Estimating Decadal Climate Variability Effects on Crop Yields: A Bayesian Hierarchical Approach

Pei Huang
Ph.D. Candidate, Department of Agricultural Economics
Texas A&M University
Email: peihuang@tamu.edu

Bruce A. McCarl
Professor, Department of Agricultural Economics
Texas A&M University
Email: mccarl@tamu.edu


Copyright 2014 by Pei Huang, and Bruce A. McCarl. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies. This work was funded by a USDA NSF project.
Abstract: Longer run ocean phenomena can influence climate and crop yields. This paper considers the effects of decadal climate variability (DCV) phase combinations on crop yields in the Missouri River Basin (MRB). The study uses a hierarchical linear mixed-effects model (LMM), in which parameters are assumed to follow a specific distribution. The LMM are estimated with Bayesian methods, which allows us to examine heterogeneity in DCV effects across counties and prevent extreme estimates for counties with small number of observations. The model considers a higher order moment, skewness, in the distribution assumption. Immediate findings are that the DCV phase combinations have heterogeneous impacts on yields across both counties and crops. This has two potentially important implications for relevant stakeholders. First, the mechanism of determining crop insurance may wish to take into account the DCV effects on crop yields. Second, information on DCV effects may be distributed to farmers, providing insights into adaptation to DCV.
1 Introduction

The growth of crops relies highly on the climate and many empirical studies have found significant effects of climate on crop yields (Mendelsohn et al. 1994; Chen and McCarl 2000; Chen et al. 2004; Schlenker et al. 2006; McCarl et al. 2008; Schlenker and Roberts 2009; Yu and Babcock 2010; Tack et al. 2012). These studies have addressed both climate change and the effects of ocean phenomena like El Niño Southern Oscillation (ENSO). This paper studies the effects of a longer term ocean phenomena, decadal climate variability, on crop yield outcomes in the context of a major U.S. river basin.

Ocean related climate variability is a natural appearance that results in systematic differences in climate patterns over time. The impacts of such phenomena on climate include fluctuating temperature, precipitations, oceanic current, and extreme events. The ENSO and the North Atlantic Oscillation (NAO) are well-known cases of this, and their climatic and societal impacts have been widely studied regionally and globally (Adams et al. 1995; Solow et al. 1998; Torrence and Webster 1999; Chen et al. 2005; Kim and McCarl 2005).

Compared to the ENSO and the NAO, another category of ocean related climate variability is the called decadal climate variability (DCV) which lasts a longer period of time (Hurrell et al. 2010). There are multiple DCV phenomena, which manifest varying impacts on climate patterns. Several of these phenomena have been the subject of climate and yield studies: the Pacific decadal oscillation (PDO) (Mantua et al. 1997; Mantua and Hare 2002), the tropical Atlantic sea-surface temperature gradient (TAG) (Mehta 1998; Rajagopalan et al. 1998), and the west Pacific warm pool (WPWP) (Wang and Mehta 2008). More and more evidence show that climate and crop yield anomalies or variations are associated with the DCV (McCabe et al. 2004; McCabe et al. 2008).
Since the DCV phenomena impact climate conditions, there should be implications of linkages between DCV phenomena and agricultural productivity. Mehta et al. (2011) reveal that the three DCV phenomena, PDO, TAG, and WPWP, significantly influence water yields in the Missouri River Basin (MRB). In the follow-up study (2012), they conduct a crop simulation study further suggesting that crop yields in this area are widely impacted mainly due to precipitation variations across different phases of the DCV phenomena; and the effects show significant variation across locations.

In this study, we examine the effects of DCV phenomena on crop yields using observed crop yield data. The DCV effects are decomposed into two parts: indirect and direct effects. The former depicts the DCV effects on climate and in turn on crop yields. The later part describes the effects that are unobserved or exerted on crops directly. Although unobserved effects are incorporated in the part, we use the term “direct effects” as they are directly estimated from the crop yield equation.

The essential questions addressed in this paper: can we observe an effect of DCV phenomena on crop yields using historical data and does this effect vary across counties? To do this, we will use the hierarchical linear mixed-effects model (LMM) introduced by Laird and Ware (1982). This model has become a widely used tool for analyzing data with panel data. It is able to express both within-group variations in one level and between-group heterogeneity in another level (Hoff 2009).

A key assumption for the hierarchical model is the distributions used to represent the within-group and between-group variations. Most hierarchical LMM frequently use a normal distribution. The normality assumption makes the model easy to implement and esti-
mate. The assumption, however, has been challenged by many studies (Verbeke and Lesaffre 1997; Zhang and Davidian 2001; Rosa et al. 2003; Ghidey et al. 2004; Jara et al. 2008; Lachos et al. 2009). A model with normal distribution may generate inaccurate estimates when data show non-normal patterns.

In the crop yield equation, we will assume a more flexible distribution that allows skewness to be considered since its presence in crop yield distributions has been found in the literature (Ramirez et al. 2003; Sherrick et al. 2004; Hennessy 2009; Tack et al. 2012). In particular, we employ a class of skew-normal distribution proposed by Sahu et al. (2003), which admits both positive and negative skewness. It is worth noting that imposing the skew-normal distribution allows for a more general distributional form, even if the skewness is not an issue for a particular dataset (Balcombe et al. 2011).

In this study, we implement Bayesian approaches to estimate the system of equations in the model. There are a number of reasons why Bayesian methods are ideal for estimating the LMM (Layton and Levine 2003; Leon-Gonzalez and Scarpa 2008; Balcombe et al. 2009; Balcombe et al. 2011). First, the Bayesian approach allows us to obtain varying parameters in a model, in which the heterogeneity is considered. Second, the Bayesian approach can make good estimations with small data size. Third, the Bayesian methods update the posterior distribution after observing the data for each parameter of interest, providing more information to make inference than classic methods. Finally, with a large data set, Bayesian estimates asymptotically converge to maximum likelihood estimates.

The sampling results for the estimated parameters are used to estimate the marginal total effects of DCV on yields of eight dryland crops in the MRB and predict posterior yield distributions.
2 Background

2.1 DCV phenomena and phase combinations

There are multiple ocean-atmospheric DCV phenomena. The ones examined in this study include: the Pacific decadal oscillation (PDO), the tropical Atlantic gradient (TAG), and the west Pacific warm pool (WPWP). Each phenomenon is classified as exhibiting either a positive (+) or a negative (−) phase based on sea surface temperatures (SST). Across these three, there are eight total possible combinations of the positive and negative phases of the three DCV phenomena. Each DCV phenomenon will be discussed below.

The Pacific decadal oscillation

The PDO is a pattern of climate variability similar to El Niño phenomenon but with longer time persistence. It is defined by SST variability in North Pacific (Mantua et al. 1997; Mantua and Hare 2002). The duration of a PDO phrase could be longer than 20-30 years (Deser et al. 2004; Lee et al. 2012). For instance, a negative PDO phase persisted 15 years from 1961 to 1975.

Although the mechanisms causing PDO remain unknown, there is evidence showing that PDO variations have had substantial impacts on climate in the North America and the Pacific, coinciding with periods of prolonged dryness and wetness (Miller and Schneider 2000; Mantua and Hare 2002). This in turn can result in substantial impacts on agricultural productivities.
The tropical Atlantic gradient

The TAG is a type of long-term climate variability with time scale 12 to 13 years that is characterized by SST gradient changes in tropical areas of the Atlantic Ocean (Rajagopalan et al. 1998). In addition to the magnitude of SST changes, the spatial gradients of SST variability are also important in determining TAG climate effects (Knutson and Manabe 1995).

The tropical Atlantic SST variability has been found to be causing significant changes of atmospheric and surface climate, such as winds in the lower troposphere, intense hurricanes, severe drought in the Sahel in Africa, and abundant rainfall in a wide area of the United States (Good et al. 2009; Murphy et al. 2010). Evidence shows that the climate fluctuations due to TAG variability exert major impacts on agricultural production in northeast Brazil and in the Sahel (Rowell et al. 1995; Sutton and Hodson 2005).

The west Pacific warm pool

The WPWP refers to the region located in the western tropical Pacific containing some of the warmest water in the open oceans (Webster and Lukas 1992; Picaut et al. 1996). The water in this area has a SST consistently higher than 28°C, approximately two to five degrees higher than that of other equatorial waters (Yan et al. 1992; Wang and Mehta 2008). The WPWP varies through the effects of water salinity and ocean-atmosphere heat flux (Lukas and Lindstrom 1991; Huang and Mehta 2004). Some studies have revealed the interactions between the WPWP and other climate variability phenomena, such as the ENSO, the PDO, and the Atlantic Multidecadal Oscillation (AMO) (Clarke et al. 2000; Solomon and Jin 2005).
The WPWP is crucial to the global climate since it runs the most intensive atmospheric circulation in the world and constitutes an indispensable sink of atmospheric freshwater (Picaut et al. 1996). Wang and Mehta (2008) address that there is close relationship between WPWP variability and persistent rainfall anomalies in the U.S., implying the potential impacts on U.S. water resources and agriculture.

**DCV phase combinations**

The three DCV phenomena each with two phases together constitute eight mutually exclusive combinations. We define each combination a DCV phase combination similar to Fernandez (2013). In this fashion, each year is characterized by a DCV phase combination. For example, in 1977, the positive PDO, the negative TAG, and the negative WPWP make the year fit in the phase combination (PDO+ TAG− WPWP−). Mehta et al. (2011) stress that the far-reaching and persistent climate variations such as drought or flood events are jointly caused by combinations of phases of multiple DCV phenomena instead of a single one. In this study, the DCV phase combinations in existence for the years from 1950 to 2010 are shown in Table 1.

Considering the frequency of each DCV phase combination occurring among the 61 years, we construct a frequency-based probability for each phase combination. The results are presented in Figure 1. These probabilities are then used to calculate expected crop yields under the eight DCV phase combinations.
Table 1: The DCV phase combinations and associated years from 1950 to 2010.

