Climate Multi-model Regression Using Spatial Smoothing

Karthik Subbian* Arindam Banerjee*

Abstract

There are several Global Climate Models (GCMs) reported by various countries to the Intergovernmental Panel on Climate Change (IPCC). Due to the varied nature of the GCM assumptions, the future projections of the GCMs show high variability which makes it difficult to come up with confident projections into the future. Climate scientists combine these multiple GCMs to minimize the variability and the prediction error. Most of these model combinations are specifically for a location, or at a global scale. They do not consider regional or local smoothing (including the IPCC model). In this paper, we address this problem of combining multiple GCM outputs with spatial smoothing as an important desired criterion. The problem formulation takes the form of multiple least squares regression for each geographic location with graph Laplacian based smoothing amongst the neighboring locations. Unlike existing Laplacian regression frameworks, our formulation has both inner and outer products of the coefficient matrix, and has Sylvester equations as its special case. We discuss a few approaches to solve the problem, including a closed-form by solving a large linear system, as well as gradient descent methods which turn out to be more efficient. We establish the superiority of our approach in terms of model accuracy and smoothing compared to several popular baselines on real GCM climate datasets.

1 Introduction

The rise in global temperatures, frequent natural disasters, and rising sea levels have made the problem of predicting these global climate phenomena[1] a matter of prime importance. Geophysists, climate scientists and meteorologists create complex mathematical models to represent such nonlinear climate systems. They are run as computer simulations, to predict climate variables such as temperature, pressure, and precipitation over multiple centuries. These models have different uncertainties associated with them and are grouped into three major categories: initial conditions, boundary conditions, parameter uncertainties [41]. Resolving all these uncertainties and defining a true climate system is highly complex and nearly impossible.

Every group of climate modelers come up with a different representation of the climate processes that explains the choice of initial, boundary and parameter condition assumptions. Each of these models, based on their input variations, produce different Global Climate models (GCM)[2] and are reported to the Intergovernmental Panel on Climate Change (IPCC) [4]. These models are shown to have high variance in their forecast of the future climate parameters [33] (like pressure, temperature, precipitation, etc.). The main reason for such uncertainty in the response of the GCMs is due to the model variability and can be greatly reduced by combining multiple GCM outputs [33].

There are variety of areas where ensemble of models have shown to be useful in increasing the overall predictive skill, in addition to reliability and consistency of predictions. Some examples are models forecasts in the field of public health [44] and agriculture [15] where combined approaches are shown to be much superior compared to any single forecast model. In El Nino Southern Oscillation (ENSO) weather and climate related predictions are found to perform better using multi-model ensembles than using single model for forecast [34].

Currently, there are several ways in which climate scientists combine these GCM model outputs. For example, IPCC [4] uses a global model average, where each model output is given an equal weightage. Another popular approach for combining model outputs, especially popular in the context of weather forecasting, is *Superensemble* [22]. In this approach, an ordinary least squares regression is performed on the bias removed model outputs for each geographic location. The regression coefficients are allowed to be both positive and negative in this case, unlike the model average.

An issue with this approach [22], since there is no regularization, is that the combined model tends to overfit the training data, and hence may not be general enough to characterize a new test data point. Due to this reason, the climate variables in each location may vary significantly for the new test data point, and may result in significant deviation from its neighbors. $Superensemble$ like methods [23, 22, 24] do not guarantee the regression parameters in neighboring locations to be smooth. This results in climate variables varying abruptly amongst the neighboring locations for the test data.

In this paper, we address such problem of *combining*

*University of Minnesota, Minneapolis, MN 55455. {karthik,banerjee}@cs.umn.edu

[1] The term weather often relates to short term period of 7 to 10 days and the term climate refers to long term phenomena averaged over a long period.

[2] Also referred as Global Circulation Models or Atmospheric Ocean General Circulation Models (AOGCMs).

the GCM model outputs with spatial smoothing as a multiple least squares regression problem, where each geographic location has its own least squares model. To achieve the smoothing across the neighboring locations we use a graph Laplacian based regularizer. Unlike the existing Laplacian regression frameworks [10], our formulation has both inner and outer products of the coefficient matrix, and has Sylvester equations as its special case. We discuss a few approaches to solve the problem, including a closed-form by solving a large linear system, as well as gradient descent methods which turn out to be more efficient. We also propose a distributed approach using Alternating Direction Method (ADM). Finally, we empirically evaluate our approach using various error and smoothing measures on GCM climate datasets. We show that our iterative model outperforms the baseline methods in terms of root mean square error, while maintaining excellent smoothing across the geographical neighborhood.

The rest of the paper is organized as follows. In Section 2, we give a brief overview of the related work. in Section 3, we explain our notations, model and solution to the climate multi-model regression problem. In Section 4, we discuss the climate data sets, evaluation measure, and the experimental results to compare our proposed approach against several baselines, and conclude in Section 5.

2 Related Work

The problem of spatial smoothing using regression models is well studied, especially in spatial data mining, using two popular approaches [39]: (a) Spatial Auto Regression (SAR) [7] and (b) Geographically Weighted Regression [14, 29]. In SAR models, the spatial dependencies of the dependent variable is directly modeled in the regression objective. The original regression equation $y = \mathbf{X}\beta$ is modified as $y = \rho\mathbf{W}y + \mathbf{X}\beta + \epsilon$, where $\mathbf{W}$ is the contiguity matrix denoting the neighborhood relationship and ρ is a constant parameter that decides the strength of spatial dependency amongst the elements of y, the dependent variable. Note that, the SAR approach corresponds to a single regression model with smoothing over the entire space. In our case, we have separate models for each spatial point and we need to ensure smoothness amongst multiple model parameters. The SAR approach does not address this problem.

One of the limitations of SAR approach is that, it does not account for the underlying spatial heterogeneity. The Geographically Weighted Regression (GWR) [14, 29] was proposed to address this issue using a spatially varying model parameter $\beta(s)$. They used a weighted linear regression model where the parameters in geographic location are weighed using a distance decay function. The regression model at a specific location s is given by $y = \mathbf{X}\beta(s) + \epsilon$. In this approach, the calibration of model parameters in a neighborhood is pre-defined, such as a distance decay function, and not general enough as in our case.

