Distribution correction techniques to bias correct climate simulations: hypotheses and a multivariate extension

Sylvie Parey\(^{(1)}\), Léonard Dekens\(^{(1,2)}\), Mathilde Grandjacques\(^{(3)}\), Didier Dacunha-Castelle\(^{(4)}\)

International Conference Energy & Meteorology
Bari, 26-29 June 2017

\(^{(1)}\) EDF R&D, France  -  \(^{(2)}\) Ecole Normale Supérieure Lyon et université Claude Bernard Lyon - \(^{(3)}\) CEA LIST LIANES Le Bourget le lac - \(^{(4)}\) Université Paris XI Orsay
European Climatic Energy Mixes (ECEM): A C3S Sectoral Information Service for the Energy Sector

ECEM is a 27-month project to develop a proof-of-concept climate service – or Demonstrator – for the European energy sector. It will enable the energy industry and policy makers assess how well different energy supply mixes in Europe will meet demand, over different time horizons (from seasonal to long-term decadal planning), focusing on the role climate has in the mixes.
• Bias adjustment based on distribution

• Hypotheses

• Generalization to multivariate distributions

• Results
  – Simulation study
  – Real climate data

• Next steps
• Climate models **simulate large scale** climate variables evolution under different climate change scenarios

• Impact studies =>
  – Observed meteorological variables
  – Small area or precise location

• Need for
  – Bias adjustment
  – Downscaling

• Distribution based techniques (quantile mapping, CDFt)
  – Adjust the distribution of the modelled variable by that of the observed one (at the same spatial scale or at a finer spatial scale)
General setting and hypothesis

- We have 3 distributions, from which we would like to infer a 4th one:

\[
\begin{align*}
&\text{Observations} & & \text{Model} \\
&Y(F) & & X(G) \\
&[Y_f(F_1)] & & X_f(G_1) \\
\end{align*}
\]

\(H_{3-1}\): transfer from Model to Observations stationary in time (quantile mapping)

\(H_{3-2}\): transfer from present to future similar for Model and Observations (CDFt)
Multivariate Distribution Correction

- Generalization of quantile mapping approaches
- Hypotheses
  - H1: \( \forall k, \exists \) process \( X_{k,t} \), restriction of \( X_t \) period \( P_k \) stationary and weakly mixing
  - H2: the distributions are continuous with densities strictly positive on their support
- Mathematical background

\[
\begin{align*}
U^1 &= H_1(Z^1) \\
U^k &= H_{k_1 \cdots k_{k-1}}(Z^1, \ldots, Z^k) \\
U^p &= H_{p_1 \cdots p_{p-1}}(Z^1, \ldots, Z^p)
\end{align*}
\]

Then \( U \) is a uniform distribution on \( \mathbb{R}^p \) with independent marginals

- Rosenblatt theorem: if \( U \) is defined as

\[
\begin{align*}
T_H: \mathbb{R}^p &\rightarrow \mathbb{R}^p, \quad T_H(H) = U^p \\
T_H^{-1}(U^p) &= H
\end{align*}
\]

- Consequence: \( T_{G,F} = T_F^{-1} \circ T_G \) transfers \( G \) onto \( F \) because \( T_G(G) = U^p \) and \( T_F^{-1}(U^p) = F \)
Simulation study

- Simulation of bivariate Gaussian distributions
  - Parameters (mean and variance/covariance) taken from daily mean temperature in Hamburg and Orly in July 1979-2014
    - $F$: $m_1 = 17.5$; $m_2 = 20.0$; $\nu_{11} = 10$; $\nu_{22} = 9$; $\nu_{12} = \nu_{21} = 6$
  - Different hypotheses for:
    - Model errors ($G$):
      - Moderate: $em_1 = 1$; $em_2 = 1.5$; $es_{11} = 0.8$; $es_{22} = 0.75$; $es_{12} = 0.9$
      - Large: $em_1 = 4$; $em_2 = 5$; $es_{11} = es_{22} = es_{12} = 0.5$
    - Climate change ($G_1$ and $F_1$):
      - Moderate: $dm_1 = 2$; $dm_2 = 1.5$; $ds_{11} = ds_{22} = 1.2$; $ds_{12} = 1$
      - Large: $dm_1 = 4$; $dm_2 = 5$; $ds_{11} = ds_{22} = 1.5$; $ds_{12} = 1$

