Statistical analysis of regional climate models.

Douglas Nychka

National Center for Atmospheric Research
Boulder, Colorado

National Science Foundation
Seattle, August 2012
Outline

- Simulating Regional Climate
- NARCCAP design
- A digression on Kriging
- *Functional ANOVA for changes in surface temperature*

Challenges:
Spatial and functional data, computation for large data sets.
Acknowledgements

NARCCAP

Linda Mearns (NCAR), Cari Kaufman (Berkeley), Stephan Sain (NCAR)
The global models on their own do not give enough detailed information at regional and local scales.
A single global model grid box

All the features in the atmosphere are reduced to a single one dimensional column!

Many physical processes and features cannot be modeled explicitly! E.g. thunderstorms or extreme weather events.
An approach to Regional Climate

- Nest a fine-scale weather model in part of a global model’s domain.
Regional model simulates higher resolution weather based on the global model for boundary values and fluxes.

A snapshot from the 3-dimensional RSM3 model (right) forced by global observations (left)

- Consider different combinations of global and regional models to characterize model uncertainty.
NARCCAP – the design

4GCMS \times 6RCMs:
12 runs – balanced half fraction design
Global observations \times 6RCMs
High resolution global atmosphere

<table>
<thead>
<tr>
<th>GLOBAL FORCING</th>
<th>MM5I</th>
<th>WRF</th>
<th>HADRM</th>
<th>REGCM</th>
<th>RSM</th>
<th>CRCM</th>
<th>time slice</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFDL</td>
<td></td>
<td></td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>HADCM3</td>
<td>○</td>
<td></td>
<td>●</td>
<td></td>
<td></td>
<td>○</td>
<td></td>
</tr>
<tr>
<td>CCSM</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>CGCM3</td>
<td></td>
<td>●</td>
<td></td>
<td>●</td>
<td></td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>Reanalysis</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
</tr>
</tbody>
</table>

A designed experiment is amenable to a statistical analysis and can contain more information.

But just 2-d temperatures fields are 72Gb of data.
A 2 x 2 subset of NARCCAP.

The Factors (levels):

- **Downscaling**: (timeslice, regional model)
  - GFDL climate model with GFDL high resolution atmosphere
  - UC Santa Cruz Regional Climate Model.

- **Time period**: (current climate, future scenario)
  - [1971, 2000] Present
  - [2040, 2071] A2 scenario of high GHG emissions

The Responses:

- Mean summer temperature (JJA) on a common 134 x 83 grid.
- Mean total summer precipitation (JJA)

Question:

*How does the future projection depend on the method of downscaling?*
2×2 Summer temperature

Tslicе Current

Future

RCM3 Current

Future

D. Nychka Inference for Regional Climate
A digression on some relevant spatial statistics
Data is never simple ...

Model output is on different grids and requires regridding.

Regional Climate Models from NARCCAP

ECPC and HRM3

We regrid using a compact Wendland covariance ...

See fastTps from the fields package.
Useful covariance models

- $\Sigma$ modeled as

$$\Sigma = \sigma^2 R^{-1}(\phi) \otimes C^{-1}(\phi)$$

$R$ and $C$ are 1-D “stationary” Markov field covariances depending on parameters $\phi$.

- Computationally efficient: sparse precision matrices with Kronecker construction.

- Approximates exponential form of spatial covariance.

**Figure:** Graphical representation of the covariance models with visual examples.
Analysis of the RCM data
A Functional ANOVA Model

Each response is an entire spatial field!

$Y_{M,T,k}$ Summer rainfall at grid boxes based on:
- Model (timeslice vs RCM)
- Time Period (current vs future)
  for 30 years ($k = 1, \ldots, 30$).

$Y_{M,T,k} = \text{Baseline climate} + \text{Model effect} + \text{Time Period} + \text{Interaction} + \text{future trend} + "noise_k"$
\[ Y_{M,T,k} = \text{Baseline climate} + \text{Model effect} + \text{Time Period} + \text{Interaction} + \text{future trend} + \text{"noise}_k\text{"
}\]

- Interaction is the nonlinear part of response

- "noise" = sum of a trend component and Markov random field.

- Effects are assumed to be correlated spatial fields: Baseline and Time Period have prior mean from the global model, Interaction and Model effect have prior mean of zero.
A Bayesian implementation

Additive model

\[ Y_{M,T,k} = \alpha_0 + \alpha_M + \alpha_T + \alpha_I + \gamma_T(k) + \eta_{M,T,k} \]

Priors

- \( \alpha_0 \sim \mathcal{N}(\mu_{\text{current}}, \Sigma(\theta_0)) \)
- \( \alpha_M \sim \mathcal{N}(0, \Sigma(\theta_M)) \)
- \( \alpha_T \sim \mathcal{N}(\mu_{\text{future}}, \Sigma(\theta_T)) \)
- \( \alpha_I \sim \mathcal{N}(0, \Sigma(\theta_I)) \)

Spatial dependence parameters from Markov random field have independent priors \( U[-1,1] \)

Marginal variance for fields have uniform priors.
A Bayesian implementation

• Prior distributions are chosen to be largely non-informative.
• A Gibbs sampler was constructed with some Metropolis-Hastings steps.
• A single chain of length 5000 was run; convergence suggested after a few hundred iterations.
Inference for temperature

Baseline (275,310)  Downscaling (-10,10)

95 % posterior probability of nonzero interaction

Time Period (-4,4)  Interaction (-1,1)

D. Nychka Inference for Regional Climate
Inference for precipitation

Baseline (0,1.2)  Downscaling (-.6,.6)

Time Period (-.15,.15)  Interaction (-.07,.07)

95% posterior probability of nonzero interaction
Conclusions

Numerical models for climate projections or prediction produce complicated output that benefits from statistical summaries.

Spatial statistics needs to address functional data in the form of curves and surfaces.

Large data sets are the norm.

Some of the largest benefits will be in interaction with the modelers about significant spatial differences and their dynamical sources.
Thank you!