In a parallel thread, there are several approaches considered by climate scientists to address the GCM combination problem. These models can be broadly grouped in to three categories: (1) Weighted averaging models [22, 23, 47, 18, 26] (2) Bayesian models [42, 43, 18, 35] and (3) Online learning models [31, 27, 21, 8].

In the weighted averaging models, the approaches vary in the way the weights are chosen. For instance, the weights in [24] are chosen to fit a least squares model, and in [26] the weights are exponentially decaying over the neighboring locations. The farther the point, the lesser the significance of weighting. The main challenge in this approach is choosing the decay function corresponding to each location is quite tedious [26]. In Bayesian models [42, 43], the current and future climate parameters are treated as random variables with a prior probability distribution. The likelihood component of the model specifies the conditional distribution of the data, given the parameters. The posterior is obtained by combining the prior and likelihood of the model. In the online learning models [31], each GCM is modeled as an expert and for every time instance experts give predictions. The online learning approach takes one data point at a time and updates its confidence (weights) for each expert based on the accuracy of the recent prediction.

We note that, *none* of these models address the problem of *combining multiple GCM model outputs* at each location, with *spatial smoothing* across the model parameters in neighboring locations.

3 Spatially Smooth Multi-Model Regression

In this section, we will describe our model Spatially Smooth Multi-Model Regression (S^2M^2R) and discuss the closed form and iterative solutions to the problem. We briefly discuss the alternating direction method (ADM) to scale the solution to large data sets. Finally, we show that our model is widely applicable as it is a generalization of well-known Sylvester equations which is finding increasing applications in multivariate ML models [11] in control theory.

3.1 Notation We use bold capital letters to denote a matrix, such as $\mathbf{X}$. Any vector is denoted by a bold small letter $\mathbf{x}$. A scalar is denoted by lower or upper case letter, x or X. If $\mathbf{X}$ is a matrix, a bold face $\mathbf{x_i}$ denotes the i-th column of matrix $\mathbf{X}$. x_{ij} denotes the (i,j)-th element of matrix $\mathbf{X}$. The iterations are denoted using a superscript in braces, such as $x^{(t)}$ denotes the value of x in iteration t.

3.2 Model Let m be the number of climate variables of interest, e.g., temperature, pressure, precipitation, wind speed, etc., T be the number of periods of observations indexed by t, e.g., at monthly resolution, $t = 1$ corresponds to Jan 1960, $t = 2$ corresponds to Feb 1960, etc., and k be

the number of climate models under consideration. From each model, we have a value for each climate variable at each period at a number of geographical locations. Let N be the total number of such locations indexed by i, so that the model outputs at location i can be represented as a matrix $\mathbf{A}_i \in \mathbb{R}^{mT\times k}$. We also have the true observation vector for the m variables for T periods at each location i, represented by $\mathbf{b}_i \in \mathbb{R}^{mT\times 1}$.

The goal is to develop a regression model which combines the model outputs to get a prediction which is close to the true value. One option is to consider a separate regression model for each location i. Let the regression parameters for each location i be denoted by $\theta_i \in \mathbb{R}^{k\times 1}$, so that $\mathbf{b}_i \approx \mathbf{A}_i\theta_i$. The simplest approach to accomplish this is to consider a least-squares linear regression formulation, also known as the ordinary least squares (OLS) [25]. For OLS, at each location i, we consider the following problem

$$\min_{\theta_i} \frac{1}{2}\left\|\mathbf{A}_i\theta_{\mathbf{i}} - \mathbf{b}_i\right\|_2^2, i = 1, ..., N. \tag{3.1}$$

There are two potential issues with the OLS formulation: (i) Since there is no regularization, the parameters can over-fit the training data and not perform well on the test data, and (ii) Regression parameters in neighboring locations can be different, which is not meaningful for a spatial regression problem. Our formulation addresses both issues based on a spatially smooth regularization, which we now describe.

Let $G = (V, E)$ be a weighted graph, where there is a vertex v_i corresponding to every location i so that $|V| = n$. The edges are between neighboring locations, for a suitable definition of (spatial) neighborhood. The weights w_{ij} on the edges $(i, j) \in E$ captures the proximity of the two locations, i.e., w_{ij} is high for immediate neighbors and low/zero for far away locations. Let $\mathbf{D}$ be a diagonal matrix with $D_{ii} = \sum_j w_{ij}$. Then, the unnormalized graph Laplacian corresponding to the graph is given by $\mathbf{L} = \mathbf{D} - \mathbf{W}$ [11][10]. One of the key properties of the Laplacian in the context of regression on graph is as follows: for $\mathbf{f} \in \mathbb{R}^n$, where f_i is the value for vertex v_i, we have [11][10]

$$\sum_{(i,j)\in E} w_{ij}(f_i - f_j)^2 = \mathbf{f}^T\mathbf{L}\mathbf{f}\ . \tag{3.2}$$

For our current purposes, we consider a spatial regularization term which makes sure regression parameters in neighboring locations are similar. In particular, we consider the following regularizer:

$$\sum_{(i,j)\in E} w_{ij}\|\theta_i - \theta_j\|_2^2 = \mathrm{Tr}(\Theta\mathbf{L}\Theta^T)\ , \tag{3.3}$$

where $\Theta \in \mathbb{R}^{k\times N}$ is the combined regression parameter matrix so that

$$\Theta = [\theta_1\ \cdots\theta_N]\ . \tag{3.4}$$

The neighborhoods for regression can be obtained from a simple 4-nearest neighbor, as shown in Figure 1, or in general, one consider an appropriate neighborhood structure based on domain knowledge. With $\lambda \geq 0$ as the regularization constant, the objective function for S^2M^2R is given by

$$\min_{\Theta} \frac{1}{2}\sum_{i\in V}\|\mathbf{A}_i\theta_i - \mathbf{b}_i\|_2^2 + \frac{\lambda}{2}\sum_{(i,j)\in E} w_{ij}\left\|\theta_i - \theta_j\right\|_2^2 \tag{3.5}$$

$$= \min_{\Theta} \frac{1}{2}\sum_{i\in V}\|\mathbf{A}_i\theta_i - \mathbf{b}_i\|_2^2 + \frac{\lambda}{2}\mathrm{Tr}(\Theta\mathbf{L}\Theta^T)\ . \tag{3.6}$$

The model is general enough to include multiple climate variables, and capable of doing spatial smoothing using any given neighborhood graph. For example, in Figure 2, the blue and yellow circles represent land and sea surface temperatures, and the figure illustrates the neighborhood structure among the locations.

