require the assumption of stationarity. Therefore, the spatio-temporal Kriging is developed based on the covariance function in discrete time. With the general form of spatio-temporal model given in the previous section, the simple-Kriging predictor $\eta(\mathbf{s}_0, t_0)$ takes the form of a linear combination of the data (Cressie and Wikle, 2015):

$$\eta(\mathbf{s}_0, t_0) = \sum_{i=1}^{N} \sum_{j=1}^{T} \lambda_{ij} Y(\mathbf{s}_i, t_j) + c, \tag{2.23}$$

where $\lambda$ and $c$ are optimized to minimize the prediction MSE. We also assume the mean function $\mu(\mathbf{s}, t)$ is known in this case. Define $\mathbf{Y}_{tobs}$ and $\boldsymbol{\mu}_{tobs}$ as the observation data and mean value of the spatio-temporal process at observed locations, and $\boldsymbol{\mu}_{tunobs}$ and $\boldsymbol{\eta}_{tunobs}$ as the mean value of the spatio-temporal process and the log precipitation at unobserved locations. Consider the Gaussian assumption for the random spatio-temporal process and the error process in general spatio-temporal. Thus, we summarize the model as:

$$\begin{pmatrix} \boldsymbol{\eta}_{tunobs} \\ \mathbf{Y}_{tobs} \end{pmatrix} \sim N\left( \begin{pmatrix} \boldsymbol{\mu}_{tunobs} \\ \boldsymbol{\mu}_{tobs} \end{pmatrix} + \begin{pmatrix} \mathbf{C}_{0,0} & c_0' \\ c_0 & \mathbf{C}_\eta \end{pmatrix} \right), \tag{2.24}$$

where $\mathbf{C}_\eta = \Sigma_\eta + \Sigma_Y$, $\mathbf{C}_{0,0} = var(\boldsymbol{\eta}_{tunobs})$, $c_0 = cov(\mathbf{Y}_{tobs}, \boldsymbol{\eta}_{tunobs})$. According to Rencher (2002), the posterior distribution is:

$$\boldsymbol{\eta}_{tunobs} | \mathbf{Y}_{tobs} \sim \mathrm{N}\Big( \boldsymbol{\mu}_{tunobs} + c_0' \mathbf{C}_\eta^{-1} (\mathbf{Y}_{tobs} - \boldsymbol{\mu}_{tobs}), \mathbf{C}_{0,0} - c_0' \mathbf{C}_\eta^{-1} c_0 \Big). \tag{2.25}$$

Under the Gaussian assumption, the simple Kriging predictor is the posterior mean and the simple Kriging variance is the posterior variance.

### 2.3. Application of Spatio-Temporal Model to Tree Ring Data

Now we have the basic knowledge about the general form of the spatio-temporal model. Consider the application data of the relationship between tree ring width and log precipitation de-

scribed in Chapter 1. We introduce the spatio-temporal model with same HM structure in this section.

### 2.3.1. Data Model

#### Tree Ring Data Model

We work with tree ring chronology data where we assume the observed tree ring chronology $Y(.,.)$ is a noisy measurement of latent precipitation value $Z(.,.) = \{Z(\mathbf{s},t) : \mathbf{s} \in \mathcal{D}_s, t = 1, \cdots, T\}$ measured at locations $\mathbf{s} \in \mathcal{D}_s$ and time $t$. Then, we assume the observation model:

$$Y(\mathbf{s},t) = \gamma_0 + \gamma_1 Z(\mathbf{s},t) + \varepsilon_y(\mathbf{s},t), \tag{2.26}$$

where $\varepsilon_y(\mathbf{s},t)$ is a white noise process with mean zero and variance $\sigma_y^2 \geq 0$, $\gamma$ is unknown coefficients representing the weight of precipitation impact. For simplicity, in simulation we assume $\gamma_0 = 0$ and $\gamma_1 = 1$. Equivalently, we can write:

$$Y(\mathbf{s},t)|\gamma, Z(\mathbf{s},t), \sigma_y^2 \overset{iid}{\sim} \mathrm{N}(\gamma_0 + \gamma_1 Z(\mathbf{s},t), \sigma_y^2). \tag{2.27}$$

#### Climate Data Model

Because precipitation is also a measured value, the data model for log annual precipitation is:

$$W(\mathbf{s},t) = Z(\mathbf{s},t) + \varepsilon_w(\mathbf{s},t), \tag{2.28}$$

where $Z(\mathbf{s},t)$ stands for the spatio-temporal process of true log precipitation that is Gaussian distributed and $\varepsilon_w(\mathbf{s},t)$ is the white noise process with zero mean and variance $\sigma_w^2$, which is the measurement error. Equivalently, we can write:

$$W(\mathbf{s},t)|Z(\mathbf{s},t), \sigma_w^2 \overset{iid}{\sim} \mathrm{N}(Z(\mathbf{s},t), \sigma_w^2). \tag{2.29}$$

### 2.3.2. Process Model

The independent Gaussian process $Z(.,.)$ is the geostatistical process. Continuing with both the tree ring data model and climate data model introduced in the previous section, we define our process model as

$$Z(\mathbf{s},t) = X(\mathbf{s},t)'\boldsymbol{\beta} + \eta(\mathbf{s},t) \quad \mathbf{s} \in \mathcal{D}, t \in [1,T], \tag{2.30}$$

where for each time step $t$, $\boldsymbol{\beta} = (\beta_0,\beta_1,...,\beta_{p-1})'$ is a $p$-dimensional fixed effect, $X(\mathbf{s},t) = (1,X_1(\mathbf{s},t),...,X_{p-1}(\mathbf{s},t))$ is a known $N \times p$ matrix (p < N) of covariates, containing the constant 1 and $p-1$ columns explanatory covariates potentially observed at location $\mathbf{s}$, and $\eta(\mathbf{s},t)$ is the spatio-temporal random effect at location $\mathbf{s}$ for time step $t$ which is assumed to be an AR(1) time series process that

$$\eta(\mathbf{s},t) \sim \mathrm{N}(\mathbf{M}_t\eta(\mathbf{s},t-1),\Sigma), \tag{2.31}$$

where $\Sigma = \tau^2\mathbf{R}_\mathbf{s}(\phi)$ and $\mathbf{M}_t$ is the diagonal matrix of $\rho$. We use $\mu(\mathbf{s},t) \equiv X(\mathbf{s},t)'\beta$ to represent the mean of the geostatistical spatio-temporal process. Equivalently, we can write:

$$Z(\mathbf{s},t) - \mu(\mathbf{s},t)|\phi,\rho,\tau^2 \overset{iid}{\sim} \mathrm{N}(\mathbf{M}_t(Z(\mathbf{s},t-1) - \mu(\mathbf{s},t-1)),\Sigma). \tag{2.32}$$

