Statistics and Climate Models

Cari Kaufman

Statistical and Applied Mathematical Sciences Institute (SAMSI) and National Center for Atmospheric Research (NCAR)

Note: Much of this talk comes from A Climate modelling Primer by McGuffie and Henderson-Sellers, including all uncited figures.

SAMSI Undergraduate Workshop, May 2007
Outline

- An overview of climate modelling
  - Components of the climate system
  - Types of models

- Energy balance models (EBMs)
  - How they work
  - Some experiments with a simple model

- General circulation models (GCMs)
  - Components + coupling
  - More complexity or more model runs?

- Modelling spatial dependence (time permitting)
Components of the Climate System

- Atmosphere
- Ocean
- Cryosphere
- Land surface, biosphere
- Atmospheric chemistry
Climate vs. Weather

- We can think of climate as characterizing a distribution, while weather is a particular realization. (Lorenz: “Climate is what we expect. Weather is what we get.”)

- Some ways that climate might change:

(Source: Andrew Gettelman, NCAR)
Models in Climate Change Assessment

(Source: IPCC, Climate Change 2007: Summary for Policymakers)
Models Vary in Complexity

Components of climate models:

- Radiation - input, absorption, and emission
- Dynamics - movement of energy around the globe
- Surface processes - ice, snow, vegetation
- Chemistry - e.g. carbon exchanges

Some types of climate models:

- Energy balance models (EBMs)
- Single column models (SCMs)
- General circulation models (GCMs)
Energy Balance Models

- Balance Earth’s planetary radiation budget
- Can run on a desktop computer
- Predict surface temperature
- Usually are 0-D or 1-D (zonal)
0-D EBM

- Assume Earth is in radiative equilibrium:

\[
\text{(shortwave) radiation in} = \text{(longwave) radiation out}
\]

\[
(1 - \alpha) \pi R^2 S = 4 \pi R^2 \sigma T_e^4
\]
0-D EBM

- Assume Earth is in radiative equilibrium:

\[(\text{shortwave}) \text{ radiation in} = (\text{longwave}) \text{ radiation out}\]
\[(1 - \alpha)\pi R^2 S = 4\pi R^2 \sigma T_e^4\]

- \(S = \text{solar constant} = 1370\text{W/m}^2 \text{ for Earth}\)
- \(R = \text{radius} \Rightarrow \pi R^2 S = \text{incident solar radiation}\)
- \(\alpha = \text{albedo} = \text{fraction reflected} = 0.3 \text{ for Earth}\)
0-D EBM

- Assume Earth is in radiative equilibrium:

\[(\text{shortwave}) \text{ radiation in} = (\text{longwave}) \text{ radiation out}\]

\[(1 - \alpha)\pi R^2 S = 4\pi R^2 \sigma T_e^4\]

- \(S = \) solar constant = 1370\(W/m^2\) for Earth
- \(R = \) radius \(\Rightarrow \pi R^2 S = \) incident solar radiation
- \(\alpha = \) albedo = fraction reflected = 0.3 for Earth
- \(\sigma = \) Stefan-Boltzmann constant = \(5.67 \times 10^{-8} Wm^{-2}K^{-4}\)
- \(T_e = \) effective temperature (equivalent blackbody temperature)

\[\Rightarrow T_e = \left(\frac{(1 - \alpha)S}{4\sigma}\right)^{1/4} = 255K \text{ for Earth}\]
Greenhouse Effect

- Water vapor and gases like \( CO_2 \) absorb radiation more strongly in the longwave part of the spectrum, so surface temperature \( T_s \) is greater than effective temperature \( T_e \).

- \( \Delta T = T_s - T_e \) is the greenhouse increment.