The formulation in (3.6) looks quite similar to the widely studied and by now mature literature in learning with Laplacian regularization [10]. In spite of the apparent similarities, as we now show, the structure of the problem is rather different. In fact, the problem can be seen as a generalization of ideas which have appeared in the literature on control theory, in particular involving the Sylvester (or Lyapunov) equations [9][20][45]. In the context of data analysis, Sylvester equations naturally show up while trying to capture covariance structures along both rows and columns of a data matrix. Note that for a $d \times n$ data matrix, for n data points in d dimensions, classical methods such as PCA only focus on $d \times d$ covariances over the rows (dimensions) and kernel methods focus on $n \times n$ covariances over the columns (data points). Models based on Sylvester equations [5][6] can account for both covariance structures simultaneously.

For ease of exposition, assume $\mathbf{A}_i = \mathbf{A}, \forall i$. Then, (3.6) takes the form

$$\min_{\Theta} \frac{1}{2}\mathrm{Tr}(\Theta^T\mathbf{K}\Theta) + \frac{1}{2}\mathrm{Tr}(\Theta\mathbf{L}\Theta^T) - \mathrm{Tr}(\Theta^T\mathbf{C})\ , \tag{3.7}$$

where

$$\mathbf{K} = \mathbf{A}^T\mathbf{A}\ , \qquad \mathbf{C} = [\mathbf{c}_1\ \cdots\mathbf{c}_N]\ , \quad \mathbf{c}_i = \mathbf{A}^T\mathbf{b}_i\ . \tag{3.8}$$

Taking gradient w.r.t. Θ, we have the following matrix equation:

$$\mathbf{K}\Theta + \Theta\mathbf{L} = \mathbf{C}\ , \tag{3.9}$$

which is the Sylvester equation [9][20]. It is important to note that, unlike simple Laplacian regularization, in the Sylvester equation, the rows of Θ get coupled by the (local) kernel $\mathbf{K}$ derived from the design matrix, and the columns of

Θ get coupled by the spatial regularization from the graph Laplacian $\mathbf{L}$. In our formulation, the design matrices $\mathbf{A}_i$ are different in every location i, and hence (3.6) is more general than (3.7). Now, we will focus on some approaches for solving the optimization problem in (3.6).

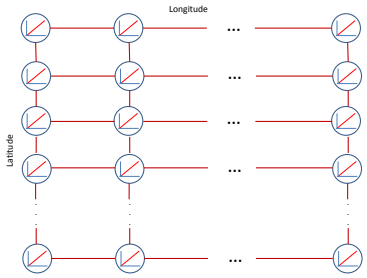


Figure 1: A graph representing the regional relationships. Each circle represent a geographic location with a linear regression model. The edges of the graph represents the relationship between the learned parameters of each region.

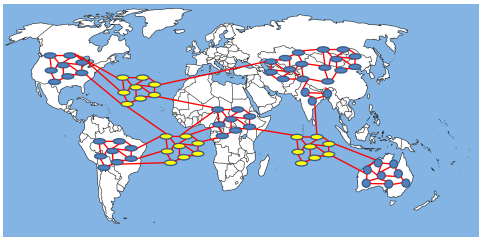


Figure 2: An illustrative example of world map depicting the models of the land and sea surface temperature in each geographic location in blue and yellow circles respectively. The edge connecting the circles corresponds to the required smoothing between the model parameters in two regions.

3.3 S^2M^2R-CF: A Closed Form Solution Since (3.6) can be written as a quadratic form in $vec(\Theta) \in \mathbb{R}^{Nk\times 1}$, the problem can in principle be solved as a large linear system. First, we construct the following matrices:

$$\text{vec}(\Theta) = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_N \end{bmatrix} \quad \text{vec}(\mathbf{C}) = \begin{bmatrix} \mathbf{A_1}^T\mathbf{b_1} \\ \mathbf{A_2}^T\mathbf{b_2} \\ \vdots \\ \mathbf{A_N}^T\mathbf{b_N} \end{bmatrix}$$

$$\bar{\mathbf{A}} = \begin{bmatrix} \mathbf{A_1}^T\mathbf{A_1} & 0 & 0 & 0 \\ 0 & \mathbf{A_2}^T\mathbf{A_2} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \mathbf{A_N}^T\mathbf{A_N} \end{bmatrix}.$$

Then, (3.6) can be written as:

$$\min_{\Theta} \frac{1}{2}\text{vec}(\Theta)^T\bar{\mathbf{A}}\,\text{vec}(\Theta) - \text{vec}(\Theta)^T\,\text{vec}(\mathbf{C}) + \frac{\lambda}{2}\text{vec}(\Theta)^T\mathbf{P}(\mathbf{L}\otimes\mathbb{I}_k)\mathbf{P}^T\,\text{vec}(\Theta)\,, \tag{3.10}$$

where $\otimes$ denotes the Kronecker product, $\mathbb{I}_k$ is the $k \times k$ identity matrix, and $\mathbf{P}$ is a permutation matrix that converts the column stacked arrangement of $\text{vec}(\mathbf{\Theta})$ to a row stacked arrangement. Taking derivatives w.r.t. $\text{vec}(\Theta)$, we obtain the following linear system:

$$(\bar{\mathbf{A}} + \lambda\mathbf{P}(\mathbf{L}\otimes\mathbb{I}_k)\mathbf{P}^T)\,\text{vec}(\Theta) = \text{vec}(\mathbf{C})\,. \tag{3.11}$$

