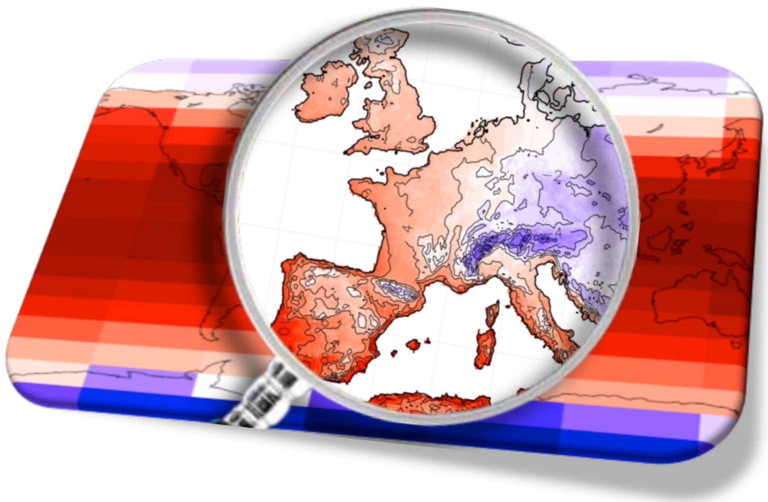
LSCE
LABORATOIRE DES SCIENCES DU CLIMAT & DE L'ENVIRONNEMENT

Conditional Stochastic Weather Generators for precipitation downscaling



Mathieu Vrac

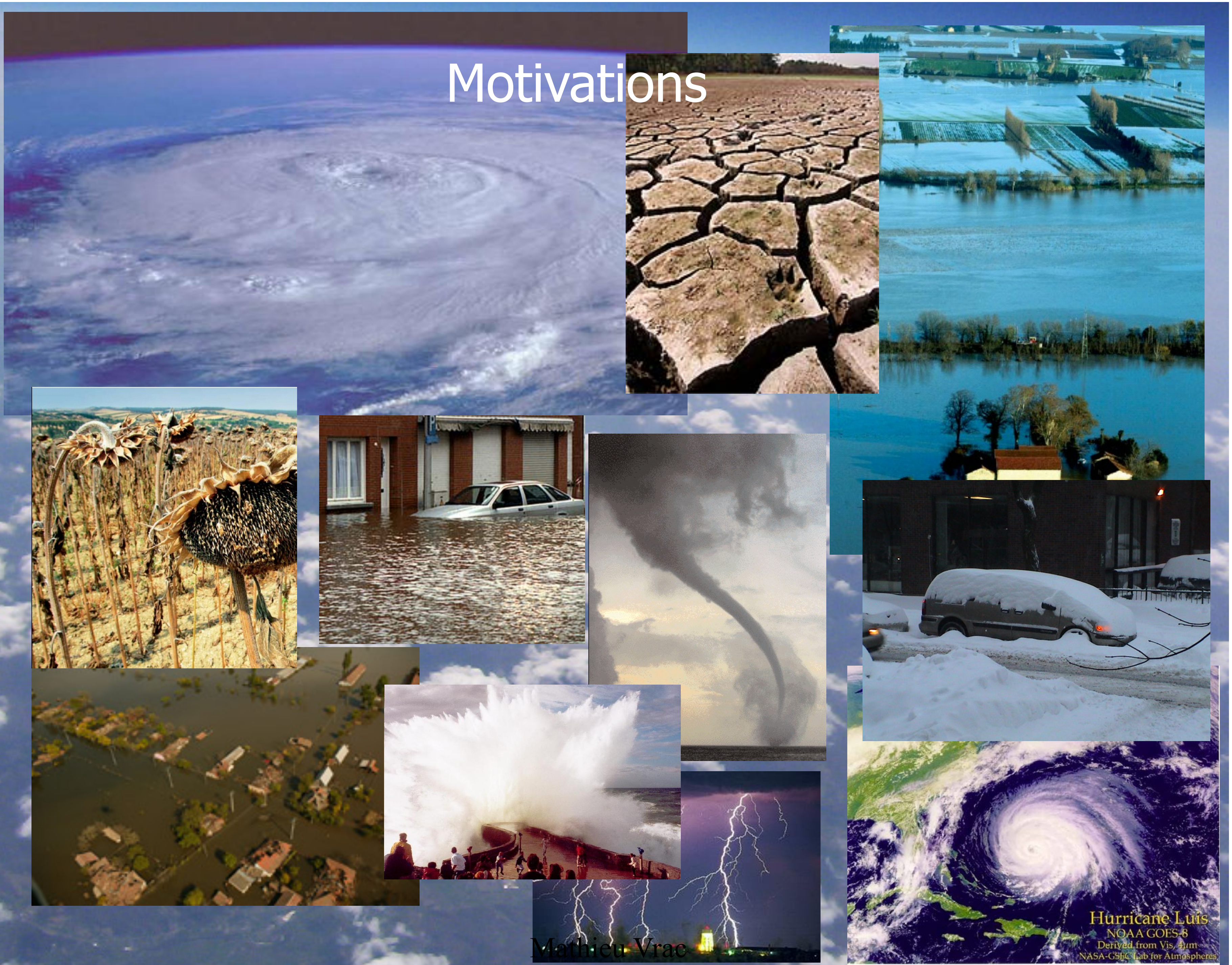
Joint works with: Many people…
(J. Carreau, J. Eden, D. Maraun, M. Widmann, G. Wong, etc.)

ANR StaRMIP

SWGEN2014 Workshop
Avignon, September, 16-19, 2014

PLEIADES

Motivations

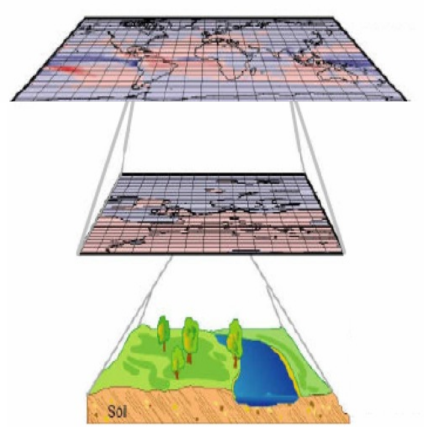


Motivations

- 30% of the world economic activities are affected by meteorological conditions (source: IPCC)

- **IPCC** scenarios of climate change have a **coarse spatial resolution !!** Not adapted to ecological, social, economic scales of impact studies
  - Social, environmental and economic impacts: water resources, hydrology, agriculture, air pollution, human health, etc.
  - How will climate change interact with existing environmental features 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)



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

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 approaches

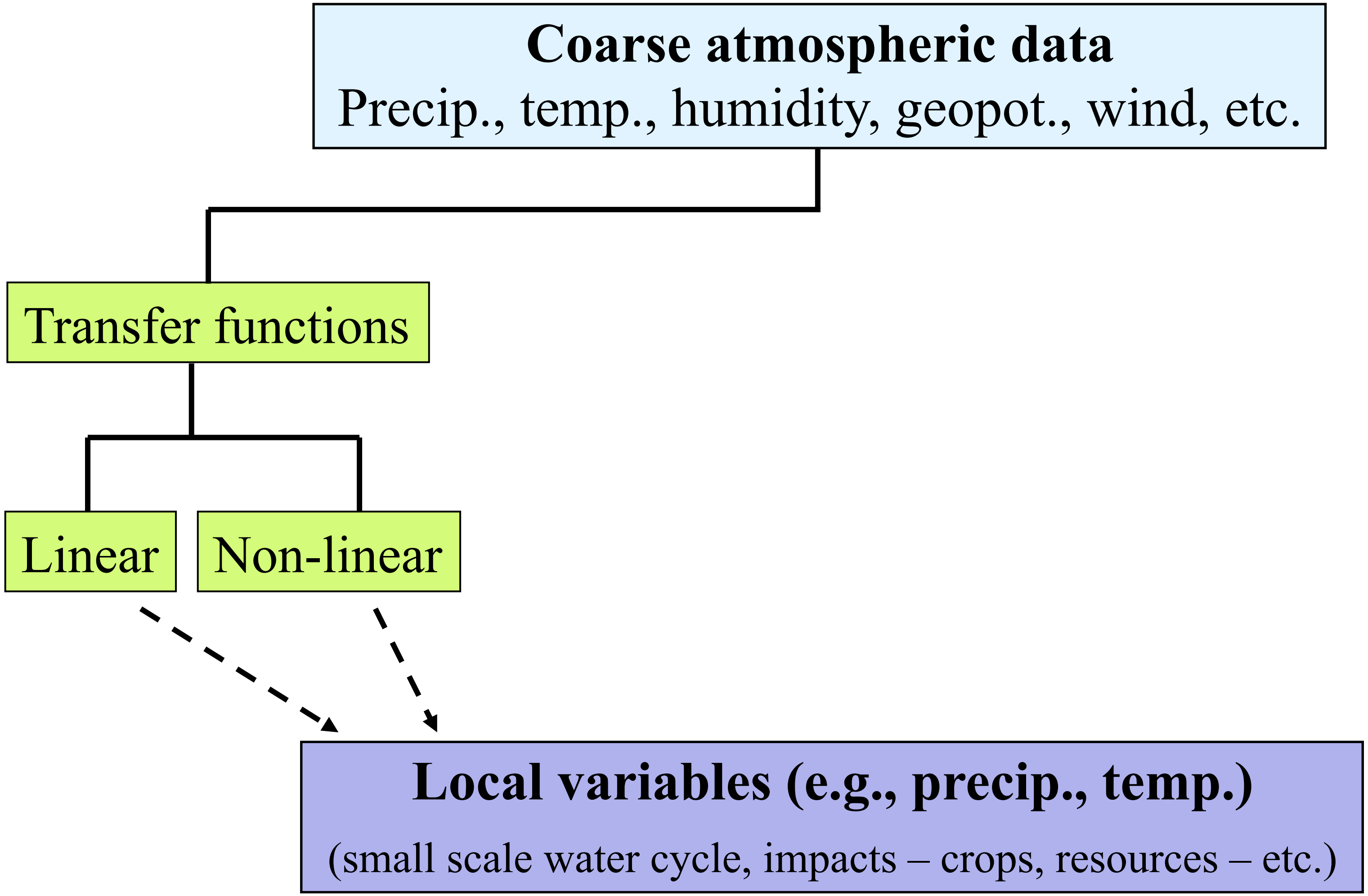
Coarse atmospheric data

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

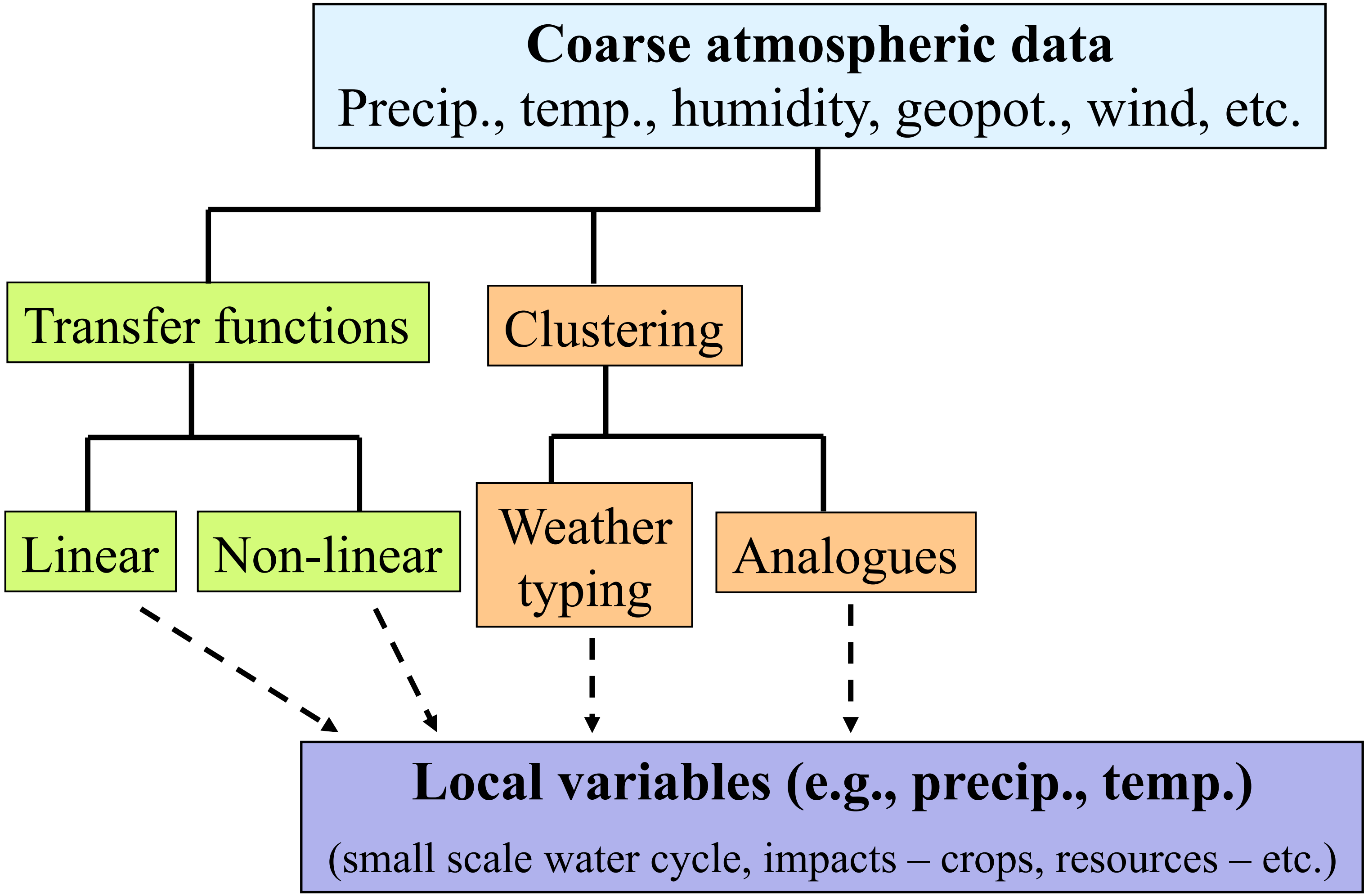
Local variables (e.g., precip., temp.)

