Modélisation statistique de données climatiques : Descente d’échelle (aka downscaling) & Correction de biais - extrêmes inclus -

Mathieu Vrac

Journée SAMA (IPSL), 13 mars 2018
Motivations

• Many environmental/hydro/agro/human/economic activities and studies are directly affected by meteorological conditions

• **IPCC** scenarios of climate change have a **coarse spatial resolution**!! Not adapted to the spatial scales of impact studies
  - Environmental, human, social and economic impacts
  - How will climate change interact with environmental features existing at a regional/local scale?

  ➢ **Downscaling**: To derive sub-grid scale (regional or local) weather or climate using General Circulation Models (GCMs) outputs or reanalysis data (e.g. NCEP)

  ➢ **Statistical Bias Correction** also often needed!!

  ➢ Need to improve also the **modelling of extreme events**
Outline

- Downscaling: main statistical approaches (including bias correction)
- Illustration of a SWG approach
  - One extension to extremes
- Illustration of a BC approach
  - One extension to extremes
- Conclusions & perspectives
What is downscaling???

Definition:
Downscaling is the action of generating climatic or meteorological values and/or characteristics at a local scale, based on information (from GCM/reanalyses) given at a large scale.

Coarse atmospheric data
Precipitation, temperature, humidity, geopotential, wind, etc.

~ 250 km

Region, city, fields, station

Local variables (e.g., precip., temp.)
(small scale water cycle, impacts – crops, resources – etc.)
How to downscale?: The basics

- Coarse atmospheric data
  - Precipitation, temperature, humidity, geopotential, wind, etc.

How to use the **coarse simulations** to produce **regional/local** climate features?

- Local variables (e.g., precip., temp.)
  - (small scale water cycle, impacts – crops, resources – etc.)
How to downscale?: The basics

Coarse atmospheric data
Precipitation, temperature, humidity, geopotential, wind, etc.

Dynamical downscaling (RCMs):
- GCMs to drive regional models (5-50km) determining atmosphere dynamics
- Requires a lot of computer time and resources => Limited applications

Statistical downscaling:
- Based on statistical relationships between large- and local-scale variables
- Low costs and rapid simulations applicable to any spatial resolution
- Uncertainties (results, propagation, etc)

Local variables (e.g., precip., temp.)
(small scale water cycle, impacts – crops, resources – etc.)
Main statistical **downscaling** approaches

**Predictors**

\[(X_i)_{i=1,...,p}\]

**Coarse atmospheric data**

Precip., temp., humidity, geopot., wind, etc.

**Downscaling function**

\[f(X_1,...,X_p)\]

**Predictands**

\[Y\]

**Local variables (e.g., precip., temp.)**

(smaller scale water cycle, impacts – crops, resources – etc.)
Main statistical downscaling approaches

Coarse atmospheric data
Precip., temp., humidity, geopot., wind, etc.

Transfer functions
Linear models:
- Multi-linear regression (MLR)
  - Rarely on “raw” climate values
  - More often on PCs from a PCA
  - Or on CVs from a CCA

Local variables (e.g., precip., temp.)
(small scale water cycle, impacts – crops, resources – etc.)

Wigley et al. (1990); Huth (2002); Busuioc et al. (2007); Holloway et al. (2008); etc.
Main statistical downscaling approaches

Coarse atmospheric data
Precip., temp., humidity, geopot., wind, etc.

Transfer functions

Linear Non-linear

Non-linear models:
- Polynomial regression (e.g., \( f \) is cubic)
- Artificial neural networks (ANN)
- Spline functions, GAM, etc.
  - raw values, PCs, or CVs

Local variables (e.g., precip., temp.)
(small scale water cycle, impacts – crops, resources – etc.)

Snell et al. (2000); Cannon and Whitfield (2003); Vrac et al. (2007); Salameh et al. (2009); Burke et al. (2014)
**Main statistical downscaling approaches**

- **Coarse atmospheric data**
  - Precip., temp., humidity, geopot., wind, etc.