Although this is a large linear system, since $(\bar{\mathbf{A}} + \lambda\mathbf{P}(\mathbf{L}\otimes\mathbb{I}_k)\mathbf{P}^T)$ is a sparse matrix, one can take advantage of solvers for sparse linear systems to obtain $\text{vec}(\Theta)$ [19]. Another approach is to use a low-rank SVD of $(\bar{\mathbf{A}}+\lambda\mathbf{P}(\mathbf{L}\otimes\mathbb{I}_k)\mathbf{P}^T)$, in order to get an approximation of $\text{vec}(\Theta)$. In particular, if $\mathbf{U}\Sigma\mathbf{V}^T$ is a low-rank SVD, then we get:

$$\mathbf{U}\Sigma\mathbf{V}^T\,\text{vec}(\Theta) = \text{vec}(C) \tag{3.12}$$

$$\Rightarrow \quad \text{vec}(\Theta) = \mathbf{V}\Sigma^\dagger\mathbf{U}^T\,\text{vec}(C)\,, \tag{3.13}$$

where $\Sigma^\dagger$ is the pseudo-inverse of Σ, i.e., since Σ is a diagonal matrix, $\Sigma^\dagger$ has $1/\sigma_{ii}$ for the non-zero singular values, and 0 for the others. We experiment with this approximate solution in Section 4.

3.4 S^2M^2R-GD: A Gradient Descent Approach We can also solve the problem by a direct gradient descent approach. In particular, for a suitable learning rate γ, the updates are of the form:

$$\text{vec}(\Theta)^{(t+1)} = \text{vec}(\Theta)^{(t)} - \gamma((\bar{\mathbf{A}}+\lambda\mathbf{P}(\mathbf{L}\otimes\mathbb{I}_k)\mathbf{P}^T)\,\text{vec}(\Theta)^{(t)} - \mathbf{C})\,. \tag{3.14}$$

Since the objective is quadratic, the gradient descent approach is guaranteed to converge to the global minimum at a linear rate [11]. With proper choice of (adaptive) learning rate, the method can be quite fast in practice.

The gradient update can be slow for really large scale problems since once has to perform a matrix-vector multiplication in each iteration between a $Nk \times Nk$ matrix and $Nk \times 1$ vector. For such scenarios, one can consider the Alternating Direction Method of Multipliers (ADMM) [12][13]. First, note that the problem in (3.6) can be written in the following form:

$$\min_{\substack{\Theta\in\mathbb{R}^{k\times N}, \mathbf{V}\in\mathbb{R}^{N\times k} \\ \Theta=\mathbf{V}^T}} \frac{1}{2}\sum_{i=1}^{N}\|\mathbf{A}_i\theta_i - \mathbf{b}_i\|^2 + \frac{\lambda}{2}\text{Tr}(\mathbf{V}^T L\mathbf{V})\,, \tag{3.15}$$

where $\mathbf{V} = [\mathbf{v}_1\ \cdots\ \mathbf{v}_k]$, where $\mathbf{v}_j \in \mathbb{R}^N$. With Lagrange multipliers $\Lambda \in \mathbb{R}^{k\times N}$, the augmented Lagrangian of the

above problem is:

(3.16)
$$\mathcal{L}(\Theta, \mathbf{V}, \Lambda) = \frac{1}{2}\sum_{i=1}^{N} \|\mathbf{A}_i\theta_i - \mathbf{b}_i\|^2 + \frac{\lambda}{2}\operatorname{Tr}(\mathbf{V}^T\mathbf{L}\mathbf{V}) + \langle \Lambda, \Theta - \mathbf{V}^T\rangle + \frac{\rho}{2}\|\Theta - \mathbf{V}^T\|_F^2 ,$$

where $\rho > 0$ is a constant, and $\langle \cdot, \cdot \rangle$ denotes inner product, i.e., $\langle \Lambda, \Theta - \mathbf{V}^T\rangle = \operatorname{vec}(\Lambda)^T \operatorname{vec}(\Theta - \mathbf{V}^T)$. The inner product notation is convenient as it makes the notation compact. Then, the ADM updates are given by:

$$\theta_i^{(t+1)} = \operatorname{argmin}_{\theta_i} \{\frac{1}{2}\|\mathbf{A}_i\theta_i - \mathbf{b}_i\|^2 + \langle \Lambda_{:,i}^{(t)}, \theta_i\rangle + \frac{\rho}{2}\|\theta_i - \mathbf{V}_{i,:}^{T\,(t)}\|^2\} \tag{3.17}$$

$$\mathbf{v}_j^{(t+1)} = \operatorname{argmin}_{\mathbf{v}_j} \{\frac{\lambda}{2}\mathbf{v}_j^T\mathbf{L}\mathbf{v}_j - \langle \Lambda_{j,:}^{(t)}, \mathbf{v}_j\rangle + \frac{\rho}{2}\|\Theta_{j,:}^{(t+1)^T} - \mathbf{v}_j\|^2\} \tag{3.18}$$

$$\Lambda^{(t+1)} = \Lambda^{(t)} + \rho(\Theta^{(t+1)} - \mathbf{V}^{(t+1)^T}). \tag{3.19}$$

The update in (3.17) is straightforward, and (3.19) is closed form. For the update in (3.18), one can linearize the objective, and do a proximal update, i.e., in other words, do a gradient descent to get $\mathbf{v}_j^{(t+1)}$ where one computes the gradient of the objective in (3.18), evaluated at $\mathbf{v}_j^{(t)}$. One can show that the inexact ADMM updates, with linearization of one or more updates, will have the same rate of convergence as full ADMM [12]. Perhaps more importantly, (3.17) allows parallel update of the columns of Θ, (3.18) allows parallel updates of the columns of $\mathbf{V}$ (equivalent to rows of Θ), and (3.19) is elementwise parallel.

