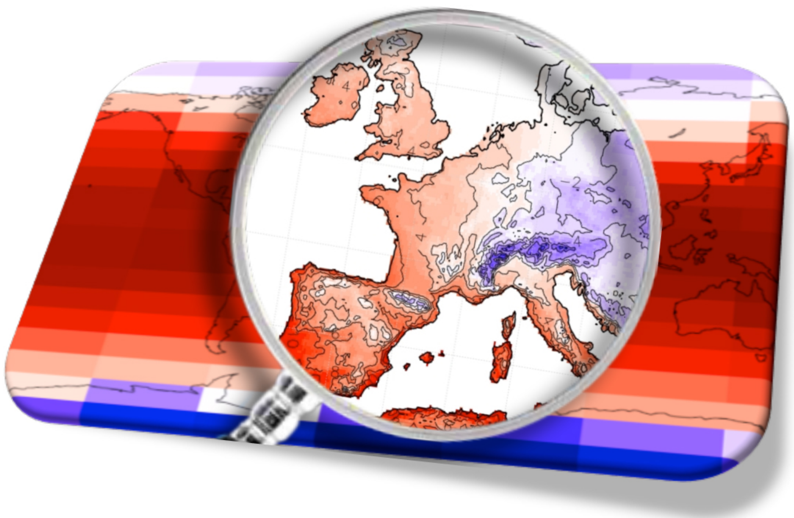
LSCE

Laboratoire des sciences du climat & de l'environnement

Sciences de l'environnement

Institut Pierre Simon Laplace

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**

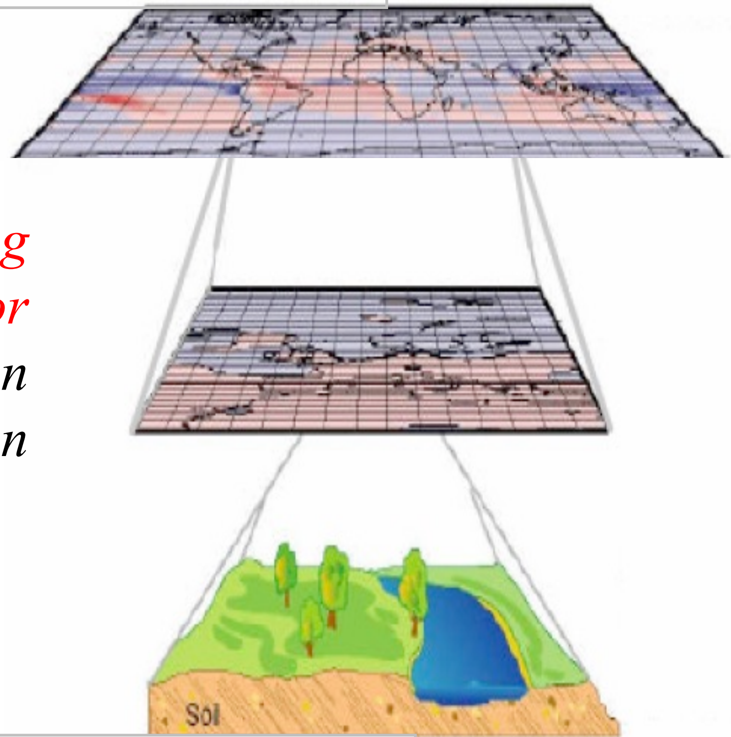
Outline

- 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 ???

≈ 250 km

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



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.

Region, city,
fields, station

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

How to downscale?: The basics

≈ 250 km

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

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

?

Region, city, fields, station

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

How to downscale?: The basics

≈ 250 km

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)

Region, city, fields, station

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

Main statistical downscaling approaches

Predictors
$(X_i)_{i=1,\ldots,p}$

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

Donwscaling
function
$f(X_1,\ldots,X_p)$

Predictands
Y

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.

Transfer functions

Linear | Non-linear

Linear models:

- Multi-linear regression (MLR)
  - ✓ Rarely on "raw" climate values
  - ✓ More often on PCs from a PCA
  - ✓ Or on CVs from a CCA

Error

$$\vec{Y} = AX + \varepsilon$$

Matrix of linear **coefficients**

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

More complex *f*

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

Linear | Non-linear

- ✓ 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.

Preliminary step before SDM

Same large-scale conditions

same local-scale phenomena

Clustering

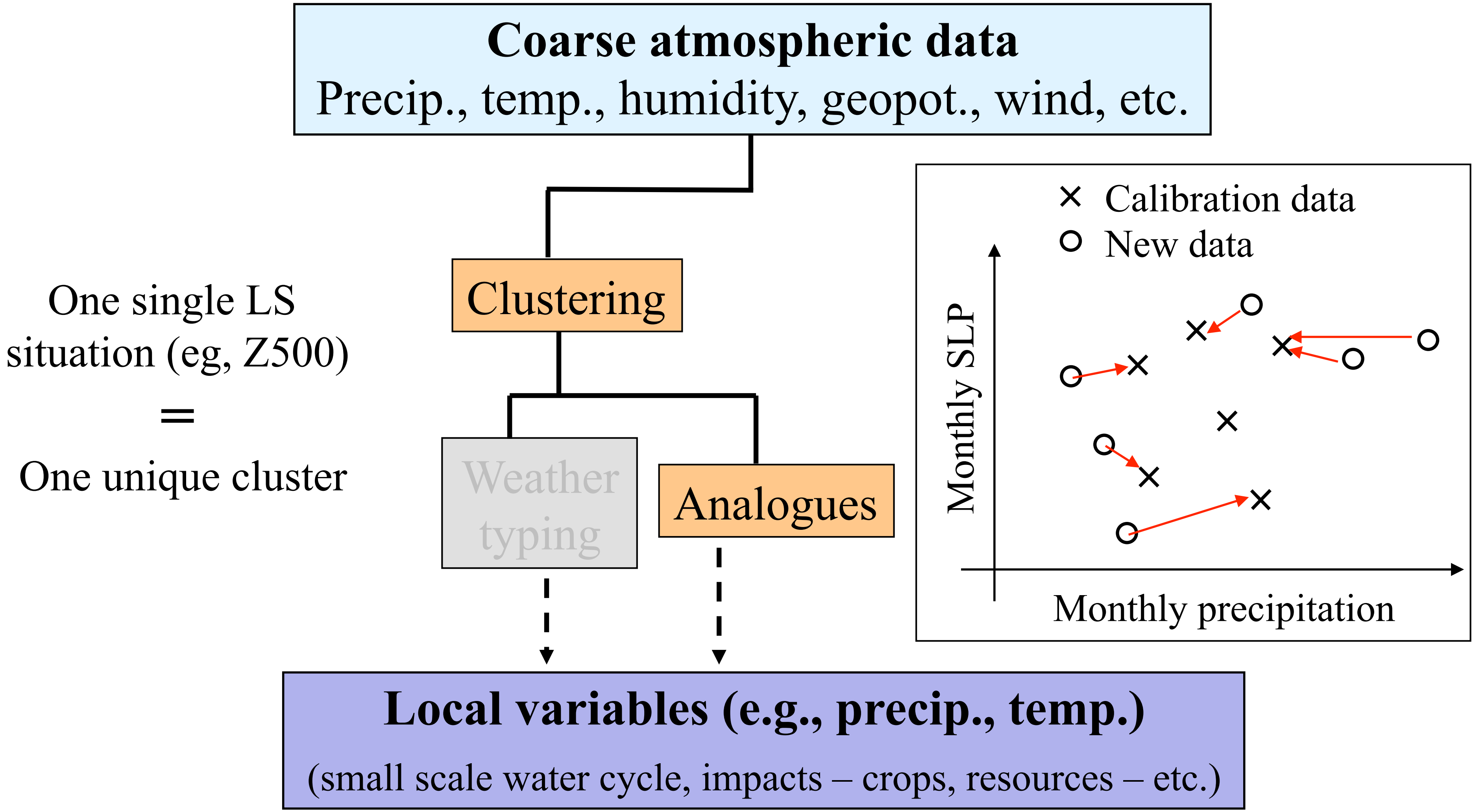
Weather typing

- *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



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.

Clustering

- Needs some "tuning":
  - ✓ Time and spatial window (similarity)
  - ✓ Choice of the distance
  - ✓ PCs or raw data

Weather typing

Analogues

- 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 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)

Stoch. Weather Generators

Stat.

Initially (Richardson, 1981, WGEN):
1. rain occurrences as Markov chain $Po_i = P(O_i | O_{i-1})$
2. rain intensity as Gamma pdf cond'l on O_i
3. Other var. as pdf cond'l on O_i

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

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.