Consequently, the simplified spatio-temporal model in Eq.2.26 can be written as:

$$Y(\mathbf{s},t) = \gamma_0 + \gamma_1(\mu(\mathbf{s},t) + \eta(\mathbf{s},t)) + \varepsilon_y(\mathbf{s},t), \tag{2.33}$$

or, equivalent,

$$Y(\mathbf{s},t)|\gamma,\mu(\mathbf{s},t),\eta(\mathbf{s},t),\sigma_y^2 \overset{iid}{\sim} \mathrm{N}(\gamma_0 + \gamma_1\mu(\mathbf{s},t) + \gamma_1\eta(\mathbf{s},t),\sigma_y^2). \tag{2.34}$$

## 2.4. Markov Chain Monte Carlo

After modeling the geospatial process with geostatistical model, it is necessary to estimate parameters of the desired model distribution to make prediction of the random effects at unknown locations. In a Bayesian framework, the posterior distribution of parameter $\theta$ given observations $y$ is

$$[\theta|y] = \frac{[y|\theta][\theta]}{\int_{\theta}[y|\theta][\theta]d\theta} \tag{2.35}$$

where the integral in the denominator of Eq.2.35 is to ensure that the Bayesian posterior distribution is integrated to 1. For some models (for example, when both likelihood and prior are normal distributions), the integral can be solved analytically. Unfortunately, it is intractable to integrate the Bayesian posterior distributions for most target models in a closed form. Alternatively, we approximate the integral using another approach called Markov Chain Monte Carlo (MCMC).

In statistics, MCMC is one of the most commonly used methods to draw samples from posterior distributions and estimate quantities of interest from a Bayesian Hierarchical model. To solve large hierarchical models, it is required to integrate over lots of unknown parameters. In order to approximate the integral numerically, we generate samples from the marginal posterior distribution of each parameter, instead of estimating the probability distribution and calculating the normalizing constant. Recent developments of MCMC have made the computation easier. The joint distribution of the samples from the marginal distributions of each parameter is equivalent to samples from the posterior distribution, up to Monte Carlo error. However, MCMC typically can only approximate the target distribution since there are always some residual effects of starting position. To create a good MCMC chain with the desired distribution, it is also important to determine how many steps that are needed to converge to the stationary distribution from initial position within an acceptable error. Typically, we run MCMC for a large number of iterations and obtain samples after a ”burn-in” period when observe the chain, by reason of that samples at early iterations may not characterize the target posterior distribution, that are often discarded

(Gilks et al., 1995b). A poor starting value will cause the increase of burn-in steps. After the Markov chain converges to the equilibrium distribution, the distribution estimated from samples gets close to the actual desired distribution as the number of iterations increase. Therefore, when we estimate parameters using MCMC, we normally draw samples from $K$ iterations of the marginal posterior and use mean value for easy representation, because if we choose a large $K$, the mean of samples gets close to the posterior mean.

### 2.4.1. Metropolis Algorithm

A general framework to generate a Markov chain is the Metropolis algorithm. Define the parameter of interest as $\theta$. The Metropolis algorithm uses a proposal distribution to generate a new value $\theta^*$ for the Markov chain based on the value of the parameter $\theta$ at the current iteration and computes a probability for accepting the proposed move. The proposal distribution can either be symmetric (Metropolis algorithm) or asymmetric (Metropolis-Hasting algorithm).

#### Metropolis Algorithm

The Metropolis algorithm was proposed by Metropolis et al. (1953). The proposal distribution of the Metropolis algorithm can be independent or dependent of the parameter at current state of the Markov chain. The most commonly used proposal is the random walk, which is dependent on the current state of Markov chain. The random walk proposal distribution $[\theta^*|\theta^{(k-1)}]$, is generated based on the value of $\theta$ at the latest iteration $k-1$, where $\theta^* \sim N(\theta^{(k-1)}, \sigma^2_{tune})$, with the mean is the value of the parameter at previous iteration $\theta^{(k-1)}$ and the variance is $\sigma^2_{tune}$, a tuning parameter that is chosen to adjust the performance of the MCMC samples. The tuning parameter has the property that a small $\sigma^2_{tune}$ leads to a relatively close value to the previous accepted value, meaning the acceptance rate of the Metropolis sampler increases, while choosing a larger $\sigma^2_{tune}$ results in decrease of the acceptance rate as it makes bigger transitions for the Markov chain. Hence, one should note that it is important to balance generating accepted proposals and exploring the range of the posterior. Another distinct property of the Metropolis proposal is symmetry:

if $[\theta^*|\theta^{(k-1)}] = [\theta^{(k-1)}|\theta^*]$, the conditional probability of $\theta^*$ given $\theta^{(k-1)}$ and the conditional probability of $\theta^{(k-1)}$ given $\theta^*$ are equal. The Metropolis acceptance rate is

$$\alpha(\theta^*, \theta^{(k-1)}) = min(1, \frac{[\mathbf{y}|\theta^*][\theta^*]}{[\mathbf{y}|\theta^{(k-1)}][\theta^{(k-1)}]}). \tag{2.36}$$

**Metropolis-Hastings Algorithm**

Metropolis-Hastings was implemented by Hastings (1970) on more general cases based on the Metropolis algorithm. The most important property of a Metropolis-Hasting proposal distribution is asymmetry with the most compelling reason that it can guarantee that the proposed value of the parameter is in the support of the parameter range. Say, for a variance parameter, a symmetric normal proposal may generate negative proposals that must be rejected, resulting in inefficient MCMC samplers with low acceptance rate. Instead, we can use an asymmetric proposal to guarantee a positive proposal leading to higher acceptance rate. With an asymmetric proposal $[\theta^*|\theta^{(k-1)}]$, the Metropolis-Hastings acceptance rate is

$$\alpha(\theta^*, \theta^{(k-1)}) = min(1, \frac{[\mathbf{y}|\theta^*][\theta^*]}{[\mathbf{y}|\theta^{(k-1)}][\theta^{(k-1)}]} \frac{[\theta^{(k-1)}|\theta^*]}{[\theta^*|\theta^{(k-1)}]}). \tag{2.37}$$

where $\frac{[\theta^{(k-1)}|\theta^*]}{[\theta^*|\theta^{(k-1)}]}$ is a correction factor for the asymmetric proposal. Specifically, the Metropolis algorithm can be treated as a special case of Metropolis-Hastings algorithm with the factor $\frac{[\theta^{(k-1)}|\theta^*]}{[\theta^*|\theta^{(k-1)}]}$ equal to 1. By comparing $u$, which is drawn from $uniform(0, 1)$, with $\alpha(\theta^*, \theta^{(k-1)})$, we use this rate to decide whether the proposed value will be accepted as the new state or not (Yildirim, 2012).