Thus, the proposed model can be scaled to really large datasets. For GCM model outputs, the standard gradient descent was sufficiently fast. Finally, note that one can do block coordinate descent, but each descent step has to solve a Sylvester equation. Further, it will not be parallel and hence may be slow for large scale problem.

4 Experimental Results

In this section, we will present a number of experimental results illustrating the effectiveness of the proposed technique. We will verify the accuracy and the spatial smoothness achieved by our method compared to the baselines using 10-fold cross validation.

4.1 GCM Datasets The Global Climate Model (GCM) CMIP3 [28] data set can be downloaded from [2]. We downloaded the surface temperature monthly runs for the following eight GCM model outputs: (1) BCCR_BCM2, (2) CCMA_CGCM3, (4) MIROC3-2-HIRES, (4) CNRM_CM3, (5) GFDL_CM2_1, (6) GISS_E_H, (7) INGV_ECHAM4, and (8) IPSL. In Table 1, we have listed the GCM models used, their origin and reference. The true observations for surface temperature from 1901 to 2000 is downloaded from Climatic Research Unit (CRU) [1]. The first challenge in combining these models is to synchronize the models to have *same spatial and temporal resolution*. Each model provides the climate measurements such as pressure, temperature, precipitation, etc. in different frequencies (monthly, weekly, daily) and different spatial resolution (latitude and longitude). We used spatial interpolation techniques [46] to first synchronize the data to same spatial frequency. For the experiments described in this paper, we used a spatial resolution of $2.5^\circ \times 2.5^\circ$ in latitude and longitude respectively, resulting in 73 x 144 points on the globe. The temporal frequency of the data downloaded was monthly for all 8 data sets and we offset the models appropriately to match the start date (December 31, 1901) of the observational (CRU) data set. We found there were 549 months of data available after the offset procedure that were common amongst all 8 GCM outputs.

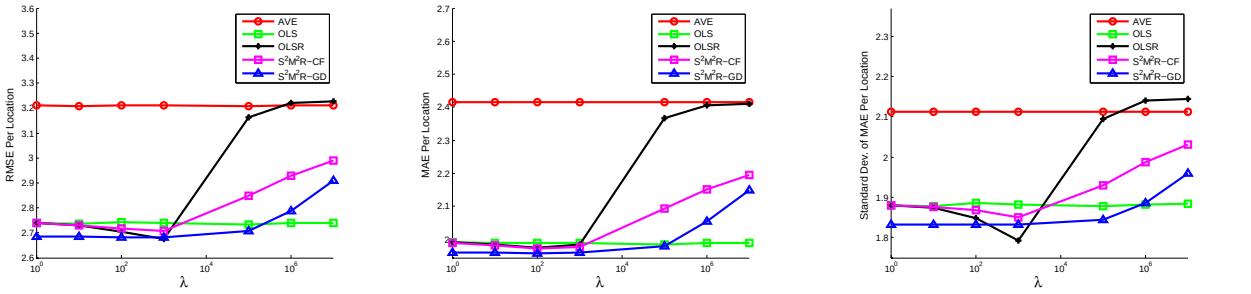
4.2 Baselines and Evaluation Measures We use the following three baselines for our evaluation:

- **Average Model:** Our first baseline is a popular technique used by IPCC [4]. This model is a simple average of GCM model outputs, where all models are given equal weightage. We refer to this model as *AVE* in our experiments.
- **Linear Regression:** This baseline is a Ordinary Least Squares (OLS) regression for each geographic location with no regularization. This model is a special case of Eqn. (3.11), where $\lambda = 0$. We refer to this model as *OLS* in our experiments.
- **Linear Regression with Ridge Regularization:** For this baseline, we have added a squared 2-norm ridge regularizer to the objective of *OLS*. This is also a special case of as (3.11), where $\mathbf{L} = \mathbf{I}$. We refer to this model as *OLSR*.

We have two sets of evaluation measures, one is to measure the model accuracy and the other is to measure the spatial smoothness. We use the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) along with the standard deviation of MAE to measure the accuracy of our model. All these measures were computed per location per variable. We evaluate the smoothness achieved by our model using the Laplacian regularization parameter $\operatorname{Tr}(\mathbf{\Theta}\mathbf{L}\mathbf{\Theta}^T)$ and correlation coefficients. We use Kendall Tau and Spearman correlation coefficient for this purpose. We denote the average Kendal Tau and Spearman correlation [32] scores across all geographic neighbors by *KT* and *SC*

GCM	Origin	Reference
BCCR_BCM2	Bjerknes Centre for Climate Research, Norway	[17]
CCMA_CGCM3	Canadian Centre for Climate Modelling and Analysis, Canada	[38]
MICRO3-2-Hires	Center for Climate System Research, Univ. of Tokyo, Japan	[30]
CNRM_CM3	Center for National Weather Research, France	[37]
GFDL_CM2_1	Geophysical Fluid Dynamics Laboratory, USA	[16]
GISS_E_H	Goddard Institute for Space Studies, USA	[36]
INGV_ECHAM4	European Center for Medium-Range Weather Forecasts, UK	[3]
IPSL	Institut Pierre Simon Laplace, France	[40]

Table 1: Various Global Climate Models (GCMs), their origin and reference.



(a) RMSE value per location (b) MAE value per location (c) MAE standard deviation per location

Figure 3: Model accuracy of our proposed approach S^2M^2R-GD measured in terms of RMSE and MAE and compared against the baseline methods. Our gradient descent version of the solution does better than *AVE* and *OLS*.

respectively. We use correlation coefficients, in order to measure the correlation of weights (θ_i and θ_j) between the neighboring points (i and j). This measure gives us the preferential ordering similarity of the eight GCM models in each location compared to their neighbors, a correlation of 1 corresponds to perfect similarity and -1 corresponds to complete dissimilarity.