- Transfer functions
  - Easy to implement; fast but... often lots of tuning
  - Linear methods tend to underestimate the variance.
  - Both methods can give **absurd values** when applied to predictors out of calibration range.

- Local variables (e.g., precip., temp.)
  - (small scale water cycle, impacts – crops, resources – etc.)
Main statistical downscaling approaches

Coarse atmospheric data
Precip., temp., humidity, geopot., wind, etc.

Same large-scale conditions
same local-scale phenomena

Clustering

Preliminary step before SDM
- **Subjective** methods (Lamb, 1972, UK; Hess & Brezowsky, 1976, EU; etc.)
- **Objective** methods
  - K-means (Diday, 1977)
  - Hierarchy (trees, e.g., Ward, 1963)
  - Mixture of pdf (“EM” algorithm, Dempster et al., 1977)

Local variables (e.g., precip., temp.)
(small scale water cycle, impacts – crops, resources – etc.)

Akkur et al. (1992); Bárdossy et al. (1993); Huth (2001); Sheridan & Kalkstein (2004); etc.
Main statistical downscaling approaches

Coarse atmospheric data
Precip., temp., humidity, geopot., wind, etc.

Clustering

One single LS situation (e.g., Z500)
= One unique cluster

Weather typing

Analogues

Local variables (e.g., precip., temp.)
(small scale water cycle, impacts – crops, resources – etc.)

Calibration data
New data

Monthly SLP

Monthly precipitation

Barnett & Preisendorfer (1978); Zorita & von Storch (1998); Yiou et al. (2007, 2013,...), etc.
Main statistical downscaling approaches

Coarse atmospheric data
Precip., temp., humidity, geopot., wind, etc.

- Not able to predict values outside the calibration range
  e.g.: Extremes may be higher in the future. Analogs won’t.
- Needs some “tuning”:
  - Time and spatial window (similarity)
  - Choice of the distance
  - PCs or raw data

Clustering

Weather typing

Analogue

- Spatial and inter-var dependences kept

Local variables (e.g., precip., temp.)
(small scale water cycle, impacts – crops, resources – etc.)

Barnett & Preisendorfer (1978); Zorita & von Storch (1998); Yiou et al. (2007, 2013,...), etc.
Main statistical \textbf{downscaling} approaches

\begin{itemize}
  \item \textbf{Coarse atmospheric data}
  \begin{itemize}
    \item Precip., temp., humidity, geopot., wind, etc.
  \end{itemize}

  \begin{itemize}
    \item \textit{Model simulating daily weather statistically similar to observations, based on parameters determined by historical records (Wilks and Wilby, 1999)}
  \end{itemize}

  \begin{itemize}
    \item Initially (Richardson, 1981, WGEN):
      \begin{enumerate}
        \item rain occurrences as Markov chain $P_{O_i} = P(O_i | O_{i-1})$
        \item rain intensity as Gamma pdf cond’l on $O_i$
        \item Other var. as pdf cond’l on $O_i$
      \end{enumerate}
  \end{itemize}

  \begin{itemize}
    \item \textbf{Stoch. Weather Generators}
  \end{itemize}

  \begin{itemize}
    \item Many variants e.g., Poisson process for occ. sequences, LARS-WG, Racsko et al. (1991)
  \end{itemize}

  \begin{itemize}
    \item \textbf{Local variables (e.g., precip., temp.)}
      \begin{itemize}
        \item (small scale water cycle, impacts – crops, resources – etc.)
      \end{itemize}
  \end{itemize}
\end{itemize}
Main statistical **downscaling** approaches

**Coarse atmospheric data**
Precip., temp., humidity, geopot., wind, etc.

- Model simulating daily weather statistically similar to observations, based on parameters determined by historical records (Wilks and Wilby, 1999)
- Initially (Richardson, 1981, WGEN):
  1. rain occurrences as Markov chain
  2. rain intensity as Gamma pdf cond’l on $O_i$
  3. Other var. as pdf cond’l on $O_i$

- **NOT Downscaling !!**

**Local variables (e.g., precip., temp.)**
(small scale water cycle, impacts – crops, resources – etc.)

**Stoch. Weather Generators**

Many variants
- e.g., Poisson process for occ. sequences, LARS-WG, Racsko et al. (1991)
- Stat.
Main statistical downscaling approaches

Coarse atmospheric data
Precip., temp., humidity, geopot., wind, etc.