(small scale water cycle, impacts – crops, resources – etc.)

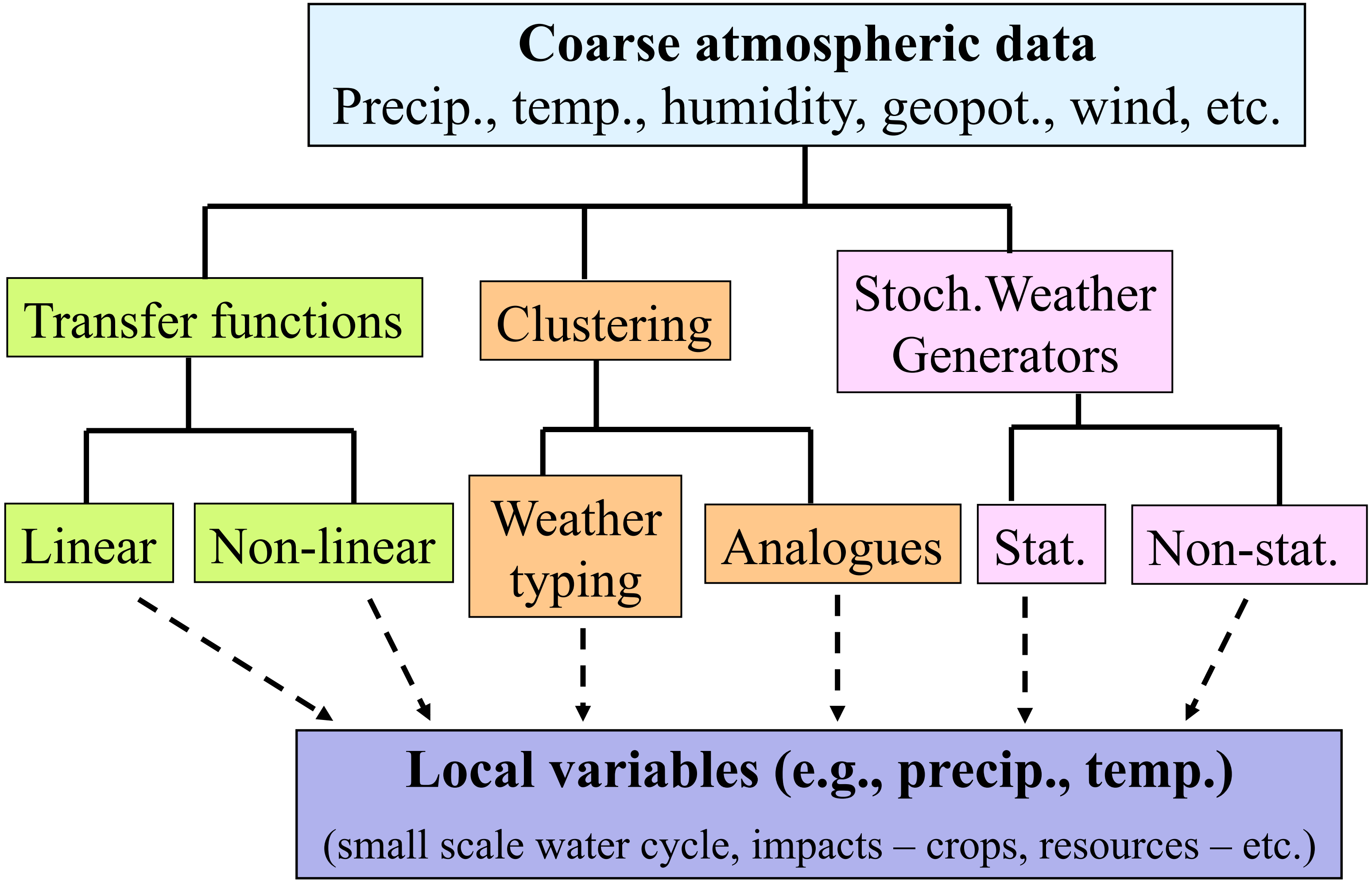
Main statistical approaches



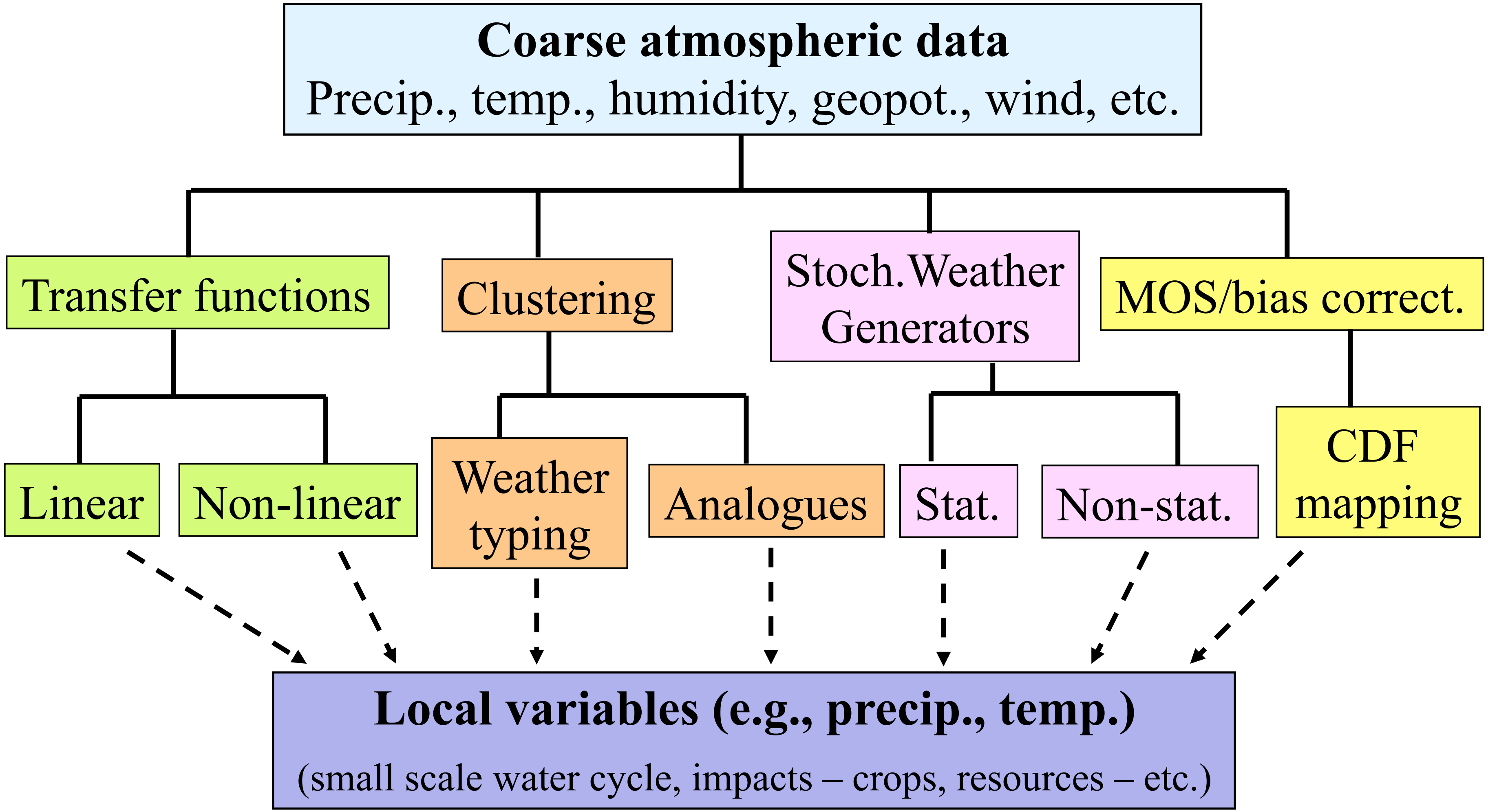
Main statistical approaches



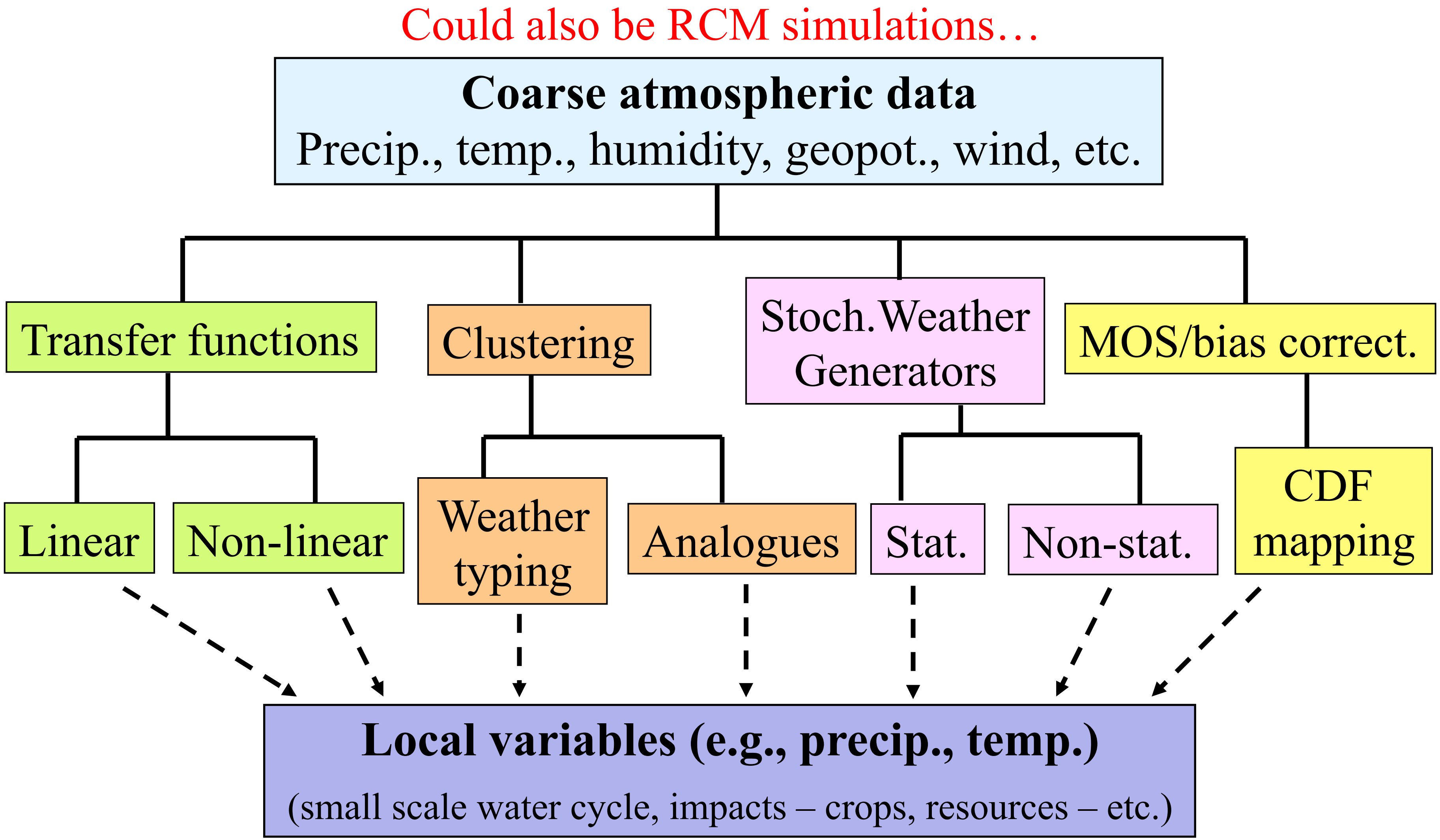
Main statistical approaches



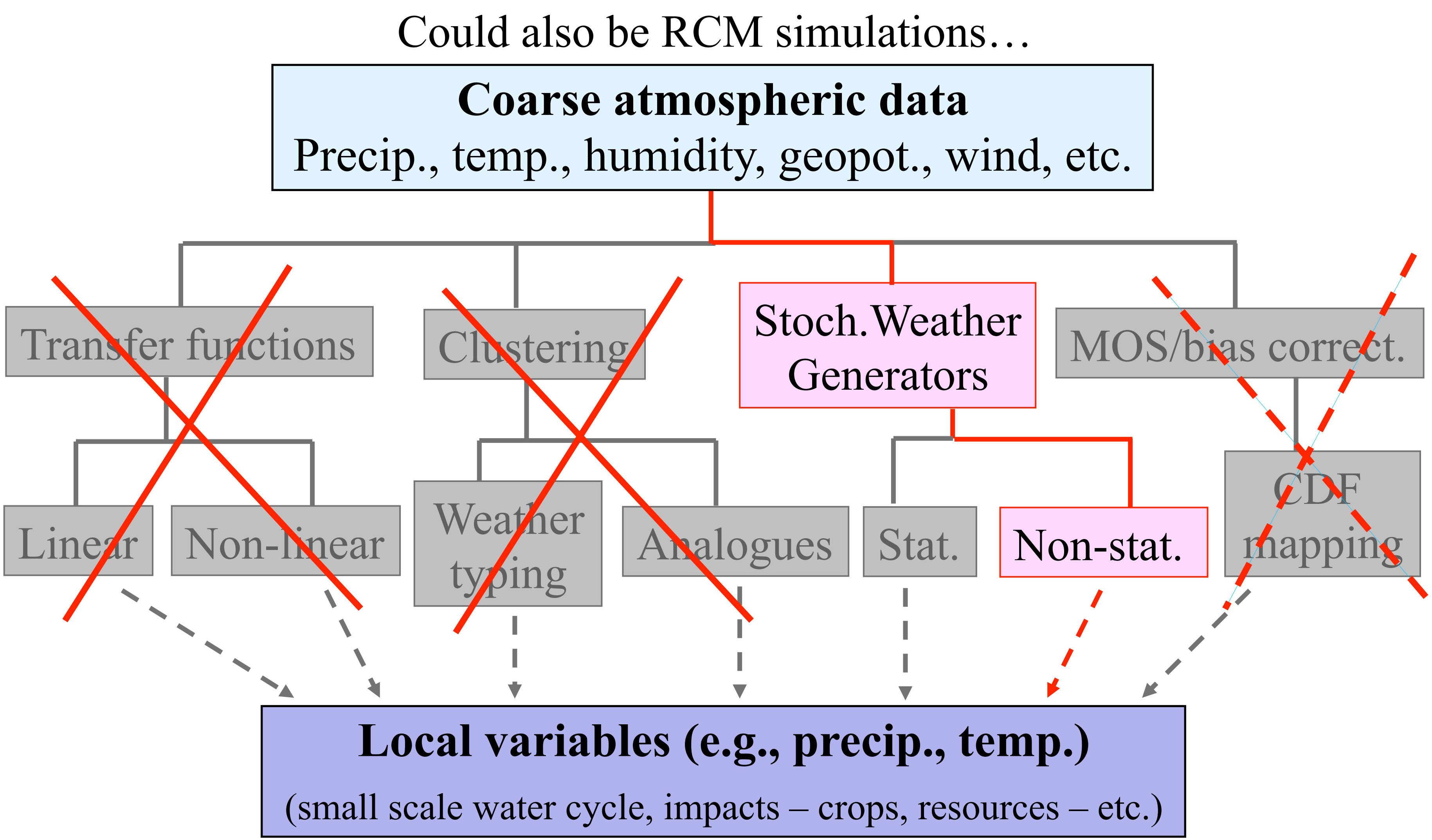
Main statistical approaches



Main statistical approaches



Main statistical approaches



Outline of the talk

- Two conditional SWGs for precipitation downscaling
- Some illustrations
- Inclusion of /Extension to extreme values distributions
- Conclusions & Perspectives

Outline of the talk

- Two conditional SWGs for precipitation downscaling
- Some illustrations
- ~~Inclusion of /Extension to extreme values distributions~~
- Conclusions & Perspectives

Stochastic Weather Generators (WGs)

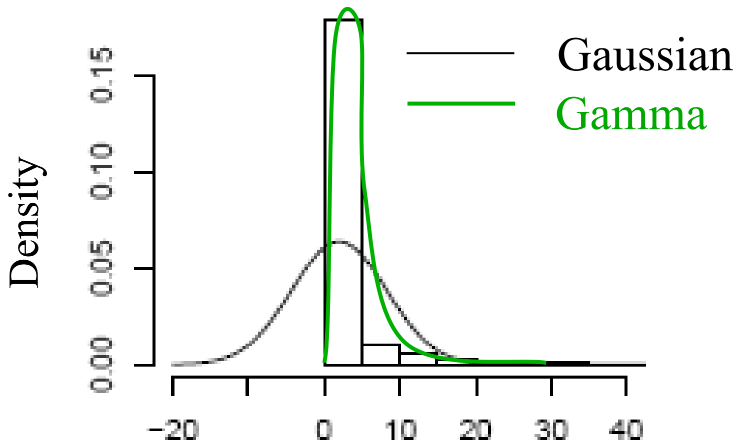
Principle: *A WG is a stochastic model simulating daily weather statistically similar to observations, based on parameters determined by historical records (Wilks and Wilby, 1999).*

Stochastic Weather Generators (WGs)

<u>Principle</u>: *A WG is a stochastic model simulating daily weather statistically similar to observations, based on parameters determined by historical records (Wilks and Wilby, 1999).*

Stochastic:

- The rainfall occurrence of *today* is conditional on the one of *yesterday* => Historical key-tool = **Markov Chains**
- **Simulations** are performed according to **pdfs**



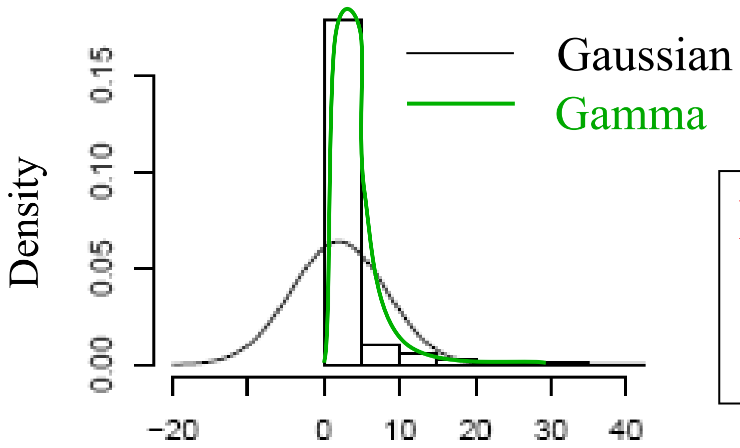
Histo of observed precip. (mm/day)

Stochastic Weather Generators (WGs)

Principle: *A WG is a stochastic model simulating daily weather statistically similar to observations, based on parameters determined by historical records (Wilks and Wilby, 1999).*

Stochastic:

- The rainfall occurrence of *today* is conditional on the one of *yesterday* => Historical key-tool = **Markov Chains**
- **Simulations** are performed according to **pdfs**



Histo of observed precip. (mm/day)

For rain intensity, most of the WGs **simulate** values in $(0,+\infty)$ according to a **Gamma distribution** (here in green)

"Non-homogeneous" WGs

(i.e., SWGs for downscaling)

- Recently: WGs for downscaling: **large-scale info is included**
  - Pryor et al. (2006) for wind: Weibull param. = GLM(GCM features)
  - Furrer & Katz (2007) for prec: Gamma param. = GLM(GCM data)

"Non-homogeneous" WGs
(i.e., SWGs for downscaling)

- Recently: WGs for downscaling: **large-scale info is included**
  - Pryor et al. (2006) for wind: Weibull param. = GLM(GCM features)
  - Furrer & Katz (2007) for prec: Gamma param. = GLM(GCM data)