Model simulating daily weather statistically similar to observations, based on parameters determined by historical records (Wilks and Wilby, 1999)

Stoch. Weather Generators

Initially (Richardson, 1981, WGEN):

NOT Downscaling !!

1.
2. rain intensity as Gamma pdf cond'l on O_i
3. Other var. as pdf cond'l on O_i

Stat.

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

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.

Introducing <u>LS covariates X_i</u>:

$$Po_i = P(O_i \mid O_{i-1}, X_i)$$

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

Stoch. Weather Generators

Non-stat.

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.)

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.

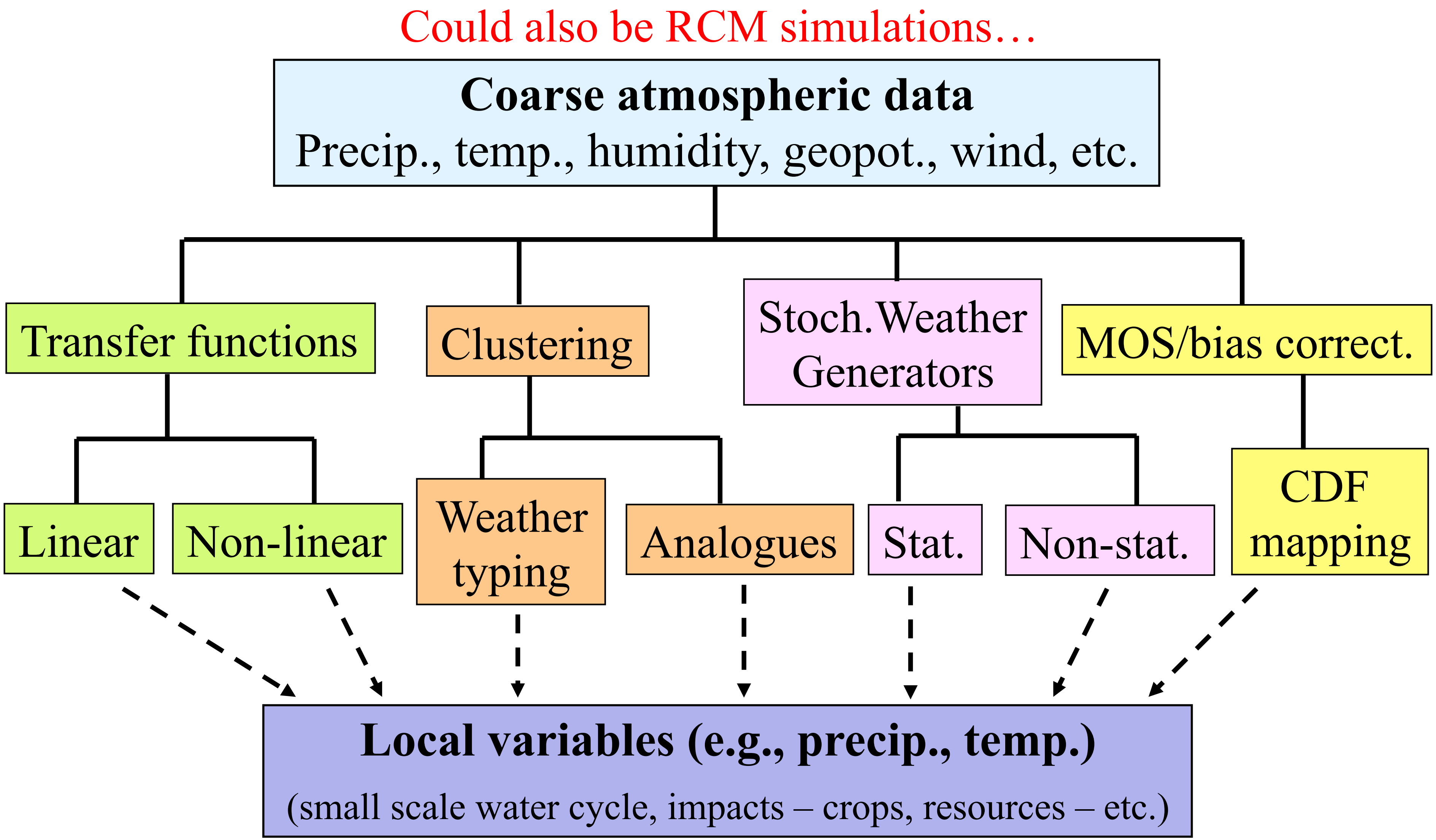
MOS/bias correct.

CDF mapping

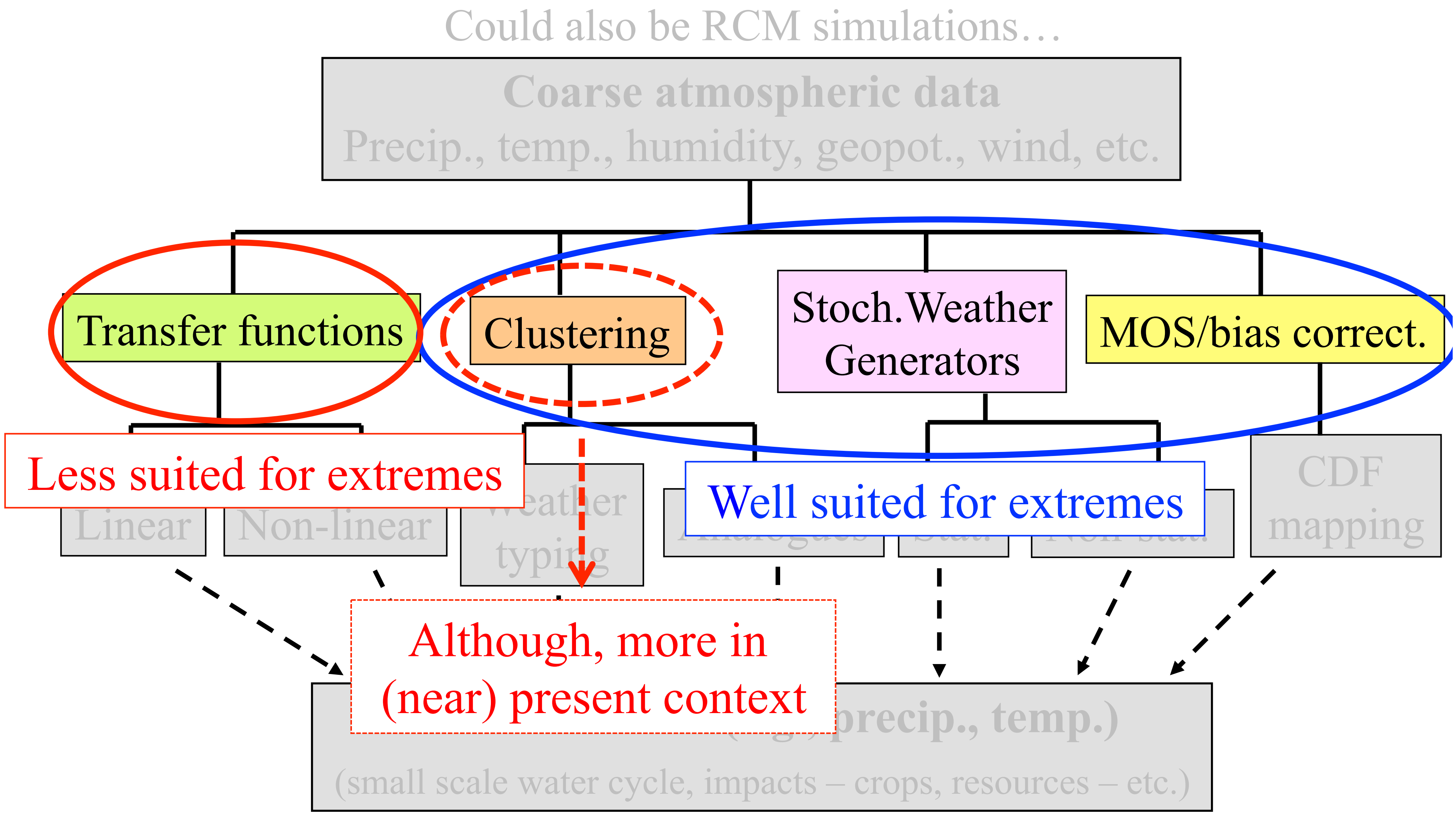
- Initially, a NWP issue: **forecasts have biases…**
  - **Not only the mean** but higher moments (variance…) and extremes
  - How to correct (i.e., remove the bias) the weather forecasts?
  - Now, **extended to climate models**:
    How to correct the GCM or RCMs of their statistical biases?