4.3 Accuracy Results The RMSE measured per location per variable for different values of λ is shown in Figure 3(a). In this figure, our proposed method S^2M^2R-GD performs extremely well compared to all other baselines. The RMSE per location for the popular *AVE* model is $3.2°C$, where as for our model it is around $2.7°C$. However, as λ increases to a large value, we see that the squared error term increases gradually towards the *AVE* baseline. Note that *OLSR* model performs well as λ increases as the generalization of the model helps until a certain value of λ. As the λ value increases further the *RMSE* converges towards the *AVE* model. In case of our model, when the λ tends to a large value, the spatial smoothing constraint ensures that the weights are uniform across the entire Earth's surface resulting in an increase in the RMSE value. Even with large smoothing our RMSE errors are low compared to the *AVE* model. Note that our approximate closed form solution in (3.11) is not performing as good as our iterative solution. The reason for this is the numerical instability of the low-rank SVD approximation approach in solving the large linear system described in (3.11). The RMSE per location for *AVE* and S^2M^2R-GD is shown in Figure 4. One can clearly see, that our model has much lower RMSE across the entire globe, compared to the *AVE* model, almost a difference of $0.5°C$.

In Figure 3(b) and 3(c), we show the MAE measured per location and its standard deviation for various methods. It is consistent with the RMSE measurements, discussed earlier. For relatively small values of λ the change in MAE and standard deviation is less and our method is quite stable. As λ increases to larger values, the error increases along with standard deviation. One possibility is that the spatial relationship of these climate parameters covers only small geographic regions, and performing smoothing to large regions (say an entire continent) may actually reduce the accuracy of the model combination. This spatial smoothing coverage is controlled by the λ parameter in our model.

4.4 Smoothing Results To evaluate the smoothing of our model we first measure the spatial regularizer term $\text{Tr}(\mathbf{\Theta L \Theta^T})$ for our model. We find that as λ increases, the smoothing regularizer term decreases in value, i.e. the co-

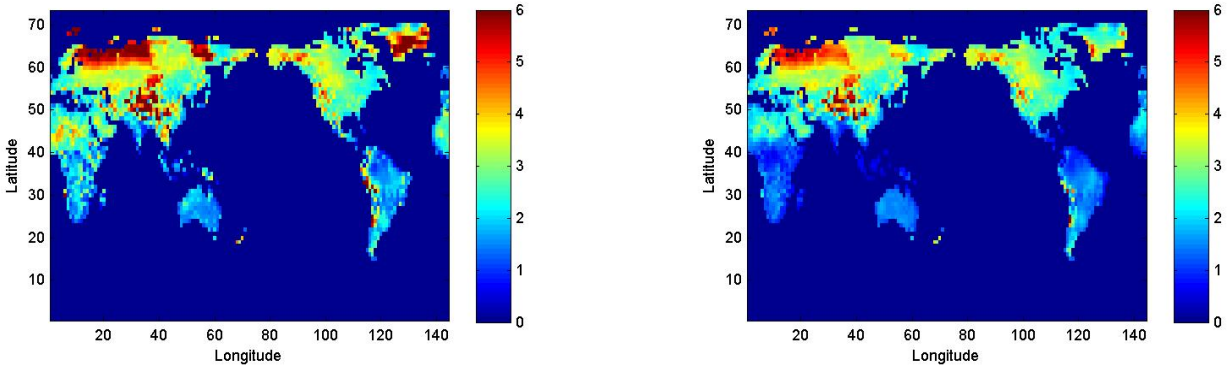


Figure 4: The RMSE value for surface temperature for all of Earth's surface locations measured for the AVE model (*left*) and S^2M^2R-GD (*right*). The S^2M^2R-GD model has smaller errors in comparison to the AVE model.

efficients become more smooth spatially. This is in order to balance the overall minimization objective and ensuring there is a high spatial correlation amongst the model parameters of neighboring locations.

$\log(\lambda)$	1	2	3	4	5
$\text{Tr}(\mathbf{\Theta} L \mathbf{\Theta}^T)$	473.6801	281.5237	89.7391	2.4754	0.1694

Table 2: Smoothness measured using $\text{Tr}(\mathbf{\Theta} L \mathbf{\Theta}^T)$.

Next, we measure the average correlation for 4363 edges found in the neighborhood graph ($\mathbf{W}$). The higher the correlation score, better the smoothness with the neighboring location in terms of the relative model preference order. From Table 3, we see that as λ increases, the smoothness increases in terms of the KT correlation. In similar lines, we also measured the SC correlation in Table 4, and consistently, we see that as the regularization parameter λ increases, the smoothness in terms of SC increases. The correlations for *OLS* stay low for the entire range as expected, whereas that for *OLSR* improves with λ.

$\log(\lambda)$	OLS	OLSR	S^2M^2R-CF	S^2M^2R-GD
0	0.6185	0.6187	0.6215	0.5605
1	0.6140	0.6160	0.6366	0.5590
2	0.6132	0.6197	0.7047	0.5602
3	0.6122	0.6232	0.8213	0.5604
5	0.6155	0.7244	0.9727	0.6328
6	0.6156	0.7483	0.9933	0.8121
7	0.6153	0.7504	0.9984	0.9360
8	0.6144	0.7502	0.9998	0.9820

Table 3: Average Kendall Tau correlation coefficient (KT) measured over 4363 neighboring locations in the neighborhood graph.

Finally, we plot the best model in each location which has the highest model weight for OLS and S^2M^2R-GD in Figure 5. It is evident that OLS does not have sufficient spatial smoothing, as the neighboring locations have completely different best models, especially in Asia and the Americas. In contrast, our approach creates smooth regions denoting the best GCM models for each spatial region as shown Figure 5 while retaining a low RMSE value compared to the *AVE* model. This clearly shows that our approach not only produces highly correlated orderings of the GCM models across neighboring locations, it also produces the best top model in each spatial region. This is a clear benefit for climate scientists, as they do not have to rely on multiple GCM models for each location, instead use a single or a small set of good models for a specific neighborhood as prescribed by our analysis.