⇒ Non-stationary model (Vrac et al., 2007; Carreau & Vrac, 2011, Wong et al., 2014):

$$Po_t = P(O_t | O_{t-1}, X_t)$$

Large-scale info
(e.g., simulations, WR, statistics)

"Non-homogeneous" WGs

(i.e., SWGs for downscaling)

- Recently: WGs for downscaling: **large-scale info is included**
  - Pryor et al. (2006) for wind: Weibull param. = GLM(GCM features)
  - Furrer & Katz (2007) for prec: Gamma param. = GLM(GCM data)

⇒ Non-stationary model (Vrac et al., 2007; Carreau & Vrac, 2011, Wong et al., 2014):

$$Po_t = P(O_t | O_{t-1}, X_t)$$

Large-scale info (e.g., simulations, WR, statistics)

& $\underbrace{\text{PDF of intensity with parameters cond'l on (=function of) large-scale data } X_t}_{f\left(\cdot \middle| \alpha\left(X_t\right)\right)}$

"Non-homogeneous" WGs
(i.e., SWGs for downscaling)

- Recently: WGs for downscaling: **large-scale info is included**
  - Pryor et al. (2006) for wind: Weibull param. = GLM(GCM features)
  - Furrer & Katz (2007) for prec: Gamma param. = GLM(GCM data)

⇒ Non-stationary model (Vrac et al., 2007; Carreau & Vrac, 2011, Wong et al., 2014):

$$Po_t = P(O_t | O_{t-1}, X_t)$$

Large-scale info
(e.g., simulations, WR, statistics)

& PDF of intensity with parameters cond'l on (=function of) large-scale data X_t

$$f\left(. \middle| \alpha\left(X_t\right)\right)$$

Take-home story about Stochastic WGs:
Local-scale data are **simulated from conditional pdf**

⇒ If X evolves with time => $f(.|\alpha(X))$ evolves too
⇒ Uncertainty assessment (e.g., Semenov, 2007)

Introduction | 2 Stoch. WGs | Illustrations | Conclusions & Persp.

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

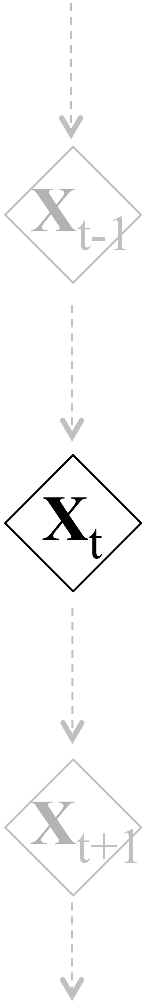
Wong et al. (2014, J. of Climate)
Eden et al. (2014, JGR, in press)

Carreau & Vrac (2011, WRR)

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

Wong et al. (2014, J. of Climate)
Eden et al. (2014, JGR, in press)

Carreau & Vrac (2011, WRR)

VGLM & NN-CMM

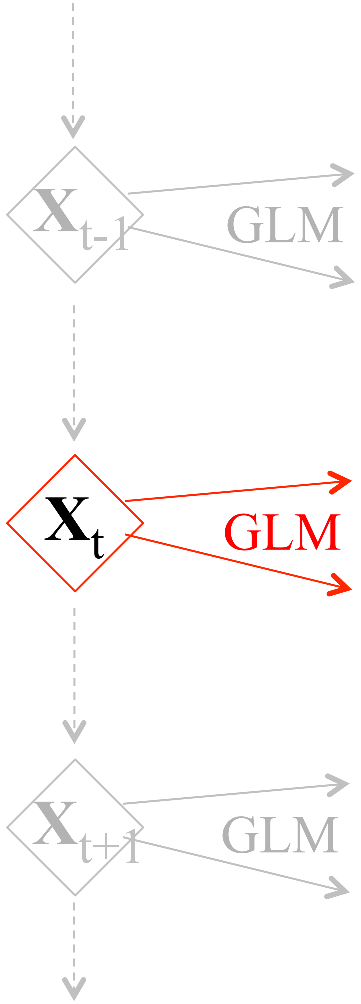
Vector Generalized Linear Model

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



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

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

Wong et al. (2014, J. of Climate)
Eden et al. (2014, JGR, in press)

Carreau & Vrac (2011, WRR)

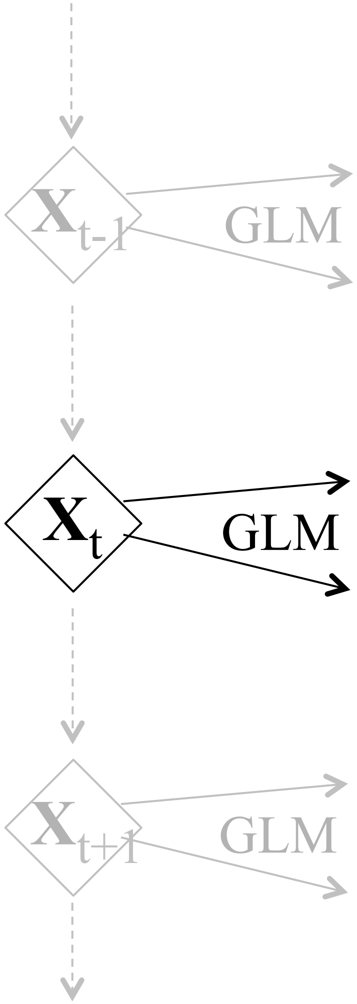
VGLM & NN-CMM

Vector Generalized Linear Model — Neural Network – Conditional Mixture Model

Precipitation pdf (at one station)

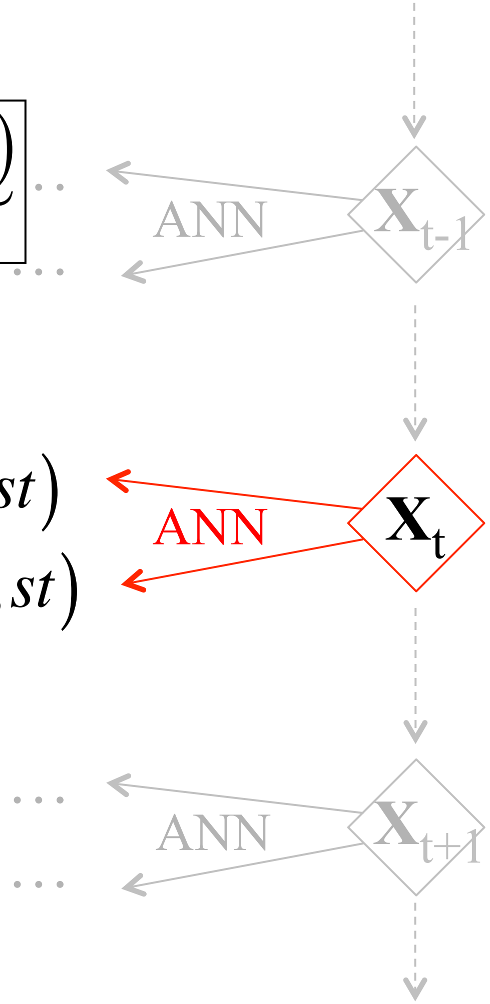
$$\boxed{\phi(y;\psi) = \underbrace{(1-\alpha)\,\delta_0(y)}_{\text{no rain}} + \underbrace{\alpha\phi_0(y;\psi_0)}_{\text{rain>0}}}$$

