Variability in Annual Temperature Profiles
A Multivariate Spatial Analysis of Regional Climate Model Output

Tamara Greasby, Stephan Sain

Institute for Mathematics Applied to Geosciences,
National Center for Atmospheric Research

August 7, 2012
Introduction

Climate Models
North American Regional Climate Change Assessment Program (NARCCAP)

A Bayesian Hierarchical Spatial Model
An Overview
Periodic Basis Functions
The Hierarchy
Quick Details

Results
Temperature Profiles
Changing Seasonality
Impacts

Conclusions
The Goal

Using climate model output from the North American Regional Climate Change Assessment Program (NARCCAP), summarize annual temperature profiles spatially, identify how it may change in the future.
The Goal

Questions include:

- Are some regions more vulnerable to change than others, and how does one quantify this?
- In this case, we are defining vulnerable in terms of impacts:
  - mosquito breeding season
  - length of “hot” season, etc.
General Circulation Models (GCMs)

- Based on scientists’ current understanding of the Earth’s climate system.
- Climate processes are represented by a series of differential equations.
- When modeling future climate, also attempt to incorporate changes in the forcings that influence the climate system.
General Circulation Models (GCMs)

- Models are complex and computationally expensive.
- Grid boxes are typically on the scale of 200 to 500 km.
  - All of Colorado would likely be contained in four grid boxes
Regional Climate Models

- Created by downscaling GCMs.
- Focus on a limited spatial domain with higher resolution, typically 20-100 km.
- Particularly of interest for impacts studies.
  - Heat stress and heat island effects in a certain metropolitan area.
  - Vector-borne disease and length of breeding season.
Regional Climate Models, Global Climate Models, and Uncertainty

Dynamically downscaled regional climate models use boundary conditions provided by a global climate model.

Image courtesy of Stephan Sain
Climate Models and Uncertainty

In the both global and regional climate models there are several sources of uncertainty.

- Future emissions
- The climate system’s response to those emissions
- Natural climate variability
- The physical processes and implementation

Collections of climate models, called ensembles, are used to explore these uncertainties.
North American Regional Climate Change Assessment Program (NARCCAP)

NARCCAP’s goal is to explore uncertainties in climate change projections from RCMs.

- All models are run under the A2 emissions scenario
- Use a balanced design
  - Each RCM is run with boundary conditions from 2 GCMs
  - Each GCM provides inputs to 3 RCM model.

<table>
<thead>
<tr>
<th>RCM</th>
<th>GFDL</th>
<th>CGCM3</th>
<th>HADCM3</th>
<th>CCSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRCM</td>
<td>finished</td>
<td>finished</td>
<td></td>
<td>finished</td>
</tr>
<tr>
<td>ECP2</td>
<td>finished</td>
<td>running</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HRM3</td>
<td>running</td>
<td></td>
<td>finished</td>
<td></td>
</tr>
<tr>
<td>MM5I</td>
<td>finished</td>
<td>finished</td>
<td></td>
<td>finished</td>
</tr>
<tr>
<td>RCM3</td>
<td>finished</td>
<td>finished</td>
<td></td>
<td>finished</td>
</tr>
<tr>
<td>WRFG</td>
<td>finished</td>
<td>finished</td>
<td></td>
<td>finished</td>
</tr>
</tbody>
</table>
North American Regional Climate Change Assessment Program (NARCCAP)

In particular, NARCCAP is looking to answer

- What is the impact of GCM on model output?
- What is the impact of RCM on model output?
- Is there an interaction between the two?
The Model: Overview

Using the NARCCAP models we aim to summarize annual temperature profiles spatially and identify how they may change in the future.
The Model: Overview

Using the NARCCAP models we aim to summarize annual temperature profiles spatially and identify how they may change in the future.

Secondary goals include:

- Understanding how those results change based on GCM, RCM and the combination.
- Understanding these results in contexts of interest to impacts researchers.
The Model: Overview

Using the NARCCAP models we aim to summarize annual temperature profiles spatially and identify how they may change in the future. Secondary goals include:

- Understanding how those results change based on GCM, RCM and the combination.
- Understanding these results in contexts of interest to impacts researchers.