<table>
<thead>
<tr>
<th>PDO</th>
<th>TAG</th>
<th>WPWP</th>
<th>Year</th>
</tr>
</thead>
</table>


Figure 1: The frequency-based probability distribution for eight DCV phase combinations.

2.2 The Missouri River Basin

The study area for the empirical work is the Missouri River Basin. The MRB is the largest river basin in the U.S., partially or fully covering Colorado, Iowa, Kansas,
Minnesota, Missouri, Montana, Nebraska, North Dakota, South Dakota, and Wyoming.
The MRB is also one of the major crop and livestock producing regions in the U.S., producing approximately 46% of wheat, 22% of grain corn, and 34% of cattle (Mehta et al. 2012).

The agriculture in the MRB region is vulnerable to climate variations, since almost 90% of the crops are planted on dryland. The decadal SST variability in the Pacific and the Atlantic is associated with major droughts, for example, the “Dust Bowl” striking the Great Plains in the 1930s (McCabe et al. 2004; Nigam et al. 2011). Specifically in the MRB, precipitation and surface air temperature variability have been found to be highly correlated with PDO, TAG, and WPWP phases (Cayan et al. 1998; Mehta et al. 2011). Hence, given the importance of agriculture and its vulnerability to DCV, it is worth attention to studying the effects of DCV on major crops in the MRB.

3 Model Specification

3.1 A general model

Figure 2 depicts the direct and indirect DCV effects on crop yields. First, the DCV phenomena have an impact on regional and global climate, which in turn affect crop yields. These effects are categorized as indirect effects. Second, DCV may directly impact crop yields or influence yields via other factors that are unobserved. For example, in addition to precipitation and temperature variability, DCV is correlated with winds, storminess, air circulation, and pest distributions. These climate effects are not included in the model herein, but also alter crop yields. We name this part of effects as direct effects, since they are estimated directly from the crop yield equation.
The standard approach for modeling both direct and indirect effects such as those depicted in Figure 2 involves two levels of equations (Baron and Kenny 1986). We adapt that approach and apply it to studying direct and indirect DCV effects on crop yields. The model involves the following general equations:

\[ w_k = g_k(D, T; \theta^k) + \epsilon^k, \quad k = 1, ..., K, \]  

(1)  

\[ y = f(D, W, T; \theta^y) + \epsilon^y, \]  

(2)

where \( g_k \) is the function describing the mean of the \( k \)th climate variable, and \( f \) is the function giving the mean of crop yield. \( D \) represents dummy variables identifying the DCV phase combinations. \( D \) appears in both the climate equations (1) and the crop yield equation (2), implying that the DCV phase combination affects both climate and crop yields. \( T \) is a time trend variable, capturing in yield and temperature across times. \( w_k \) is
the $k^{th}$ climate variable in the climate variable set $W$ such that $W = (w_1, \ldots, w_K)$. $\theta^k$ and $\theta^y$ for all $k$ are the parameters associated with the two equations, respectively. $\epsilon^k$ and $\epsilon^y$ are the error terms.

The marginal direct effect (MDE) for the $l^{th}$ DCV phase combination is then calculated by taking derivative of the crop yield equation with respect to the $l^{th}$ DCV dummy:

$$MDE_l = \frac{\Delta f(D, W, T; \theta^y)}{\Delta D_l},$$

(3)

and the marginal indirect effect (MIE) is described by the interactions of the derivatives of the climate equations and the crop yield equation:

$$MIE_l = \sum_{k=1}^{K} \frac{\Delta g_k(D, T; \theta^k)}{\Delta D_l} \times \frac{\partial f(D, W, T; \theta^y)}{\partial w_k}.$$  

(4)

The marginal total effect (MTE) of DCV is the summation of MDE and MIE.

Equations (1) and (2) are general forms. Next we describe specific forms and assumptions for both the climate and crop yield equations.

### 3.2 Continuous climate equation

Suppose there are $j = 1, \ldots, m$ counties in the data set, and each county $j$ consists of a sequence of $n_j$ observations. The sample size of each county is allowed to differ.

The dependent variable $w_k$ for climate in equation (1) may take continuous or discrete values. For example, the variables associated with the temperature and precipitation are continuous, while the variable associated with the number of days that precipitation is
greater than one inch is discrete. In such cases, different functional forms are needed to represent the continuous and discrete climate variables.

For a continuous climate variable $w^C$, where superscript $C$ represent “continuous”, we use a linear functional form to capture the effects of DCV on climate for each county $j$:

$$w^C_j = X^w \theta^C_j + \epsilon^C_j, \quad j = 1, \ldots, m,$$

where $w^C_j$ is a $n_j \times 1$ continuous climate variable in the county $j$; $X^w$ is a $n_j \times q$ matrix of explanatory variables such that $X^w = (D, T)$, including DCV phase combination dum-

mies, $D$, and time trend and its square term, $T$. In the climate equation, $X^w$ are constant across counties.

There are some caveats associated with estimated parameters in a model with the panel data. If $X^w$ can fully explain the variations of dependent variables among the subjects in the panel, the estimated parameters $\theta^C_j$ in the equation (5) are constant across counties such that $\theta^C_j = \theta^C$ for all $j$, implying that all counties have the same mean effects of explanatory variables. In this case, we can pool the data from all counties, and estimate the model together. If the model specifications are proper, it results in consistent estimates with greater degrees of freedom.

The model with pooled data, although a good starting point, is not suitable for this application, since the coefficients in the model are not allowed to vary across counties. In

---

1 To avoid singularity problem, we select the phase combination (PDO− TAG+ WPWP+) as the base case that is not included in the equation.
fact, there are likely to be unobserved factors associated with the DCV that potentially make the marginal effects on climate formation vary. In this sense, it may be unreasonable to assume that marginal effects of explanatory variables are identical among locations. Pooling the data would generate biased results. This concern leads to the model (5) that allows the parameters $\theta_j^C$ vary across counties.

A common but simple method is to estimate the equation (5) county by county individually using OLS. The separate estimation assumes that the marginal effects of explanatory variables are completely independent across counties. The information in county $j$ does not contribute to estimations in other $m-1$ counties. This assumption may be inappropriate with the panel structure of data used in the model. Neighboring counties have similar geographic characteristics such that climate patterns under a DCV phase combination may be correlated among them. The individual estimation cannot accommodate that correlation. In addition, the results from individual estimation may be biased for some counties with small sample sizes.

The pooled and individual estimations represent the two extreme cases for data with multiple subjects. The pooled estimation assumes that information is completely transferable among subjects, while the individual estimation assumes that information is only valid within the subject. In reality, most situations lie between the two extremes.

The linear mixed-effects model does a better job of accommodating across county heterogeneity and allowing information to be common across counties. It can describe both within-county variation and between-county heterogeneit. The within-county variation is expressed in equation (5), where the error term $\epsilon_j^C$ is assumed to independently follow a
distribution. Specifically in the continuous climate equation, we assume the error term follows a multivariate normal distribution:

\[ \epsilon_j^C \mid \Psi_j^C \sim \text{ind. Multi-Normal}_{n_j} \left( 0, \Psi_j^C \right), \]  

(6)

where \( \Psi_j^C \) is a \( n_j \times n_j \) covariance matrix. For the remainder of this paper, we assume \( \Psi_j^C \) takes the form \( \sigma_j^2 I_{n_j} \), indicating that the standard errors are the same across counties. This is a common assumption in studies with longitudinal data (Arellano-Valle et al. 2007).

Along with the linear functional form (5), the within-county variation can be expressed with a form such that the dependent variable follows the multivariate normal distribution as well:

\[ w_j^C \mid \theta_j^C, X^w, \sigma_j^2 \sim \text{ind. Multi-Normal}_{n_j} \left( X^w \theta_j^C, \sigma_j^2 I_{n_j} \right). \]  

(7)

The between-county variations are modeled by allowing the parameters in the equation to be independently and identically distributed according to a specific distribution. The variations among the parameters represent the between-county heterogeneity. In the continuous climate equation, the sampling distribution for parameters is assumed as multivariate normal:

\[ \theta_j^C \mid \beta^C, \Sigma^C \sim \text{i.i.d. Multi-Normal}_{q} \left( \beta^C, \Sigma^C \right), \]  

(8)

where \( \beta^C \) is a \( q \times 1 \) mean vector and \( \Sigma^C \) is a \( q \times q \) covariance matrix.

Equations (5), (7), and (8) together form the normal linear mixed-effects model for the continuous climate variables. There are two levels of parameters in the model. The first level parameters are \( \theta_j^C = (\theta_1^C, \ldots, \theta_m^C) \), representing the heterogeneity of marginal effects.
across counties; the other level parameters, \( (\beta^c, \Sigma^c) \), express information common across counties.

### 3.3 Discrete climate equation

In the equation (1), some of the climate variables take on discrete values, and the model presented in (5) is not suitable for such situations. The Poisson regression model is an appropriate choice for within-county variation if the dependent variable is countable (Hoff 2009). In this approach, we take the log-mean of the discrete variable \( w^D_j \) in a regression setting, where superscript \( D \) represents “discrete”:

\[
\log \left( \mathbb{E} \left( w^D_j \mid X^w \right) \right) = X^w \theta^D_j, \quad j = 1, \ldots, m,
\]

where \( \theta^D_j \) are the \( q \times 1 \) estimated parameters in the discrete climate equation; \( X^w \) is the explanatory variables as defined for the continuous climate equation, including both DCV variables and time trend variables. Equation (9) is equivalent to:

\[
\mathbb{E} \left( w^D_j \mid X^w \right) = \exp \left( X^w \theta^D_j \right),
\]

which denotes the mean of the discrete dependent variable. By assuming the Poisson distribution for \( w^D_{ij} \) for the \( i^{th} \) observation in the county \( j \), the within-county variation can be displayed in the following form:

\[
w^D_{ij} \mid x^w_i, \theta^D_j \sim \text{i.i.d. Poisson} \left( \exp \left( (\theta^D_j)^T x^w_i \right) \right).
\]

The between-county variation representing heterogeneity is captured by allowing \( \theta^D_j \) to follow a multivariate normal distribution as well:
\theta_j^D | \beta^D, \Sigma^D \sim \text{i.i.d. Multi-Normal} \left( \beta^D, \Sigma^D \right), \quad (12)  

where \( \beta^D \) and \( \Sigma^D \) are parameters associated with the multivariate normal distribution. Hence, equations (9), (11), and (12) constitute the hierarchical LMM for discrete dependent variables.