Introducing LS covariates $X_i$:

$P_{O_i} = P(O_i|O_{i-1}, X_i)$

& pdf of intensity = Gamma w/ param. cond’l on (=f° of) covariates $X_i$

Local-scale data are simulated from (cond’l) pdf
=> If twice the same large-scale input, two different results
=> Uncertainty assessment (Semenov, 2007)

Local variables (e.g., precip., temp.)
(small scale water cycle, impacts – crops, resources – etc.)

Stoch. Weather Generators

Non-stat.

Vrac et al. (2007); Furrer & Katz (2007); Carreau & Vrac (2011); Wong et al. (2014); Eden et al. (2014); etc.
Main statistical downscaling approaches

Coarse atmospheric data
Precip., temp., humidity, geopot., wind, etc.

- Initially, a NWP issue: forecasts have biases…
  - Not only the mean but higher moments (variance…) and extremes
  - How to correct (i.e., remove the bias) the weather forecasts?
  - Now, extended to climate models:
    How to correct the GCM or RCMs of their statistical biases?

Local variables (e.g., precip., temp.)
(small scale water cycle, impacts – crops, resources – etc.)

MOS/bias correct.

CDF mapping
Main statistical downscaling approaches

Could also be RCM simulations…

Coarse atmospheric data
Precip., temp., humidity, geopot., wind, etc.

Transfer functions
- Linear
- Non-linear

Clustering
Weather typing
Analogues
Stat.
Non-stat.

Stoch. Weather Generators

MOS/bias correct.

CDF mapping

Local variables (e.g., precip., temp.)
(small scale water cycle, impacts – crops, resources – etc.)
Main statistical downscaling approaches

Could also be RCM simulations…

Coarse atmospheric data
Precip., temp., humidity, geopot., wind, etc.

Transfer functions
- Linear
- Non-linear
- Weather typing

Clustering

Stoch. Weather Generators

MOS/bias correct.

Well suited for extremes

Less suited for extremes

CDF mapping

Although, more in (near) present context

(Local variables (e.g., precip., temp.))
(small scale water cycle, impacts – crops, resources – etc.)
Stochastic weather generators
One illustration with 2 models

VGLM & NN-CMM
Parameters are functions of (atmospheric, etc.) predictors

\[
\phi(y; \psi) = (1 - \alpha) \delta_0(y) + \alpha \phi_0(y; \psi_0)
\]

Precipitation pdf (at one station)

VGLM & NN-CMM

Vector Generalized Linear Model

Neural Network – Conditional Mixture Model

Same philosophy, different implementations

Vrac et al. (2007, WRR)
Eden et al. (2014, JGR)

Carreau & Vrac (2011, WRR)
The modelling part of VGLM

- **Precipitation probability density function** (N stations):

\[
\phi_{Y_i|X_t}(y) = \prod_{i=1}^{N} \left[ \phi(y_i; \psi_i(X_t)) \right] \\
= \prod_{i=1}^{N} \left[ (1 - \alpha_i(X_t)) \delta_0(y_i) + \left( \alpha_i(X_t) \phi_0(y_i; \psi_{0,i}(X_t)) \right) \right]
\]

with \( \alpha_i(X_t) = \text{Logistic regression}(X_t) \)

\[
\exp(X_t ' \lambda_i) \\
= \frac{\exp(X_t ' \lambda_i)}{1 + \exp(X_t ' \lambda_i)}
\]

and \( \phi_0 = \text{Gamma pdf with parameters} \)

\[
\psi_{0,i}(X_t) = \begin{cases} 
  k_i(X_t) = a_0 + a_1X_1 + \ldots + a_pX_p = a_0 + AX_t \\
  \beta_i(X_t) = b_0 + b_1X_1 + \ldots + b_pX_p = b_0 + BX_t
\end{cases}
\]
The modelling part of **NN-CMM**