$\log(\lambda)$	OLS	OLSR	S^2M^2R-CF	S^2M^2R-GD
0	0.7282	0.7284	0.7314	0.6372
1	0.7244	0.726	0.7473	0.6362
2	0.7226	0.7278	0.8098	0.6364
3	0.7220	0.7249	0.9031	0.6371
5	0.7256	0.8224	0.9897	0.713
6	0.7261	0.8437	0.9977	0.8882
7	0.7247	0.8453	0.9995	0.9726
8	0.7236	0.8453	0.9998	0.9936

Table 4: Average Spearman correlation coefficient (SC) measured over 4363 neighboring locations in the neighborhood graph.

5 Conclusion

The problem of understanding climate change is of utmost importance. For this reason the Intergovernmental Panel on Climate Change (IPCC) [4] is informed by several countries on various Global Climate Models and the changing climate phenomena. These climate models have significant variance in predicting the future which makes it to have a single reliable adjudicator, which is a combination of multiple GCM outputs.

In this paper, we modeled the problem of combining multiple GCM outputs using spatial smoothing as an impor-

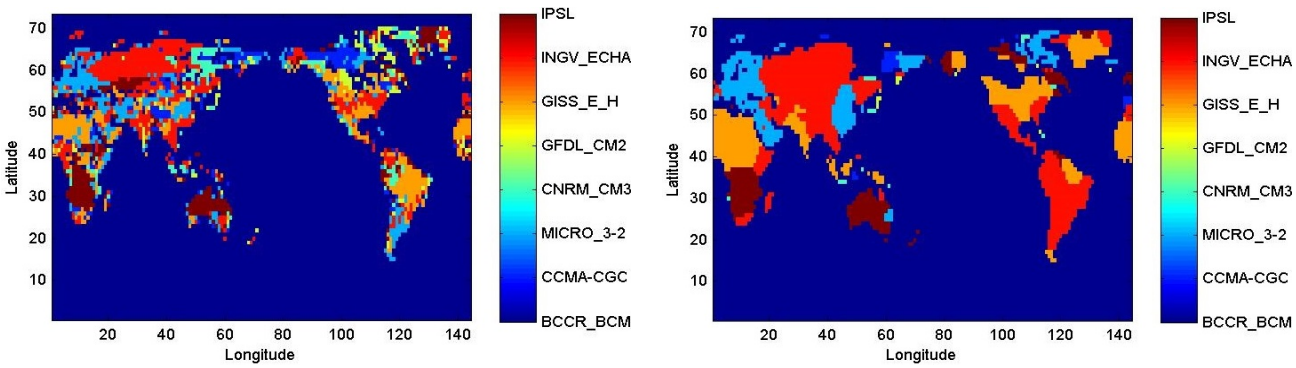


Figure 5: The world map showing the best models for each geographic location for OLS model (*left*) and S^2M^2R-GD (*right*). The nature of the OLS model clearly shows that there is very less spatial smoothing compared to the S^2M^2R-GD model.

tant design criterion. We formulated the problem as multiple OLS regression with spatial smoothing amongst model parameters in neighboring locations. We developed a closed form and an iterative solution for this problem. Further, we show that our solution is applicable to wide range of problems, as our formulation is a generalization of Sylvester equation. Finally, we empirically establish the superiority of our method compared to several popular baselines in terms of model accuracy and smoothness.

Acknowledgments

The research was supported in part by NSF grants IIS-1029711, IIS-0916750, and IIS-0812183, NSF CAREER award IIS-0953274, and NASA grant NNX12AQ39A. The authors thank the anonymous reviewers for their valuable comments.

References

[1] Climatic Research Unit. http://www.cru.uea.ac.uk/.

[2] Global Climate Model Datasets. https://sites.google.com/site/1stclimateinformatics/.

[3] INGV ECHAM4 Model. http://www.ecmwf.int/research/demeter/general/docmodel/ingv.html.

[4] Intergovernmental Panel on Climate Change. http://www.ipcc.ch/.

[5] A. Agovic and A. Banerjee. Gaussian process topic models. In *UAI*, pages 10–19, 2010.

[6] A. Agovic, A. Banerjee, and S. Chatterjee. Probabilistic matrix addition. In *ICML*, pages 1025–1032, 2011.

[7] L. Anselin. *Spatial Econometrics: Methods and Models*. Kluwer, Dordrecht, 1988.

[8] S. Arora, E. Hazan, and S. Kale. The multiplicative weights update method: A meta-algorithm and applications. *Theory Of Computing*, 8:121–164, 2012.

[9] R. H. Bartels and G. W. Stewart. Solution of the matrix equation AX + XB = C. In *Communications of the ACM*, volume 15, pages 820–826, 1972.

[10] M. Belkin, I. Matveeva, and P. Niyogi. Regularization and semi-supervised learning on large graphs. In *COLT*, pages 624–638, 2004.

[11] C. M. Bishop. Pattern recognition and machine learning. Springer-Verlag New York, Inc., 2006.

[12] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

[13] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[14] C. Brunsdon, S. Fotheringham, and M. Charlton. Geographically weighted regression-modelling spatial non-stationarity. pages 431–443. The Statistician, 1998.

[15] P. Cantelaube and J.M. Terres. Seasonal weather forecasts for crop yield modeling in europe. In *Tellus A*, volume 57, pages 476–487, 2005.

[16] T. L. Delworth and T. L. Broccoli. GFDLs CM2 global coupled climate models. Part I: formulation and simulation characteristics. volume 19, pages 643–674. Journal of Climate, 2006.

[17] M. Deque, C. Dreveton, A. Braun, and D. Cariolle. The ARPEGE/IFS atmosphere model: a contribution to the french community climate modelling. pages 249–266. Climate Dynamics, 1994.

[18] F. Giorgi and L. Mearns. Calculation of average, uncertainty range, and reliability of regional climate changes from aogcm simulations via the reliability ensemble averaging (REA) method. In *Journal of Climate*, pages 1141–1158, 2002.

[19] G. Golub and C. Van Loan. *Matrix Computations*. John Hopkins, 3rd edition, 1996.

[20] G. Golub, S. Nash, and C. V. Loan. A hessenberg-schur method for the problem ax+ xb= c. *Automatic Control, IEEE Transactions on*, 24(6):909–913, 1979.