### 3.4 Crop Yield equation

For the crop yield equation (2), we consider the following linear regression form:

\[ y_j = X_j^Y \theta_j^Y + \epsilon_j^Y, \quad j = 1, \ldots, m, \quad (13) \]

where \( y_j \) is a \( n_j \times 1 \) vector of crop yield observations for the county \( j \); \( \epsilon_j^Y \) is a \( n_j \times 1 \) vector of standard errors; \( X_j^Y \) is a group-specific \( n_j \times p \) matrix corresponding to \( p \) explanatory variables for the county \( j \), where \( X_j^Y = \left( D, W_j, T \right) \). Similar to the climate equation, \( D \) represents DCV phase combination dummies. \( W_j \) includes county-specific climate variables and their squared terms. The quadratic terms for climate variables are included because the effects of climate on crop yields are rarely linear (Mendelsohn et al. 1994; Schlenker and Roberts 2009). \( T \) include time trend and its squared term to capture technology development associated with crop yields. \( \theta_j^Y \) is a \( p \times 1 \) vector of estimated parameters.

We also model the yield equation in a fashion that the model can both display heterogeneity and share information among counties. Previous studies have shown that climate and DCV effects on crop yields vary across counties (Izaurralde et al. 2003; Mehta et al.
On the other hand, the effects in different counties should not be treated independently in that crops planted in neighboring counties grow under similar climate conditions and may also have access to close water reservoirs.

The within-county variation is also expressed in a manner that the error term $\epsilon_{j}^{Y}$ is assumed to independently follow a distribution. In most studies, the multivariate normal assumption is commonly used. However, the standard multivariate normal distribution is, to some extent, lack of robustness against deviations from the normality assumption. Hence, the normality assumption may not be a good fit for the crop yield equation here, because crop yields usually display skewness in their distributions.

The skew-normal distribution is a flexible parametric family, which is able to accommodate the situation that the true distribution deviates from a normal one (Arellano-Valle et al. 2007). In this study, we assume the error term $\epsilon_{j}^{Y}$ in equation (13) follows a skew-normal distribution:

$$\epsilon_{j}^{Y} \mid \Psi_{j}^{Y}, \Delta_{j} \sim \text{ind. Skew-Normal}_{n_{j}} \left(0, \Psi_{j}^{Y}, \Lambda_{j} \right),$$

(14)

where $\Psi_{j}^{Y}$ is a $n_{j} \times n_{j}$ covariance matrix, and $\Lambda_{j}$ is a $n_{j} \times n_{j}$ diagonal skewness matrix.

The probability density function of a skew-normal distribution and its linkage with the standard normal distribution are discussed in Appendix A. For the remainder of this paper, we assume $\Psi_{j}^{Y} = \sigma_{Y}^{2}I_{n_{j}}$ and $\Lambda_{j} = \delta I_{n_{j}}$.

The within-county variation can be expressed with a form such that crop yields also follow a skew-normal distribution:

$$y_{j} \mid X_{j}^{Y}, \theta_{j}^{Y}, \sigma_{Y}^{2}, \delta \sim \text{ind. Skew-Normal}_{n_{j}} \left(X_{j}^{Y} \theta_{j}^{Y}, \sigma_{Y}^{2}I_{n_{j}}, \delta I_{n_{j}} \right).$$

(15)
The between-county heterogeneity is modeled by allowing the parameters $\theta_j^Y$ in the equation (13) to be independently and identically distributed as a skew-normal distribution:

$$\theta_j^Y | \beta^Y, \Sigma^Y, \Pi \sim \text{i.i.d. Skew-Normal}_p \left( \beta^Y, \Sigma^Y, \Pi \right),$$

(16)

where $\beta^Y$ is a $p \times 1$ mean vector; $\Sigma^Y$ is a $p \times p$ covariance matrix; $\Pi$ is a $p \times p$ diagonal skewness matrix such that $\Pi = \text{diag}(\pi)$, where $\pi = (\pi_1, \ldots, \pi_p)$. Hence, equations (13), (15), and (16) constitute a hierarchical LMM with a skew normal distribution assumption.

### 4 Estimation with Bayesian methods

Now given the specified equations, we need to estimate the unknown parameters. We assume that the error terms and parameters in both the climate equations (1) and the yield equation (2) are pairwise uncorrelated, which is a strategy for solving a recursive system of simultaneous equations (Wooldridge 2010). This assumption allows the system to be solved equation by equation. In this study, the three hierarchical models discussed in the previous section are estimated with Bayesian methods.

A Bayesian approach states that the posterior distributions for unknown parameters are updated after observing the data, given the prior beliefs of such parameters. The posterior distributions are then approximated using sampling algorithms. The general procedures of estimating hierarchical LMM in a Bayesian framework are discussed in the following subsections, with further details presented in Appendices B-D.
4.1 Joint posterior distribution

In general, the Bayesian approach derives and approximates the joint posterior distribution for all unknown parameters in a model. It is basically derived from prior distributions, \( p(\Theta) \), and sampling models, \( p(y|\Theta) \), according to Bayes’ rule:

\[
p(\Theta|y) \propto p(\Theta) \times p(y|\Theta),
\]

where \( \Theta \) is a vector that contains all unknown parameters. The prior distributions are pre-specified, and the sampling distributions are assumed such as (15) and (16). The posterior distributions represent belief updates from the priors after obtaining information from the data. Specifically, the joint posterior distributions for the models in this study, if applicable, are presented in Appendices B-D.

4.2 Prior distributions

An essential part of Bayesian analysis is to assign prior distributions to all unknown parameters. These priors represent a researcher’s initial beliefs about the problem they are studying. In most cases, however, we do not have enough information to assign proper priors to a large number of parameters. Hoff (2009) argues that if the prior distribution is likely to deviate from the true prior information, it should be selected to be as less informative as possible, which will generate more objective results. This choice of non-informative prior distributions tells how posterior distributions would change with weak prior information.

We use two types of non-informative priors in this study. The first is the diffuse prior distribution, which assigns approximately equal probability to large areas of the support by
selecting specific values for parameters in the prior distributions. The other type is the unit information prior, which uses small amount of information in a single observation from the data (Kass and Wasserman 1995).

In order to obtain proper posterior distributions for calculation convenience, the diffuse prior distributions and unit information prior distributions are selected to be conjugate (Arellano-Valle et al. 2007; Hoff 2009). Conjugate priors are defined in a manner that the posterior distribution derived from the prior and sampling model for each parameter has the same form with its prior distribution. The details of the prior distributions used in the three hierarchical LMM are discussed in Appendices B-D.

4.3 Sampling algorithms

Sampling algorithms are used to approximate the joint posterior distribution. For models with multiple parameters, it is difficult to directly sample from the joint posterior distribution that has a non-standard form. However, it is often viable to sample from the posterior distribution of each parameter conditional on all other parameters, i.e., the full conditional distribution. The joint and marginal posterior distributions can be approximated with the Gibbs sampler, a Markov Chain Monte Carlo (MCMC) algorithm that iteratively generates a sequence of samples for all relevant parameters (Hoff 2009). The samples are then used to construct the joint posterior distribution.

In cases where the full conditional distributions for some parameters are not in standard forms, the Gibbs sampler is not practical. We use the Metropolis-Hastings algorithm to approximate the posterior distribution for such parameters (Hoff 2009). It is a general
MCMC method, sampling a sequence of values using a proposal distribution and an acceptance criterion. The details of how to implement the two MCMC algorithms are discussed in the Appendices B-D.

4.4 Bayesian inference

We make inferences for parameters based on the approximated posterior distributions, which identifies subsets of the parameter space that are likely to include the true value of the parameter. This can be done by constructing an interval from the MCMC samples that converge to the target posterior distribution. In Bayesian analysis, the highest posterior density (HPD) interval is often used. In a specific HPD interval, the probability of the true value contained in the interval is the selected confidence level; the posterior density of all values in the interval is greater than values outside the interval (Hoff 2009).

5 Deriving Results

5.1 Marginal total effects of DCV

After estimating the model with Bayesian approaches, we calculate the county-specific marginal effects of DCV. The marginal total effects of DCV are the summation of MDE and MIE that are estimated by taking derivatives of the equations according to (3) and (4). The marginal effects are evaluated at the means of explanatory variables. In practice, we exclude insignificant mean estimates from both the yield and climate equations at the 95% confidence level, in which the associated HPD intervals for parameters contain zero.

The marginal effects from equations (3) and (4) only represent relative numbers compared to the base DCV phase combination that we choose for the model. The MTE cannot
be displayed for the base DCV phase combination from the estimated parameters. It is often of interest to display the MTE for all DCV phase combinations. We change the baseline and make the MTE compare to the predicted mean yields. The adjusted MTE are calculated as the following steps, for each county $j = 1, \ldots, m$:

1. Calculate the MTE based on the estimated parameters according to the equations (3) and (4).

2. Predict climate variables along with mean of the MCMC samples for $\theta_j^C$ and $\theta_j^D$ under each DCV phase combination.

3. Incorporate the predicted DCV-specific climate values into the equation (13), in conjunction with the estimated parameters $\theta_j^F$, to predict crop yield under each DCV phase combination.

