Regional climate predictions: Uncertainty and biases

Christoph Buser, Hans R. Künsch and Christoph Schär

Seminar für Statistik and Institute for Atmosphere and Climate
ETH Zürich

Cherrybud workshop, March 2008

Contents

Starting point

- Predicting the future climate is important for policy making, but difficult because of the complexity of the processes in the ocean, the atmosphere and on the land surface.
- Global models for atmosphere and ocean have a coarse resolution. Regional models allow downscaling by using the output of global models for initial and boundary conditions.
- The number of global and regional models in use is increasing. Each model is run under different emission scenarios. The number of different answers becomes confusing.

Model selection vs. model combination

- Not all models are equal. A good model should be able to reproduce the current climate and be within the range of other models with its prediction.
- Selecting a single “best” model is not adequate in view of the complexity of the climate.
- Weighted averaging seems intuitive plausible, but choice of weights is not clear.
- Bayesian methods allow model combinations in a coherent and transparent way.

Data and distributions

See: Tebaldi et al., J. Climate 18 (2005).

They onsider 4 seasons, 22 regions and different scenarios separately. For each season, region and scenario

- Mean of observed temperatures 1961-1990 $\sim \mathcal{N}(\mu, \sigma_0^2)$.
- Mean of temperatures 1961-1990 from model $i \sim \mathcal{N}(\mu, \sigma_i^2)$ $(i = 1, \ldots, 9)$.
- μ is present mean temperature.
- σ_0^2 is variance of current temperature, assumed to be known.
- σ_i^2 is uncertainty of model i for predicting present temperature. We expect $\sigma_i^2 > \sigma_0^2$.

Data and Distributions, ctd.

- Mean of temperatures for 2071-2100 from model i $\sim \mathcal{N}(\mu + \Delta\mu, (q\sigma_i)^2)$ ($i = 1, \ldots, 9$).
- $\Delta\mu$ is climate change.
- q is increase of uncertainty of model i about future.

All variables are independent given the parameters in the basic model. A second model introduces a correlation between the mean for present and future prediction by the same model.

Put a noninformative prior on all parameters (except σ_0) and compute posterior by MCMC. See results on a separate figure.

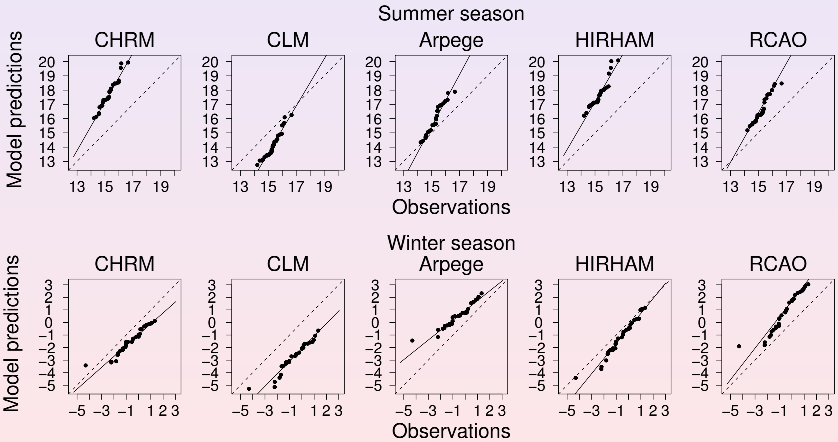
Criticism

- Interannual variability (for a fixed season and region) is not considered.
- Independence of different models is questionable.
- Biases of the models are not explicitely estimated, but subsumed under the variances σ_i^2.
- Regional models should give better predictions.

Our approach

- Do not average over 30 years. Then need to include possible trends in the model.
- Consider regional models for Europe and results for the alpine region only: $44° - 48°$N, $5° - 15°$E.
- Observations and model output are transformed to the same grid of $0.5°$ ($\approx$ 56km) in both directions.
- Use only 5 models which are based on different global models (or at least different runs of the same model). Otherwise need a hierarchical model because of high correlation between GCM and RCM.
- Consider only the A2 scenario (moderately high emissions).

Biases of control runs



Model assumptions

- Observed data for $t = 1961, \ldots, 1990$

$$\sim \mathcal{N}(\mu + \gamma(t - 1975.5), \sigma^2).$$

- Outputs from control run of model i for the same years

$$\sim \mathcal{N}(\mu + \beta_i + \gamma(t - 1975.5), b_i^2\sigma^2).$$

  (β_i is additive bias, b_i multiplicative bias)

- All variables are independent: RCM's attempt to reproduce the climate, not the weather of a specific year.
- Unobserved data for $t = 2071, \ldots, 2100$

$$\sim \mathcal{N}(\mu + \Delta\mu + (\gamma + \Delta\gamma)(t - 2085.5), q^2\sigma^2).$$

Assumptions about scenario runs

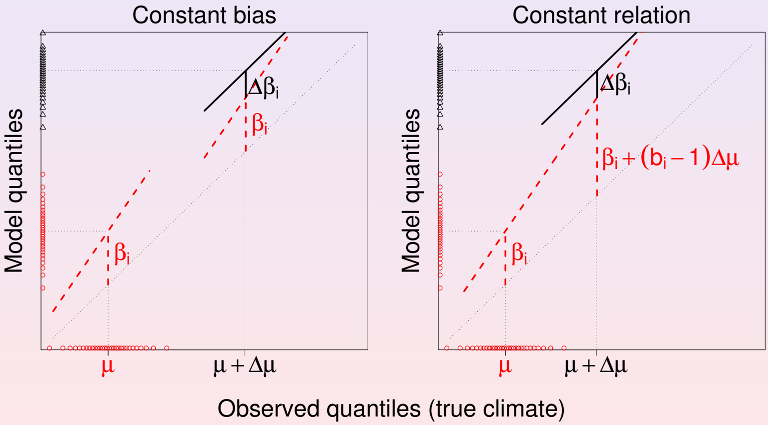
Two key questions

- How to extrapolate biases to the future?
- Can we allow for changes in the future biases?