Parameters are fonctions
of (atmospheric, etc.) predictors



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

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



Wong et al. (2014, J. of Climate)
Eden et al. (2014, JGR, in press)

Carreau & Vrac (2011, WRR)

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

$\mathbf{X}_{t-1}$ GLM … ANN $\mathbf{X}_{t-1}$

$$\mathbf{X}_t \xrightarrow{\text{GLM}} \alpha(\mathbf{X}_t, st),\ \psi_0(\mathbf{X}_t, st) \qquad \alpha(\mathbf{X}_t, st),\ \psi_0(\mathbf{X}_t, st) \xleftarrow{\text{ANN}} \mathbf{X}_t$$

$\mathbf{X}_{t+1}$ GLM … ANN $\mathbf{X}_{t+1}$

Same philosophy, different implementations

Wong et al. (2014, J. of Climate)
Eden et al. (2014, JGR, in press)

Carreau & Vrac (2011, WRR)

The modelling part of **NN-CMM**

- Precipitation probability density function (One station):

$$\phi\left(y;\psi_i\right)=\left(1-\alpha_i\right)\delta_0(y)+\alpha_i\phi_0\left(y;\psi_{0,i}\right)$$

For one station i

The modelling part of **NN-CMM**

- Precipitation probability density function (N Stations):

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

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

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

The modelling part of **NN-CMM**

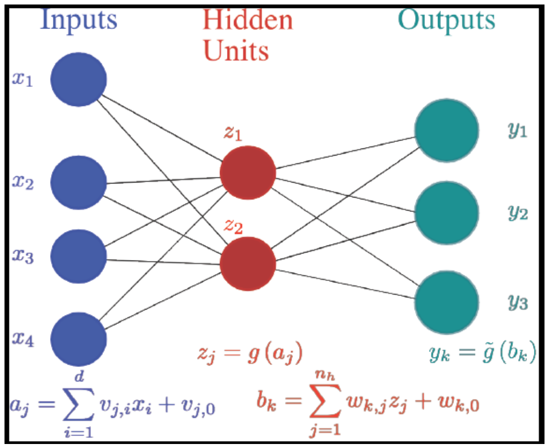
- Precipitation probability density function (N Stations):

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

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

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

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



The modelling part of **NN-CMM**

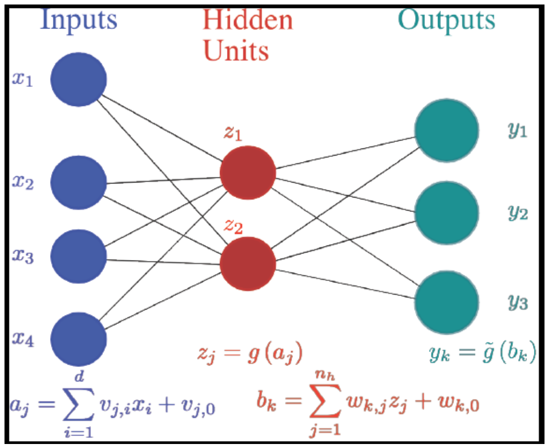
- Precipitation probability density function (N Stations):

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

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

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

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



$$f = \begin{cases} \text{Gaussian} \\ \quad\text{or} \\ \text{Log-Normal} \\ \quad\text{or} \\ \text{Hybrid Pareto} \end{cases}$$

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

The modelling part of **NN-CMM**

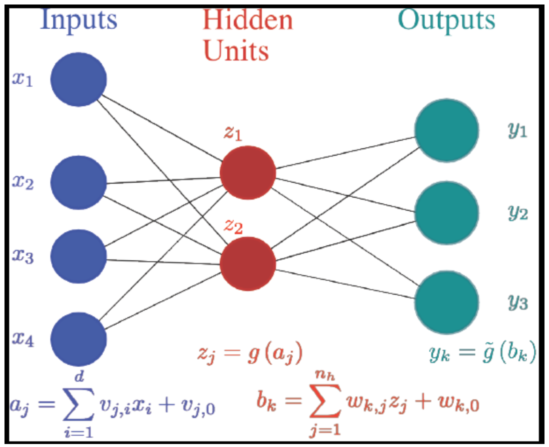
- Precipitation probability density function (N Stations):

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

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

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

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



$f = \Bigg\{$
- ➢ Gaussian
- or
- ➢ **Log-Normal**
- or
- ➢ Hybrid Pareto

✓ Carreau & Vrac (2011)

✓ Carreau & Bengio (2009a,b)

The modelling part of **VGLM**

- Precipitation probability density function (One station):

$$\phi\left(y;\psi_i\right)=\left(1-\alpha_i\right)\delta_0(y)+\alpha_i\phi_0\left(y;\psi_{0,i}\right)$$

For one station *i*

The modelling part of **VGLM**

- Precipitation probability density function (N stations):

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

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

The modelling part of **VGLM**

- Precipitation probability density function (N stations):

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

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

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

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

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 $\phi_0 =$ Gamma pdf with parameters

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

Illustration 1: Daily pdfs with **NN-CMM**-2L

from Carreau and Vrac (2011)

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

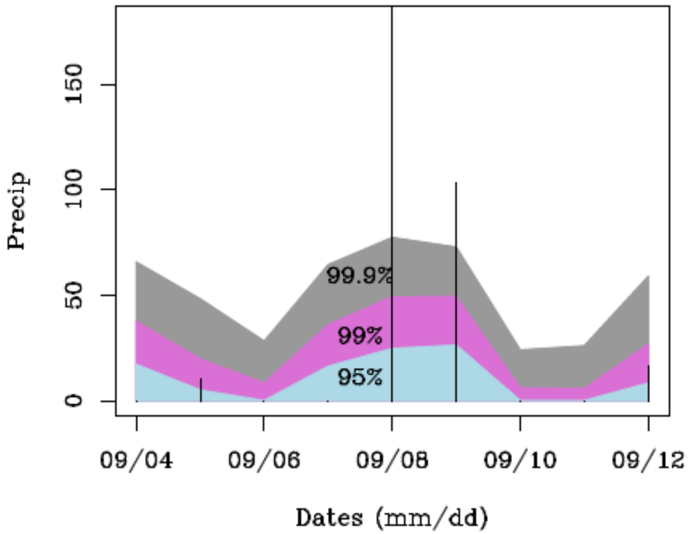
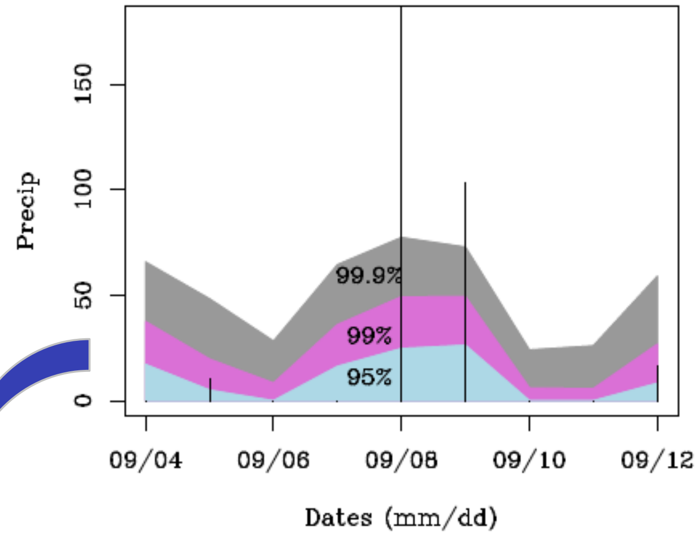


Illustration 1: Daily pdfs with **NN-CMM**-2L

from Carreau and Vrac (2011)

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



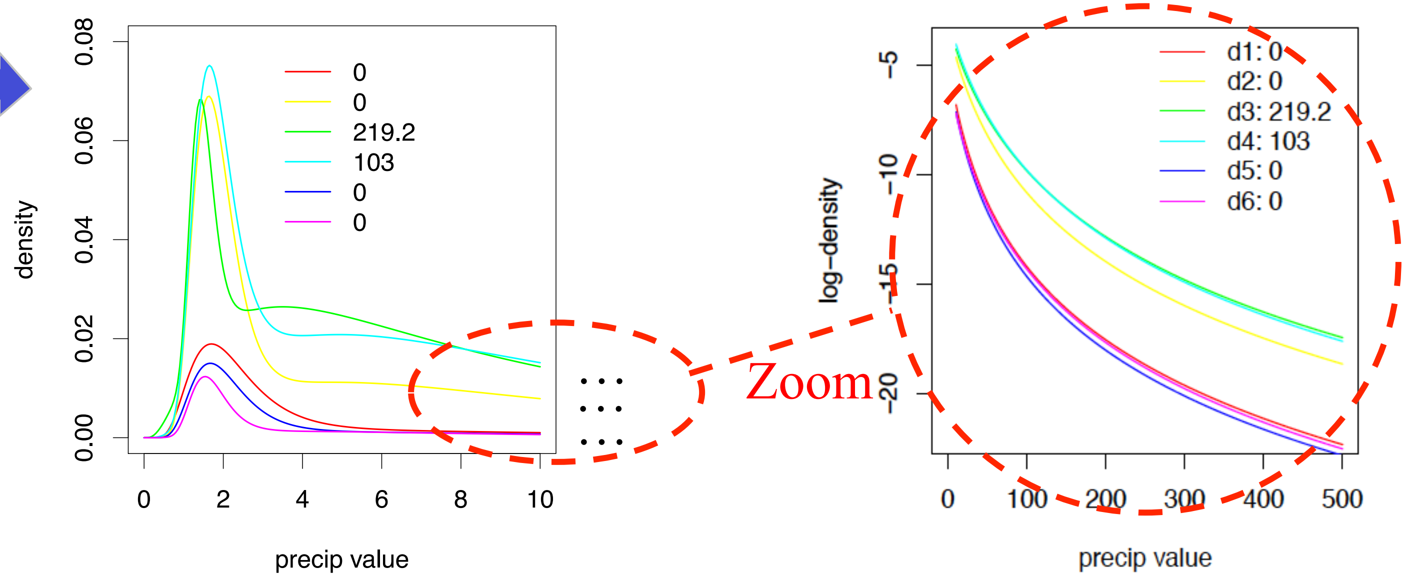


Illustration 1: Daily pdfs with **NN-CMM**-2L

from Carreau and Vrac (2011)

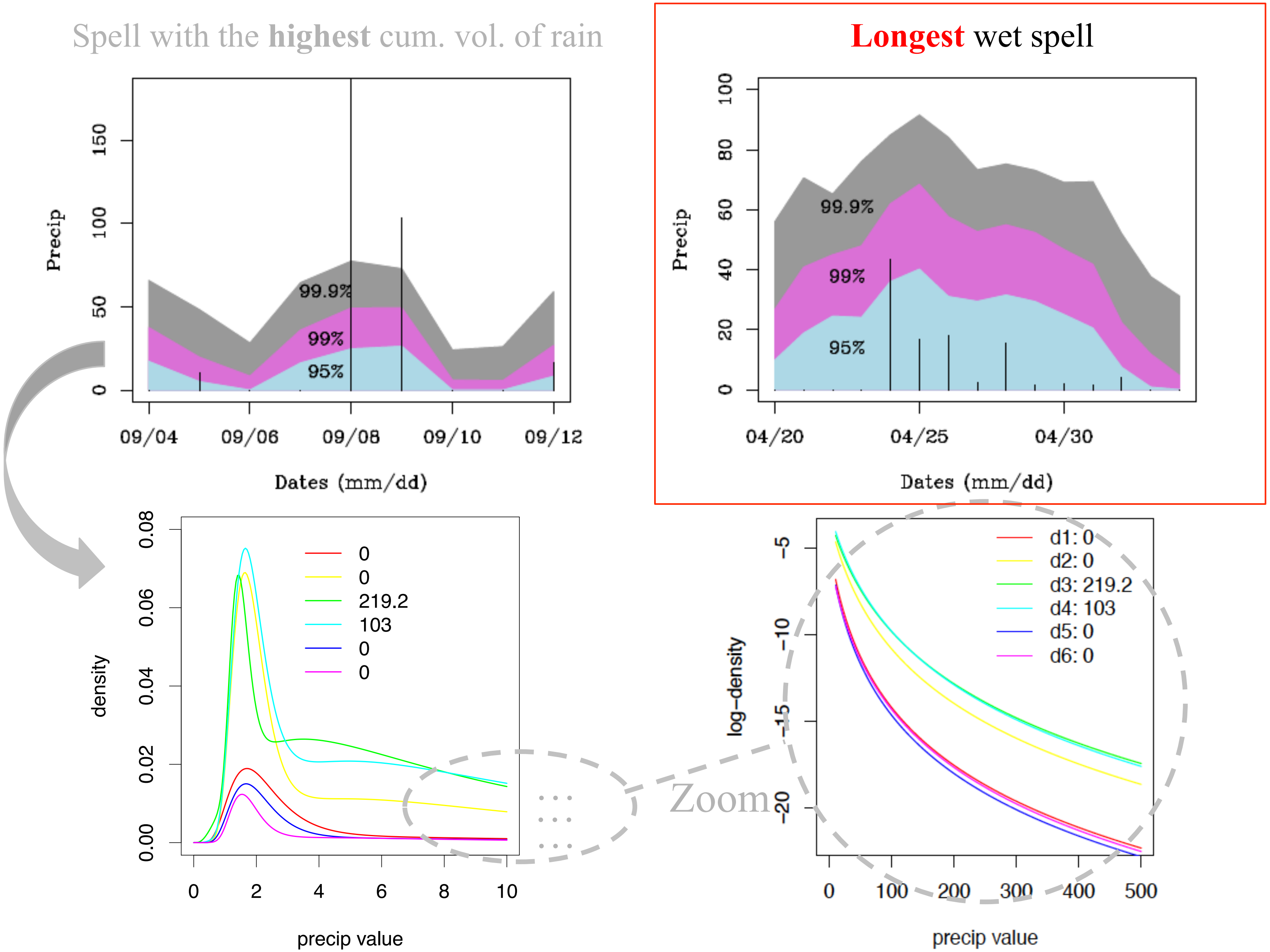


Illustration 2: **VGLM-Gamma**

from Eden et al. (2014)