- On Earth, \( T_s \approx 288K \Rightarrow \Delta T \approx 33K \)

- Contrast with Venus: \( S = 2619W/m^2 \) (closer to the sun) and \( \alpha = 0.7 \) (lots of clouds) \( \Rightarrow T_e = 242K \), but \( T_s \approx 730K \Rightarrow \Delta T = 488K \)
1-D EBM

- In a particular latitude zone $i$,

\[
(\text{Shortwave in}) = (\text{Transport out}) + (\text{Longwave out})
\]

\[
S_i(1 - \alpha(T_i)) = F(T_i) + R \uparrow (T_i)
\]
1-D EBM

- In a particular latitude zone $i$,

\[
\text{(Shortwave in)} = \text{(Transport out)} + \text{(Longwave out)}
\]

\[
S_i (1 - \alpha(T_i)) = F(T_i) + R \uparrow (T_i)
\]

- Albedo now varies by temperature, say $\alpha(T_i) = \begin{cases} 0.6 & T_i \leq T_c \\ 0.3 & T_i > T_c \end{cases}$ where $T_c$ is the temperature at snowline.
1-D EBM

- In a particular latitude zone $i$,

\[
(S\text{hortwave in}) = (T\text{ransport out}) + (L\text{ongwave out})
\]

\[
S_i(1 - \alpha(T_i)) = F(T_i) + R \uparrow (T_i)
\]

- Albedo now varies by temperature, say $\alpha(T_i) = \begin{cases} 0.6 & T_i \leq T_c \\ 0.3 & T_i > T_c \end{cases}$ where $T_c$ is the temperature at snowline.

- $F(T_i)$, the horizontal flux, is modeled as $k_t(T_i - \bar{T})$, where $\bar{T}$ is the global mean temperature and $k_t$ is an empirically defined constant.
1-D EBM

- In a particular latitude zone $i$,

\[
S_i(1 - \alpha(T_i)) = F(T_i) + R \uparrow (T_i)
\]

- Albedo now varies by temperature, say $\alpha(T_i) = \begin{cases} 0.6 & T_i \leq T_c \\ 0.3 & T_i > T_c \end{cases}$ where $T_c$ is the temperature at snowline.

- $F(T_i)$, the horizontal flux, is modeled as $k_t(T_i - \bar{T})$, where $\bar{T}$ is the global mean temperature and $k_t$ is an empirically defined constant.

- $R \uparrow (T_i)$, radiation emitted to space, is proportional to $T_i^4$, but over our range of interest we can approximate it by a linear function $A + BT_i$. 
1-D EBM, Continued

- Putting it all together, we have

\[ T_i = \frac{S_i(1 - \alpha_i) + k_t \bar{T} - A}{B + k_t} \]

- We can solve iteratively for the \( T_i \)'s.

- Let's try some experiments:
  - What change in the solar constant will glaciate the whole Earth?
  - What is the effect of changing the transport term \( k_t \)?
  - How about changing the albedo of ice?
General Circulation Models

- Characterize the full 3-D behavior of the climate system
- Consist of individual programs for atmosphere, oceans, cryosphere, and land, linked by a ‘coupler’
Simplifications in GCMs

- Continuous processes are discretized into grids of varying resolutions.

- Some processes are described by physical relationships, others must be “parameterized,” for example cloud cover.

(Source: climateprediction.net)
Modelling the Atmosphere

- Solve meteorological primitive equations, but some differences compared to weather prediction:

<table>
<thead>
<tr>
<th>Weather prediction</th>
<th>Climate modelling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial state is critical</td>
<td>‘Spin up’ removes dependence on initial state</td>
</tr>
<tr>
<td>Don’t need whole PDF, just probable region</td>
<td>Want whole PDF, especially tails (extreme events)</td>
</tr>
<tr>
<td>Non-conservation to match observations</td>
<td>Conservation is critical</td>
</tr>
</tbody>
</table>
Increases in Computing Power

As computing power increases, so can:

- Complexity
- Resolution
- Length of simulations
- Number of simulations
High Resolution Models

Higher resolution can be helpful because

- We can resolve smaller scale processes.
- Definitions of extreme events depend on spatial scale.
- Impact studies require finer, local resolution.

(NCAR AGCM for the Earth Simulator, Japan)
Regional Models (RCMs)

- Operate on finer resolutions at smaller spatial scales, but require boundary conditions from observations or GCMs.

- Plot shows a subset of data from the Prudence project in Europe.

- A 2 RCM x 2 GCM x 2 scenario experiment.

