The Bayesian Processor of Forecasts on the probabilistic forecasting of long-range dependent variables using General Circulation Models

Hristos Tyralis, and Demetris Koutsoyiannis

National Technical University of Athens

Asia Oceania Geosciences Society
14th Annual Meeting
Singapore | 06-11 August 2017
Session HS20: Modeling and Analysis of Hydrologic Processes in the Context of Climate Change
Quantification of GCMs uncertainty

- Quantification of the uncertainties of the GCMs projections is a mainstream subject.
- Discussion on the potential of the reduction of uncertainties (Hawkins and Sutton 2009, 2011).
- Knutti and Sedláček (2013) conclude that the progress in terms of narrowing uncertainties is too limited.
- An overview of methods to evaluate uncertainty of deterministic models, not only in the climate science, is presented in Uusitalo et al. (2015).
- Quantification of uncertainty with simulation of the local weather (e.g. Groves et al. 2008), combination of multiple models (Smith et al. 2009, Chowdhury and Sharma 2011, Strobach and Bel 2015), bias corrections.
- Methods are criticized.
Proposed framework

• The Bayesian Processor of Forecasts (BPF) is based on the concept of conditional stochastic independence (de Finetti 1974, Krzysztofowicz 1985).

• The BPF “combines a prior distribution, which describes the natural uncertainty about the realization of a hydrologic process, with a likelihood function, which describes the uncertainty in categorical forecasts of that process, and outputs a posterior distribution of the process, conditional upon the forecasts” (Krzysztofowicz 1985).

How the BPF works

The BPF combines historical observations $x_2$ and deterministic forecasts $x_1$ to make predictions.

### Fitting

- $y_3$:
  - Historical observations $x_2$
  - Deterministic forecast $x_1$

### Prediction

- $y_4$:
  - Fitting
  - Prediction

Stochastic model:

\[
h(y_4|y_3, x_1) = f(x_1|y_3, y_4) g(y_3, y_4) / \xi(y_3, x_1)
\]

Combining information from observations and deterministic model outputs.

Conditional independence:

\[
f_i(x_1|21, x_{22}, ..., x_{2n}) = f_i(x_1|x_i)
\]

\[
f_n(x_{11}, x_{12}, ..., x_{1n}|x_{21}, x_{22}, ..., x_{2n}) = \prod_{i=1}^{n} f_i(x_1|x_1, x_2, ..., x_{2n})
\]
Distinct fitting periods

- The estimation of the stochastic model parameters should better be performed using only data that were not used in the GCM fitting/tuning, i.e. for the period after 2006.
- This would correspond to the so-called split-sample technique (Klemeš 1986), which avoids possible model overfitting on the available data and thus artificially good performance.
- This corresponds to model fitting period after 2006.
Case study

• The observations are modelled using the Hurst-Kolmogorov process (HKp, also known as fractional Gaussian noise, fGn, Koutsoyiannis 2002, 2003).

• However, the modelling can be performed using any normal stationary stochastic process.

• A linear model is used to represent the relation between the observations and the deterministic model output.

• Estimates of the parameters are obtained using the Maximum Likelihood Estimator for both the HKp (Tyralis and Koutsoyiannis 2011, Tyralis 2016) and the linear cases.

• Uncertainty in the estimation of the parameters is not considered (See also Tyralis and Koutsoyiannis 2014).
Examples using simulations

Deterministic model of poor quality

Perfect deterministic model
Examples using simulations

Deterministic model of good quality

Deterministic model of good quality, moved up
### Application of the methods – Deterministic models

<table>
<thead>
<tr>
<th>Model Name</th>
<th>Temperature</th>
<th>Precipitation</th>
<th>Institute ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>GISS-E2-H</td>
<td>✓</td>
<td>✓</td>
<td>NASA GISS</td>
</tr>
<tr>
<td>GISS-E2-R</td>
<td>✓</td>
<td></td>
<td>NASA GISS</td>
</tr>
<tr>
<td>HadGEM2-AO</td>
<td>✓</td>
<td>✓</td>
<td>NIMR/KMA</td>
</tr>
<tr>
<td>IPSL-CM5A-LR</td>
<td>✓</td>
<td>✓</td>
<td>IPSL</td>
</tr>
<tr>
<td>IPSL-CM5A-MR</td>
<td>✓</td>
<td>✓</td>
<td>IPSL</td>
</tr>
<tr>
<td>MIROC5</td>
<td>✓</td>
<td>✓</td>
<td>MIROC</td>
</tr>
<tr>
<td>MIROC-ESM</td>
<td>✓</td>
<td>✓</td>
<td>MIROC</td>
</tr>
<tr>
<td>MIROC-ESM-CHEM</td>
<td>✓</td>
<td>✓</td>
<td>MIROC</td>
</tr>
<tr>
<td>MRI-CGCM3</td>
<td>✓</td>
<td>✓</td>
<td>MRI</td>
</tr>
<tr>
<td>NOAA GFDL GFDL-CM3</td>
<td>✓</td>
<td>✓</td>
<td>NOAA GFDL</td>
</tr>
<tr>
<td>NOAA GFDL GFDL-ESM2G</td>
<td>✓</td>
<td>✓</td>
<td>NOAA GFDL</td>
</tr>
<tr>
<td>NOAA GFDL GFDL-ESM2M</td>
<td>✓</td>
<td>✓</td>
<td>NOAA GFDL</td>
</tr>
<tr>
<td>NorESM1-M</td>
<td>✓</td>
<td>✓</td>
<td>NCC</td>
</tr>
<tr>
<td>NorESM1-ME</td>
<td>✓</td>
<td>✓</td>
<td>NCC</td>
</tr>
</tbody>
</table>
General Circulation Models
Area of interest and Thiessen polygons

Temperature

Precipitation

GISS-E2-H

Stations
Temperature, fitting period 1916-2005

GISS-E2-H

prediction quantiles refer 95% confidence regions

MRI-CGCM3
Temperature, fitting period 2006-2015

Prediction quantiles refer 95% confidence regions

GISS-E2-H

MRI-CGCM3
Precipitation, fitting period 1916-2005

Prediction quantiles refer 95% confidence regions

GISS-E2-H

MRI-CGCM3
Precipitation, fitting period 2006-2015

Prediction quantiles refer 95% confidence regions.
Temperature 95% envelopes for all examined GCMs

fitting period 1916-2005

Looks like a Bayesian thistle! See for naming Tyralis and Koutsoyiannis (2017)

fitting period 2006-2015
Precipitation 95% envelopes for all examined GCMs

fitting period 1916-2005

fitting period 2006-2015
Conclusions

• Proofs and results can be found in Tyralis and Koutsoyiannis (2017).
• The BPF can be applied to any normal stationary stochastic process. Examples so far included the case of Markovian processes.
• The framework quantifies the uncertainty of the GCMs predictions.
• Large uncertainties are observed.
• The inclusion of the uncertainty in a fully Bayesian setting, also considering the uncertainty of parameters, would result in even higher uncertainties of the forecasted variables.
References