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

Main statistical downscaling approaches



Main statistical downscaling approaches



Stochastic weather generators

One illustration with 2 models

VGLM & NN-CMM

VGLM & NN-CMM

Vector Generalized Linear Model | Neural Network – Conditional Mixture Model

Precipitation pdf (at one station)

$$\boxed{\phi(y;\psi) = \underbrace{(1-\alpha)\,\delta_0(y)}_{\text{no rain}} + \underbrace{\alpha\phi_0(y;\psi_0)}_{\text{rain}>0}}$$

Parameters are fonctions of (atmospheric, etc.) predictors

X_{t-1} GLM | ANN X_{t-1}

X_t GLM

$$\alpha(\mathbf{X}_t, st)$$
$$\psi_0(\mathbf{X}_t, st)$$

$$\alpha(\mathbf{X}_t, st)$$
$$\psi_0(\mathbf{X}_t, st)$$

ANN X_t

X_{t+1} GLM | ANN X_{t+1}

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):

$$\phi_{\mathbf{Y}_t|\mathbf{X}_t}(\mathbf{y}) = \prod_{i=1}^{N}\left[\phi\left(y_i;\psi_i\left(\mathbf{X}_t\right)\right)\right]$$

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

with $\alpha_i\left(\mathbf{X}_t\right) = \mathit{Logistic\ regression}\left(\mathbf{X}_t\right)$

$$= \frac{\exp\left(\mathbf{X}_t{}'\lambda_i\right)}{1+\exp\left(\mathbf{X}_t{}'\lambda_i\right)}$$

and $\phi_0 = \text{Gamma pdf with parameters}$

$$\psi_{o,i}\left(\mathbf{X}_t\right) = \begin{cases} k_i\left(\mathbf{X}_t\right) = a_0 + a_1 X_1 + \ldots + a_p X_p = a_0 + \mathbf{A}\mathbf{X}_t \\ \beta_i\left(\mathbf{X}_t\right) = b_0 + b_1 X_1 + \ldots + b_p X_p = b_0 + \mathbf{B}\mathbf{X}_t \end{cases}$$

The modelling part of **NN-CMM**

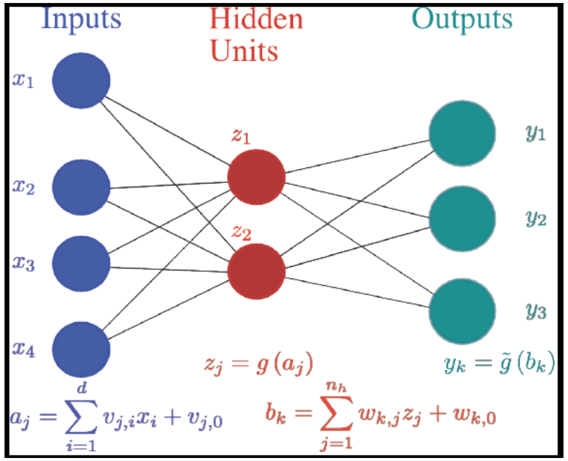
- Precipitation probability density function (N Stations):

$$\phi_{\mathbf{Y}_t|\mathbf{X}_t}(\mathbf{y}) = \prod_{i=1}^{N}\left[\phi\left(y_i;\psi_i\left(\mathbf{X}_t\right)\right)\right]$$

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

with $\phi_0\left(y;\psi_{0,i}\left(\mathbf{X}_t\right)\right) = \sum_{j=1}^{m}\pi_{i,j}\left(\mathbf{X}_t\right)f\left(y;\theta_{i,j}\left(\mathbf{X}_t\right)\right)$

$$\psi_i\left(\mathbf{x}\right) = \left(\alpha_i\left(\mathbf{x}\right),\left(\pi_{i,j}\left(\mathbf{x}\right)\right)_{j=1,\dots,m},\left(\theta_{i,j}\left(\mathbf{x}\right)\right)_{j=1,\dots,m}\right)$$

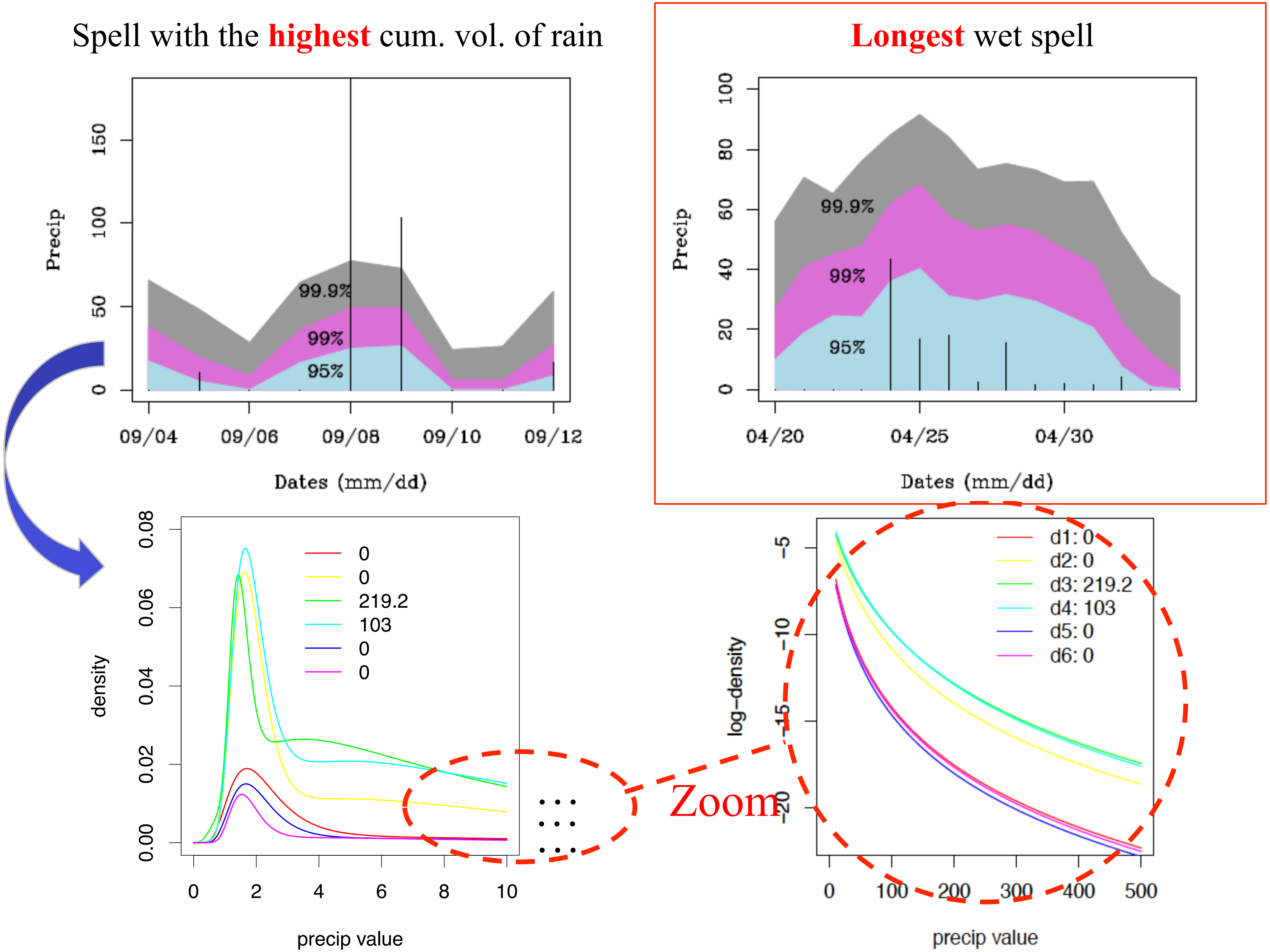


$f =$
- Gaussian
  or
- **Log-Normal**
  or
- Hybrid Pareto

- ✓ Carreau & Vrac (2011)
- ✓ Carreau & Bengio (2009a,b)