4. Calculate the mean predicted yield as the weighted average of the predicted yields under different DCV phase combinations given the historical probabilities of DCV phase combinations as shown in Figure 1.

5. Compare the yield in the base phase combination to the mean predicted yield calculated from the Step 4.

6. Adjust MTE estimates from Step 1 based on the comparison results from Step 5.

5.2 Predicted yield distribution

The Gibbs or Metropolis-Hastings sampling algorithms iteratively generate a sequence of samples for each unknown parameter in the model, as elaborated in Appendices B-D. These samples can be used to predict the posterior yield distributions. We predict the crop
yields based on the MCMC samples for $\theta_j^Y$ and predicted climate variables from the Step 2 in the previous section.

Given the value of $\theta_j^Y$ at scan $s$, the predicted crop yield associated with the DCV phase combination $l$ is generated in the following steps:

1. For each $j=1,\ldots,m$, predict crop yields $\hat{y}_j^{(s)}$ by plugging in the predicted climate variables associated with the DCV phase combination $l$ and taking the DCV dummy $D_l = 1$ according to the equation (13).

2. Take average of $\hat{y}_j^{(s)}$ over $j$ to generate average crop yield $\bar{y}_l^{(s)}$ for the DCV phase combination $l$.

Each parameter scan $\theta_j^Y$ can be used to predict a crop yield $\hat{y}_j^{(s)}$. For a large scan $s$, the predicted sample of $\hat{y}_l$ approximate the yield distribution under the DCV phase combination $l$.

6 Data

The data used in this study were obtained from multiple sources. The county-specific yield data were obtained from the U.S. Department of Agriculture Quickstats. The yield data are longitudinal such that each county that plants a specific crop has a sequence of annual observations. In general, the time span of the data is from 1950 to 2010. There are some counties missing some observations for some crops. There are eight major dryland crops examined: barley, corn, alfalfa hay, oats, sorghum, soybeans, spring wheat, and winter wheat.
The climate data were gathered from the National Climatic Data Center, National Oceanic and Atmospheric Administration (NOAA). The climate data were selected as growing season variables for crops. In terms of growing season, the eight crops can be divided into two groups. The crops in the first group grow in summer, including barley, corn, alfalfa hay, oats, sorghum, soybeans, and spring wheat. The corresponding climate variables for summer crops are: monthly mean temperature, total precipitation, number of days with maximum temperature greater than or equal to 90°F, and number of days with precipitation greater than or equal to one inch. The winter wheat is the only crop growing in winter. The associated climate variables are monthly mean temperature, total precipitation, number of days with precipitation greater than or equal to one inch, and number of days with minimum temperature less than or equal to 0°F. The DCV phase specification data by year were obtained from Fernandez (2013), and are presented in Table 1. These data are yearly dummies and constant across counties in a year.

7 Results and Discussions

7.1 Model justification

Before implementing the Bayesian procedures to estimate the model, we first preliminarily examine the data. Figure 3 shows the levels of the corn yield and summer mean temperature over time. Figure 3(a) suggests an increasing trend for corn yield over time in most counties but with evident between-county variation. Figure 3(b) also shows inter-
county variations in the climate variable. Figure 3 indicates that the between-group heterogeneity needs to be considered in both climate and crop yield equations.\textsuperscript{2} Hence, the LMM estimated in a Bayesian framework are suitable tools for studying such county-specific DCV effects.

![Figure 3: Crop yield and climate data in every ten years. (a) Corn yields for counties that have planted such crop (black), with trajectories (gray) for five randomly selected counties. (b) A county-specific continuous climate variable (black), mean temperature in growing season, and trajectories (gray) for five random counties.](image)

The distributions for crop yield and parameters in (13) are assumed to be skew-normal. Figure 4 presents data regarding the applicability of this assumption. Figure 4(a) presents a plot of the county-specific skewness index for crop yield data. The gray dots indicate that most skewness indices indicate skewed yield distributions. We also calculate skewness

\textsuperscript{2} We have generated the same type of graphs for other crop yields and climate variables. They all have the same pattern displaying inter-county variation.
indices for pooled crop yield data across counties and find positive skewness for all crops as shown in black dots.

To investigate potential skewness for the parameters in the yield equation, we fit a simple OLS regression model for each county individually, and pool the county-specific parameters to calculate the skewness indices. The results for parameters in eight crops are plotted in Figure 4(b). It suggests that estimated parameters in most crop yield equations have significant skewness indices.

Figure 4: Skewness index of crop yields and OLS estimates, with zero representing no skewness. (a) County-specific skewness indices of yields for eight crops (gray) and skewness index for each crop calculated from pooled yield data (black). (b) Skewness indices for estimated parameters in eight yield equations from the individual OLS fits.
7.2 Bayesian estimates

A type of MCMC algorithm has been implemented for each equation in the model system. The Gibbs sampler with normal distributions is applied to the continuous climate equations as discussed in Appendix B. The integrated Gibbs and Metropolis-Hastings algorithm is implemented for the discrete climate equations as elaborated in Appendix C. The Gibbs sampler with skew-normal distributions discussed in Appendix D is used for the crop yield equation.

Each sampler ran 100,000 scans and saved every 100th scan with the purpose of avoiding autocorrelation in the MCMC samples. Hence, each equation produces a sequence of 1,000 values for each unknown parameter in the model. We then use the samples to make Bayesian inference. The resulting parameter estimates are taken as posterior means of the MCMC samples.

A good feature of the Bayesian method is that it shrinks extreme estimates from individual OLS estimations towards the between-group average, while keeping heterogeneity across groups. Figure 5 represents a graph of estimates associated with DCV variables in the corn yield equation under the OLS and LMM Bayes approaches.3 The individual OLS results exhibit considerable disparity, as shown in gray diamonds. The Bayes point estimates from LMM are shown as black dots, indicating less disparity across counties.

---

3 We have generated the same graphs for rest of the crops. They all display the same pattern.
7.3 County-specific marginal total effects of DCV

The marginal total effects for each crop under eight DCV phase combinations are calculated according to the procedures in Section 5.1.

**Barley**

The county-specific impacts of DCV on barley yields are presented in Figure 6. It shows substantial heterogeneity across DCV phase combinations and counties. Overall, the

![Graph showing comparison of estimates associated with DCV variables in the corn yield equation between OLS fits (gray) and Bayesian estimation (black).](image)

Figure 5: The comparison of estimates associated with DCV variables in the corn yield equation between results from the individual OLS fits (gray) and from the Bayesian estimation (black).

phase combination (PDO+ TAG+ WPWP+) has the most positive effects as shown in the panel (g), while the phase combination (PDO+ TAG− WPWP+) has the most negative effects as displayed in the panel (e). Figure 6(a) suggests that under the phase combination (PDO+ TAG− WPWP−), barley yields in most counties of Montana, Colorado, and Nebraska decrease by up to 25% from the predicted mean yields; while in most counties of
North Dakota and South Dakota, barley yields increase by up to more than 10%. Barley yield changes under the DCV phase combinations (b) (PDO− TAG+ WPWP−), (c) (PDO− TAG− WPWP+), and (f) (PDO− TAG+ WPWP+) have similar spatial patterns but with different magnitudes of variation. Figure 6(d) implies that the phase combination (PDO+ TAG+ WPWP−) has small impacts on most counties in northern MRB, but slight negative effects with less than 10% yield loss on counties in the middle part of MRB. It is interesting to note that in response to the DCV phase combination (e) (PDO+ TAG− WPWP+) barley yields decrease up to 30% in almost all counties in the MRB. The negative effects of this combination of DCV phenomena are greater in northern counties than in southern counties.

Under the phase combination (g) (PDO+ TAG+ WPWP+), barley yields in most counties of Montana, Wyoming, Colorado, and North Dakota increase by 10-30%. The phase combination (h) (PDO− TAG− WPWP−) has positive effects on northern counties by more than 5% but negative effects on some southern counties by 10%.

**Corn**

The DCV impacts on corn yields are presented in Figure 7. In the western MRB, there are no effects showing due to very small corn planting in the area. In general, Figure 7 illustrates that most phase combinations have positive effects on corn yields except for the phase combination (e) (PDO+ TAG− WPWP+) and (h) (PDO− TAG− WPWP−). Particularly for the phase combination (e) (PDO+ TAG− WPWP+), the corn yields decrease by 25% in some counties. Figure 7(a) and (d) have similar spatial effects but with different magnitudes, where corn yield decreases appear in the eastern part of the MRB, while corn yield increases occur in the middle part of the MRB. This spatial pattern is opposite in the phase combination (g) (PDO+ TAG+ WPWP+), where corn yield reductions mainly result
in the southwestern part of the MRB. In the phase combination (b) (PDO− TAG+ WPWP−), there are negative changes in corn yields, and most severe corn yield losses appear in South Dakota. The phase combinations (c) (PDO− TAG− WPWP+) and (f) (PDO− TAG+ WPWP+) suggest that corn yields increase in most MRB counties by 4% except for a few decreases.

*Alfalfa hay*

Figure 8 shows yield changes of alfalfa hay under different DCV phase combinations. The DCV phenomena have more significant effects on alfalfa hay yields in the western MRB compared to the southeastern part. In response to the DCV phase combination (a) (PDO+ TAG- WPWP−), alfalfa hay yields decrease by 10-30% in most counties of Montana and Wyoming. This combination of DCV phenomena has the most significant and widespread negative effects on alfalfa hay yields. In contrast to the phase combination (a), there are yield increases in Montana and Wyoming under the phase combinations (b) (PDO− TAG+ WPWP−), (d) (PDO+ TAG+ WPWP−), and (f) (PDO− TAG+ WPWP+). It is interesting to note that the impacts of the phase combination (g) (PDO+ TAG+ WPWP+) are positive or insignificant in most counties but significantly negative in one county in Wyoming. There are either positive or negative effects distributing in Montana, Wyoming, and Colorado under (c) (PDO− TAG− WPWP+), (e) (PDO+ TAG− WPWP+), and (h) (PDO− TAG− WPWP−).