- **Precipitation probability density function** (N Stations):

\[
\phi_{Y_i|X_t}(y) = \prod_{i=1}^{N} \left[ \phi(y_i; \psi_i(X_t)) \right]
\]

\[
= \prod_{i=1}^{N} \left[ (1 - \alpha_i(X_t)) \delta_0(y_i) + \left( \alpha_i(X_t) \phi_0(y_i; \psi_{0,i}(X_t)) \right) \right]
\]

with \( \phi_0(y; \psi_{0,i}(X_t)) \in \sum_{j=1}^{m} \pi_{i,j}(X_t) f(y; \theta_{i,j}(X_t)) \);

\[
\psi_i(x) = \left( \alpha_i(x), (\pi_{i,j}(x)), (\theta_{i,j}(x)) \right)_{j=1,...,m}
\]

![Neural Network Diagram]

\[ f = \begin{cases} 
\text{Gaussian} & \\
\text{Log-Normal} & \\
\text{Hybrid Pareto} & 
\end{cases} \]

✓ Carreau & Vrac (2011)
✓ Carreau & Bengio (2009a,b)
One illustration: Daily pdfs with **NN-CMM-2L**
from Carreau and Vrac (2011)

Spell with the **highest** cum. vol. of rain

Longest wet spell
**NN-CMM-2L vs. (NN-Cond’l) Gamma**

Illustration on the Orange station

QQ-plot (log) **CMM-2L**

QQ-plot (log) **Ber-Gamma**

Williams (1998)

**VGLM w/ Gamma: good but... not always enough!**
Peaks over threshold (POT): Generalized Pareto Distribution (GPD)

- Not simply values higher than the threshold but **excesses**
  - Excess $V$ of the variable $Z$ above threshold $u$ is defined as $Z-u$, given that $Z>u$: $V = Z-u \mid Z>u$
  - **EVT**: If $u$ is large enough, $F_u(v)$ can be approximated by the Generalized Pareto Distribution (GPD)

$$P(Z - u \leq y \mid Z > u) = 1 - \left(1 + \frac{\xi y}{\sigma_u} \right)^{-1/\xi}$$

- $u =$ selected threshold
- $\sigma_u =$ scale parameter ($>0$)
- $\xi =$ shape parameter

- $\xi < 0 \Rightarrow$ bounded tail (e.g., from uniform, Weibull, Beta)
- $\xi = 0 \Rightarrow$ light tail (e.g., from exponential, Gaussian, Gumbel)
- $\xi > 0 \Rightarrow$ heavy tail (e.g., from Fréchet, Student t, Cauchy)
Modelling the **whole** precipitation distribution

**Vrac & Naveau (2007)**


**Wong, Maraun, Vrac, Widmann, Eden, Kent (2014)**

*Stochastic model output statistics for bias correcting and downscaling precipitation including extremes*, Journal of Climate, 27, 6940–6959, doi: http://dx.doi.org/10.1175/JCLI-D-13-00604.1
Merging classical and EV distributions in VGLM

- Based on Frigessi et al. (2002):
  “Dynamic mixture model for unsupervised tail estimation without threshold”

\[
\phi_0(y | \psi_0) = c_{\psi_0} \left[ (1 - w(y|m, \tau)) \Gamma(y | \gamma, \lambda) \right. \\
\left. + w(y|m, \tau) GPD(y | \xi, \sigma, u = 0) \right]
\]

Gamma pdf

Generalized Pareto Distribution (GPD) pdf

with \( w(y|m, \tau) = \frac{1}{2} + \frac{1}{\pi} \arctan \left( \frac{y - m}{\tau} \right) \)

Value where transition from \( \Gamma \) to GPD

Transition rate
Illustration on two stations

Quincy: VGLM w/ **Gamma**

Quincy: VGLM w/ **Gamma & GPD**

Aledo: VGLM w/ **Gamma & GPD**

Bivariate extension

(See Vrac, Naveau, Drobinski, 2007, NPG)
Main statistical downscaling approaches

Coarse atmospheric data
Precip., temp., humidity, geopot., wind, etc.