Observations: 465 UK stations with daily PR in 1961-2000 from the Meteorological Office Integrated Data Archive System (MIDAS)

Illustration 2: **VGLM-Gamma**

from Eden et al. (2014)

Observations: 465 UK stations with daily PR in 1961-2000 from the Meteorological Office Integrated Data Archive System (MIDAS)

Predictors = 3x3 grid-cells average precipitation from 2 RCMs (driven by ERA-40):

- COSMO-CLM: **spectrally-nudged** simulations, 18x18 km (Geyer and Rockel, 2013; Geyer, 2014)
- KNMI-RACMO2: **no-nudging**, 25x25 km (van Meijgaard et al., 2008)

Illustration 2: **VGLM-Gamma**

from Eden et al. (2014)

Observations: 465 UK stations with daily PR in 1961-2000 from the Meteorological Office Integrated Data Archive System (MIDAS)

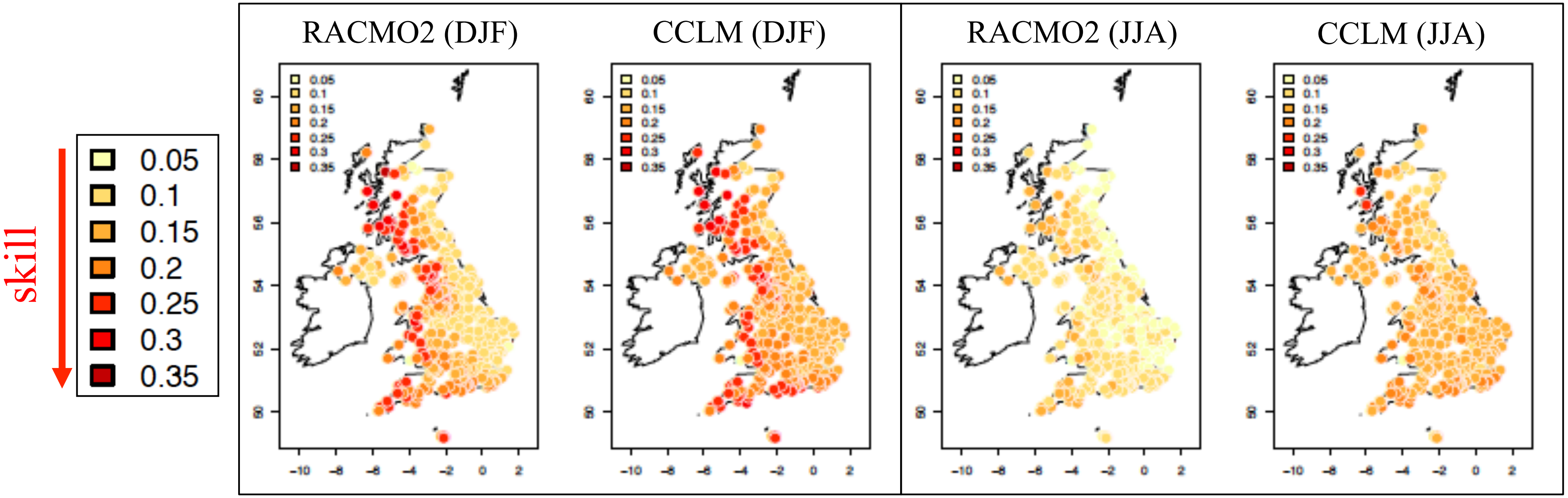
Predictors = 3x3 grid-cells average precipitation from 2 RCMs (driven by ERA-40):

- COSMO-CLM: **spectrally-nudged** simulations, 18x18 km (Geyer and Rockel, 2013; Geyer, 2014)
- KNMI-RACMO2: **no-nudging**, 25x25 km (van Meijgaard et al., 2008)

4-fold cross-validation: Calibration=30 yrs ; Projection=10yrs

Illustration 2a: **Brier skill scores** for VGLM-Γ

from Eden et al. (2014)



Brier skill scores (1961-2000) for VGLM-G fitted on PR from RACMO2 and CCLM for winter (DJF) and summer (JJA)

$$BS = \frac{1}{T}\sum_{t=1}^{T}(\alpha_t - o_t)^2 \qquad BSS = 1 - \frac{BS}{BS_{ref}}$$

Proba(rain) 0 (dry) or 1 (wet) Stationary climatology

$BSS =$ Improvement of the model to make accurate probabilistic predictions (here, rain occurrence), with respect to a reference model

Illustration 2b: **Quantile skill scores** for VGLM-Γ

from Eden et al. (2014)

$$QSS_p = 1 - \frac{QS_p}{QS_{p,ref}}$$

where $$QS_p = \sum_{t=1}^{T} \rho_p(o_t - q_p(X_t))$$

with $$\rho_p(u) = \begin{cases} pu & for\ u \geq 0 \\ (p-1)u & for\ u < 0 \end{cases}$$

Illustration 2b: **Quantile skill scores** for VGLM-Γ

from Eden et al. (2014)

$$QSS_p = 1 - \frac{QS_p}{QS_{p,ref}}$$

where $$QS_p = \sum_{t=1}^{T} \rho_p(o_t - q_p(X_t))$$

with $$\rho_p(u) = \begin{cases} pu & for\ u \geq 0 \\ (p-1)u & for\ u < 0 \end{cases}$$

QS_p quantifies the capability of the model to predict quantiles associated to a specific probability p.

Illustration 2b: **Quantile skill scores** for VGLM-Γ

from Eden et al. (2014)

$$QSS_p = 1 - \frac{QS_p}{QS_{p,ref}}$$

QSS_p = the increase (or decrease) of the capability of the model to predict quantiles associated to a specific probability p, with respect to a reference model.

where $$QS_p = \sum_{t=1}^{T} \rho_p(o_t - q_p(X_t))$$

with $$\rho_p(u) = \begin{cases} pu & \mathit{for}\ u \geq 0 \\ (p-1)u & \mathit{for}\ u < 0 \end{cases}$$

QS_p quantifies the capability of the model to predict quantiles associated to a specific probability p.

Illustration 2b: **Quantile skill scores** for VGLM-Γ

from Eden et al. (2014)

$$QSS_p = 1 - \frac{QS_p}{QS_{p,ref}}$$

Quantile skill scores for VGLM-Γ fitted on PR from **RACMO2** and **CCLM** for winter (DJF) 1961-2000.

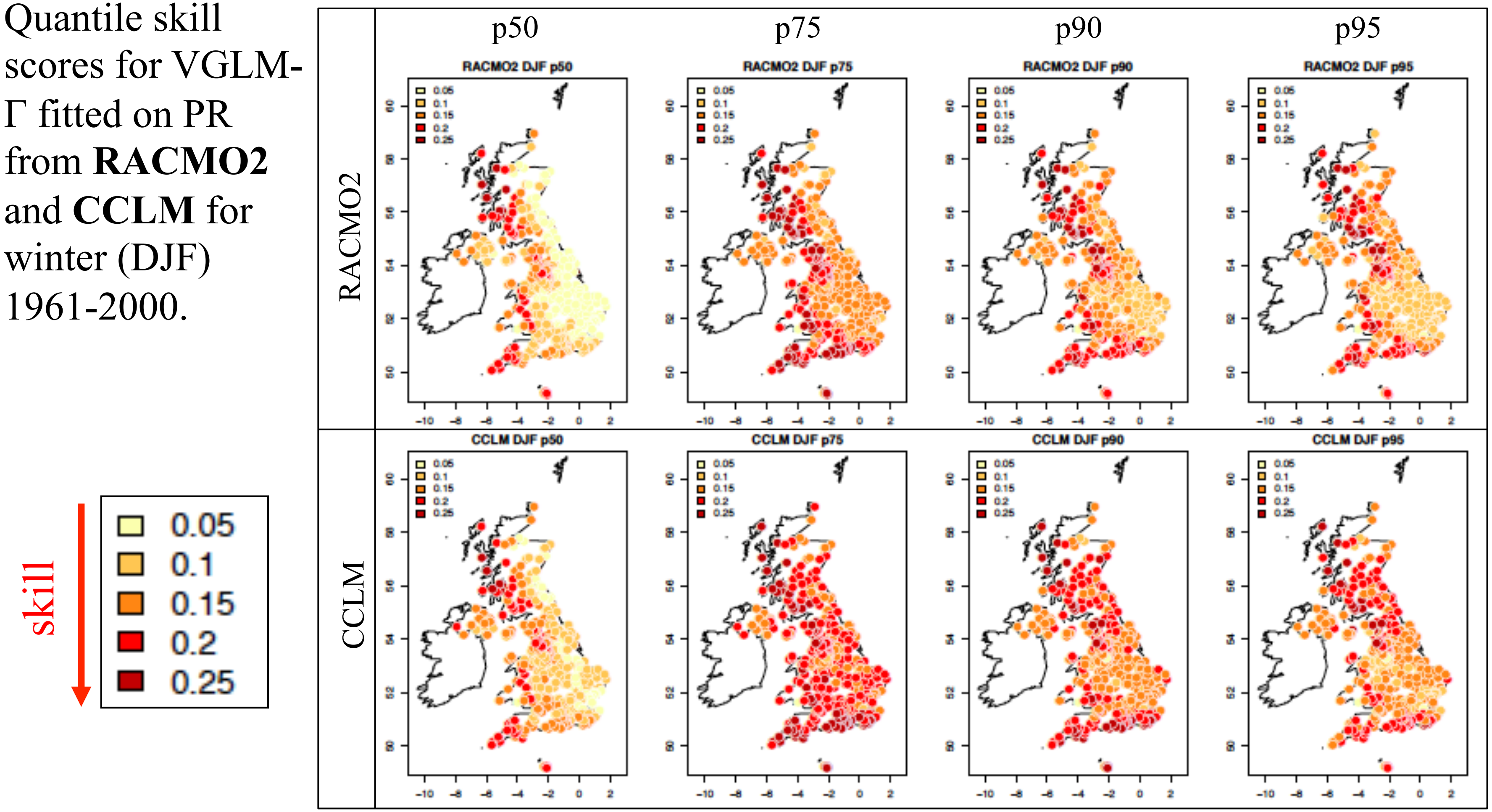


Illustration 2: VGLM-Gamma

from Eden et al. (2014)

<u>Observations</u>: 465 UK stations with daily PR in 1961-2000 from the Meteorological Office Integrated Data Archive System (MIDAS)

<u>Predictors</u> = 3x3 grid-cells average precipitation from 2 RCMs (driven by ERA-40):

- COSMO-CLM: **spectrally-nudged** simulations, 18x18 km (Geyer and Rockel, 2013; Geyer, 2014)
- KNMI-RACMO2: **no-nudging**, 25x25 km (van Meijgaard et al., 2008)

<u>4-fold cross-validation</u>: Calibration=30 yrs ; Projection=10yrs

Illustration 2: VGLM-Gamma

from Eden et al. (2014)

Observations: 465 UK stations with daily PR in 1961-2000 from the Meteorological Office Integrated Data Archive System (MIDAS)

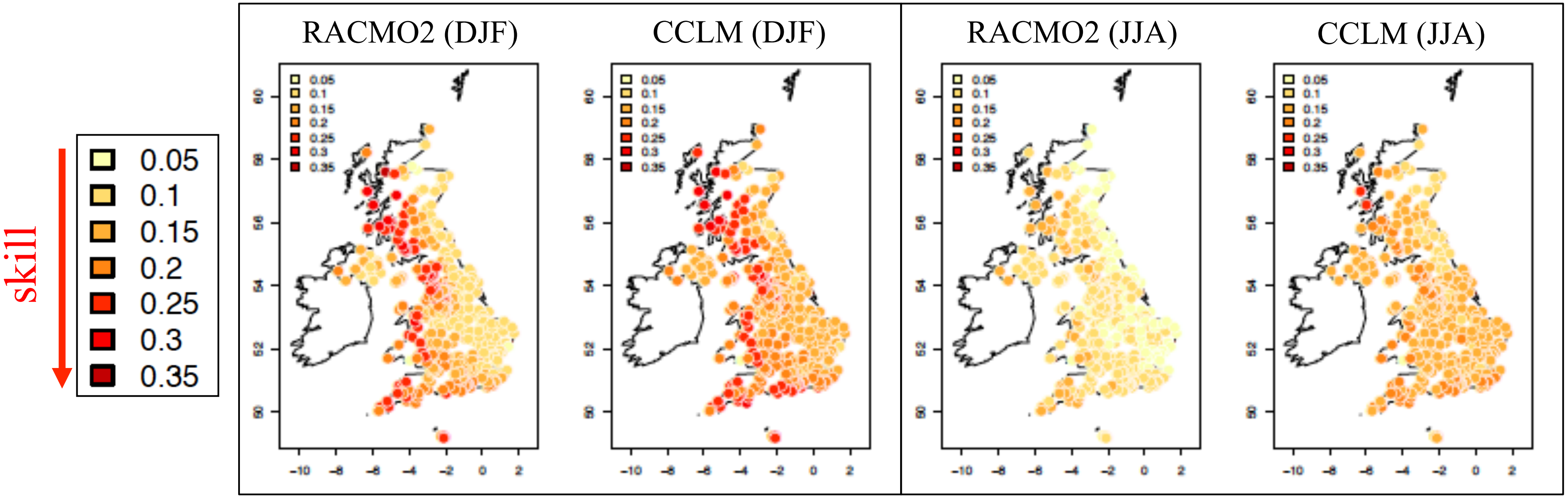
Predictors = 3x3 grid-cells average precipitation from 2 RCMs and 1 GCM (driven by ERA-40):