[21] M. Herbster and M. K. Warmuth. Tracking the best expert. In *Machine Learning*, volume 32, pages 151–178, 1998.

[22] T. N. Krishnamurti, C. M. Kishtawal, T. E. LaRow, D. R. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran. Improved weather and seasonal climate forecasts from multimodel superensemble. In *Science*, volume 285, pages 1548–1550, 1999.

[23] T. N. Krishnamurti, C. M. Kishtawal, Z. Zhang, T. Larow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendaran. Multimodel ensemble forecasts for weather and seasonal climate. In *Journal of Climate*, volume 12, pages 4196–4216, 2000.

[24] T. N. Krishnamurti, S. Surendran, D. W. Shin, R. J. Correa-Torres, T. S. V. Vijaya Kumar, E. Williford, C. Kummerow, R. F. Adler, J. Simpson, R. Kakkar, W. S. Olson, and F. J. Turk. Real-time multianalysis/multimodel superensemble forecasts of precipitation using TRMM and SSM/I products. In *Mon. Wea. Rev*, volume 129, pages 2861–2883, 2001.

[25] C. L. Lawson and R. J. Hanson. *Solving Least Squares Problems*. NJ: Prentice-Hall, 1974.

[26] D. Masson and R. Knutti. Spatial-Scale Dependence of Climate Model Performance in the CMIP3 Ensemble. In *Journal of Climate*, volume 24, pages 2680–2692. American Meteorological Society, 2011.

[27] S. McQuade and C. Monteleoni. Global climate model tracking using geospatial neighborhoods. In *AAAI Conference on Artificial Intelligence*, 2012.

[28] G. A. Meehl, C. Covey, K. E. Taylor, T. Delworth, R. J. Stouffer, M. Latif, B. McAvaney, and J. F. Mitchell. The WCRP CMIP3 Multimodel Dataset: A New Era in Climate Change Research. In *Bulletin of the American Meteorological Society*, volume 88, pages 1383–1394. American Meteorological Society, September 2007.

[29] H. J. Miller and J. Han. *Geographic data mining and knowledge discovery*. CRC, 2009.

[30] K-1 model developers. K-1 coupled model (MIROC) description, k-1 technical report. Eds. H. Hasumi and S. Emori, Center for Climate System Research, University of Tokyo, 2004.

[31] C. Monteleoni, G. A. Schmidt, S. Saroha, and E. Asplund. Tracking climate models. In *Statistical Analysis and Data Mining*, volume 4, pages 372–392, 2011.

[32] J. L. Myers and A. D. Well. *Research Design and Statistical Analysis*. Lawrence Erlbaum, 2002.

[33] R. K. Pachauri and A. Reisinger. IPCC fourth assessment report 4: Climate change 2007. Intergovernmental Panel on Climate Change, 2007.

[34] T. N. Palmer, F. J. Doblas-Reyes, R. Hagedorn, and A. Weisheimer. Probabilistic prediction of climate using multi-model ensembles: from basics to applications. In *Philosophical Transactions of the Royal Society*, volume 360, pages 1991–1998, 2005.

[35] B. Rajagopalan, U. Lall, and S. E. Zebiak. Categorical climate forecasts through regularization and optimal combination of multiple gcm ensembles. In *Monthly Weather Review*, volume 130, pages 1792–1811, 2002.

[36] G. L. Russell, J. R. Miller, D. Rind, R. A. Ruedy, G. A. Schmidt, and S. Sheth. Comparison of model and observed regional temperature changes during the past 40 years. volume 105, pages 14891–14898. Jour. Geophys. Res., 2000.

[37] D. Salas-Melia, F. Chauvin, M. Deque, H. Douville, J. F. Gueremy, P. Marquet, S. Planton, J. F. Royer, and S. Tyteca. Description and validation of the CNRM-CM3 global coupled model. CNRM working note 103: available at: http://www.cnrm.meteo.fr/scenario2004/papercm3.pdf, 2005.

[38] J. F. Scinocca, N. A. McFarlane, M. Lazare, J. Li, and D. Plummer. Technical Note: The CCMA third generation AGCM and its extension into the middle atmosphere. pages 7055–7074. Atmospheric Chemistry and Physics, 2008.

[39] S. Shekhar, M. R. Evans, J. M. Kang, and P. Mohan. Identifying patterns in spatial information: A survey of methods. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(3):193–214, 2011.

[40] IPSL Team. The new IPSL climate system model: IPSL-CM4. Institut Pierre Simon Laplace des Sciences de lEnvironnement Global, Paris, France, 2005.

[41] C. Tebaldi and R. Knutti. The use of the multimodel ensemble in probabilistic climate projections. In *Philosophical Transactions of Royal Society A*, volume 365, pages 2053–2075, 2007.

[42] C. Tebaldi, L. Mearns, D. Nychka, and R. Smith. Regional probabilities of precipitation change: a beysian analysis of multimodel simulations. In *Geophysical Research Letters*, volume 31, 2004.

[43] C. Tebaldi, R. Smith, D. Nychka, and L. Mearns. Quantifying uncertanity in projections of regional climate change: a Bayesian approach to the analysis of multi-model ensembles. In *Journal of Climate*, volume 18, pages 1524–1540, 2005.

[44] M. C. Thomson, F. J. Doblas-Reyes, S. J. Mason, R. Hagedorn, S. J. Connor, T. Phindela, A. P. Morse, and T. N. Palmer. Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. In *Nature*, volume 439, pages 576–579, 2006.

[45] E. L. Wachspress. Iterative solution of the Lyapunov matrix equation. In *Applied Mathematics Letters*, volume 1, pages 87–90, 1988.

[46] N. M. Waters. Spatial interpolation. NCGIA Core Curriculum in GIS, Eds. M. F. Goodchild and K. K. Kemp, 1990.

[47] W. T. Yun, L. Stefanova, and T. N. Krishnamurti. Improvement of the multimodel superensemble technique for seasonal forecasts. *Journal of Climate*, 16(22):3834–3840, 2003.