Transfer functions
Clustering
Stoch. Weather Generators
MOS/bias correct.

Linear

"Perfect-prognosis" (PP) methods

"Model Output Statistics" (MOS) methods

Local variables (e.g., precip., temp.)
(small scale water cycle, impacts – crops, resources – etc.)

Could also be RCM simulations…
Bias correction

1. ONE predictor

   PR
   ▼
   Bias Correction Model
   ▼
   Corrected PR

Statistical downscaling

vs.

Several predictors

SLP  Z850  T2  Hum
▽  ▽  ▽  ▽
Statistical Downscaling Model
▽
Local-scale (e.g., PR) simulations
Bias correction vs. Statistical downscaling

1. ONE predictor vs. Several predictors
2. Not necessarily local scale vs. Local scale

Model outputs

Corrected model outputs

Model outputs

Downscaled model outputs
Bias correction ↔ Statistical downscaling

1. ONE predictor vs. Several predictors
2. Not necessarily local scale vs. Local scale

Bias correction:
- Directly calibrated to link model outputs & observations

Statistical downscaling:
- Assumes “perfect” predictors
  - calibration needs temporal matching between (large-scale) predictors and (local-scale) observations
  - Projections based on predictors from GCMs
Bias correction ⇔ Statistical downscaling

1. ONE predictor vs. Several predictors
2. Not necessarily local scale vs. Local scale

Directly calibrated to link model outputs & observations

Assumes “perfect” predictors
- calibration needs temporal matching between (large-scale) predictors and (local-scale) observations

No daily synchronicity
Bias correction ↔ Statistical downscaling

1. ONE predictor vs. Several predictors
2. Not necessarily local scale vs. Local scale

Directly calibrated to link model outputs & observations
⇒ Link between the distributions
⇒ No need of temporal matching

Main idea: Work on / link CDFs
Bias correction ↔ Statistical downscaling

1. ONE predictor vs. Several predictors
2. Not necessarily local scale vs. Local scale
4. Model-dependant vs. Same for any model

One climate model ⇒ One BC model
(N climate models ⇒ N BC models)

N climate models ⇒ One SD model (calibrated once, e.g., on reanalyses)
Bias correction: main methods

“Delta”-like methods:

1. Characterize how climate model outputs change (between calibration & projection periods)

   \[ \text{Present model output} \] → Compute $\Delta$ = evolution → \[ \text{Future model output} \]

2. Transform (“observed (present) time series with the same change”)

   \[ \text{Present model output} \] → Apply $\Delta$ = same evolution → \[ \text{Futurized observations} \]

“Quantile-quantile”- or “anomaly”-like methods:

1. Present model output

   Compute bias

   \[ \text{Present observations} \]

2. Future model output

   Apply (de-) bias

   \[ \text{Corrected model outputs} \]
"Quantile-quantile"-like methods

- Classical approach: Quantile-mapping
  - Gridcell G; \( X_G \sim F_G \); Station S; \( X_S \sim F_S \)
  
  \[
  F_S(x_S) = F_G(x_G) \iff x_S = F_S^{-1}F_G(x_G)
  \]

You want this
You know this
You obtain this
You know this

One gridcell
\[
\begin{array}{c}
X_G \\
\sim F_G
\end{array}
\]

Reference
\[
\begin{array}{c}
X_S \\
\sim F_S
\end{array}
\]