*Oats*

Figure 9 shows that the oats yield data were available for most MRB counties. The DCV effects exhibit substantial variation both across counties and DCV phase combinations. In the phase combination (a) (PDO+ TAG− WPWP−), the positive effects
Figure 6: County-specific total DCV effects on barley yields (bushel/acre). (a)-(h) represent comparisons between crop yields in different DCV phase combinations and the predicted mean yields.
Figure 7: County-specific total DCV effects on corn yields (bushel/acre). (a)-(h) represent comparisons between crop yields in different DCV phase combinations and the predicted mean yields.
concentrate on the northern part of MRB, while the negative effects appear in the southern MRB. The phase combinations (b) (PDO− TAG+ WPWP−) and (e) (PDO+ TAG− WPWP+) have broad negative impacts on oat yields in the MRB with small positive effects located in a few counties. Figure 9 also suggests that positive DCV effects cluster in North Dakota and South Dakota in the phase combinations (a) (PDO+ TAG− WPWP−), (c) (PDO− TAG− WPWP+), (f) (PDO− TAG+ WPWP+), (g) (PDO+ TAG+ WPWP+), and (h) (PDO− TAG− WPWP−).

Sorghum

Figure 10 presents the sorghum yield changes under various DCV phase combinations. In the phase combination (a) (PDO+ TAG− WPWP−), the positive effects appear in western South Dakota and some counties of Nebraska and Colorado; the negative effects concentrate on eastern South Dakota, most areas of Nebraska and Missouri. The phase combination (b) (PDO− TAG+ WPWP−), (c) (PDO− TAG− WPWP+), and (f) (PDO− TAG+ WPWP+) have widespread positive effects on counties that planted sorghum. In contrast, under the other three phase combinations (e) (PDO+ TAG− WPWP+), (g) (PDO+ TAG+ WPWP+), and (h) (PDO− TAG− WPWP−), the sorghum yields decrease by 10-20% in most counties. Figure 10(d) suggests that the DCV phase combination (PDO+ TAG+ WPWP−) has positive effects in the central part while negative effects in the southeastern part of MRB.

Soybeans

The areas where soybean yield data were available are similar to the case of corn, which concentrate on the southern and eastern parts of MRB, as shown in Figure 11. The yield effects of DCV phenomena on soybeans are mild compared to other crops except for the
Figure 8: County-specific total DCV effects on alfalfa hay yields (tons/acre). (a)-(h) represent comparisons between crop yields in different DCV phase combinations and the predicted mean yields.
Figure 9: County-specific total DCV effects on oat yields (bushel/acre). (a)-(h) represent comparisons between crop yields in different DCV phase combinations and the predicted mean yields.
phase combination (e) (PDO+ TAG− WPWP+) that causes the soybean yields decrease by 10-20% in most MRB counties. In the phase combinations (c) (PDO− TAG− WPWP+), (f) (PDO− TAG+ WPWP+), and (g) (PDO+ TAG+ WPWP+), there are soybean yield increases in most counties. For the other DCV phase combinations (a) (PDO+ TAG− WPWP−), (b) (PDO− TAG+ WPWP−), (d) (PDO+ TAG+ WPWP−), and (h) (PDO− TAG− WPWP−), both positive and negative effects scatter in the area planting soybeans.

Spring wheat

The impacts of DCV phenomena on spring wheat yields in the MRB are displayed in Figure 12. In phase combination (a) (PDO+ TAG− WPWP−), the negative effects decrease spring wheat yields up to 20% in Montana, Wyoming, and Colorado; the positive effects scatter in Montana, North Dakota, and South Dakota. Figure 12(b) suggests that under the phase combination (PDO− TAG+ WPWP−), the significant negative effects widely distribute in Montana, which decrease yields by 20-30% compared to expected mean yields, while the spring wheat increases concentrate in North Dakota and South Dakota. The phase combination (c) (PDO− TAG− WPWP+) has a similar pattern with the phase combination (f) (PDO− TAG+ WPWP+), in which the spring wheat yields decrease by 5-10% in most MRB counties except for some counties in North Dakota and South Dakota. Under the phase combination (d) (PDO+ TAG+ WPWP−), the DCV phenomena combination makes the yields increase by 10-20% in western counties of Montana. It is interesting to note that the phase combination (e) (PDO+ TAG− WPWP+) has negative effects on crop yields in all counties that planted spring wheat. In contrast, in the phase combination (g) (PDO+ TAG+ WPWP+), yields increase in most counties.
Figure 10: County-specific total DCV effects on sorghum yields (bushel/acre). (a)-(h) represent comparisons between crop yields in different DCV phase combinations and the predicted mean yields.
Figure 11: County-specific total DCV effects on soybean yields (bushel/acre). (a)-(h) represent comparisons between crop yields in different DCV phase combinations and the predicted mean yields.
Figure 12: County-specific total DCV effects on spring wheat yields (bushel/acre). (a)-(h) represent comparisons between crop yields in different DCV phase combinations and the predicted mean yields.
Winter wheat

Figure 13 represents the estimated DCV effects on winter wheat in the MRB. Winter wheat yields decrease in most counties except for those in the northeastern MRB under the phase combination (a) (PDO+ TAG− WPWP−). Figure 13(b) shows that under the phase combination (PDO− TAG+ WPWP−), the positive effects distribute in the eastern side of MRB, while the positive effects appear in the western and middle MRB. There is also clear clustering DCV effects in the phase combination (c) (PDO− TAG− WPWP+), in which positive effects locate in the middle MRB and negative effects concentrate on the southern MRB. In the phase combinations (d) (PDO+ TAG+ WPWP−), (f) (PDO− TAG+ WPWP+), and (h) (PDO− TAG− WPWP−), the negative effects are more wide and significant. As noted in previous crop parts, the phase combination (e) (PDO+ TAG− WPWP+) has the most negative effects on winter wheat yields up to 25% decreases. In phase combination (g) (PDO+ TAG+ WPWP+), the negative effects locate in the middle MRB, while most negative effects locate along the eastern MRB.

7.4 Predicted posterior yield distributions

The posterior yield distributions for eight crops under different DCV phase combinations are approximated based on the parameter samples generated from the MCMC schemes. The detailed steps are discussed in Section 5.2. The predicted yield distributions are presented in Figure 14. It shows that the phase combination (e) (PDO+ TAG− WPWP+) pushes the location of distributions to the left for all crops. This suggests that the phase combination (e) causes lower yields compared to other phase combinations. Particularly for corn, soybean, and spring wheat, the distributions under the phase combination (e) (PDO+ TAG− WPWP+) are separated from the ones under other phase combinations.
The figure also indicates that the phase combination (g) (PDO+ TAG+ WPWP+) results in higher yields for most crops except for corn and sorghum than other phase combinations. In terms of the shape of distribution, barley, alfalfa hay, and oat display skewness even though there are potential multi-modes in their distributions. The results in Figure 14 suggest that insurance schemes and acreage choices could be altered given DCV phase combination information.

8 Conclusions

This study examined the effects of three decadal climate variability phenomena on crop yields in the Missouri River Basin. The DCV phenomena are thought to exert impacts on crop yields through both climate conditions and other unobserved factors that are labeled as indirect and direct effects, respectively. The model is constructed with two levels of equations, the climate equations and the yield equation, to estimate the total DCV effects on eight crops.

Previous studies address that the DCV effects on climate and crop yields display disparity across locations. Thus, heterogeneity in DCV effects are taken into account in linear mixed-effects models. The estimated parameters show that DCV effects demonstrate heterogeneity across counties and DCV phase combinations. For example, a DCV phase combination may have positive effects on some counties, while having negative effects on counties in other parts of the MRB.

The posterior samples for the parameters can be used to approximate the predicted posterior yield distributions under different DCV phase combinations. These could be useful information for farmers, policy makers, and insurance companies in the area. The
Figure 13: County-specific total DCV effects on winter wheat yields (bushel/acre). (a)-(h) represent comparisons between crop yields in different DCV phase combinations and the predicted mean yields.
Figure 14: The kernel densities of predicted crop yields under different DCV phase combinations for eight crops. The DCV phase combinations are represented by different colors: (a) PDO+ TAG− WPWP−, (b) PDO− TAG+ WPWP−, (c) PDO− TAG− WPWP+, (d) PDO+ TAG+ WPWP−, (e) PDO+ TAG− WPWP+, (f) PDO− TAG+ WPWP+, (g) PDO+ TAG+ WPWP+, and (h) PDO− TAG− WPWP−.
results indicate how the distribution of crop yields would be altered when we face a specific DCV phase combination. We also could gain some insights of DCV effects on crop yields from the predictive yield distributions. For example, the phase combination (PDO+ TAG−WPWP+) leads to the lowest mean for all crops compared to other phase combinations, while most crops excepting corn and sorghum perform better under the phase combination (PDO+ TAG+ WPWP+) than under other DCV phase combinations. We also notice that the predicted posterior distributions for barley, alfalfa hay, and oat yields exhibit evident skewness.

Methodologically, the Bayesian framework can be easily applied to other studies, in which the heterogeneity is present among different groups of the study subjects and the frequently used normal distribution assumption may be inapplicable.