- COSMO-CLM: **spectrally-nudged** simulations, 18x18 km (Geyer and Rockel, 2013; Geyer, 2014)
- KNMI-RACMO2: **no-nudging**, 25x25 km (van Meijgaard et al., 2008)
- ECHAM5: **nudged** simulations, ~200x150 km (T63) (Eden et al., 2012)

4-fold cross-validation: Calibration=30 yrs ; Projection=10yrs

Illustration 2a: **Brier skill scores** for VGLM-Γ

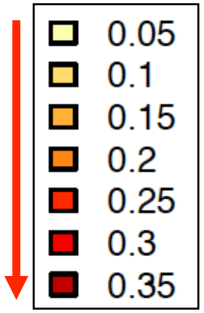
from Eden et al. (2014)

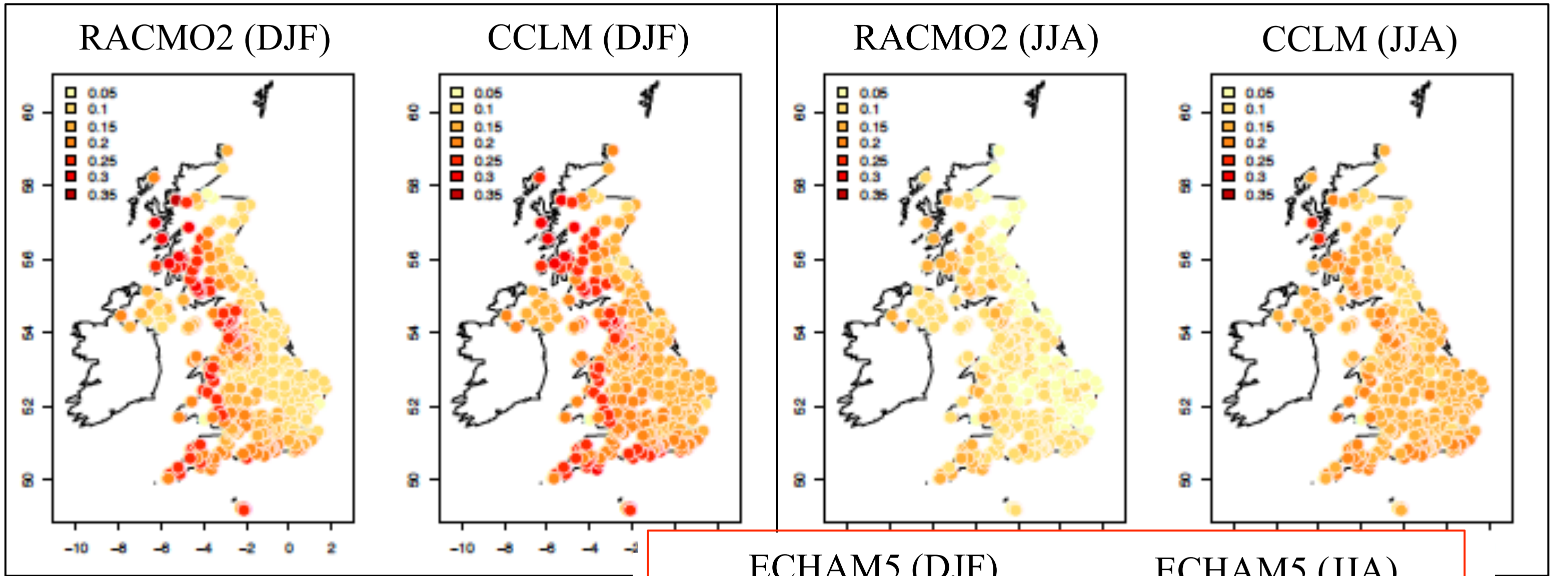


Brier skill scores (1961-2000) for VGLM-G fitted on PR from RACMO2 and CCLM for winter (DJF) and summer (JJA)

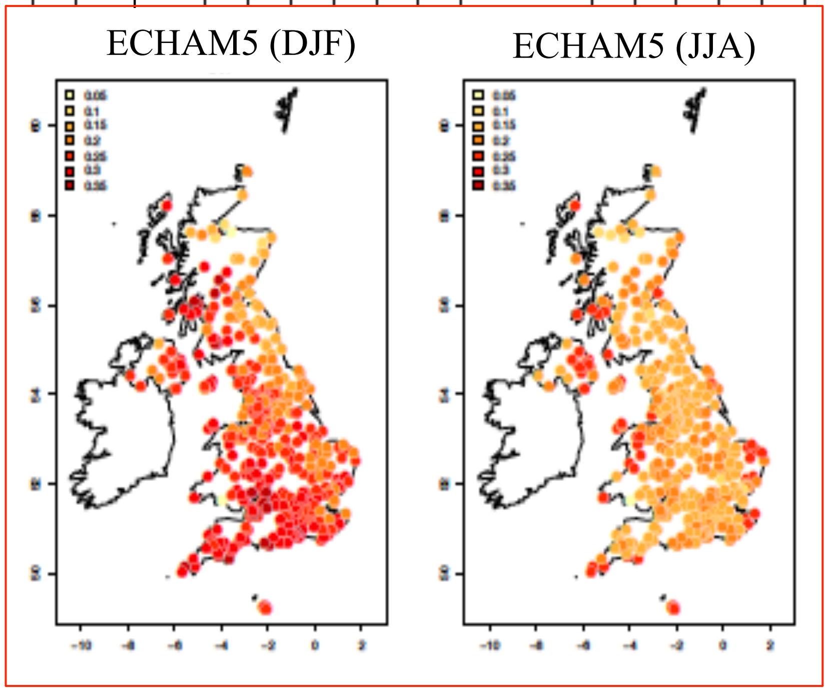
Illustration 2a: **Brier skill scores** for VGLM-Γ

from Eden et al. (2014)





Brier skill scores (1961-2000) for VGLM-
for winter (DJF) and summer (JJA)

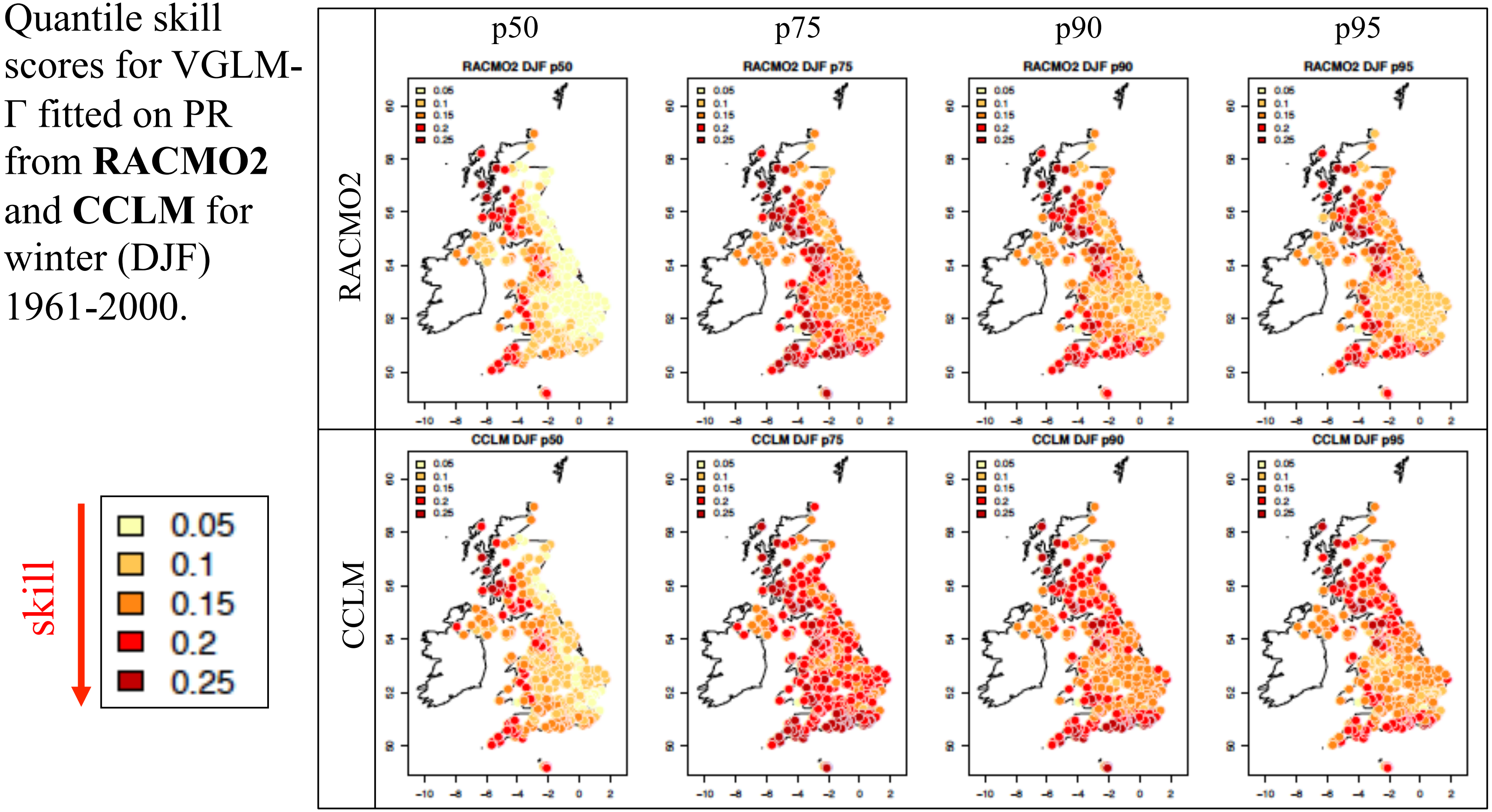


Brier skill scores (1961-2000) for VGLM-G fitted on PR **from ECHAM5** for winter (DJF) and summer (JJA)

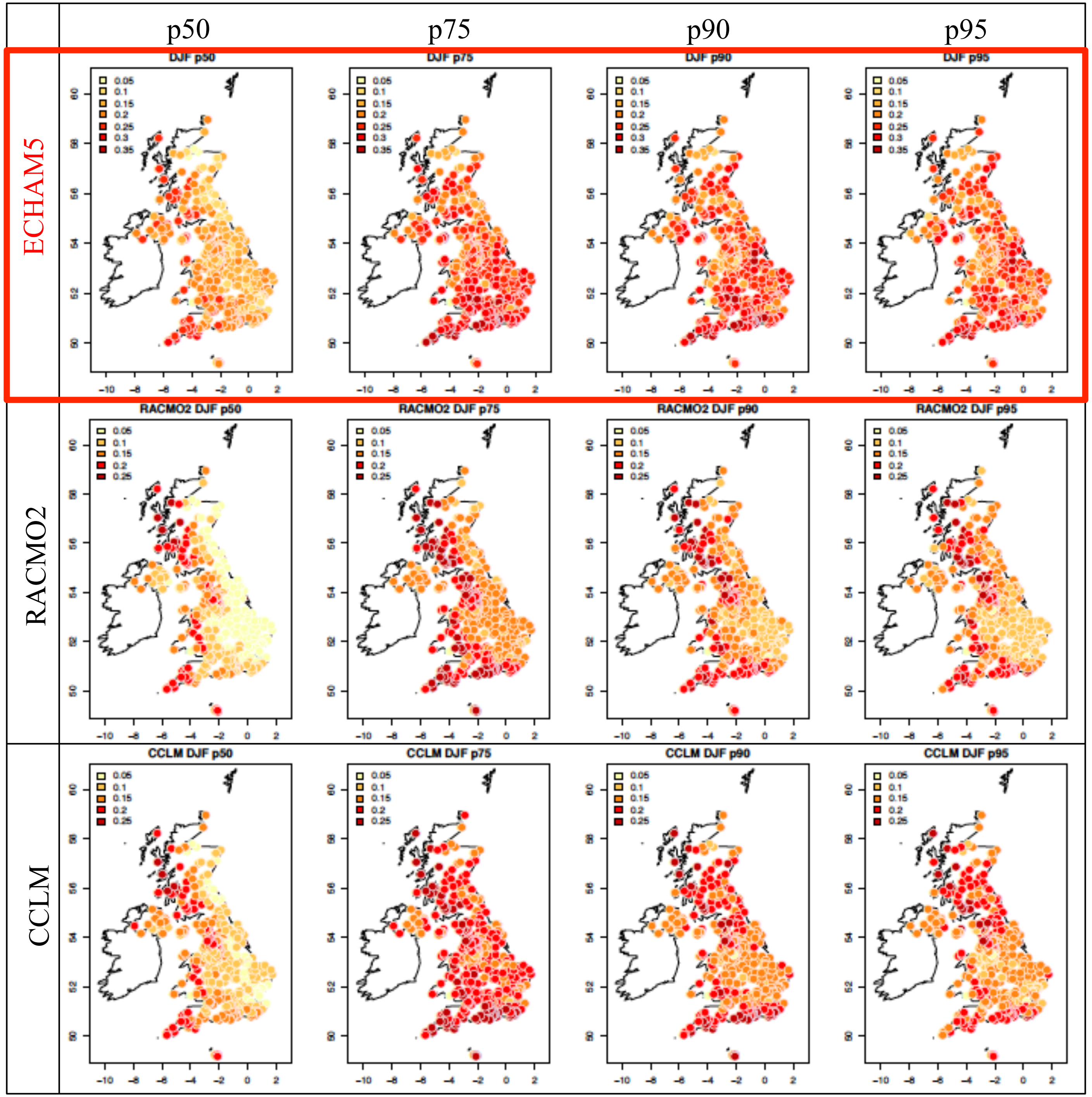
Illustration 2b: **Quantile skill scores** for VGLM-Γ

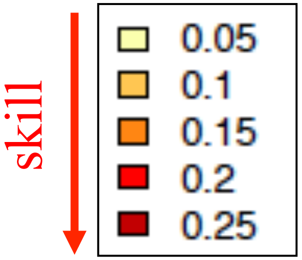
from Eden et al. (2014)

Quantile skill scores for VGLM-Γ fitted on PR from **RACMO2** and **CCLM** for winter (DJF) 1961-2000.