Equality of distributions
"Quantile-quantile"-like methods

- Classical approach: Quantile-mapping
  - Gridcell G; $X_G \sim F_G$; Station S; $X_S \sim F_S$
    \[
    F_S(x_S) = F_G(x_G) \iff x_S = F_S^{-1}(F_G(x_G))
    \]
  - Visual interpretation: QQplot (between $F_S$ and $F_G$)

First paper(s):
Panofsky and Brier (1958);
Haddad and Rosenfeld (1997)

Many variants:
Wood et al. (2004);
Déqué (2007);
Shabalova et al. (2003);
Piani et al. (2010);
etc.
"Quantile-quantile"-like methods

- **Classical approach**: Quantile-mapping
  - Gridcell $G; X_G \sim F_G$; Station $S; X_S \sim F_S$
  - $F_S(x_S) = F_G(x_G) \iff x_S = F_S^{-1} F_G(x_G)$
  - Visual interpretation: QQplot (between $F_S$ and $F_G$)

- **2 main limitations**
  - Limited to min and max
    - ? What if $x_G$ out of the calib. range ??
  - Implicit assumption:
    - CDF(proj. period) = CDF(calib. period)
    - ? What if the CDF changes ??
"Quantile-quantile"-like methods

Cumulative Distribution Function - transform (CDF-t)

<table>
<thead>
<tr>
<th></th>
<th>Present</th>
<th>Future</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GCM</strong></td>
<td>$F_{Gp}$</td>
<td>$F_{Gf}$</td>
</tr>
<tr>
<td>(1 gridcell)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Station</strong></td>
<td>$F_{Sp}$</td>
<td>$F_{Sf}$</td>
</tr>
</tbody>
</table>

Classical QQ:

$$X_{Gf} \text{ projected on } F_{Gp}$$

$$x_{Sf} = F_{Sp}^{-1} F_{Gp}(x_{Gf})$$

QQ from CDF-t:

$$X_{Gf} = \{x_{Gf, i}\}_{i=1, \ldots, N}$$

$$x_{Sf} = F_{Sf}^{-1} F_{Gf}(x_{Gf})$$

Vrac et al. (2012)
"Quantile-quantile"-like methods
*Cumulative Distribution Function - transform* (CDF-t)

- **Formulation** (from Vrac et al., 2012):
  - Based on mathematical transformation $T$ applied to LS CDF
  - $F_{Sp}$ Verifies eq.(1) by definition
  - Assumption: Eq. (2) remains valid in the future

<table>
<thead>
<tr>
<th></th>
<th>Present</th>
<th>Future</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GCM</strong> (1 gridcell)</td>
<td>$F_{Gp}$</td>
<td>$F_{Gf}$</td>
</tr>
<tr>
<td><strong>Station</strong></td>
<td>$F_{Sp}$</td>
<td>$F_{Sf}$</td>
</tr>
</tbody>
</table>

$$T(F_{Gp}(x)) = F_{Sp}(x) \quad (1)$$

Let $x = F_{Gp}^{-1}(u)$ with $u \in [0,1]$

$$\Rightarrow T(u) = F_{Sp}\left(F_{Gp}^{-1}(u)\right) \quad (2)$$

$$F_{Sf}(x) = T(F_{Gf}(x)) \iff F_{Sf}(x) = F_{Sp}\left(F_{Gp}\left(F_{Gf}(x)\right)\right) \quad (3)$$
Formulation (from Vrac et al., 2012):

- Based on mathematical transformation $T$ applied to LS CDF
- $F_{Sp}$ Verifies eq.(1) by definition
- Assumption: Eq. (2) remains valid in the future

**Advantage:**

Changes of the large-scale CDF directly accounted for

$$T(F_{Gp}(x)) = F_{Sp}(x) \quad (1)$$

Let $x = F_{Gp}^{-1}(u)$ with $u \in [0,1]$

$$\Rightarrow T(u) = F_{Sp} \left( F_{Gp}^{-1}(u) \right) \quad (2)$$

$$F_{Sp}(x) = F_{Gp} \left( F_{Gf}^{-1} \left( F_{Gf}(x) \right) \right) \quad (3)$$
✓ **Methodology & evaluations:**
Michelangeli et al. (2009, wind)
Kallache et al. (2011, extreme PR)
Vrac et al. (2012, T&PR)
Vrac & Vaittinada Ayar (2016, combined BC/DS)
Vrac et al. (2016, stochastic/PR, SSR)
Volosciuk et al. (2017, combined BC/DS), … etc.