One illustration: Daily pdfs with **NN-CMM**-2L

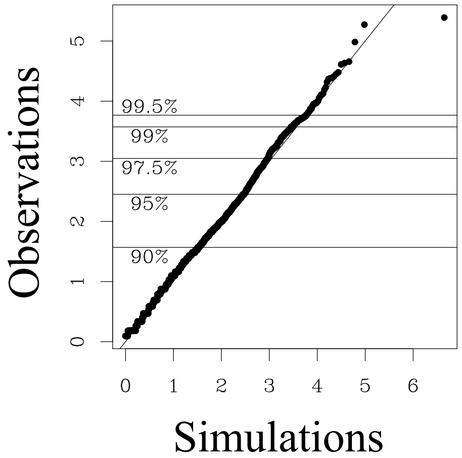
from Carreau and Vrac (2011)



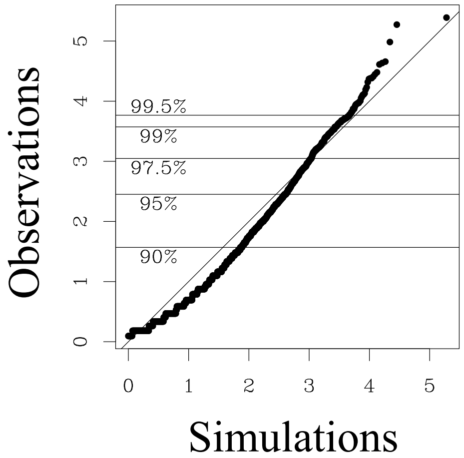
NN-CMM-2L vs. (NN-Cond'l) Gamma

Illustration on the Orange station

QQ-plot (log) CMM-2L



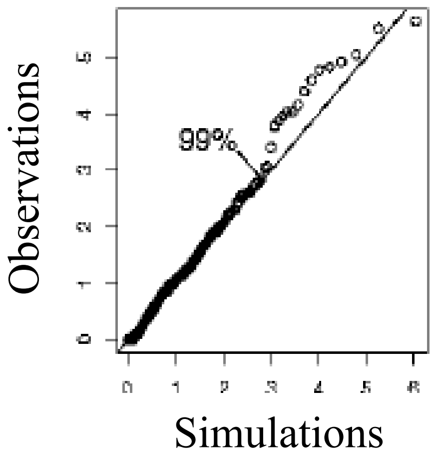
QQ-plot (log) Ber-Gamma



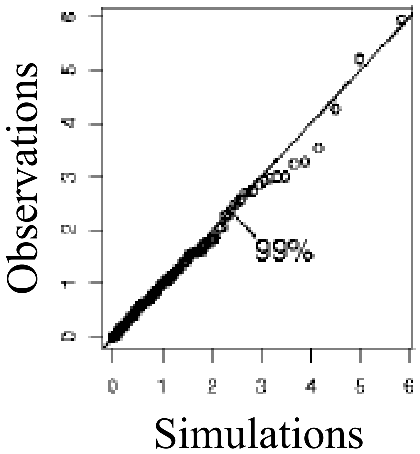
Williams (1998)

VGLM w/ Gamma: good but… not always enough!

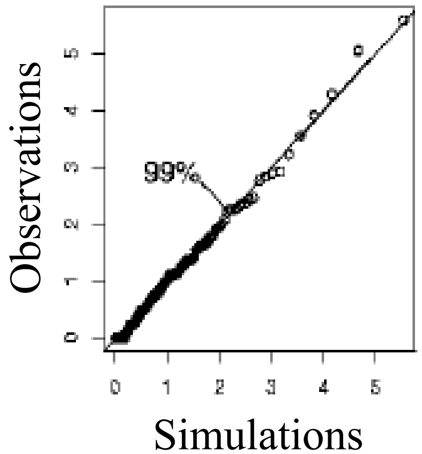
QQplot Charleston



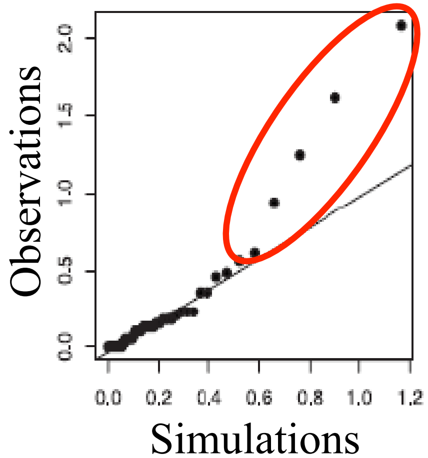
QQplot Galva



QQplot Walnut



QQplot Quincy



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)

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

- ✓ u = selected threshold
- ✓ σ_u = scale parameter (>0)
- ✓ ξ = shape parameter

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

Modelling the **whole** precipitation distribution

Vrac & Naveau (2007)
Stochastic downscaling of precipitation: From dry events to heavy rainfalls, Water Resources Research, 43, W07402, doi: 10.1029/2006WR005308

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"

$$\phi_0\left(y|\psi_O\right) = c_{\psi_0}\left[\left(1 - w\left(y|m,\tau\right)\right)\overbrace{\Gamma\left(y|\gamma,\lambda\right)}^{\text{Gamma pdf}} + w\left(y|m,\tau\right)\underbrace{GPD\left(y|\xi,\sigma,u=0\right)}_{\text{Generalized Pareto Distribution (GPD) pdf}}\right]$$