Appendix A: Multivariate skew-normal distribution

Arellano-Valle et al. (2007) state the density function for a random vector $y$ that has a multivariate skew-normal distribution is:

$$p(y|\mu, \Sigma, \Delta) = 2^n \phi_n(y|\mu, \Sigma + \Delta \Delta^T) \Phi_n \left( \Delta^T \left( \Sigma + \Delta \Delta^T \right)^{-1} (y - \mu) \right) \left( 1 + \Delta^T \Sigma^{-1} \Delta \right)^{-1},$$

where $\phi_n$ and $\Phi_n$ are the probability density function (pdf) and the cumulative density function (cdf) of a multivariate normal distribution with dimension $n$; $\mu$ is a $n \times 1$ location vector; $\Sigma$ is a $n \times n$ positive definite scale matrix; $\Delta$ is a $n \times n$ diagonal skewness matrix with elements $\delta = (\delta_1, \ldots, \delta_n)$. We denote the skew-normal distribution for a random vector $y|\mu, \Sigma, \Delta \sim \text{Skew-Normal}_n(\mu, \Sigma, \Delta)$. Compared to the normal distribution, the skew-
normal has one more component of parameters, the skewness matrix. It can be shown that the multivariate skew-normal distribution becomes a multivariate normal distribution when \( \Lambda = 0 \).

By Proposition 1 of Arellano-Valle et al. (2007), the multivariate skew-normal distribution for \( y \) can be represented by a stochastic form:

\[
y = \Lambda |U_1| + U_2,
\]

where \( |U_1| \sim \text{Trunc-Normal}_n \left( \mathbf{0}, \mathbf{I} \right) \{ u_1 > 0 \} \), which is a truncated normal distribution with positive values \( u_1 \); \( U_2 \sim \text{Multi-Normal}_n (\mu, \Sigma) \); \( \Lambda \) is the diagonal skewness matrix. The skew-normal random variable is the combination of two random variables with different normal distributions. The stochastic form of the skew-normal distribution is convenient for later Bayesian analysis.

Appendix B: Bayesian algorithm for the continuous climate equation

General form of the joint posterior distribution

The joint posterior distribution in a continuous climate equation is derived according to the Bayes’ rule, combining the sampling distributions and prior distributions. The general form of the joint posterior distribution is represented as:

\[
p\left( \theta^C_1, \ldots, \theta^C_m, \beta^C, \sigma^2_c, \Sigma^C \mid w^C_1, \ldots, w^C_m, X^w \right) \\
\propto p\left( w^C_1, \ldots, w^C_m \mid \theta^C_1, \ldots, \theta^C_m, \sigma^2_c, X^w \right) \rightarrow \text{Multi-Normal sampling} \\
\propto p\left( \theta^C_1, \ldots, \theta^C_m \mid \beta^C, \Sigma^C \right) \rightarrow \text{Multi-Normal sampling} \\
\propto p\left( \beta^C, \sigma^2_c, \Sigma^C \right) \rightarrow \text{Priors}
\] (17)
In the equation (17), components associated with $w_j^C$ and $\theta_j^C$ for all $j$ are specified by their sampling distributions (7) and (8), respectively. We plug in the sampling distributions and the prior distributions for $\beta^C$, $\sigma^2_C$, and $\Sigma^C$ that are discussed in the next subsection to yield the specific form of the joint posterior distribution.

**Prior distributions**

The unknown parameters in equation (17) that require priors are $\beta^C$, $\sigma^2_C$, and $\Sigma^C$. We assume that the priors are independent in the model. Hence, the joint prior distribution in (17) can be represented as:

$$p(\beta^C, \sigma^2_C, \Sigma^C) = p(\beta^C) p(\sigma^2_C) p(\Sigma^C).$$

To obtain a closed form posterior distribution, we need to first specify the functional forms for the prior distributions. In this study, conjugate priors are used, resulting in proper posterior distributions (Arellano-Valle et al. 2007):

$$\beta^C \sim \text{Multi-Normal}_p(\mu_0^C, L_0^C),$$

$$\sigma^2_C \sim \text{Inv-Gamma}\left(\frac{v_0^C}{2}, \frac{v_0^C \sigma_{00}^2}{2}\right),$$

$$\Sigma^C \sim \text{Inv-Wishart}\left(\eta_0^C, S_{00}^{-1}\right).$$

We need to assign values for a large set of introduced parameters in the prior distributions. Prior distributions in this model are selected to be non-informative. As noted before, if we do not have correct information regarding the priors, it is appropriate to use non-informative priors that lead to objective results. These priors represent that how much information we gain after we observe the data if we do not have information in priori.
The prior for $\mathbf{\beta}^C$ is selected to be a multivariate normal distribution. The parameters in the prior are specified similar in spirit to the unit information prior introduced in Kass and Wasserman (1995). We take $\mathbf{\mu}_0^C$, the prior expectation of $\mathbf{\beta}^C$, to be equal to the average of the OLS estimates over counties and $\mathbf{\Lambda}_0^C$ to be the sample covariance of these estimates. The choice of prior for $\mathbf{\beta}^C$ represents unbiased but weak information (Hoff 2009).

In the inverse Gamma prior for $\sigma^2_c$, we choose $\sigma^2_{c_0}$ to be equal to the average of residual sum of squares (RSS) from the OLS estimations across counties. $\nu^C_0$ is set to be one, which guarantees that the inverse Gamma distribution is diffuse.

The prior for $\mathbf{\Sigma}^C$ is an inverse Wishart distribution. Similarly, the prior parameter $S_{c_0}$ is equal to the covariance of the OLS estimates. The prior degree of freedom $\eta^C_0$ is set to be $q + 2$, where $q$ is the number of covariates in the continuous climate model, so that the prior distribution for $\mathbf{\Sigma}^C$ is somewhat diffuse as well.

**Full conditional distributions**

Although we can write the joint posterior distribution as shown in (17), it is quite difficult to obtain the marginal posterior distribution for each parameter that is convenient for sampling. However, the full conditional distributions for unknown parameters can be straightforwardly derived from the joint posterior distribution. The full conditional distributions consist of the distribution of each parameter conditional on all other parameters. Following Hoff (2009), the full conditional distributions for the linear mixed effects model with a normal distribution assumption are:
\[ \theta_j^C \mid \beta^C, \sigma_c^2, \Sigma^C \sim \text{Multi-Normal}_p \left( \mu_j^C, \Lambda_j^C \right), \quad j = 1, \ldots, m, \] 

where

\[ \Lambda_j^C = \left( \Sigma_j^C \right)^{-1} + \frac{\left( X_j^w \right)^T X_j^w}{\sigma_c^2}, \quad (18) \]

\[ \mu_j^C = \Lambda_j^C \left( \Sigma_j^C \right)^{-1} \beta^C + \frac{\left( X_j^w \right)^T w_j^c}{\sigma_c^2}; \]

\[ \beta^C \mid \theta_1^C, \ldots, \theta_m^C, \Sigma^C \sim \text{Multi-Normal}_p \left( \mu_\beta^C, \Lambda_\beta^C \right), \text{ where} \]

\[ \Lambda_\beta^C = \left( \Lambda_0^C \right)^{-1} + m \left( \Sigma^C \right)^{-1}, \quad (19) \]

\[ \mu_\beta^C = \Lambda_\beta^C \left( \Lambda_0^C \right)^{-1} \mu_0^C + \left( \Sigma^C \right)^{-1} \sum_{j=1}^m \theta_j^C; \]

\[ \sigma_c^2 \mid \theta_1^C, \ldots, \theta_m^C \sim \text{Inv-Gamma} \left( \frac{\nu_0^C + \sum_{j=1}^m n_j}{2}, \frac{\nu_0^C \sigma_{c0}^2 + \text{SSR}_c^C}{2} \right), \text{ where} \]

\[ \text{SSR}_c^C = \sum_{j=1}^m \left( w_j^c - X_j^w \theta_j^C \right) \left( w_j^c - X_j^w \theta_j^C \right); \]

\[ \Sigma^C \mid \theta_1^C, \ldots, \theta_m^C, \beta^C \sim \text{Inv-Wishart} \left( \eta_0^C + m, \left( S_{c0} + S_{cm} \right)^{-1} \right), \text{ where} \]

\[ S_{cm} = \sum_{j=1}^m \left( \theta_j^C - \beta^C \right) \left( \theta_j^C - \beta^C \right)^T. \quad (21) \]

**Gibbs sampling algorithm**

Once we obtain the full conditional distributions we can construct a Gibbs sampler to approximate the joint posterior distribution. The Gibbs sampler is a type of MCMC algorithm that iteratively samples each parameter from its full conditional distribution given the most current state of other parameters. The order of generating a new set of parameters does not affect the approximation.
Given current values of all parameters $\Theta^C(s) = \left\{ \theta^{C(s)}_1, \ldots, \theta^{C(s)}_m, \beta^{C(s)}, \sigma^2_c, \Sigma^C(s) \right\}$ at scan $(s)$, new values are generated at scan $(s+1)$ as the following steps:

1. For each $j = 1, \ldots, m$, sample $\theta^{C(s+1)}_j \sim p\left(\theta^C_j \mid \beta^C, \sigma^2_c, \Sigma^C(s)\right)$ according to the posterior distributions in (18).

2. Sample $\beta^{C(s+1)} \sim p\left(\beta^C \mid \theta^{C(s+1)}_1, \ldots, \theta^{C(s+1)}_m, \Sigma^C(s)\right)$ according to the posterior distribution shown in (19).

3. Sample $\sigma^2_c^{(s+1)} \sim p\left(\sigma^2_c^{(s+1)} \mid \theta^{C(s+1)}_1, \ldots, \theta^{C(s+1)}_m\right)$ according to the posterior distribution represented in (20).

4. Sample $\Sigma^{C(s+1)} \sim p\left(\Sigma^C \mid \theta^{C(s+1)}_1, \ldots, \theta^{C(s+1)}_m, \beta^{C(s+1)}\right)$ based on the distribution in (21).

Through the above steps, a new state of the parameters is generated. Repeating the steps multiple times produces a sequence of the parameter set. As $s \to \infty$, the sampling distribution of $\Theta^C$ converges to the target joint posterior distribution, and the sampling distribution for each component in the parameter set $\Theta^C$ converge to its marginal posterior distribution. We then can make inferences for the parameters based on the sampling distributions.