### 2.4.2. Gibbs Sampling

The sampling method used in this research is called Gibbs sampling, a MCMC algorithm that was described by Geman and Geman (1987). Gibbs sampling generates a Markov chain, where each sample of the chain is correlated with the latest value of other samples. Gibbs sampling is

commonly used in Bayesian inference, and it allows all full conditional distributions of parameters of the target distribution to be sampled exactly.

Gibbs sampling is considered a special case of the Metropolis Hastings algorithm without the requirement of any "tuning". Specifically, the proposal distribution generated for Gibbs sampling always has a Metropolis-Hastings acceptance rate of 1, meaning that the proposal is always accepted. A Markov chain generated with Gibbs sampling is stationary. The Metropolis-Hastings algorithm works for the same reasons. Therefore, Gibbs sampling is frequently used in combination with other MCMC algorithms such as Metropolis-Hastings algorithm and rejection sampling to draw from the posterior distributions to accomplish sampling procedures (Gilks et al., 1995a). The basic idea of Gibbs sampling is to split the multidimensional parameter $\boldsymbol{\theta}$ into $m$ blocks and sample each block in turn, conditional on the most recent values of other blocks. The advantage of Gibbs sampling is that it breaks a multidimensional parameter into several parameter blocks simplify a complicated problem. Algorithm 1 describes an iteration of Gibbs sampling.

**Algorithm 1** Gibbs Sampling

Initialize $\boldsymbol{\theta}^{(1)} \sim q(\theta)$
**For** iteration $k = 2, ..., K$ **do**
$\boldsymbol{\theta}_1^{(k)} \sim [\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2^{(k-1)}, \boldsymbol{\theta}_3^{(k-1)}, ..., \boldsymbol{\theta}_m^{(k-1)}]$
$\boldsymbol{\theta}_2^{(k)} \sim [\boldsymbol{\theta}_2 | \boldsymbol{\theta}_1^{(k)}, \boldsymbol{\theta}_3^{(k-1)}, ..., \boldsymbol{\theta}_m^{(k-1)}]$
$\vdots$
$\boldsymbol{\theta}_m^{(k)} \sim [\boldsymbol{\theta}_m | \boldsymbol{\theta}_1^{(k)}, \boldsymbol{\theta}_3^{(k)}, ..., \boldsymbol{\theta}_{m-1}^{(k)}]$
**return** $\boldsymbol{\theta}_1^{(1:K)}, \boldsymbol{\theta}_2^{(1:K)}, \cdots, \boldsymbol{\theta}_m^{(1:K)}$

In Algorithm 1, the distribution $[\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2^{(k-1)}, \boldsymbol{\theta}_3^{(k-1)}, \cdots, \boldsymbol{\theta}_m^{(k-1)}]$ is known as the full conditional distribution of $\boldsymbol{\theta}_1$. The multidimensional parameter $\boldsymbol{\theta}$ to be sampled consists of $m$ blocks $\theta_1, \theta_2, \cdots, \theta_m$. For the first iteration, we initialize the $m$-dimensional parameter $\boldsymbol{\theta}^{(1)}$ by assigning each block with an initial distribution, and draw samples from these distributions. The initial values of the parameter blocks are often assigned randomly, but one can determine these value using other algorithms such as EM. From the second iteration onward, new samples of each parameter are drawn from the conditional posterior distribution of each parameter block based on the current values of all

other parameter blocks. We draw samples for all $m$ blocks at each iteration to produce one new $m$-dimensional sample $\boldsymbol{\theta}^{(k)}$. The above process need to be repeated $K$ iterations. Typically, the whole MCMC algorithm must run for a large number of iterations so that the stationary distribution generated under the above algorithm will converge to the target posterior distribution.

Because the samples of all parameters obtained from the Gibbs sampler are from probability distributions, we can calculate any quantity of interest directly. For example, the estimate for the mean of the parameter is simply the mean of the MCMC samples.

The BUGS model language is now widely used in Bayesian statistics as it has simpler code structure and easier to write and understand than a Gibbs sampler. BUGS is short for Bayesian inference using Gibbs sampling algorithm. It is a project that developed software to deal with Bayesian analysis of complex statistical models using MCMC methods. The R package NIMBLE (NIMBLE Development Team, 2019) that uses the BUGS model language as programmable objects is used in this research. Similar to BUGS, to implement Bayesian modeling and inference, Stan project developed a programming language using the No-U-Turn sampler (NUTS) to obtain posterior simulations given a user-defined model and data. The R package Rstan (Stan Development Team, 2018) used in this research, allows us to conveniently fit Stan models with R.

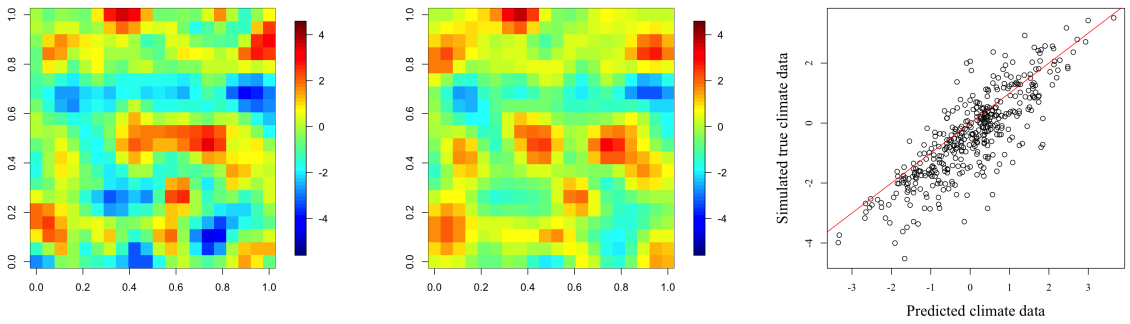
## 3. Simulation

### 3.1. Spatial Model

To simplify the tree ring chronology data example, we simulate the spatial model Eq.2.9 proposed in Chapter 2.1. We can write the integrated model as (integrating out the latent Gaussian process $\boldsymbol{\eta}$):

$$\mathbf{Y} \sim \mathrm{N}(\mathbf{0}, \sigma_y^2 \mathbf{I} + \tau^2 \mathbf{R}_{ij}(\phi)). \tag{3.1}$$

With the choice of prior distribution that: $\phi \sim \mathrm{N}(0,1)$, $\tau \sim \mathrm{N}(0,2)$, $\sigma_y \sim \mathrm{N}(0,1)$ are truncated to positive support and $\beta \sim N(0,5)$, we use RStan (Stan Development Team, 2018) to fit with the simulation data.