To do this, we are going to use a Bayesian Hierarchical Model.

- This model is designed to break apart a difficult, complex problem into smaller pieces.
- Each level of the hierarchy is designed to capture a different structure and a different source of uncertainty.
The Model: Beginnings

For each climate model, we use 30 years worth of current and 30 years worth of future temperature output to estimate an annual temperature profile for 1971-2000 and 2041-2070 respectively.

Note that each regional climate model has its own grid.
A quick reality check

- Each climate model has between 10,000 and 20,000 grid boxes.
- Including 30 years of daily data, for current and future, for one model, we have approximately $(2)(30)(365)(15,000) = 328,500,000$ observations.
- With 8 models, we have over 2 billion data points.
A quick reality check

- Each climate model has between 10,000 and 20,000 grid boxes.
- Including 30 years of daily data, for current and future, for one model, we have approximately 
  \((2)(30)(365)(15,000) = 328,500,000\) observations.
- With 8 models, we have over 2 billion data points.

This becomes a computational challenge.
Gridbox modeling

Due to the size of the climate model output, 8 periodic b-splines are fit to the data for the current and future time period for each gridbox. Only the coefficients and the standard errors are retained for spatial modeling.

\[ Y_i \sim N(X\beta_i + \theta_i t, S_i) \]

where

- \( i = 1, \ldots, N \)
- \( t = 1, \ldots, (30)(365) \)
- \( Y_i \) is \((365)(30) \times 1\)
- \( X \) is \((365)(30) \times 8\)
- \( \beta_i \) is \(8 \times 1\)
- \( S_i \) is \((365)(30) \times (365)(30)\)
- Current and future are fit separately.
A periodic spline basis is a collection of curves that when added in varying combination can be used to estimate a smooth curve.
Periodic Spline Basis

A periodic spline basis is a collection of curves that when added in varying combination can be used to estimate a smooth curve.
A periodic spline basis is a collection of curves that when added in varying combination can be used to estimate a smooth curve.
Periodic Spline Basis

A periodic spline basis is a collection of curves that when added in varying combination can be used to estimate a smooth curve.
A periodic spline basis is a collection of curves that when added in varying combination can be used to estimate a smooth curve.
A periodic spline basis is a collection of curves that when added in varying combination can be used to estimate a smooth curve.
Gridbox modeling

\[ Y_i \sim N(X\beta_i + \theta_it, S_i) \]

where

- \( i = 1, \ldots, N \)
- \( t = 1, \ldots, (30)(365) \)
- \( Y_i \) is \((365)(30) \times 1\)
- \( X \) is \((365)(30) \times 8\)
- \( \beta_i \) is \(8 \times 1\)
- \( S_i \) is \((365)(30) \times (365)(30)\)
- Current and future are fit separately.
Why heteroscedastic?
The Model: The Data Level

\[ \beta | U \sim N(Z\gamma + U, \Sigma) \]

- \( \beta \) is \((8)(2)(N) \times 1\) and represents the coefficients on the spline basis, 8 current and 8 future.
- \( Z\gamma \) is \((8)(2)(N) \times 1\) and represents the fixed effects of an intercept, latitude, elevation, and land/sea on the coefficients.
- \( U \) is a \((8)(2)(N) \times 1\) spatial random effect.
- \( \Sigma \) is block diagonal and incorporates the standard errors from the grid box fits.
The Model: The Process Level

At the process level, the spline coefficients are tied together in space.

- The basic idea: Think about one day of one year, for one location in space, for current/future.
  - Given it’s neighbors, the temperature at one gridbox is independent of all other gridboxes.
  - This neighbor structure is used to specify the precision (inverse covariance) for random effects
  - Instead of talking about days, we will be talking about the spline coefficients.
The Model: The Process Level

- In words, the spatial random effect lets each gridbox have it’s own coefficients on the spline basis, but in a systematic way.