**Appendix C: Bayesian algorithm for the discrete climate model**

In the Appendix C, the full conditional distribution for each parameter is in a standard form, so that the Gibbs sampling algorithm can be implemented. However, for the discrete climate equations (9), standard full conditional distributions only exist for $\beta^D$ and $\Sigma^D$. For
other parameters such as $\theta^D_j$, a Metropolis-Hastings algorithm can be used for sampling if a full conditional distribution is not available.

The full conditional distributions for $\beta^D$ and $\Sigma^D$ have the same form with (19) and (21), respectively. The prior distributions for $\beta^D$ and $\Sigma^D$ are specified with same strategy as for $\beta^C$ and $\Sigma^C$ discussed in the Appendix B. The prior distributions are selected to be non-informative.

The update of $\theta^D_j$ in a Markov chain can be operated with a Metropolis-Hastings algorithm (Hoff 2009). A new value $\theta^{D*}_j$ is sampled from a proposal distribution $J(\theta^{D*}_j|\theta^{D(x)}_j)$ nearby the current value $\theta^{D(x)}_j$, and then it is accepted or rejected based on an acceptance criterion. A common choice of the proposal distribution for $\theta^D_j$ would be a multivariate normal distribution such that the mean equals the current value $\theta^{D(x)}_j$ and the covariance matrix is set to be equal to $r\Sigma^{D(x)}$, where $r$ is a scale parameter and $\Sigma^{D(x)}$ is the current value of $\Sigma^D$ updated from the Gibbs sampling steps. In this study, we set $r = 0.4$ that generates a well-mixing Markov chain.

Following Hoff (2009), we use an integrated algorithm including both the Gibbs and Metropolis-Hastings samplers to approximate the joint posterior distribution. Given current values of parameters at scan $(s)$, $\Theta^{D(x)} = \{\theta^{D(x)}_1, \ldots, \theta^{D(x)}_m, \beta^{D(x)}, \Sigma^{D(x)}\}$, the new values are generated as follows:

1. Sample $\beta^{D(x+1)}$ from its full conditional distribution that has the same structure with the equation (19), that is, $\beta^{D(x+1)} \sim p(\beta^D|\theta^{D(x)}, \ldots, \theta^{D(x)}_m, \Sigma^{D(x)})$. 

2. Sample \( \Sigma_{D^{(s+1)}} \) from its full conditional distribution that is similar to the equation (21), i.e., \( \Sigma_{D^{(s+1)}} \sim p\left(\Sigma_{D}^D|\theta_{j}^{D^{(s)}}, \ldots, \theta_{m}^{D^{(s)}}, \beta_{j}^{D^{(s+1)}}\right) \).

3. For each \( j = 1, \ldots, m \),
   a. Propose a new value \( \theta_{j}^{D^{*}} \sim \text{Multi-Normal}\left(\theta_{j}^{D^{(s)}}, r\Sigma_{D^{(s+1)}}\right) \);
   b. Compute the acceptance ratio by comparing the posterior probability densities between the proposed value \( \theta_{j}^{D^{*}} \) and the current value \( \theta_{j}^{D^{(s)}} \):

   \[
   \rho = \frac{p\left(w_{j}^{D}|\theta_{j}^{D^{*}}, X_{j}^{W}\right) p\left(\theta_{j}^{D^{*}}|\beta_{j}^{D^{(s+1)}}, \Sigma_{D^{(s+1)}}\right)}{p\left(w_{j}^{D}|\theta_{j}^{D^{(s)}}, X_{j}^{W}\right) p\left(\theta_{j}^{D^{(s)}}|\beta_{j}^{D^{(s+1)}}, \Sigma_{D^{(s+1)}}\right)},
   \]

   where \( p\left(w_{j}^{D}|\theta_{j}^{D^{*}}, X_{j}^{W}\right) \) and \( p\left(w_{j}^{D}|\theta_{j}^{D^{(s)}}, X_{j}^{W}\right) \) are the Poisson density functions evaluated at \( \theta_{j}^{D^{*}} \) and \( \theta_{j}^{D^{(s)}} \), respectively, while \( p\left(\theta_{j}^{D^{*}}|\beta_{j}^{D^{(s+1)}}, \Sigma_{D^{(s+1)}}\right) \) and \( p\left(\theta_{j}^{D^{(s)}}|\beta_{j}^{D^{(s+1)}}, \Sigma_{D^{(s+1)}}\right) \) are the multivariate normal density functions;
   c. Sample a value \( u \sim \text{uniform}(0,1) \). Update \( \theta_{j}^{D^{(s+1)}} \) to \( \theta_{j}^{D^{*}} \) if \( u < \rho \), and keep \( \theta_{j}^{D^{(s+1)}} \) to \( \theta_{j}^{D^{(s)}} \) if \( u > \rho \).

The above steps generate a new set of parameters. Repeating the procedures multiple times produces a sequence of parameters. As \( s \to \infty \), the sampling distribution of the generated parameters converges to the target joint posterior distribution.
Appendix D: Bayesian algorithm for the yield equation

The procedure of conducting Bayesian analysis of the skew-normal linear mixed effects model is similar to the Appendix B but with more parameters.

Decomposed skew-normal sampling models

Since a skew-normal distribution can be represented in a stochastic form shown in the Appendix A, the sampling distribution for \( y_j \) in (15) can be decomposed into two parts, as shown in Arellano-Valle et al. (2007):

\[
y_j | X_j^y, \theta_j^y, \sigma_j^2, \delta, z_j \sim \text{ind. Multi-Normal}_{n_j} \left( X_j^y \theta_j^y + \delta z_j, \sigma_j^2 I_{n_j} \right),
\]
\[
Z_j \sim \text{i.i.d. Trunc-Normal}_{n_j} \left( 0, I_{n_j} \right) 1 \{ z_j > 0 \},
\]
\[
(22)
\]

and the sampling distribution for \( \theta_j^v \) in (16) is similarly represented as:

\[
\theta_j^v | \beta^v, \Sigma^v, \Pi, v_j \sim \text{i.i.d. Multi-Normal}_p \left( \beta^v + \Pi v_j, \Sigma^v \right),
\]
\[
V_j \sim \text{i.i.d. Trunc-Normal}_p \left( 0, I_p \right) 1 \{ v_j > 0 \},
\]
\[
(23)
\]

where \( Z_j \) and \( V_j \) are latent parameters that illustrate the skewness of a distribution.

General form of the joint posterior distribution

The joint posterior distribution for all unknown parameters in the yield equation is derived according to the Bayes’ rule:
\[
p\left(\theta_i^y, \ldots, \theta_m^y, \beta^y, \sigma_y^2, \Sigma_y^y, \delta, \pi, z, \ldots, z_m, v, \ldots, v_m\right|y_1, \ldots, y_m, X_1^y, \ldots, X_m^y)
\]

\[
\propto p\left(y_1, \ldots, y_m \right| \theta_1^y, \ldots, \theta_m^y, \beta^y, \Sigma_y^y, \delta, z, \ldots, z_m, X_1^y, \ldots, X_m^y) \rightarrow \text{Multi-Normal sampling}
\]

\[
x \times p\left(\theta_i^y, \ldots, \theta_m^y \right| \beta^y, \Sigma_y^y, \pi, v, \ldots, v_m, X_1^y, \ldots, X_m^y) \rightarrow \text{Multi-Normal sampling}
\]

\[
x \times p\left(z, \ldots, z_m\right) \rightarrow \text{Trunc-Normal sampling}
\]

\[
x \times p\left(v, \ldots, v_m\right) \rightarrow \text{Trunc-Normal sampling}
\]

\[
x \times p\left(\beta^y, \sigma_y^2, \Sigma_y^y, \delta, \pi\right) \rightarrow \text{Priors}
\]

In the equation (24), components associated with \(y_j, z_j, \theta_j^y\), and \(v_j\) for all \(j\) are specified by their sampling distributions (22) and (23). The priors are discussed in the next subsection.

**Prior distributions**

Similar to the normal linear mixed effects model, we use non-informative conjugate priors for the skew-normal model. We assume that the priors are independent in the model such that

\[
p\left(\beta^y, \sigma_y^2, \Sigma_y^y, \delta, \pi\right) = p\left(\beta^y\right) p\left(\sigma_y^2\right) p\left(\Sigma_y^y\right) p\left(\delta\right) p\left(\pi\right).
\]

Following Arellano-Valle et al. (2007), the functional forms of the conjugate priors are:

\[
\begin{align*}
\beta^y &\sim \text{Multi-Normal}_p\left(\mu_0^y, \Lambda_0^y\right), \\
\sigma_y^2 &\sim \text{Inv-Gamma}\left(\frac{\nu_0^y}{2}, \frac{\nu_0^y \sigma_{y0}^2}{2}\right), \\
\Sigma_y^y &\sim \text{Inv-Wishart}\left(\eta_0^y, \mathbf{S}_0^y\right), \\
\delta &\sim \text{Normal}\left(\gamma_0, \Lambda_0^2\right), \\
\pi &\sim \text{Multi-Normal}_p\left(\xi_0, \Gamma_0\right).
\end{align*}
\]
The prior distributions plus the sampling distributions together generate the specific form for the joint posterior distribution (24).

We need to specify the parameters in the prior distributions that guarantee that they are weakly informative. For the prior of $\beta^Y$, we set $\mu_{0}^{Y}$ to be equal to the average of the individual OLS estimates across counties and $A_{0}^{Y}$ to be the sample covariance of the OLS estimates.

For the prior for $\sigma_{Y}^{2}$, we take $\nu_{0}^{Y} = 1$ and $\sigma_{Y0}^{2}$ to be the average of residual sum of squares from the OLS estimations, which guarantees that the prior distribution is somewhat diffuse.