(a) Simulated geospatial climate process

(b) Spatial Kriging using spatial model

(c) Simulated climate data compared with predicted climate data

Figure 3.1: Compare simulated geospatial climate process with Kriging estimates in space. In (c), dots represent predicted sample data, red line means that simulation is exactly same as prediction.

Derived from Figure 3.1, a geospatial climate process was simulated over an area of $20 \times 20$ sites, then we randomly selected 100 sites as samples. Figure 3.1c shows the relationship between simulated climate data and predicted climate data, the red line represents that the simulation is same as prediction, in this figure, most data points lie along with the red line, which leads to the result shown in Figure 3.1b. We notice that the Kriging prediction in Figure 3.1b shows the same pat-

tern as the simulated climate data (Figure 3.1a), thus, our spatial model overall performs well in predictions.

### 3.2. Spatio-Temporal Model

To model tree ring width and climate data with the full spatio-temporal model, we simulate with the spatio-temporal model Eq.2.33 proposed in Chapter 2.3, with the assumption that $\gamma_0 = 0$, $\gamma_1 = 1$ and the mean value of the spatio-temporal process is 0, we can write the model as:

$$Y(\mathbf{s},t) \sim \mathrm{N}(\eta(\mathbf{s},t),\sigma_y^2) \tag{3.2}$$

$$\eta(\mathbf{s},t) \sim \mathrm{N}(\rho\eta(\mathbf{s},t-1),\tau^2\mathbf{R}_\mathbf{s}(\phi)) \tag{3.3}$$

The Posterior distribution is given by:

$$[\eta,\rho,\sigma_y^2,\phi,\tau^2|\mathbf{Y}_t] \propto \prod_{t=1}^{T}[\mathbf{Y}_t|\eta,\rho,\sigma_y^2,\phi,\tau^2][\eta][\sigma_y^2][\phi][\tau^2][\rho] \tag{3.4}$$

With the choice of prior distribution that: $\phi \sim \mathrm{Unif}(\phi_L,\phi_U)$, $\tau^2 \sim \mathrm{InverseGamma}(\alpha_\tau^2,\beta_\tau^2)$, $\sigma_y^2 \sim \mathrm{InverseGamma}(\alpha_\sigma^2,\beta_\sigma^2)$, $\rho \sim \mathrm{Unif}(\rho_L,\rho_U)$, the full conditional distributions for all parameters are calculated in the Appendix.

For the simulation study, 300 years of the tree ring chronologies over an area of $30\times30$ sites were simulated then a subset of 36 sites were randomly selected as samples. In this simulation, $\boldsymbol{\eta}$ is considered as initial simulated climate data over 300 years which is the spatio-temporal AR process. I assumed that the first 200 years of climate data are unknown which can represent that the climate data was not observed over this period. $\mathbf{Y}$ is simulated tree ring chronologies including simulated climate data and white noise over 300 years. Therefore, the goal is to predict the previous unobserved climate based on the spatio-temporal model and the current data (Figure 3.2, Figure 3.3).

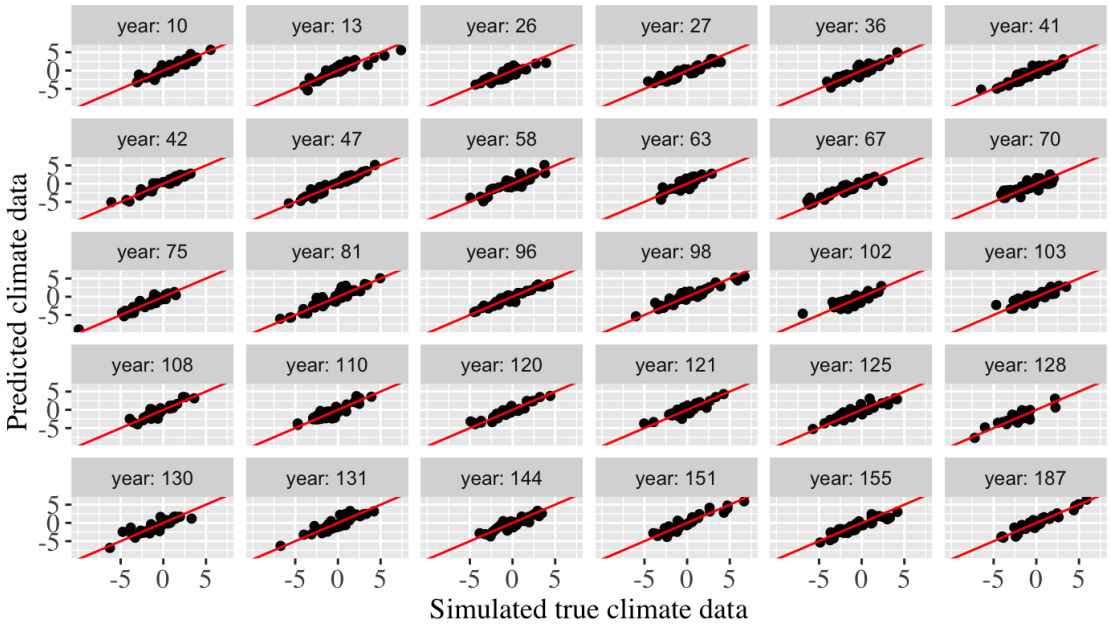



Figure 3.2: 30 years are randomly selected out of the first 200 years to show the relationship between predicted and true climate data of simulation. Dots are climate data and lines represent that the prediction is same as truth.