✓ **Applications:**
Oettli et al. (2011, T/PR/Rad/evt for crop model)
Colette et al. (2012, BC/RCM)
Tisseuil et al. (2012, river flows)
Vautard et al. (2012, DRIAS)
Vigaud et al. (2013, PR for Indian water resources)
Defrance et al. (2017, Africa PR), … etc.

✓ **Intercomparison exercises:**
Vaittinada Ayar et al. (2015, EURO/MED-CORDEX)
Guttierez et al. (2018, VALUE)
Hertig et al. (2018, extremes), … etc.
Cumulative Distribution Function - transform

Two extensions: Extremes & covariates

CDF-t \[ \Rightarrow F_{sf}(x) = F_{sp}\left(F_{gp}\left(F_{gf}(x)\right)\right) \]

XCDF-t: F’s are GPD’s (Kallache, Vrac, Michelangeli, Naveau, 2011, JGR)

\[
F_{sf}(y - c_{fac}) = 1 - \left(1 + \frac{\xi_{sp}}{\xi_{gp}} \frac{\sigma_{gp}}{\sigma_{sp}} \left(\left(1 + \frac{\xi_{gf}}{\sigma_{gf}} y\right)^{\xi_{gp}/\xi_{gf}} - 1\right)\right)^{-\left(1/\xi_{sp}\right)}
\]

More complex than a GPD…

But… If we assume \( \xi_{gf} = \xi_{gp} \), then

\[
F_{sf}(y) = 1 - \left(1 + \frac{\xi_{sp}}{\xi_{gf}} \frac{\sigma_{gp}}{\sigma_{sp}} (y + c_{fac})\right)^{-\left(1/\xi_{sp}\right)}
\]

\[\Rightarrow F_{sf} \text{ is a GPD with } \xi_{sf} = \xi_{sp} \text{ and } \sigma_{sf} = \sigma_{gf} \left(\sigma_{sp} / \sigma_{gp}\right)\]
Cumulative Distribution Function - transform

Two extensions: Extremes & covariates

CDF-t \implies F_{sf}(x) = F_{sp} \left( F_{gp}^{-1} \left( F_{gf}(x) \right) \right)

- **XCDF-t:** F’s are GPD’s (Kallache, Vrac, Michelangeli, Naveau, 2011, JGR)

  If we assume $\xi_{gf} = \xi_{gp}$, then

  \Rightarrow F_{sf} is a GPD with $\xi_{sf} = \xi_{sp}$ and $\sigma_{sf} = \sigma_{gf} \left( \sigma_{sp} / \sigma_{gp} \right)$

- **Inclusion of covariate information:**

  $F(.) = GPD(\sigma, \xi)$

  \[ \sigma_t = \exp \left( a_0 + a_1 \text{cov}_1^t + \ldots + a_n \text{cov}_n^t \right) \]

  ✓ The parameters may have different link functions

  ✓ The covariates vary with time (e.g., as in Clim.Ch.)

  \Rightarrow **one CDF** $F_{sf}^t(.)$ **per time step**!