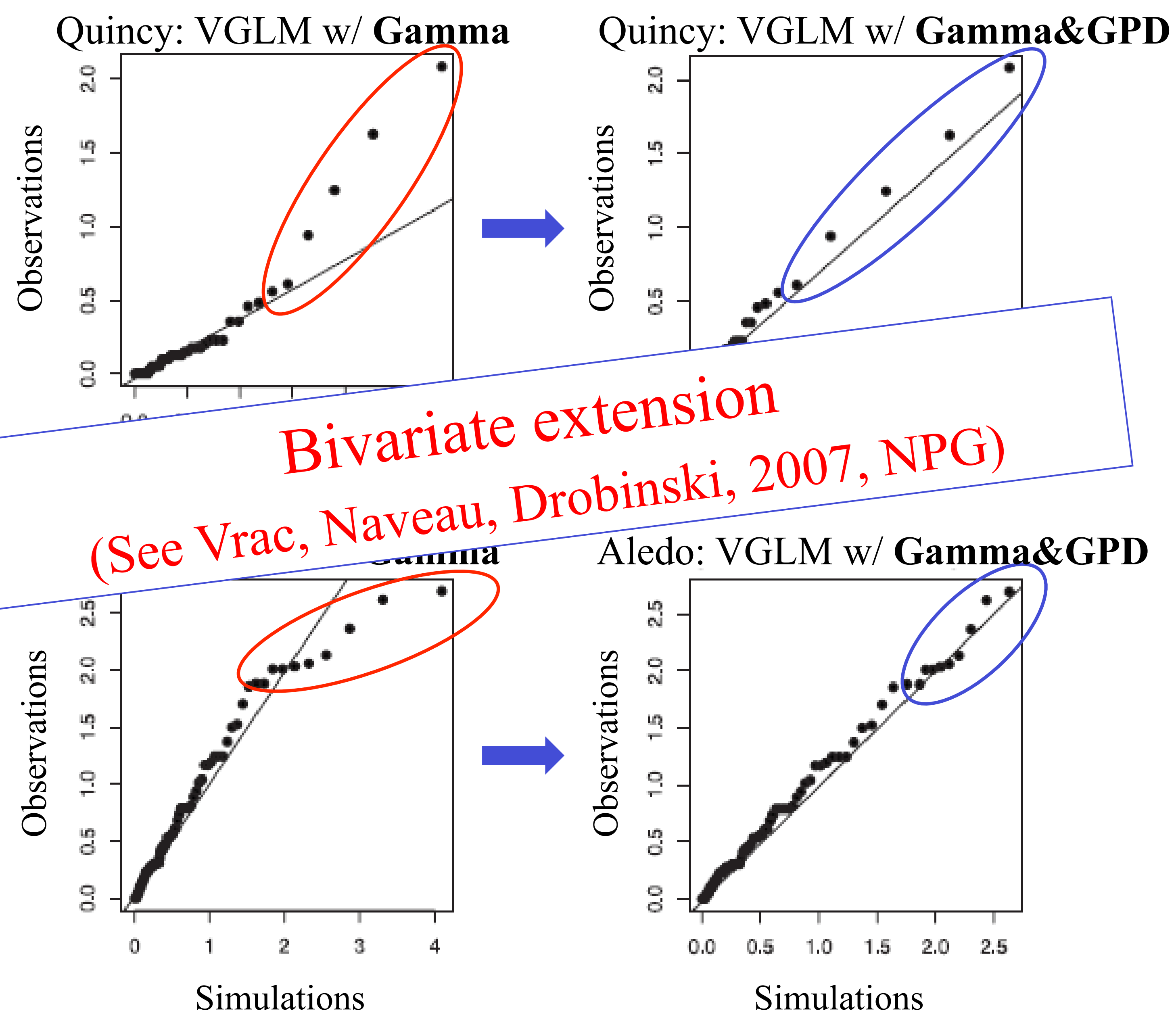
functional weight

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

Value where transition from Γ to GPD

Transition rate

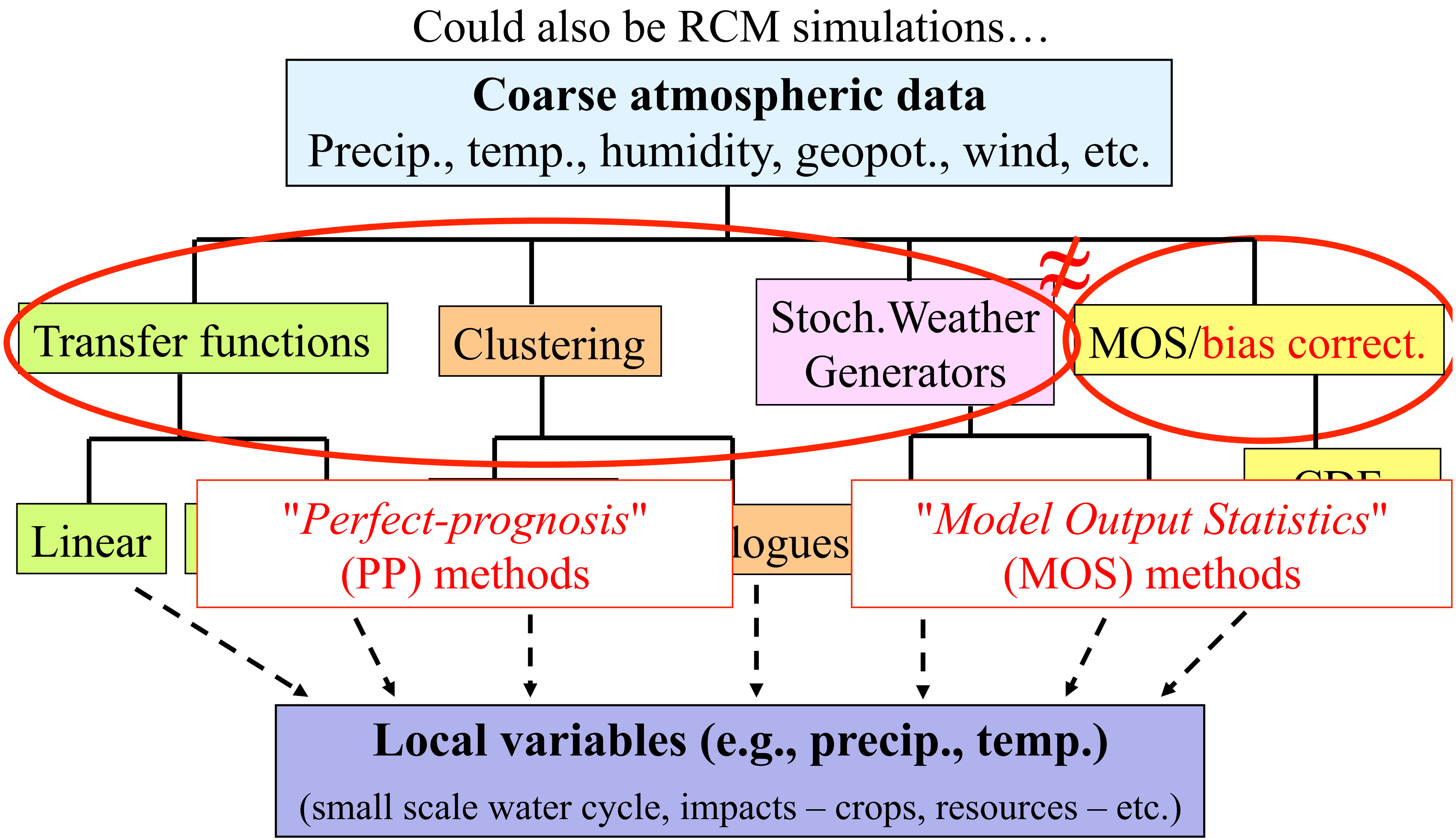
Illustration on two stations



Bivariate extension

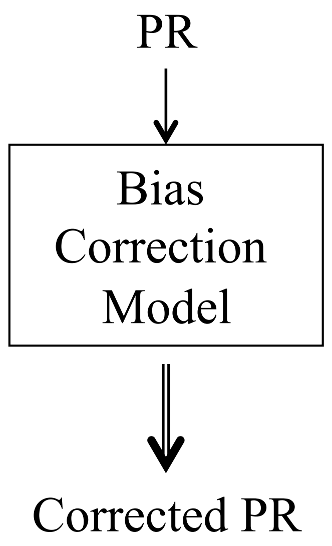
(See Vrac, Naveau, Drobinski, 2007, NPG)

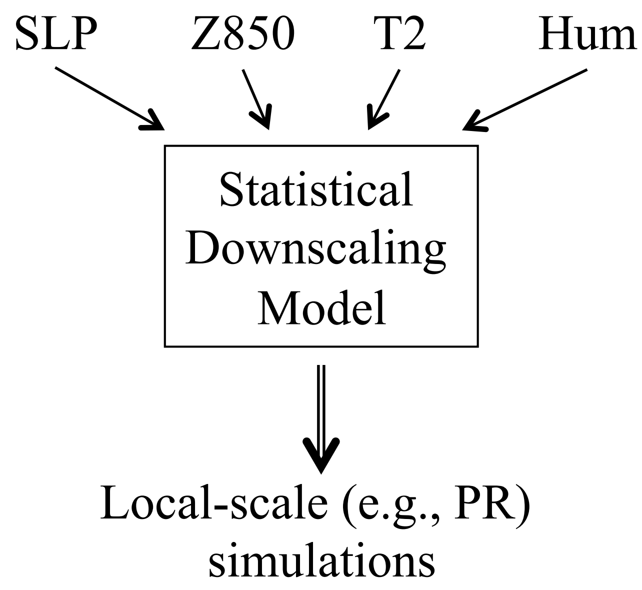
Main statistical downscaling approaches



Bias correction $\not\Leftrightarrow$ Statistical downscaling

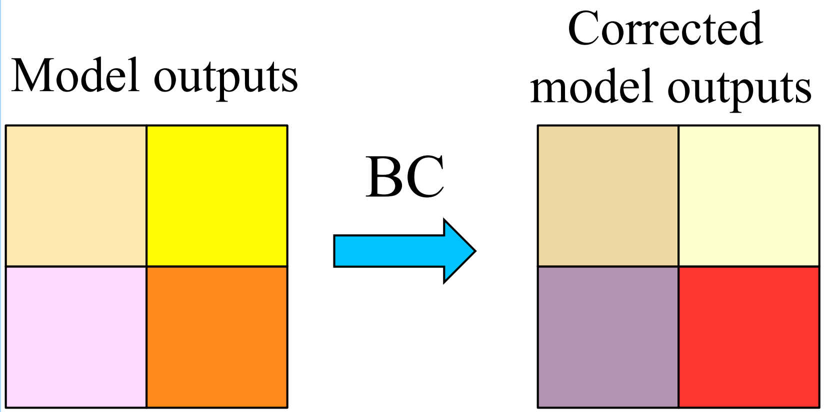
1. ONE predictor vs. Several predictors

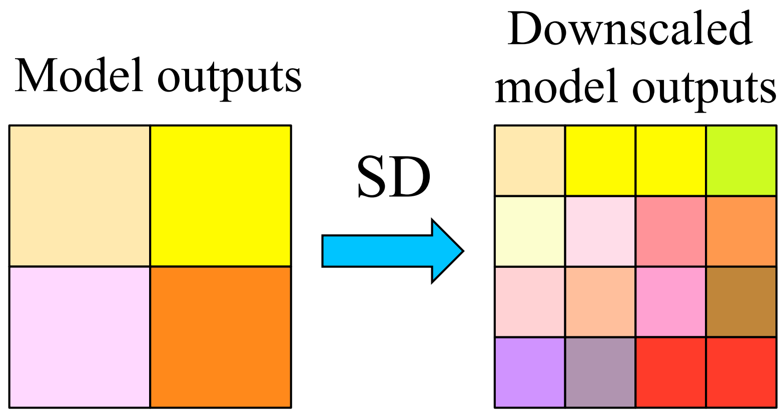




Bias correction $\not\Leftrightarrow$ Statistical downscaling

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





Bias correction ~~⇔~~ Statistical downscaling

1. ONE predictor	vs.	Several predictors
2. Not necessarily local scale	vs.	Local scale
3. “Model Output Statistics”	vs.	“Perfect prognosis”