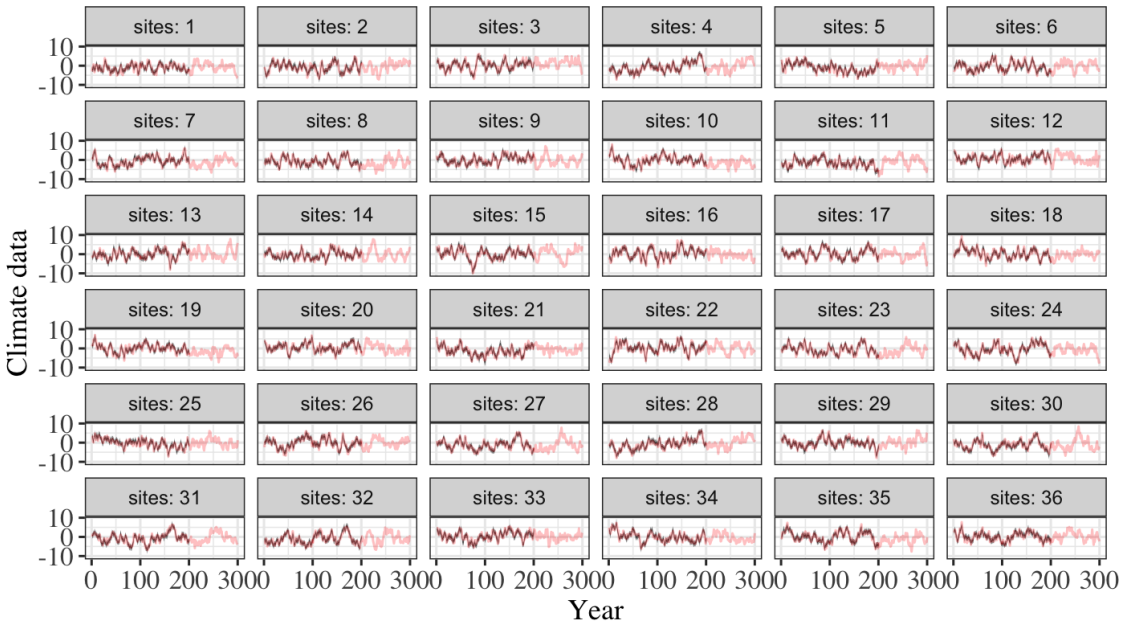



Figure 3.3: Predicted climate data of the first 200 years for each of 36 sites. Black ribbons with dark and light shading represent the predicted 50% and 95% Bayesian credible intervals respectively and red lines are true simulation data.

Figure 3.2 evaluates the fit of our spatio-temporal model. In Figure 3.2, the x-axis is the true simulation data and y-axis is the predicted data, a subset of 30 years are randomly selected to show the relationship between true climate data and predicted climate data. The lines represent that the prediction climate is exactly same as true climate. As shown in Figure 3.2, all dots lie along with the line for all 30 selected years, therefore, our model fits the simulation data well.

Based on the simulation climate data of the last 100 years, we use the proposed spatio-temporal model to predict the climate data for the first 200 years for all 36 sites (Figure 3.3). In Figure 3.3, each facet represents the climate change over 300 years at one measurement site. Black ribbons represent the predicted climate distribution, with darker shading representing 50% Bayesian credible intervals (BCIs) and lighter shading representing 95% Bayesian credible intervals. Red lines are true simulated climate data. As shown in the figure, the predicted climate change captures the pattern of true simulation data, most of predicted simulation data falls within the 95% Bayesian credible intervals.

Based on the simulation results, the model performs well to predict unobserved climate data over the first 200 years for all 36 measurement sites. Therefore, the next step is to use the developed spatio-temporal model with real tree ring chronology and precipitation data to predict climate.

## 4. Results

Because of the high goodness of fit of the proposed spatio-temporal model, I fit the model with the real tree ring chronology and log precipitation. The goal is to predict log precipitation data for years that climate information is unobserved. Figure 4.1 and Figure 4.2 show the model fitting results.

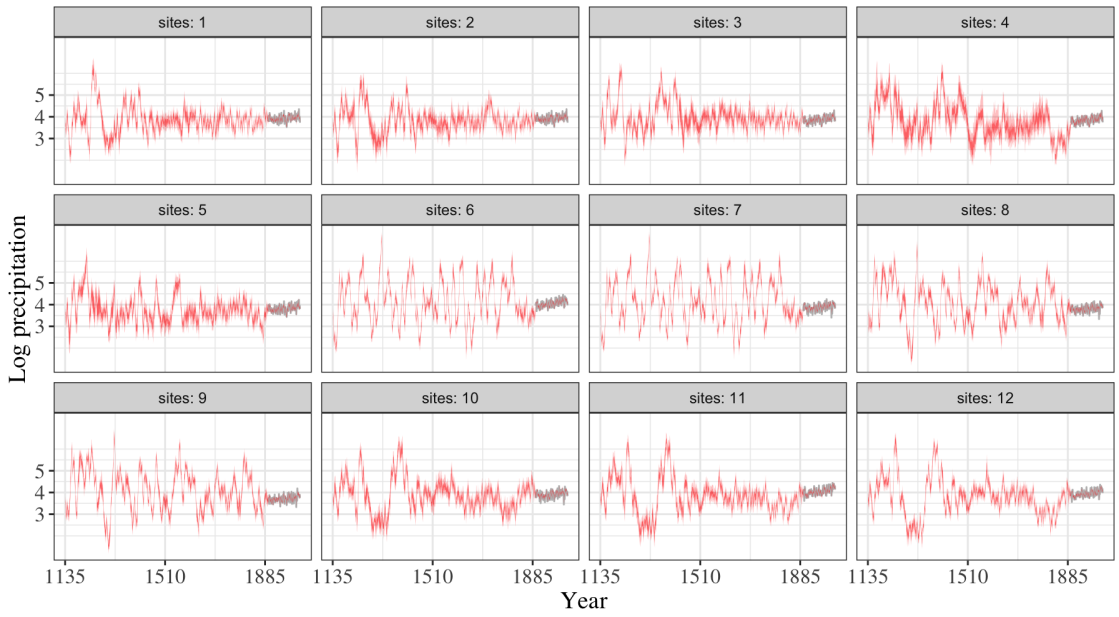



Figure 4.1: 12 measurement sites are selected to show predicted log precipitation results. In each facet, gray line is the log precipitation data observed from year 1895-2015 and red ribbons with dark and light shading are 95% and 50% Bayesian credible intervals of predicted log precipitation distribution for year 1135-1894.