Answers

- At least two extrapolations are possible, that we call constant bias and constant relation.
- Allowing bias changes leads to non-identifiability. Informative priors provide a reasonable solution.

Graphical illustration



Mathematical Formulation

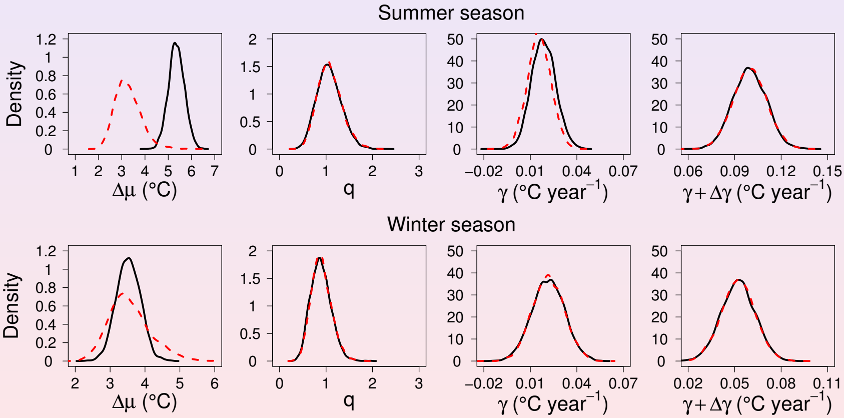
Constant bias:
Outputs from run of model i for years $t = 2071, \ldots, 2100$:

$$\sim \mathcal{N}(\mu + \Delta\mu + \beta_i + \Delta\beta_i + (\gamma + \Delta\gamma)(t - 2085.5), (qb_i q_{b_i})^2 \sigma^2).$$

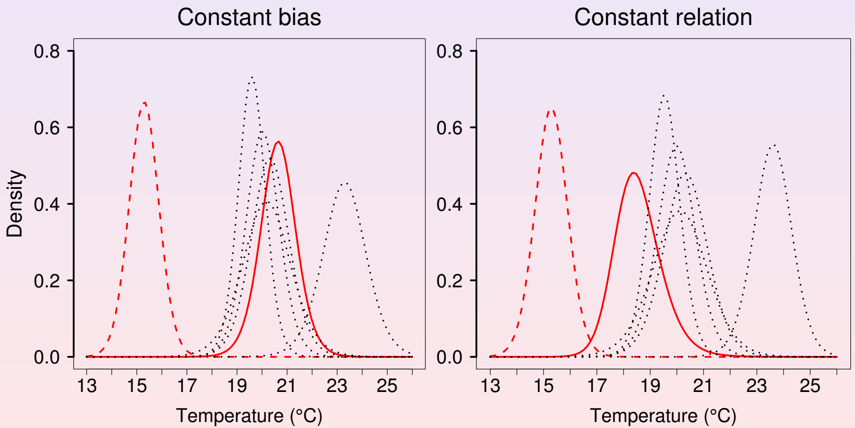
$\Delta\beta_i$ is change in additive bias, q_{b_i} change in multiplicative bias: Put an informative prior on these to keep them near 0 and 1 respectively.

Constant relation replaces $\Delta\mu$ by $b_i \Delta\mu$.

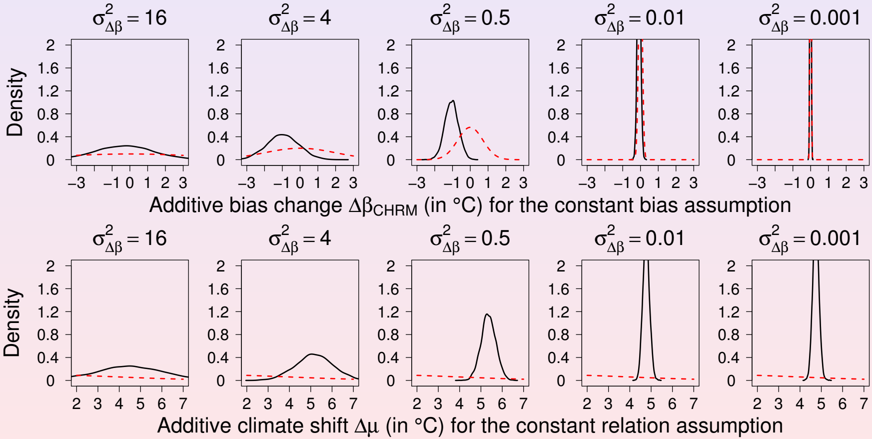
Posteriors for main parameters



Posterior predictive densities



Sensitivity to priors



Summary

- Statistics for model output from complex data raises new questions.
- Studying distributions instead of mean values gives more information.
- Correcting for biases is important, but assumptions are necessary to do this also for future predictions.
- All model overestimate variability in the control run in summer. Implications for climate predictions ? Is this a special feature of the alpine region ?

Future plans

- More than one scenario. This might help to distinguish between constant bias and constant relation.
- Cross validation for information about reasonable choice of priors for bias changes $\Delta\beta_i$ and q_{b_i}.
- Less temporal and spatial averaging (individual gridpoints, monthly means).
- Other variables than temperature; multivariate analysis.
- Hierarchical modeling for different GCM/RCM combinations (unbalanced designs).