  ✓ Feature similar to (cond’l) SWG
eXtreme CDF-t (XCDF-t)

One illustration

XCDF-t on precipitation at 5 French stations.
Calib=1951-1985, Proj= 1986-1999, x-axis in mm/d, y-axis in probability
Conclusions (some) on downscaling & BC

- Many (and many) **models and applications** of downscaling & BC
  - **My favorite ones:**
    - *Stochastic WGs*: cond’l event-wise variability/uncertainty
    - *MOS / Bias correction*: DS of CDFs from CDFs
  - Choice of the predictors is a major issue in Stat. DS
  - Non-stationarity (⚠️ the SWGs should not explode ⚠️)
  - Applying Stochastic WGs to GCMs *may be* better than to RCMs

- **RCMs vs. SDMs**: Not a conflict => complementary approaches
  - ⇒ Both have pros & cons

- There is not one good SDM for all variables and regions
  - ⇒ Different skills according to regions/variables/applications, etc.
  - ⇒ Use ensembles if possible!
Commercial break (well, it’s free)

- Some **R packages** developed for **Stochastic downscaling & BC**:
  - **NHMixt** (Vrac & Naveau, 2007, Wong et al., 2014)
    - Statistical mixture model Gamma & GPD
    - Inclusion of covariates
  - **condmixt** (Carreau & Vrac et al., 2011)
    - ANN-Conditional mixture model
    - Various distributions (Gaussian, Log-N, hybrid Pareto)
  - **McSIM** (Bechler, Vrac, Bel, 2015)
    - Spatial models for extreme (maxima) downscaling
    - Max-stable processes
  - **CDVineCopulaConditional** (Bevacqua et al., 2017)
    - Copula-based model for compound events
    - Multivariate dependence
  - **CDFt** (Vrac et al., 2012) & **XCDFt** (Kallache et al., 2011) → **Run at IPSL**
    - Bias correction
    - Also for excesses (i.e., GPD)
  - **EC–BC** (Vrac and Friederichs, 2015) → **R²D²** (Vrac, 2018)
    - For multivariate properties of BC/DS
    - Post-processing of any 1d method
Some perspectives: dependences

- **Spatial/multi-sites** dependence structure, e.g.:
  - Spatial VGLM (Chandler and Wheater, 2002)
    - Stationary spatial dependences / Limited number of locations
  - “Hybrid Spatial Downscaling” (HSD) for extreme fields (Bechler et al., 2015)
    - Combines RCM with geostatistical cond’l simulations / DS anywhere in the region
  - “Rank Resampling for Distributions and Dependences” (R2D2, Vrac, 2018)
    - 2-step approach: univariate SDM or BC + dependence reconstruction

- **Multi-variables** dependence structure, e.g.:
  - Copula-based model for compound events (Bevacqua et al., 2017)
    - Parametric model (=> can be limited in dimensions)
  - R2D2 (Vrac, 2018)
    - Non-parametric approach / Multi-sites & multi-variables / even in high-dimension

- **Multidimensional** (sites and/or variables) dependence of extremes
  - More complex / Non-parametric approach (e.g., Naveau et al., 2014)
  - Needed to improve/extend "Extreme Event Attribution", impact studies, etc.
Some perspectives: dependences

- **Spatial/multi-sites** dependence structure, e.g.:
  - Spatial VGLM (Chandler and Wheater, 2002)
    - Stationary spatial dependences / Limited number of locations
  - “Hybrid Spatial Downscaling” (HSD) for extreme fields (Bechler et al., 2015)
    - Combines RCM with geostatistical cond’l simulations / DS anywhere in the region
  - “Rank Resampling for Distributions and Dependences” (R2D2, Vrac, 2018)
    - Non-parametric approach / Multi-sites & multi-variables / even in high-dimension

- **Multi-variables** dependence structure, e.g.:
  - Copula-based model for compound events (Bevacqua et al., 2017)
    - Parametric model (=> can be limited in dimensions)
  - R2D2 (Vrac, 2018)
    - Non-parametric approach / Multi-sites & multi-variables / even in high-dimension

- **Multidimensional** (sites and/or variables) dependence of extremes
  - More complex / Non-parametric approach (e.g., Naveau et al., 2014)
  - Needed to improve/extend "Extreme Event Attribution", impact studies, etc.

- The questions are not necessarily technical anymore:
  - What do you **trust** (or not) in the model outputs?
  - What do you want to **correct/preserve**?
Thank you…

SAMA group at work (?)

“Data don’t make any sense, we will have to resort to statistics.”