In Figure 4.1, 12 measurement sites are selected to show predicted log precipitation where each facet represents a site. Gray lines are observed log precipitation at each site only available for the last 121 years (1895-2015). We predict unobserved log precipitation data for the first 760 years (1135-1894), red ribbons indicate the predicted log precipitation distribution, with dark shading representing 50% BCIs and lighter shading representing 95% BCIs, which means there is 95% posterior probability that the log precipitation falls within the light shading area and 50% posterior probability that the log precipitation values lie within darker area. Log precipitation data varies a lot for the first 350 years and becomes relatively stable at some sites. We can also notice

that for some sites the credible intervals are wider than others, likely because of the lack of tree ring width data.

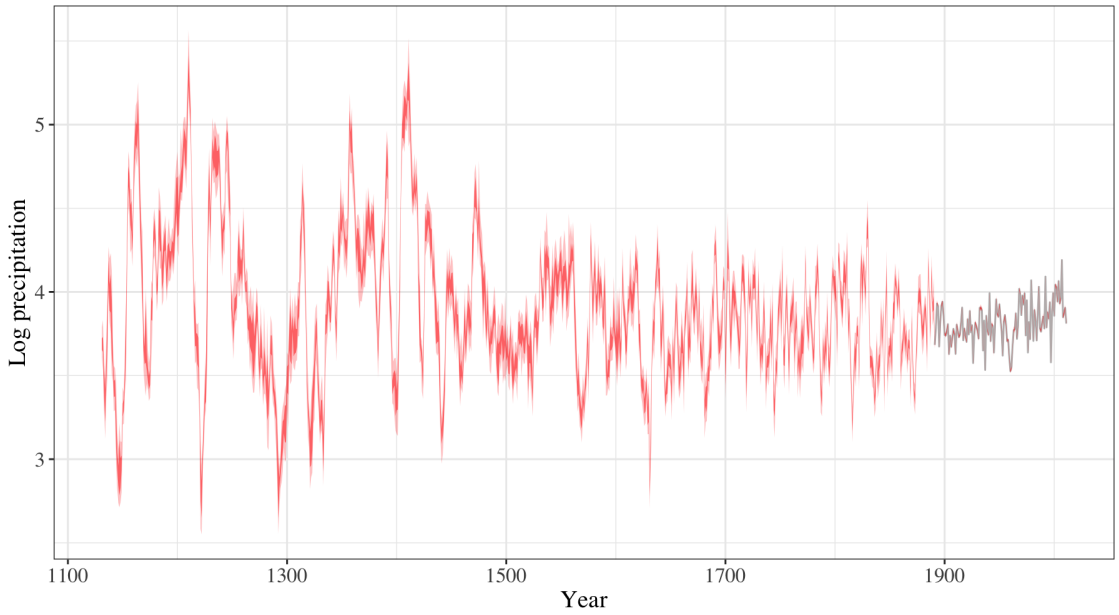



Figure 4.2: Predicted log precipitation averaged over spatial locations for years 1135-1894. Gray line is observed average log precipitation of all sites for year 1895-2015, red ribbons show predictive distribution with dark shading representing 50% BCI and light shading representing 95% BCI.

We plot the log precipitation averaged over space for years 1135- 1894 in Figure 4.2. The gray line is the observed average log precipitation data of all measurement sites from year 1895 to 2015 and red ribbons are 50% and 95% credible intervals of predicted log precipitation. The average log precipitation shows the same pattern as the predicted log precipitation for each site. For the first 350 years, precipitation has a strong variation but becomes relatively stable after year 1500. Overall, the spatial average of log precipitation over recorded years is stable around 4.

To validate the model fitting result on real tree ring data, we fit the developed spatio-temporal model with observed log precipitation data from year 1895 to year 2015. Figure 4.3 and Figure 4.4 show the validation results.

In Figure 4.3, we also selected 12 measurement sites to compare observed log precipitation data with predicted credible intervals. Each facet represents a site. Black lines are observed log pre-

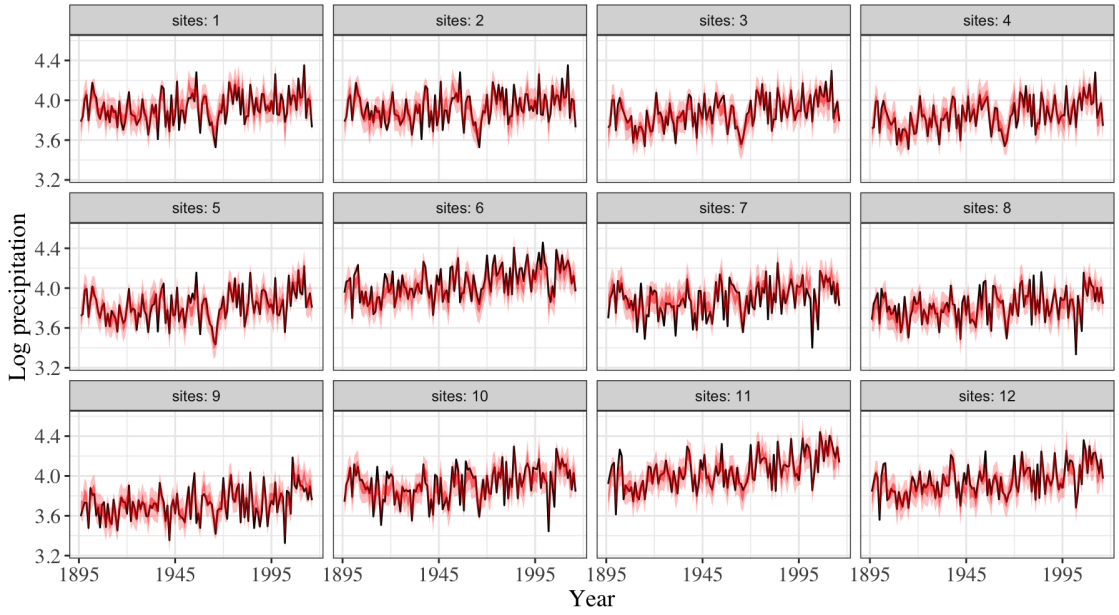



Figure 4.3: 12 measurement sites are selected to compare observed and predicted log precipitation. In each facet, black line is the log precipitation data observed from year 1895-2015 and red ribbons with dark and light shading are 95% and 50% Bayesian credible intervals of predicted log precipitation distribution over the same period.