Quantile skill scores for VGLM-Γ fitted on PR from **RACMO2** and **CCLM** and **ECHAM5** for winter (DJF) 1961-2000.





Conclusions (some) on statistical downscaling

Conclusions (some) on statistical downscaling

- **Many** (and many) **models and applications** of downscaling
  - 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

Conclusions (some) on statistical downscaling

- **Many** (and many) **models and applications** of downscaling
  - 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
  - My favorite ones:
    - ✓ *Stochastic WGs*: cond'l event-wise variability/uncertainty
    - ✓ *MOS / Bias correction*: DS of CDFs from CDFs

Conclusions (some) on statistical downscaling

- **Many** (and many) **models and applications** of downscaling
  - 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
  - My favorite ones:
    - ✓ *Stochastic WGs*: cond'l event-wise variability/uncertainty
    - ✓ *MOS / Bias correction*: DS of CDFs from CDFs
- **RCMs vs. SDMs**: Not a conflict => complementary approaches

  ⇒ Very good illustration of that in Aurélien's talk (next talk)

Conclusions (some) on statistical downscaling

- **Many** (and many) **models and applications** of downscaling
  - 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
  - <u>My favorite ones</u>:
    - ✓ *Stochastic WGs*: cond'l event-wise variability/uncertainty
    - ✓ *MOS / Bias correction*: DS of CDFs from CDFs
- **RCMs vs. SDMs**: Not a conflict => complementary approaches
  - ⇒ Very good illustration of that in Aurélien's talk (next talk)
- 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)

- **R packages** developed for Stochastic downscaling & BC:
  - `NHMixt` (Vrac & Naveau, 2007, Wong et al., 2014)
    - ✓ Statistical mixture model Gamma & GPD
    - ✓ Inclusion of covariates
    - ✓ 2D-extension in progress
  - `condmixt` (Carreau & Vrac et al., 2011)
    - ✓ ANN-Conditional mixture model
    - ✓ Various distributions (Gaussian, Log-N, hybrid Pareto)
  - Other R packages available for non-WGs downscaling

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

Still work to do…

"*There is a fine line between wrong and visionary. Unfortunately, you have to be a visionary to see it.*"

Dr. Sheldon Cooper (The Big Bang Theory)

Still work to do…

"*There is a fine line between wrong and visionary. Unfortunately, you have to be a visionary to see it.*"
Dr. Sheldon Cooper (The Big Bang Theory)

- SDMs are often univariate (although covariates):

  ⇒ Needs for **inter-sites** models (stations or grid-cells - PLEIADES)

  ? *Latent* (i.e. cond'l ind.)? Or *Complete* dependence structure?

Still work to do…

"*There is a fine line between wrong and visionary. Unfortunately, you have to be a visionary to see it.*"
Dr. Sheldon Cooper (The Big Bang Theory)

- SDMs are often univariate (although covariates):

  ⇒ Needs for **inter-sites** models (stations or grid-cells - PLEIADES)

  ? *Latent* (i.e. cond'l ind.)? Or *Complete* dependence structure?

  ⇒ Needs for **inter-variables** models (b/ climate variables – CELLO)

  ? MOS ? SWGs?

Still work to do…

"*There is a fine line between wrong and visionary. Unfortunately, you have to be a visionary to see it.*"
Dr. Sheldon Cooper (The Big Bang Theory)

- SDMs are often univariate (although covariates):

⇒ Needs for **inter-sites** models (stations or grid-cells - PLEIADES)

? *Latent* (i.e. cond'l ind.)? Or *Complete* dependence structure?

⇒ Needs for **inter-variables** models (b/ climate variables – CELLO)

? MOS ? SWGs?

⇒ Needs for **spatial models**: SD even at locations where no data

? Continuous spatial processes? Inter/Extra-polation of parameters?

Still work to do…

"There is a fine line between wrong and visionary. Unfortunately, you have to be a visionary to see it."
Dr. Sheldon Cooper (The Big Bang Theory)

- SDMs are often univariate (although covariates):

⇒ Needs for **inter-sites** models (stations or grid-cells - PLEIADES)

? *Latent* (i.e. cond'l ind.)? Or *Complete* dependence structure?

⇒ Needs for **inter-variables** models (b/ climate variables – CELLO)

? MOS ? SWGs?

⇒ Needs for **spatial models**: SD even at locations where no data

? Continuous spatial processes? Inter/Extra-polation of parameters?

➢ Especially for **dependence of extremes** (ANR McSIM)

Still work to do…

"There is a fine line between wrong and visionary. Unfortunately, you have to be a visionary to see it."
Dr. Sheldon Cooper (The Big Bang Theory)

- SDMs are often univariate (although covariates):

⇒ Needs for **inter-sites** models (stations or grid-cells - PLEIADES)

? *Latent* (i.e. cond'l ind.)? Or *Complete* dependence structure?

⇒ Needs for **inter-variables** models (b/ climate variables – CELLO)

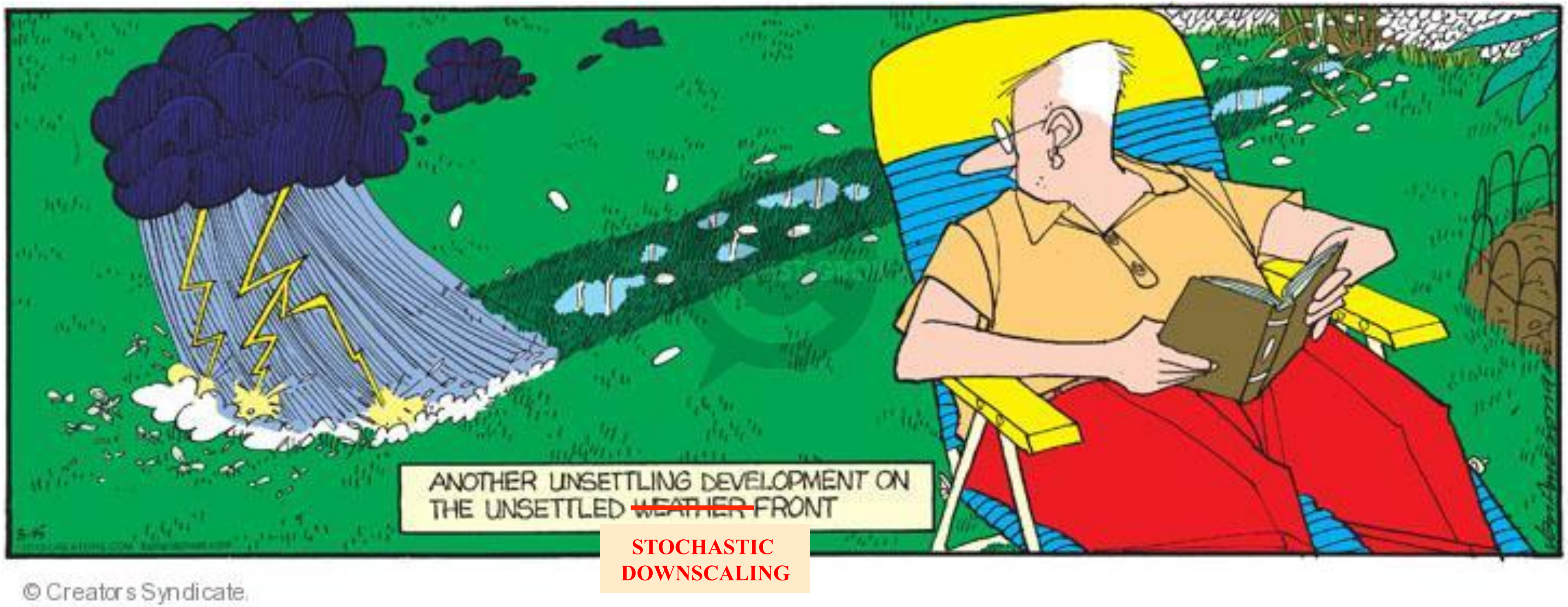
? MOS ? SWGs?

⇒ Needs for **spatial models**: SD even at locations where no data

? Continuous spatial processes? Inter/Extra-polation of parameters?

➢ Especially for **dependence of extremes** (ANR McSIM)

Goal of Aurélien's work: next talk!


ANOTHER UNSETTLING DEVELOPMENT ON THE UNSETTLED ~~WEATHER~~ FRONT
STOCHASTIC DOWNSCALING

Thank you…

Main (implicit) assumptions of SDM

For calibration under (near-) **present** climate:

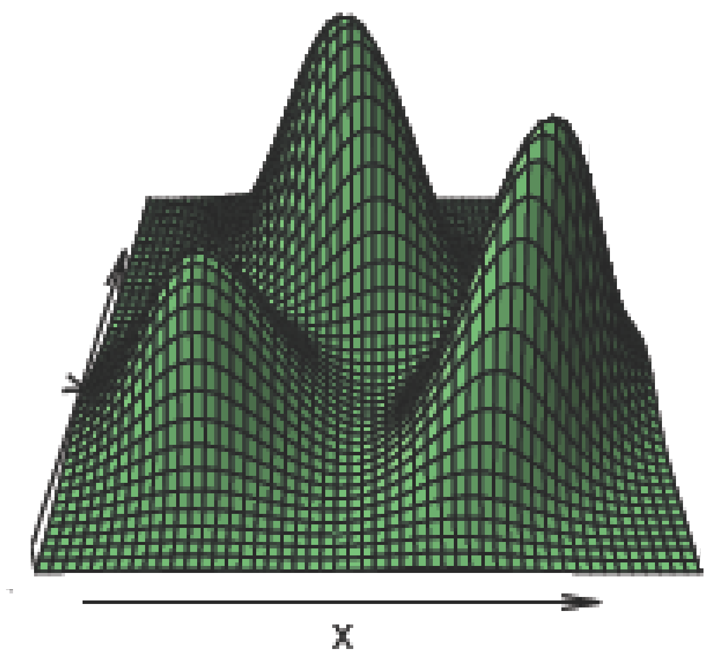
- A1: $\text{local scale} = f\left(\text{large scale, regional characteristics}\right)$
- F1: We need local-scale data!!!

Using SDMs under **climate change**:

- A2: The predictors are relevant and realistically modeled by GCM
- A3: The predictors fully represent the climate change signal
- A4: The SDM is valid also under altered climatic conditions

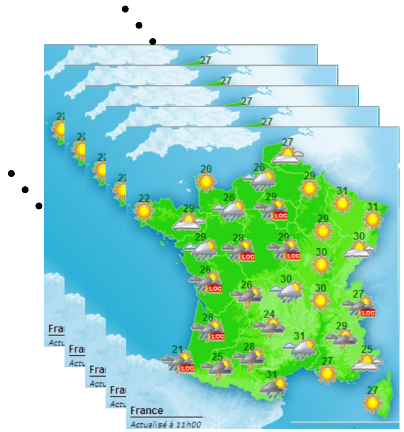
SWG and MOS

<u>Common point</u>:
Look for the **distribution (pdf)** of the phenomena/variables of interest

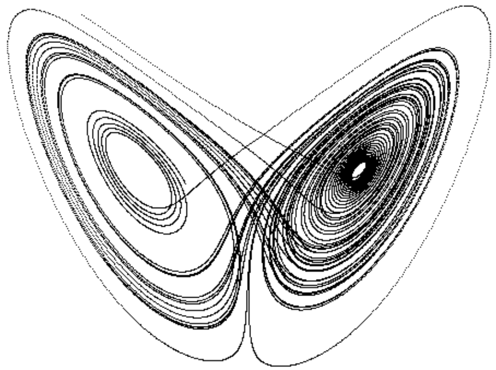


One statistical climatology point of view
(Climate ≠ meteorological events)

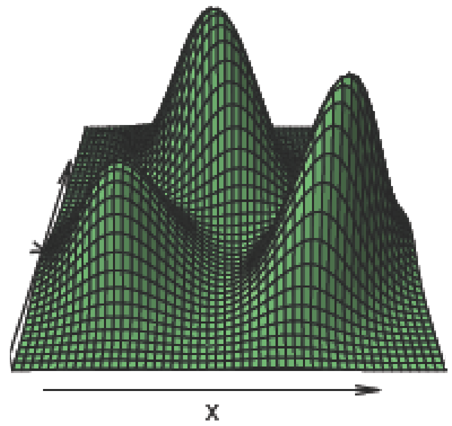
Meteorology ≠ Climate



- Time: ~1 week vs. 100 years



- Dynamics: 1 trajectory vs. the “attractor”