![Basis 4](image-url)
The Model: The Process Level

More formally,

- \( U_i \sim \mathcal{N} \left( 0, V^{-1} \right) \) \( i = 1, \ldots, 16 \).
  - \( U_i \) is a first order Intrinsic Gaussian Markov Random Field (IGMRF)
  - For identifiability each \( U_i \) is constrained so \( \sum_{j=1}^{N} U_{ij} = 0 \)
- Modeled jointly, \( U \sim \mathcal{N} \left( 0, S^{-1} \otimes V^{-1} \right) \)
- This assumption implies the correlation between two basis coefficients is consistent across space.
The Model: The Prior Level

Prior assumptions are made for two parameters, $S$ and $\gamma$ and non-informative priors were used for both.

- $\gamma \sim N(\mu, \Sigma_{\gamma})$
  - $\mu$ is set to be the least squares estimate regressing the $\beta$ coefficients on $Z$
  - $\Sigma_{\gamma}$ is assumed to be diagonal with large variance.
- $S \propto |S|^{-\frac{16-1}{2}}$. 
Model Fitting

- A Gibb’s Sampler is used to sample the conditionals for all parameters.
  - a Markov Chain Monte Carlo (MCMC) method for obtaining random samples from a multivariate probability distribution by sampling from a conditional distribution.
  - Start with an initial value, $X^{(0)}$ for each parameter.
  - Cycle through each parameter, sampling from it’s conditional distribution, $p(x_i^{(k)} | x_1^{(k)}, x_2^{(k)}, \ldots, x_{i+1}^{(k-1)}, \ldots)$
Model Fitting

- A Gibb’s Sampler is used to sample the conditionals for all parameters.
  - a Markov Chain Monte Carlo (MCMC) method for obtaining random samples from a multivariate probability distribution by sampling from a conditional distribution.
    - Start with an initial value, $X^{(0)}$ for each parameter.
    - Cycle through each parameter, sampling from it’s conditional distribution, $p(x_i^{(k)} | x_1^{(k)}, x_2^{(k)}, \ldots, x_{i+1}^{(k-1)}, \ldots)$
  - A number of chains were used to ensure convergence.
    - Trace plots were used to ensure convergence within a chain.
    - Statistics such as Gilk’s $\sqrt{R}$ were also monitored. This compares the variance between and within the chains.

- Residuals were analyzed to ensure no spatial correlation remained.
Model Fitting

- Conditional distributions are derived for all parameters: $U$, $S$, $\gamma$. The joint posterior is shown below, along with each conditional.

$$
U, \gamma, S, \beta \propto |\Sigma|^{-1/2} \exp \left( -\frac{1}{2} (\beta - Z \gamma - U)^T \Sigma^{-1} (\beta - Z \gamma - U) \right)
$$

$$
|\Psi|^{-1/2} \exp \left( -\frac{1}{2} (\gamma - \mu)^T \Psi^{-1} (\gamma - \mu) \right) |S|^{-\frac{2(8+1)-1}{2}}
$$

$$
\gamma | \beta, U, S, \Sigma \propto \mathcal{N} \left( \left( \psi^{-1} + Z^T \Sigma^{-1} Z \right)^{-1} \left( \psi^{-1} \mu - Z^T \Sigma^{-1} U - Z^T \Sigma^{-1} \beta \right), \left( \psi^{-1} + Z^T \Sigma^{-1} Z \right)^{-1} \right)
$$

$$
U_j | Y, U_{-j}, \beta, \gamma, \Sigma, S \propto \mathcal{N} \left( \left( s_{jj} V + \Sigma_{jj} \right)^{-1} \left( \beta^T_j \Sigma^{-1} j + \gamma^T \Sigma^{-1} j - V \sum_{k \neq j} \left( s_{jk} U_k + \Sigma_{jk} U_k \right) \right), \left( s_{jj} V + \Sigma_{jj} \right)^{-1} \right)
$$

$$
S | U \propto \mathcal{W} \left( N + 2(16) - 1, U^* \Sigma V U^* \right)
$$
Why not Kriging?

- Markov Random Fields are tailor made for gridded datasets.
- There are significant advantages to using Markov Random Fields including
  - no matrix inversion
  - utilization of sparse matrix algebra
- We are not interested in prediction at locations not on the native grid.
Analysis goals

Is the shape of the profile changing? Some questions include,

- Is the profile change consistent across the domain?
- What is the effect of the regional model, the global model, and is there an interaction?
Temperature profiles

Is the shape of the profile changing? Compare the results for CRCMccsm, MM5icccsm, and WRFGccsm results for San Francisco.
Too hard to compare!