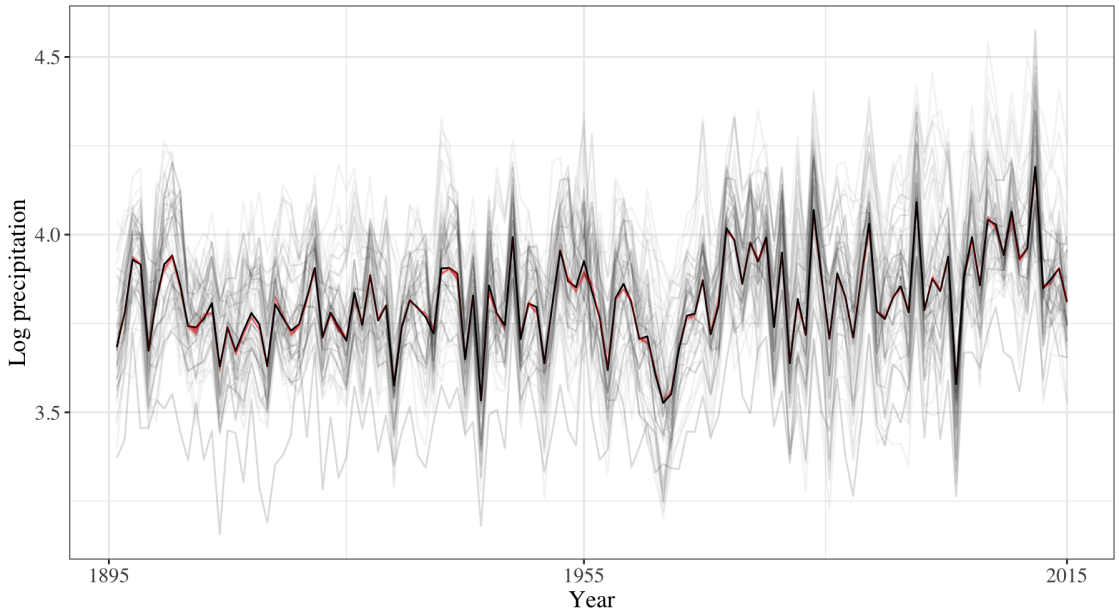



Figure 4.4: Predicted average log precipitation data over space for year 1895-2015. Gray lines are observed log precipitation at each site, black line represents the predicted average value of log precipitation of all sites. Red ribbons show predictive distribution with dark and light shadings representing 50% BCI and 95% BCI.

cipitation data at each site for the last 121 years, red ribbons indicate the predicted log precipitation distribution, with dark shading representing 50% BCIs and light shading representing 95% BCIs. As shown in the figure, the predicted log precipitation distribution capture the trend of observed data with only a few data points falls outside intervals, which indicates that the proposed spatio-model performs well on real tree ring width data.

In Figure 4.4, we plot the average of log precipitation data over space for year 1895-2015. Gray lines represent observed log precipitation grouped by each measurement site and black line is the observed average log precipitation data of all measurement sites. Red ribbons are 50% and 95% Bayesian credible intervals of predicted log precipitation. The BCIs for average value are not as wide as for each site that almost all observed log precipitation data lie within the interval, which confirms the good performance of model fitting result.

Based on predicting and validating results, the proposed Spatio-temporal model does well at predicting climate when climate data is available, but behaves not ideal when only tree ring data is available and does worse at some sites that lack of historical tree ring width. Thus, our tree ring data model is a poor fit and the results are unreliable.

## 5. Conclusion

Using tree ring chronology and annual log precipitation, the objective of this study was to develope a spatio-temporal model that aims to predict both local and regional features of unobserved log precipitation. The simulation studies demonstrate that the model is capable of accurately estimating the unobserved climate, however, the application of the model to the real data failed to produce reasonable predictions. Based on the model fitting results, reasons that cause the failure as we considered might be the improperly pre-processed original data, or the precipitation may have different weights at different locations. Thus, there is more work needed to develop an appropriate model to make us better understand historical climate change in practical point of view.

In the New England area, we predicted unobserved log precipitation at observed spatial locations for year 1135-1895. It is meaningful to use the spatio-temporal Kriging method to predict log precipitation at unobserved locations as shown in simulation. There are some challenges to face when develop the model, such as dealing with missing data. We treated log precipitation as a Gaussian process of spatial random effect, typically in a Gaussian process, we only have a small number of observations, where the goal is to estimate the latent response. A weakness of Gaussian process modeling is that the computation does not scale well with the sample sizes. Therefore, to assign a reasonable initial value to the missing log precipitation is important. What's more, the missing tree ring width data at some spatial locations were estimated from our spatio-temporal model, which caused larger variance of predicted log precipitation.

Our moisture sensitive tree ring width is influenced by mainly precipitation, that has spatial pattern over a long time period. Spatio-temporal analysis has more benefits than pure spatial or time-series analysis because it allows us to simultaneously borrow strength across space and time which gives us a more comprehensive perspective. However, there are some weakness of our model when reconstruct the paleoclimate statistically based on tree ring width. One is that with the developed spatio-temporal model and real data, although it is easier to understand the model structure, the computational complexity is large. What's more, even though the model works well

with simulation data, it is unable for us to do model validation using real data, because of the lack of historical precipitation. Based on the result, it is only confident to predict the precipitation within a short time period. Moreover, both precipitation and tree ring data are normally acquired in discrete time while climate performs a continuous influence on the tree ring growth (Tipton et al., 2016). The tree-ring width represents the integrated impact of the climate condition to the tree over a period, whereas the model we developed considers precipitation and tree ring data on a yearly time step.

For simplicity, we only model the relationship between the tree ring width and log precipitation. Tree growth is normally affected by the combined action of several factors besides precipitation, such as temperature and species. Therefore, although there is an unknown parameter representing the influence of precipitation on tree ring growth in our spatio-temporal model, it is not very efficient to estimate climate data accurately with our current model. In order to develop a spatio-temporal model better describing how climate influence tree growth, it might be necessary to consider a multivariate model, involving precipitation as well as temperature and tree species in the future work. In addition, there are spatial correlations of climate between different locations within a certain area, in the future work, applying dimension reduction theory to choose several meaningful locations that best represent regional climate feature, instead of model with original large dataset, is useful to improve the computational complexity.

## Bibliography


Al-Hasnawi, S. S., AlMaliki, J. S. B., and Nazal, Z. F. (2017). Distribution modeling of hazardous airborne emissions from industrial campuses in Iraq via GIS techniques. 227(1):12–55.