- Can use functional ANOVA to decompose model outputs.
Distributed Computing at climateprediction.net

- Simple climate model runs on individual desktops
- Each run has slightly different model parameters
- Calibration, control, and double CO₂ phases
- Gives better estimate of true uncertainty in predictions
Modelling Spatial Dependence

- *Gaussian processes* generalize the idea of multivariate normal distribution to a field. A Gaussian process is defined by its mean function $m$ and covariance function $K$.

- Some assumptions we can make
  - $K$ is **stationary**: depends only on $s - s'$
  - $K$ is **isotropic**: depends only on $d = ||s - s'||$

- Matérn: a stationary and isotropic covariance function

$$K(d; \sigma^2, \rho, \nu) = \sigma^2 \frac{(d/\rho)^\nu}{\Gamma(\nu)2^\nu-1} \mathcal{K}_\nu(d/\rho)$$
Matérn Covariance Functions: Illustration

\[ \sigma^2 = 1, \ \nu = \frac{1}{2} \]

Effective range a function of \( \rho \) and \( \nu \)

\[ \sigma^2 = \text{marginal variance}; \ \rho = \text{range}; \ \nu = \text{smoothness} \]
Sample Realizations: $\sigma^2 = 1, \nu = 1/2$

$\rho = 1/5$; Effective range $= 0.6$

$\rho = 1$; Effective range $= 3$
Likelihood Based Covariance Estimation

- Likelihood-based methods (MLE, REML, Bayesian models) are attractive for estimating a parameterized covariance function.

- Gaussian log-likelihood for observations \( z = (z(s_1), \ldots, z(s_n))' \):

\[
\ell(\theta) = -\frac{1}{2} \log \det(\Sigma(\theta)) - \frac{1}{2}(z - \mu)'\Sigma(\theta)^{-1}(z - \mu),
\]

where \( \Sigma(\theta) = \{K(s_i, s_j; \theta)\} \)

- Problem: Expensive to compute as the number of locations increases, and to estimate \( \theta \), we need to compute it many times.
  \( \Rightarrow \) Develop approximations that still give good estimates of \( \theta \).
An Application to Landcover modelling

- Feddema et al. (Climate Dynamics, 2005) ran the Department of Energy’s Parallel Climate Model (DOE-PCM) under two land use scenarios, one “modern” and one “prehuman.”

- All other input values were the same across scenarios, including pre-industrial atmospheric conditions.

- Scientific goal: Compare global temperatures under each scenario.
Temperature Data

40 years annual avg temperatures on a latitude-longitude grid (n=8192)
**40 Year Means (Degrees C)**

**Modern**

**Prehuman**

**Difference**

SAMSI Undergraduate Workshop, May 2007
Yearly Fluctuations

Modern

Prehuman

Year 1

Year 2

Year 3

Year 4

Year 5

Year 1

Year 2

Year 3

Year 4

Year 5

SAMSI Undergraduate Workshop, May 2007
Local and Long Distance Effects

Converted to agriculture

Converted to grassland

Difference
Basic Model Structure

- For each scenario $x$, two spatial Gaussian processes:

  "Climate" process:

  $\mu_x(\cdot) \mid \beta, \theta \sim \text{GP} \left( \sum_{j=1}^{J} \beta_j \phi_j(\cdot; x), K_{\mu}(\cdot, \cdot; \theta_{\mu}) \right)$

  $\phi_j(\cdot; x)$ spatial functions, eg. indicator that land use is of class $j$

  "Weather" process:

  $Z_x(\cdot, t) \mid \mu_x(\cdot), \theta \sim \text{GP} \left( \mu_x(\cdot), K(\cdot, \cdot; \theta) \right)$

- Diffuse proper priors on $\beta, \theta, \theta_{\mu}$. 
USA Data (n=253)

Modern

Prehuman

Ocean
Needleleaf evergreen
Temp mixed forest
Broadleaf decid
Trop seasonal decid tree
Savanna
Cool grassland/steppe
Warm grassland
Evergreen shrub
Decid shrub
Semi−desert
Crop

Difference
Posterior Densities: Covariance Parameters

Climate process parameters

Weather process parameters

SAMSI Undergraduate Workshop, May 2007
Posterior Mean Surfaces

Modern: Posterior Mean

Prehuman: Posterior Mean

Difference: Posterior Mean

Posterior P(Diff < −1)