- Corrections
  - Both $H_{3-1}$ and $H_{3-2}$ each variable separately or multivariate BC
  - 6 corrections:
    - Univariate: $H_{3-1}$ and $H_{3-2}$
    - Bivariate 1 first, then 2: $H_{3-1}$ and $H_{3-2}$
    - Bivariate 2 first, then 1: $H_{3-1}$ and $H_{3-2}$
2 « extreme » cases

- Large model errors and small CC
  - Stationarity of transfers => rather $H_{3-1}$
  - Results: correction ratios

- Small model errors and large CC
  - Stationarity of transfers => rather $H_{3-2}$
  - Results: correction ratios
Role of correlation strength

Model errors > climate shifts

Model errors < climate shifts
Test with climate variables

- IPSL climate model CMIP5 simulation, closest grid point
- 2 33-year sub-periods:
  - 1950-1982: calibration period
  - 1983-2015: validation period
- Bias correction
  - on a monthly basis (stationarity hypothesis)
  - Hypothesis $H_{3-1}$
  - Independent univariate / bivariate T first and wind first
### Temperature first

<table>
<thead>
<tr>
<th>Month</th>
<th>correlation</th>
<th>Bivariate correction</th>
<th>Univariate correction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>observations</td>
<td>IPSL model</td>
<td>Distance ratio</td>
</tr>
<tr>
<td>January</td>
<td>0.55</td>
<td>0.55</td>
<td><strong>0.522</strong></td>
</tr>
<tr>
<td>February</td>
<td>0.44</td>
<td>0.43</td>
<td>0.328</td>
</tr>
<tr>
<td>March</td>
<td>0.11</td>
<td>0.18</td>
<td><strong>0.758</strong></td>
</tr>
<tr>
<td>April</td>
<td>-0.08</td>
<td>-0.05</td>
<td>0.713</td>
</tr>
<tr>
<td>May</td>
<td>-0.11</td>
<td>-0.30</td>
<td><strong>0.992</strong></td>
</tr>
<tr>
<td>June</td>
<td>-0.26</td>
<td>-0.29</td>
<td>0.496</td>
</tr>
<tr>
<td>July</td>
<td>-0.31</td>
<td>-0.16</td>
<td><strong>0.179</strong></td>
</tr>
<tr>
<td>August</td>
<td>-0.20</td>
<td>-0.34</td>
<td><strong>0.600</strong></td>
</tr>
<tr>
<td>September</td>
<td>0.05</td>
<td>-0.12</td>
<td>0.855</td>
</tr>
<tr>
<td>October</td>
<td>0.15</td>
<td>0.12</td>
<td><strong>0.934</strong></td>
</tr>
<tr>
<td>November</td>
<td>0.32</td>
<td>0.36</td>
<td><strong>0.788</strong></td>
</tr>
<tr>
<td>December</td>
<td>0.49</td>
<td>0.47</td>
<td>0.671</td>
</tr>
</tbody>
</table>

**Similar results with wind first**
Discussion

• Development of a multivariate distribution correction
  – Based on Rosenblatt-Levy lemma
  – General, whatever the number of variables

• Role of stationarity hypothesis
  – Important, both for univariate and multivariate correction
  – Probably more important than use of multivariate correction

• Further work
  – Extension to consider rainfall distributions
  – Improvement of the coding (conditional distribution estimations using Kernel densities)
  – Further work on hypotheses $H_{3-1}$ and $H_{3-2}$ in the univariate context
Thanks for your attention

Dekens L., Parey S., Grandjacques M., Dacunha-Castelle D.: multivariate distribution correction of climate model outputs: a generalization of quantile mapping approaches; accepted in Environmetrics