Arab, A. (2015). Spatial and spatio-temporal models for modeling epidemiological data with excess zeros. *International Journal of Environmental Research and Public Health*, 12(9):10536–10548.

Cressie, N. and Wikle, C. K. (2015). *Statistics for Spatio-temporal Data*. John Wiley & Sons.

Diggle, P. J., Menezes, R., and Su, T.-l. (2010). Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(2):191–232.

Fritts, H. (1976). *Tree Rings and Climate*. Academic Press.

Geman, S. and Geman, D. (1987). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. In *Readings in Computer Vision*, pages 564–584.

Genton, M. G. (2007). Separable approximations of space-time covariance matrices. *Environmetrics: The official journal of the International Environmetrics Society*, 18(7):681–695.

Gilks, W. R., Best, N., and Tan, K. (1995a). Adaptive rejection Metropolis sampling within Gibbs sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 44(4):455–472.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. (1995b). *Markov chain Monte Carlo in practice*. Chapman and Hall/CRC.

Group, P. C. (2011). Prism climate data. *Oregon State University*.

Guiot, J. (1982). Response functions. In *Climate from Tree Rings*, pages 2–7. Cambridge University Press, Cambridge, UK.

Hastings, W. K. (1970). Monte carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.

Hoeting, J. A. (2009). The importance of accounting for spatial and temporal correlation in analyses of ecological data. *Ecological Applications*, 19(3):574–577.

Matheron, G. (1971). The theory of regionalized variables and its applications. *Les Cahiers du Centre de Morphologie Mathématique*, 5:212.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092.

NIMBLE Development Team (2019). NIMBLE: MCMC, particle filtering, and programmable hierarchical modeling. R package version 0.8.0.

Rencher, A. C. (2002). *Methods of Multivariate Analysis*. John Wiley & Sons.

Robinson, G. K. (1991). That BLUP is a good thing: the estimation of random effects. *Statistical Science*, 6(1):15–32.

Schweingruber, F. H. (1996). *Tree Rings and Environment: Dendroecology.* Paul Haupt AG Bern.

Searle, S. R., Casella, G., and McCulloch, C. E. (2009). *Variance components*, volume 391. John Wiley & Sons.

Sen, A., Gümüsay, M., Kavas, A., and Bulucu, U. (2008). Programming an artificial neural network tool for spatial interpolation in GIS-A case study for indoor radio wave propagation of WLAN. *Sensors*, 8(9):5996–6014.

Song, H.-R., Fuentes, M., and Ghosh, S. (2008). A comparative study of Gaussian geostatistical models and Gaussian Markov random field models. *Journal of Multivariate Analysis*, 99(8):1681–1697.

Stan Development Team (2018). RStan: the R interface to Stan. R package version 2.18.2.

Tessier, L. (1989). Spatio-temporal analysis of climate–tree ring relationships. *New Phytologist*, 111(3):517–529.

Tipton, J., Hooten, M., Pederson, N., Tingley, M., and Bishop, D. (2016). Reconstruction of late holocene climate based on tree growth and mechanistic hierarchical models. *Environmetrics*, 27(1):42–54.

Wikle, C. K. (2015). Modern perspectives on statistics for spatio-temporal data. *Wiley Interdisciplinary Reviews: Computational Statistics*, 7(1):86–98.

Williams, C. K. (1998). Prediction with Gaussian processes: From linear regression to linear prediction and beyond. In *Learning in Graphical Models*, pages 599–621.

Yildirim, I. (2012). Bayesian inference: Metropolis-hastings sampling. *Dept. of Brain and Cognitive Sciences, Univ. of Rochester, Rochester, NY*.

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465):250–261.

Zhao, Y., Lee, A. H., and Barnes, T. (2010). On application of the empirical Bayes shrinkage in epidemiological settings. *International Journal of Environmental Research and Public Health*, 7(2):380–394.

## A. Appendix

Full conditional for $\rho$:

$$\begin{aligned}[\rho|\cdot] &\propto \prod_{t=1}^{T}[\mathbf{Y}_t|\sigma^2,\phi,\tau^2,\rho][\rho] \\ &\propto \prod_{t=1}^{T}\mathrm{N}(\mathbf{Y}_t|\mathbf{0},\tau^2\mathbf{R}_\mathbf{s}(\rho)+\sigma^2\mathbf{I})\,\mathrm{Uniform}(\rho|\rho_L,\rho_U)\end{aligned} \tag{A.1}$$

which can be sampled using Metropolis-Hastings.

Full conditional for $\phi$:

$$\begin{aligned}[\phi|\cdot] &\propto \prod_{t=1}^{T}[\mathbf{Y}_t|\sigma^2,\phi,\tau^2][\phi] \\ &\propto \prod_{t=1}^{T}\mathrm{N}(\mathbf{Y}_t|\mathbf{0},\tau^2\mathbf{R}_\mathbf{s}(\phi)+\sigma^2\mathbf{I})\,\mathrm{Uniform}(\phi|\phi_L,\phi_U)\end{aligned} \tag{A.2}$$

which can be sampled using Metropolis-Hastings.

Full conditional for $\sigma^2$:

$$\begin{aligned}\left[\sigma^2|\cdot\right] &\propto \prod_{t=1}^{T}[\mathbf{Y}_t|\sigma^2,\phi,\tau^2][\sigma^2] \\ &\propto \prod_{t=1}^{T}\mathrm{N}(\mathbf{Y}_t|\mathbf{0},\tau^2\mathbf{R}_\mathbf{s}(\phi)+\sigma^2\mathbf{I})\,\mathrm{Inverse\text{-}gamma}(\sigma^2|\alpha_\sigma^2,\beta_\sigma^2)\end{aligned} \tag{A.3}$$

which can be sampled using Metropolis-Hastings.

Full conditional for $\tau^2$:

$$\begin{aligned}\left[\tau^2,|\cdot\right] &\propto \prod_{t=1}^{T}[\mathbf{Y}_t|\sigma^2,\phi,\tau^2][\tau^2] \\ &\propto \prod_{t=1}^{T}\mathrm{N}(\mathbf{Y}_t|\mathbf{0},\tau^2\mathbf{R}(\phi)+\sigma^2\mathbf{I})\,\mathrm{Inverse\text{-}gamma}(\tau^2|\alpha_\tau^2,\beta_\tau^2)\end{aligned} \tag{A.4}$$

which can be sampled using Metropolis-Hastings.