We set $\eta_{0}^{Y} = p + 2$ and $S_{Y0}$ to be equal to the covariance of the OLS estimates. In this setting, the prior distribution for $\Sigma^{Y}$ is quite diffuse but has an expectation equal to $S_{Y0}$.

Finally, for the skewness parameters, $\delta$ and $\pi$, the priors are assumed to both follow normal distributions. We set $\gamma_{0} = 0$, $\lambda_{0}^{2} = 100$, $\xi_{0} = 0$, and $\Gamma_{0} = \text{diag}(100)$. Both prior distributions have large variances, indicating they are diffuse.

**Full conditional distributions**

The full conditional distributions consist of the distribution of each parameter conditional on all other parameters. Following the Proposition 2 in Arellano-Valle et al. (2007), we derive the full posterior conditional distributions for the yield equation:
\( \theta_j^\gamma \mid \beta^\gamma, \sigma^2_y, \Sigma^\gamma, \delta, \pi, z_j, v_j \sim \text{Multi-Normal}_p \left( \mu_j^\gamma, \Lambda_j^\gamma \right), \quad j = 1, \ldots, m, \) where

\[
\Lambda_j^\gamma = \left( \Sigma^\gamma \right)^{-1} + \frac{(X_j^\gamma)^T X_j^\gamma}{\sigma^2_y},
\]

\[
\mu_j^\gamma = \Lambda_j^\gamma \left( \Sigma^\gamma \right)^{-1} (\beta^\gamma + \Pi v_j) + \frac{(X_j^\gamma)^T (y_j - \delta z_j)}{\sigma^2_y};
\]

\( \beta^\gamma \mid \theta_1^\gamma, \ldots, \theta_m^\gamma, \Sigma^\gamma, \pi, v_1, \ldots, v_m \sim \text{Multi-Normal}_p \left( \mu^\beta_\beta, \Lambda^\beta_\beta \right), \) where

\[
\Lambda^\beta_\beta = \left( \Lambda^\gamma_0 \right)^{-1} + m \left( \Sigma^\gamma \right)^{-1},
\]

\[
\mu^\beta_\beta = \Lambda^\beta_\beta \left( \Lambda^\gamma_0 \right)^{-1} \mu^\gamma_0 + \left( \Sigma^\gamma \right)^{-1} \sum_{j=1}^{m} \left( \theta_j^\gamma - \Pi v_j \right);
\]

\( \sigma^2_y \mid \theta_1^\gamma, \ldots, \theta_m^\gamma, \delta, z_1, \ldots, z_m \sim \text{Inv-Gamma} \left( \frac{v_0^\gamma + \sum_{j=1}^{m} n_j}{2}, \frac{v_0^\gamma \sigma^2_{\tau 0} + \text{SSR}^\gamma}{2} \right), \) where

\[
\text{SSR}^\gamma = \sum_{j=1}^{m} (y_j - X_j^\gamma \theta_j^\gamma - \delta z_j)^T (y_j - X_j^\gamma \theta_j^\gamma - \delta z_j);
\]

\( \Sigma^\gamma \mid \theta_1^\gamma, \ldots, \theta_m^\gamma, \beta^\gamma, \pi, v_1, \ldots, v_m \sim \text{Inv-Wishart} \left( \eta_0^\gamma + m, \left( S_{\gamma 0} + S_{ym} \right)^{-1} \right), \) where

\[
S_{ym} = \sum_{j=1}^{m} \left( \theta_j^\gamma - \beta^\gamma - \Pi v_j \right) \left( \theta_j^\gamma - \beta^\gamma - \Pi v_j \right)^T;
\]

\( \delta \mid \theta_1^\gamma, \ldots, \theta_m^\gamma, \sigma^2_y, z_1, \ldots, z_m \sim \text{Normal} \left( \gamma_m, \lambda_m^2 \right), \) where

\[
\lambda_m^2 = \frac{1}{\lambda_0^2} + \frac{\sum_{j=1}^{m} z_j^T z_j}{\sigma^2_y},
\]

\[
\gamma_m = \frac{\lambda_0^2 \left( \frac{\gamma_0}{\lambda_0^2} + \frac{\sum_{j=1}^{m} z_j^T (y_j - X_j^\gamma \theta_j^\gamma)}{\sigma^2_y} \right)}{\lambda_0^2};
\]
\[ \pi \left| \theta_1^y, \ldots, \theta_m^y, \beta^y, \Sigma^y, v_1, \ldots, v_m \right. \sim \text{Multi-Normal}_p \left( \xi_\beta, \Gamma_\beta \right), \]

where

\[
\Gamma_\beta = \left( \Gamma_0 + \sum_{j=1}^m \text{diag} \left( v_j \right) \left( \Sigma^y \right)^{-1} \text{diag} \left( v_j \right) \right)^{-1},
\]

\[
\xi_\beta = \Gamma_\beta \left( \Gamma_0^{-1} \xi_0 + \sum_{j=1}^m \text{diag} \left( v_j \right) \left( \Sigma^y \right)^{-1} \left( \theta_j^y - \beta^y \right) \right);
\]

\[ Z_j \left| \theta_1^y, \ldots, \theta_m^y, \sigma^2_y, \delta \sim \text{Trunc-Normal}_{n_j} \left( \phi_{qj}, \Omega_{qj} \right) 1 \{ z_j > 0 \}, \]

where

\[
\Omega_{qj} = \left( 1 + \frac{\delta}{\sigma^2_v} \right) \mathbf{I}_{N_j},
\]

\[
\phi_{qj} = \left( \frac{\delta \left( y_j - X_j^y \theta_j^y \right)}{\sigma^2_y} \right);
\]

\[ V_j \left| \theta_1^y, \ldots, \theta_m^y, \beta^y, \Sigma^y, \pi \sim \text{Trunc-Normal}_{n_j} \left( \phi_{qj}, \Omega_{qj} \right) 1 \{ v_j > 0 \}, \]

where

\[
\Omega_{qj} = \left( I_p + \Pi \left( \Sigma^y \right)^{-1} \Pi \right),
\]

\[
\phi_{qj} = \left( \Pi \left( \Sigma^y \right)^{-1} \left( \theta_j^y - \beta^y \right) \right).
\]

**Gibbs sampling algorithm**

Since the full conditional distributions are obtainable, we can implement Gibbs sampling algorithm to approximate the joint posterior distribution. The algorithm is similar to the one in the Appendix B but with more steps.

Given current values of the set of all parameters \( \Theta^{(s)} \) at scan \((s)\), new values are iteratively generated through the following steps:

1. For each \( j = 1, \ldots, m \), sample \( \theta_j^{(s+1)} \sim p \left( \theta_j^y \left| \beta^{(s)}, \sigma^2_y, \Sigma^y, \delta^{(s)}, \pi^{(s)}, z_j^{(s)}, v_j^{(s)} \right. \right) \) according to the posterior distributions in (25).
2. Sample $\beta^{Y_{(s+1)}} \sim p\left(\beta^{Y_{(s+1)}} \mid \theta^{Y_{(s+1)}}_{1}, \ldots, \theta^{Y_{(s+1)}}_{m}, \Sigma^{Y_{(s)}}, \pi^{(s)}, v^{(s)}_{1}, \ldots, v^{(s)}_{m}\right)$ according to the posterior distribution shown in (26).

3. Sample $\sigma^{2Y_{(s+1)}} \sim p\left(\sigma^{2Y_{(s+1)}} \mid \theta^{Y_{(s+1)}}_{1}, \ldots, \theta^{Y_{(s+1)}}_{m}, \beta^{Y_{(s+1)}}, \pi^{(s)}, v^{(s)}_{1}, \ldots, v^{(s)}_{m}\right)$ according to (27).

4. Sample $\Sigma^{Y_{(s+1)}} \sim p\left(\Sigma^{Y_{(s+1)}} \mid \theta^{Y_{(s+1)}}_{1}, \ldots, \theta^{Y_{(s+1)}}_{m}, \beta^{Y_{(s+1)}}, \pi^{(s)}, v^{(s)}_{1}, \ldots, v^{(s)}_{m}\right)$ based on the posterior distribution (28).

5. Sample $\delta^{(s+1)} \sim p\left(\delta \mid \theta^{Y_{(s+1)}}_{1}, \ldots, \theta^{Y_{(s+1)}}_{m}, \sigma^{2Y_{(s+1)}}, \Sigma^{Y_{(s)}}, \pi^{(s)}, v^{(s)}_{1}, \ldots, v^{(s)}_{m}\right)$ in accordance with (29).

6. Sample $\pi^{(s+1)} \sim p\left(\pi \mid \theta^{Y_{(s+1)}}_{1}, \ldots, \theta^{Y_{(s+1)}}_{m}, \beta^{Y_{(s+1)}}, \Sigma^{Y_{(s+1)}}, v^{(s)}_{1}, \ldots, v^{(s)}_{m}\right)$ based on the distribution shown in (30).

7. For each $j = 1, \ldots, m$, sample $Z^{(s+1)}_{j} \sim p\left(Z_{j} \mid \theta^{Y_{(s+1)}}_{1}, \ldots, \theta^{Y_{(s+1)}}_{m}, \sigma^{2Y_{(s+1)}}, \delta^{(s+1)}\right)$ based on the posterior distributions in (31).

8. For each $j = 1, \ldots, m$, sample $V^{(s+1)}_{j} \sim p\left(V_{j} \mid \theta^{Y_{(s+1)}}_{1}, \ldots, \theta^{Y_{(s+1)}}_{m}, \beta^{Y_{(s+1)}}, \Sigma^{Y_{(s+1)}}, \pi^{(s+1)}\right)$ according to (32).

Iterating the previous steps for a large number generates a sequence of parameters that approximates the joint posterior distribution. Inferences of the parameters can be made based on the sampling values.
References


Torrence, C., and Webster, P. J. (1999). Interdecadal changes in the ENSO-monsoon system. *Journal of Climate*, 12(8), 2679-2690.