Look at all climate models at once:

Detroit

Denver

Sacramento
Interactive graphics help

- Hard to tell what is going on.
- Look at it interactively.
Length of summer

- Another approach, look at the length of summer:
  - Define summer as the period with temperatures over the 75th percentile.
  - Plots show the CRCMccsm, MM5Iccsm, and WRFGccsm results for San Francisco.
Length of summer

Another approach, look at the length of summer:
- Define summer as the period with temperatures over the 75th percentile.
- Plots show the CRCMccsm, MM5Iccsm, and WRFGccsm results for San Francisco.
Distribution of summer lengths

We can get a distribution of summer lengths for each of the climate models.
Another approach to changing seasonality

- Define winter as the coldest quarter of the year, summer as the hottest quarter and spring and fall as the seasons in between.

- Other analyses define a season starting based on temperature or the first appearance of a flower, animal, etc.
Seasonal changes for 3 cities
ANOVA-style analysis

- Using the posterior draws, perform an ANOVA-style analysis to try to understand the role of RCM and GCM.
- Remember all models are on different grids. For now, compare based on metropolitan region median.
- The raw data here will be fitted temperature values ($Y$).

$$Y_{grt} = Y_{\bullet\bullet} + Y_{g\bullet} + Y_{r\bullet} + Y_{\bullet t} + Y_{g\bullet t} + Y_{r\bullet t}$$
ANOVA-style analysis: Global mean

![Graphs showing temperature profiles for Denver and Seattle over the year.](image-url)
ANOVA-style analysis: RCM vs. GCM

![Graphs showing temperature profiles for Denver and Seattle, comparing RCM and GCM models.](image-url)
ANOVA-style analysis: Future interaction

Denver

Seattle

Day of the Year

Day of the Year

Future Effect

Future Effect

Uncertainty in Climate Models
ANOVA-style analysis: 2x2x2 Experiment

Built in the NARRCAP design is a 2x2x2 experiment:

- **GCMs**: ccsm, cgcm3
- **RCMs**: CRCM, WRFG
- **time**: Current vs. Future

![Graphs showing temperature profiles in Denver and Seattle with GCMs ccsm and cgcm3](image.png)
ANOVA-style analysis: 2x2x2 Experiment

Temperature Profiles
Changing Seasonality
Impacts
Temperature and Vector-Bourne Disease

- Experiments have indicated that amongst other environmental conditions, the virus causing Dengue Fever is transmitted at temperatures over 30°C.
- At 30°C, the extrinsic incubation time is 12 days.
- At temperatures over 32°C, this decreases to 7 days.
- The typical life span of the Aedes aegypti mosquito is 2 weeks to a month.
Temperature and Dengue Fever*

* This is an over-simplification of vulnerability to Dengue. Precipitation and humidity play a crucial role in transmission, as does human mitigation strategies.

** Bottom map indicates regions currently at risk for Dengue. Map is from the CDC and Healthmap: http://www.healthmap.org/dengue/index.php
Temperature and Dengue Fever*

Median days over 32°C – leading to a decreased incubation time.

* This is an over-simplification of vulnerability to Dengue. Precipitation and humidity play a crucial role in transmission, as does human mitigation strategies.
Concluding Thoughts

- The hierarchical Bayes model allows us to deal with spatial correlation in the climate model output.
- This statistical model allows quantification of the uncertainty in the changes to the annual temperature profiles, along with the ability to examine how these profiles and changes to these profiles vary across season, space, RCM, and GCM.
- These results can be applied to many applications where temperature is an important covariate.
Future Work

- Model two curves for each grid box, the daily low and daily high.
- Include two variables, such as temperature and precipitation or relative humidity.
- Include all GCM/RCM models in the same statistical model.
Lattice Krig?

Doug’s R package:

- Move from a Bayesian format to maximum likelihood estimation.
- He promises much faster computation time and the ability to combine all climate models without interpolating to a common grid.
Thanks!