Directly calibrated to link model outputs & observations

Assumes “perfect” predictors

- calibration **needs temporal matching** between (large-scale) predictors and (local-scale) observations
- Projections based on predictors from GCMs

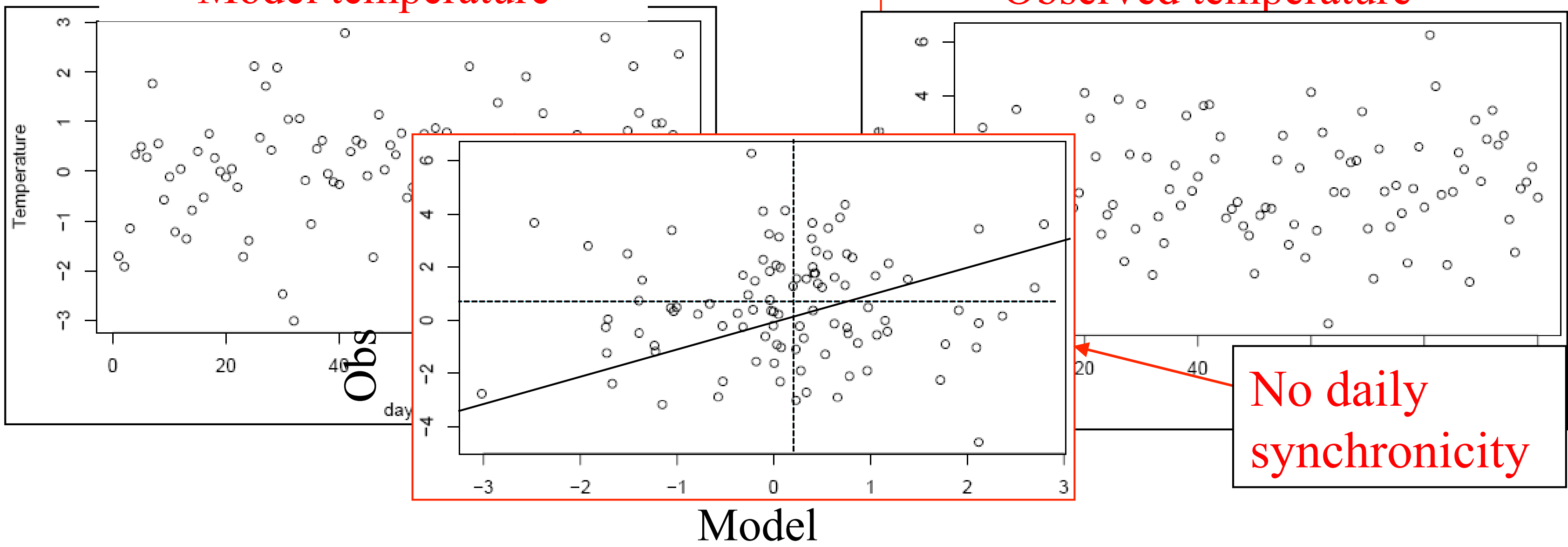
Bias correction $\not\Leftrightarrow$ Statistical downscaling

1. ONE predictor vs. Several predictors
2. Not necessarily local scale vs. Local scale
3. "Model Output Statistics" vs. "Perfect prognosis"

Model Output Statistics	Perfect prognosis
Directly calibrated to link model outputs & observations	Assumes "perfect" predictors - calibration **needs temporal matching** between (large-scale) predictors and (local-scale) rs

Model temperature

Observed temperature



Bias correction ~~⇔~~ Statistical downscaling

1. ONE predictor vs. Several predictors
2. Not necessarily local scale vs. Local scale
3. "Model Output Statistics" vs. "Perfect prognosis"

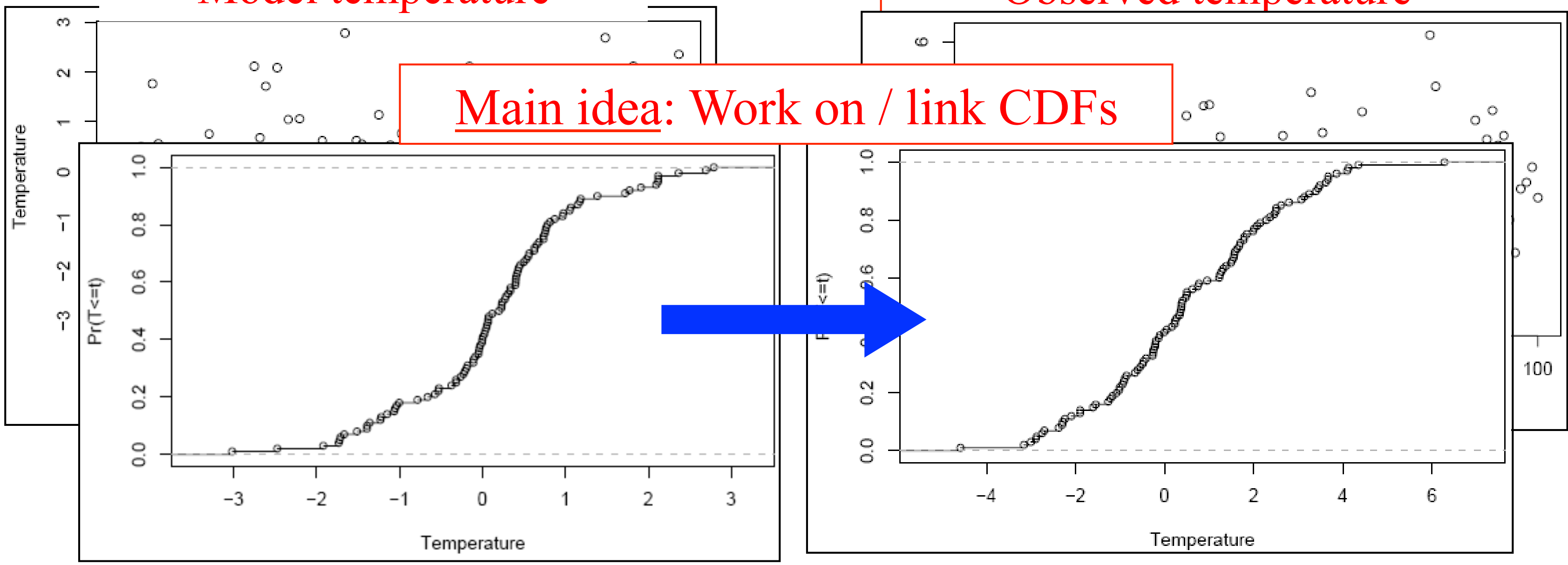
Directly calibrated to link model outputs & observations
⇒ Link between the distributions
⇒ No need of temporal matching

Assumes "perfect" predictors
- calibration **needs temporal matching** between (large-scale) predictors and (local-scale)

rs

Model temperature

Observed temperature



Main idea: Work on / link CDFs

Bias correction $\not\Leftrightarrow$ Statistical downscaling

1.	ONE predictor	vs.	Several predictors
2.	Not necessarily local scale	vs.	Local scale
3.	“Model Output Statistics”	vs.	“Perfect prognosis”
4.	Model-dependant	vs.	Same for any model

One climate model
$\Rightarrow$ One BC model
(N climate models => N BC models)

N climate models
$\Rightarrow$ One SD model (calibrated once, e.g., on reanalyses)

Bias correction: main methods

- "**Delta**"-like methods:

Horizontal approach

Present model output → Compute Δ = evolution → **Future** model output

Present observations → Apply Δ = same evolution → **Futurized** observations

- "**Quantile-quantile**"- or "**anomaly**"-like methods:

Vertical approach

Present model output ↓ Compute bias ↓ **Present** observations

Future model output ↓ Apply (de-) bias ↓ **Corrected** model outputs

"Quantile-quantile"-like methods

- Classical approach: Quantile-mapping
  - Gridcell G; $X_G \sim F_G$; Station S; $X_S \sim F_S$

$$F_S(x_S) = F_G(x_G) \quad \Leftrightarrow \quad x_S = F_S^{-1} F_G(x_G)$$

You **want** this

You **know** this

You **obtain** this

You **know** this

One gridcell

X_G

$\sim F_G$

Equality of distributions

Reference

X_S

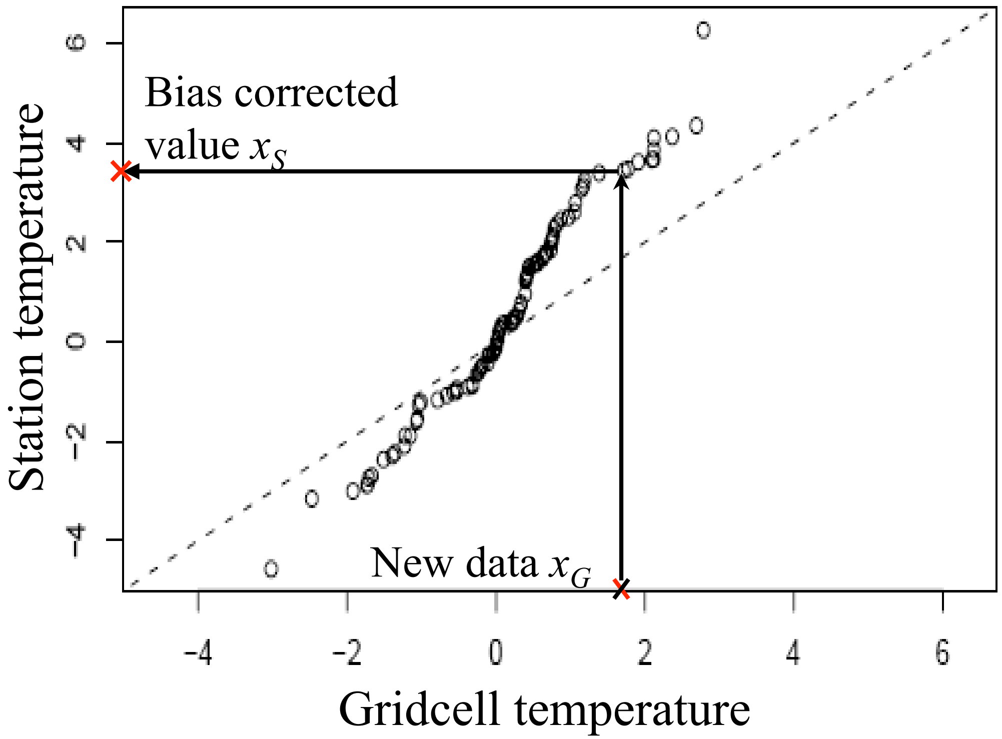
$\sim F_S$

“Quantile-quantile”-like methods

- Classical approach: Quantile-mapping
  - Gridcell G; $X_G \sim F_G$; Station S; $X_S \sim F_S$

$$F_S(x_S) = F_G(x_G) \quad \Leftrightarrow \quad x_S = F_S^{-1} F_G(x_G)$$

  - Visual interpretation: QQplot (between F_S and F_G)



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

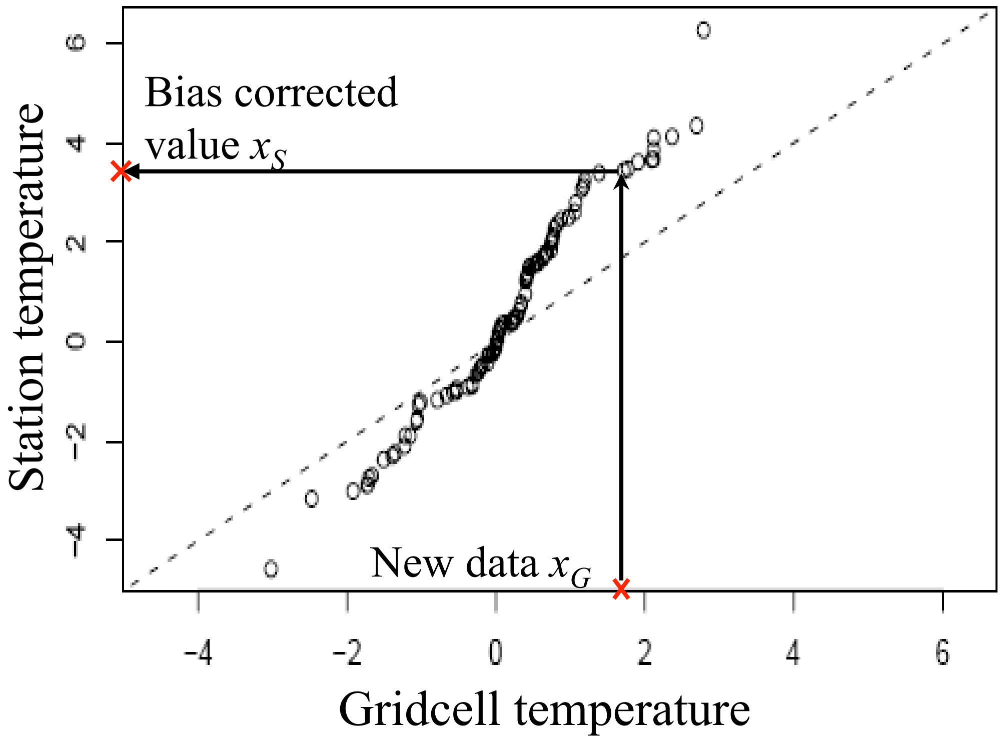
Many variants:
Wood et al. (2004);
Déqué (2007);
Shabalova et al. (2003);
Piani et al. (2010);
etc.

“Quantile-quantile”-like methods

- Classical approach: Quantile-mapping
  - Gridcell G; $X_G \sim F_G$; Station S; $X_S \sim F_S$

$$F_S(x_S) = F_G(x_G) \quad \Leftrightarrow \quad x_S = F_S^{-1} F_G(x_G)$$

  - 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)

	Present	Future
GCM (1 gridcell)	F_{Gp}	F_{Gf}
Station	F_{Sp}	F_{Sf}

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

Classical QQ:

X_{Gf} projected on F_{Gp}

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

QQ from CDF-t:

X_{Gf} projected on F_{Gf}

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

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

	Present	Future
GCM (1 gridcell)	F_{Gp} $\downarrow \boldsymbol{T}$	F_{Gf} $\downarrow \boldsymbol{T}$
Station	F_{Sp}	F_{Sf}

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

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

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

$$F_{Sf}(x) = T\left(F_{Gf}(x)\right) \quad \Leftrightarrow \quad \boxed{F_{Sf}(x) = F_{Sp}\left(F_{Gp}^{-1}\left(F_{Gf}(x)\right)\right)} \quad (3)$$

"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: F

Advantage:

Changes of the large-scale CDF directly accounted for

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

GCM (1 gridcell)	F_{Gp} $\boldsymbol{T}$	F_{Gf} $\boldsymbol{T}$
Station	F_{Sp}	F_{Sf}

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

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

$$F_{Sf}(x) = T\left(F_{Gf}(x)\right) \quad \Leftrightarrow \quad \boxed{F_{Sf}(x) = F_{Sp}\left(F_{Gp}^{-1}\left(F_{Gf}(x)\right)\right)} \quad (3)$$

✓ **<u>Methodology & evaluations</u>:**

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 => $F_{Sf}(x) = F_{Sp}\left(F_{Gp}^{-1}\left(F_{Gf}(x)\right)\right)$

- **X**CDF-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

$$F_{Sf}(y) = 1 - \left(1 + \frac{\xi_{Sp}}{\sigma_{Gf}} \frac{\sigma_{Gp}}{\sigma_{Sp}} \left(y + c_{fac}\right)\right)^{-\left(1/\xi_{Sp}\right)}$$

=> F_{Sf} is a GPD with $\xi_{Sf} = \xi_{Sp}$ and $\sigma_{Sf} = \sigma_{Gf}\left(\sigma_{Sp} / \sigma_{Gp}\right)$

Cumulative Distribution Function - transform

Two extensions : Extremes & covariates

CDF-t => $F_{Sf}(x) = F_{Sp}\left(F_{Gp}^{-1}\left(F_{Gf}(x)\right)\right)$

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

If we assume $\xi_{Gf} = \xi_{Gp}$, then

=> 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:

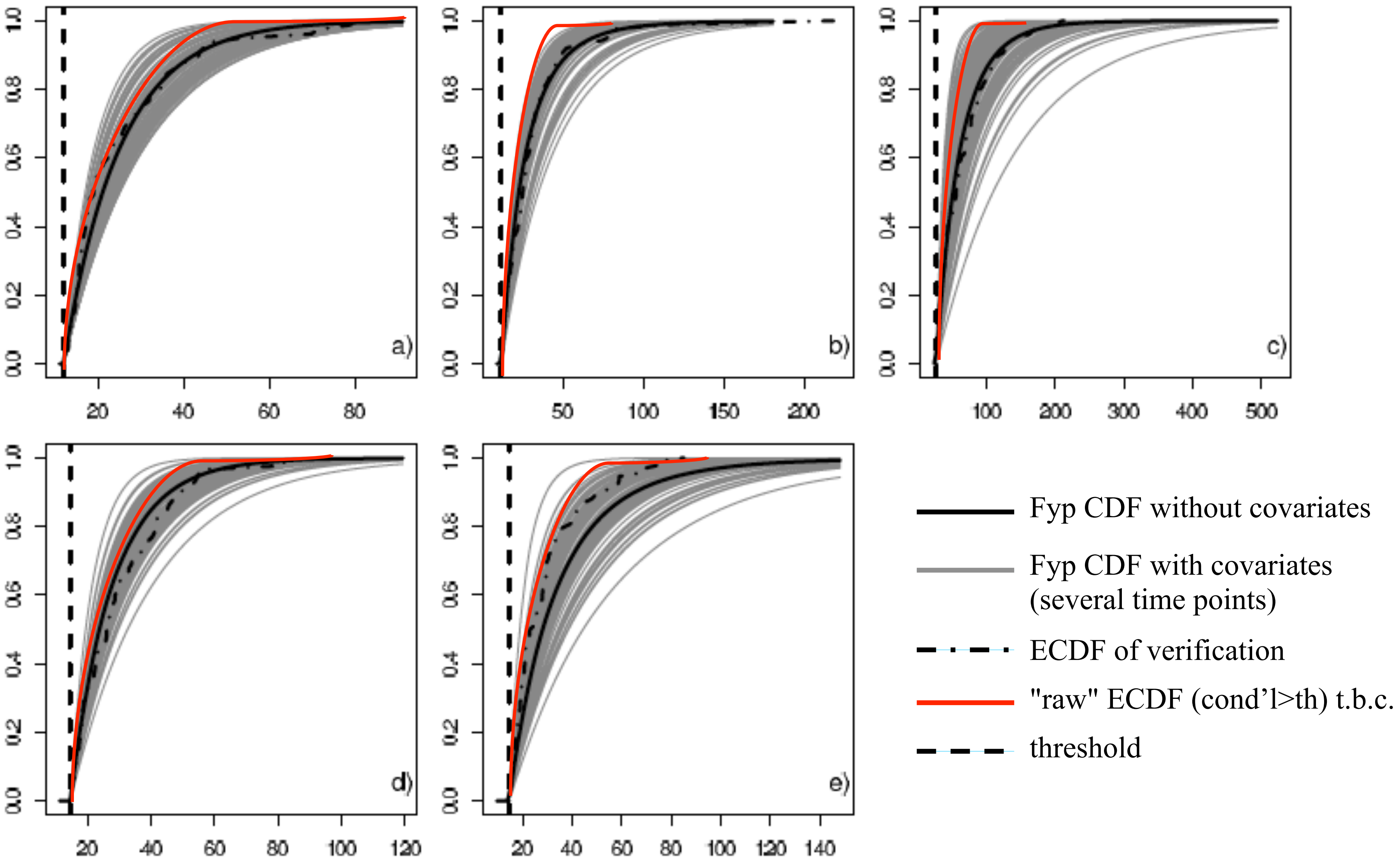
$$F(.) = GPD(\sigma, \xi)$$

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

- ✓ The parameters may have different link functions
- ✓ The covariates vary with time (e.g., as in Clim.Ch.)
  => 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) → **R^2D^2** (Vrac, 2018)
    - ✓ For multivariate properties of BC/DS
    - ✓ Post-processing of any 1d method

http://www.r-project.org
Or my website

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

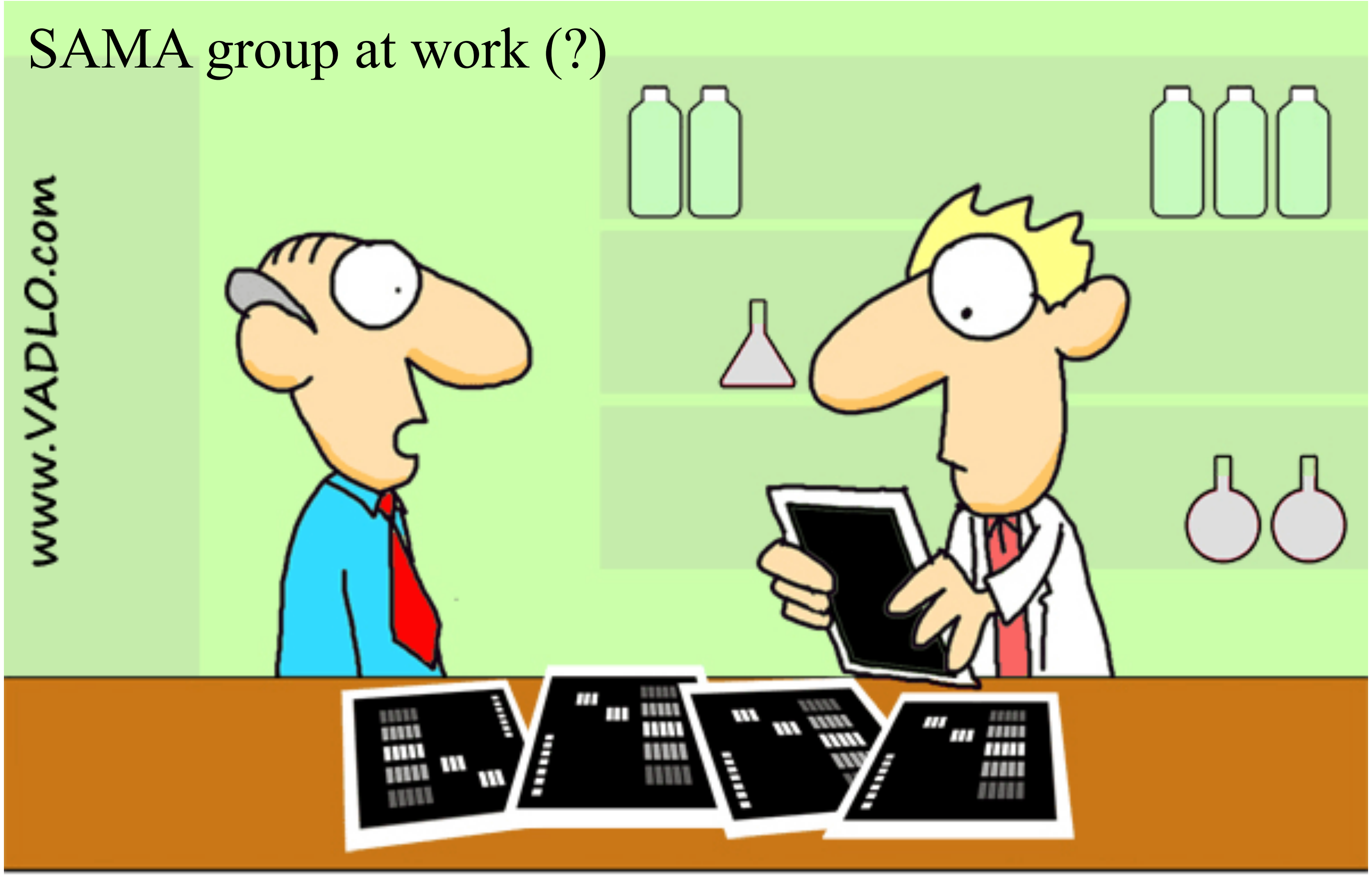
- **Spatial/multi-sites** dependence structure, e.g.:
  - Spatial VGLM (Chandler and Wheater, 2002)
    - ✗ Stationary spatial dependences / Limited number of locations
  - "*Hybrid Spatial Downscaling*" (HSD) for <u>extreme fields</u> (Bechler et al., 2015)
    - ✓ Combines RCM with geostatistical cond'l simulations / DS anywhere in the region
  - "*Rank Resampling for Distributions and Dependences*" (R2D2, Vrac, 2018)
    - ✓

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

- **Mult**
  - C
    - ✓ 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.

Thank you…



"Data don't make any sense,
we will have to resort to statistics."