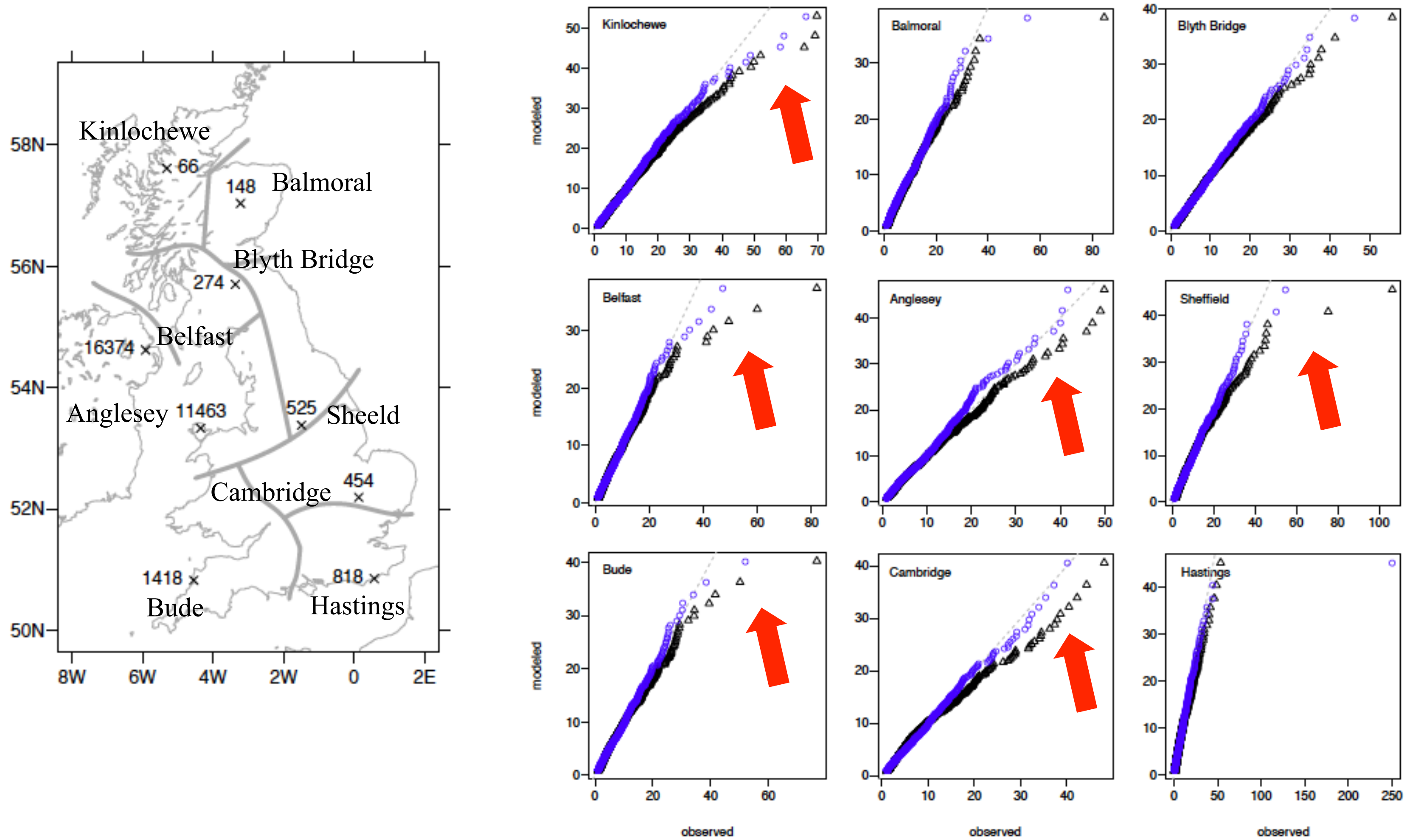
- Statistics:
  1 realization vs. its **random variable**

Main thread of **SWG and MOS** (at least in my work):

What we need is the **pdf or CDF** describing the climate variables

VGLM-Gamma: not always suited… for extremes

From Wong et al. (2014)



QQ plots of wet day intensities (mm/day) for nine example gages in Summer (JJA).

(Values have been standardized to stationary gamma distribution fitted to observed wet day intensities)

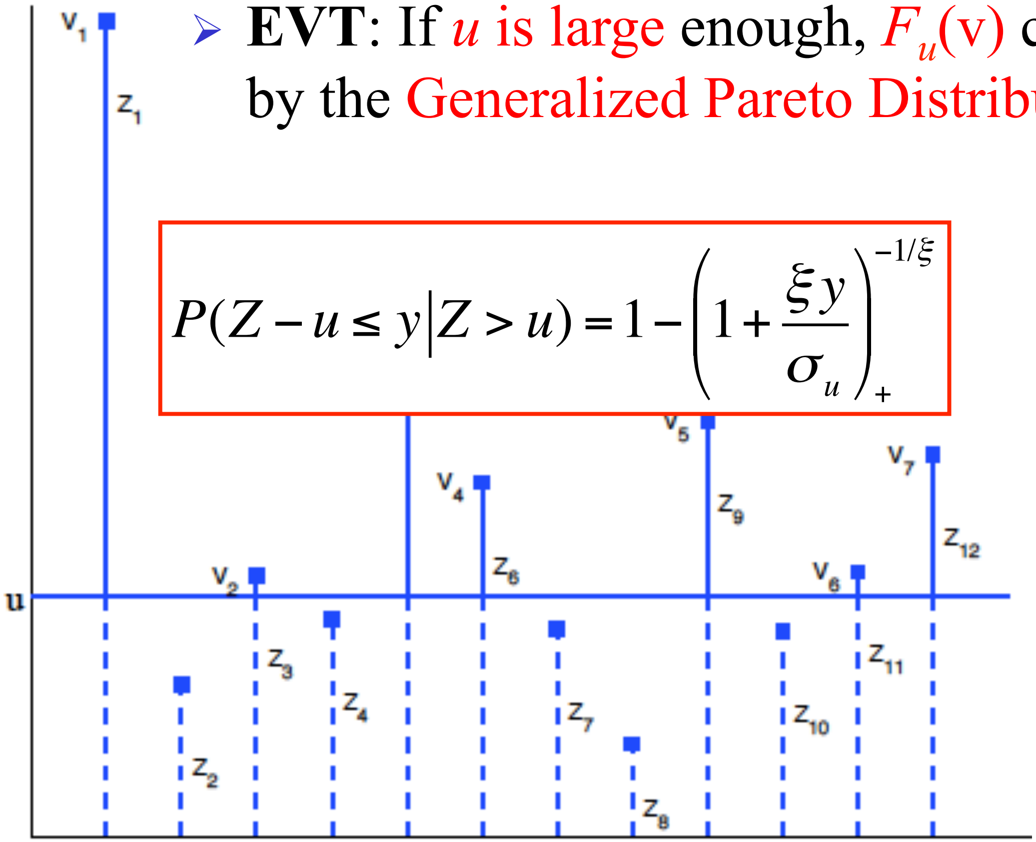
Black triangles: VGLM-Γ.

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)

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

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



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, $F_u(\mathrm{v})$ can be approximated by the Generalized Pareto Distribution (GPD)

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

- ✓ 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)

Merging classical and EV distributions

- Based on Frigessi et al. (2002):

"*Dynamic mixture model for unsupervised tail estimation without threshold*"

$$\phi_0\left(y \middle| \psi_O\right) = c_{\psi_0}\Big[\left(1 - w\left(y \middle| m,\tau\right)\right)\overbrace{\Gamma\left(y \middle| \gamma,\lambda\right)}^{\text{Gamma pdf}}$$

$$+\ w\left(y \middle| m,\tau\right)\underbrace{GPD\left(y \middle| \xi,\sigma,u=0\right)}_{\text{Generalized Pareto Distribution (GPD) pdf}}\Big]$$

functional weight

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

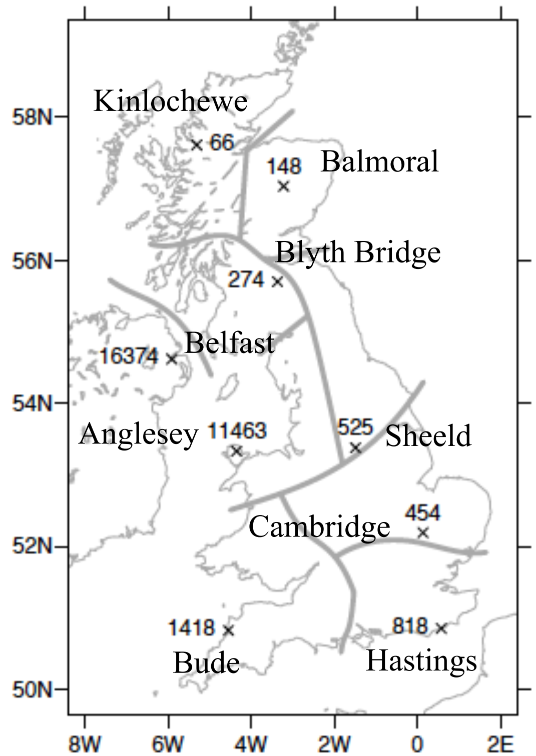
Value where transition from Γ to GPD

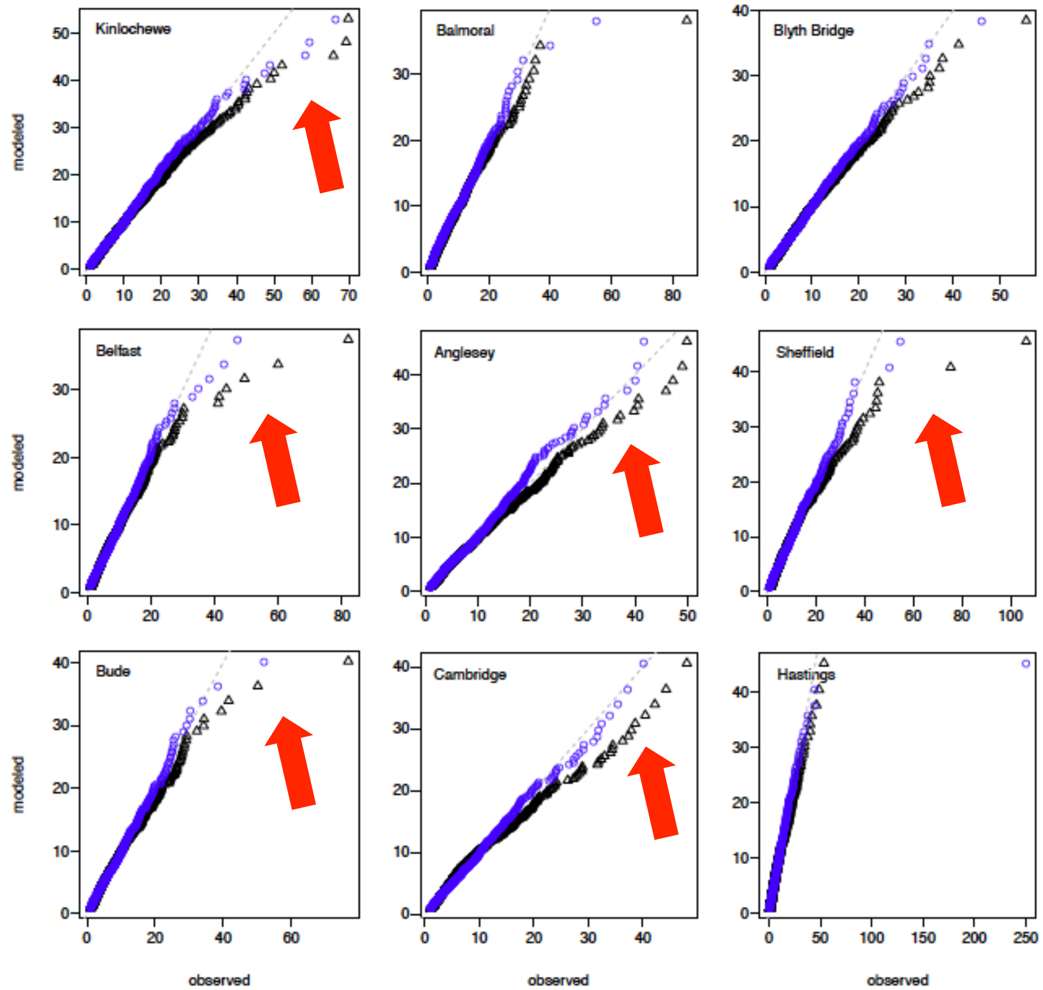
Transition rate

Vrac & Naveau (2007, WRR)

VGLM-mixture model: improving the extremes

From Wong et al. (2014)





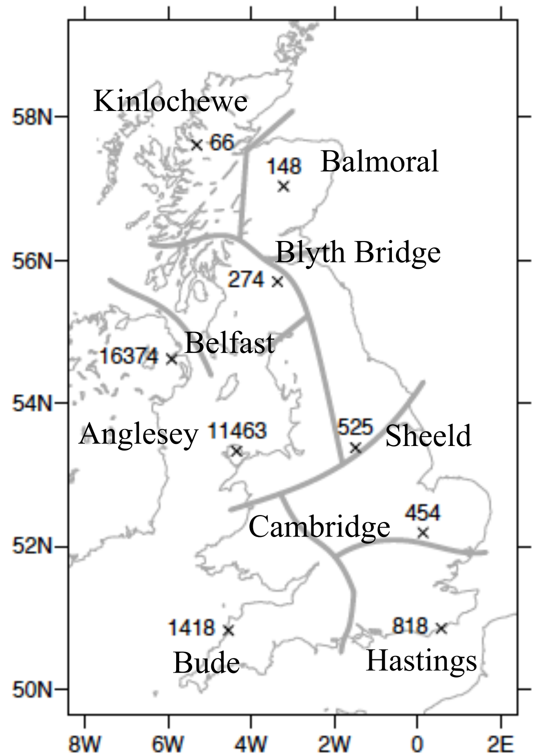
QQ plots of wet day intensities (mm/day) for nine example gages in Summer (JJA).

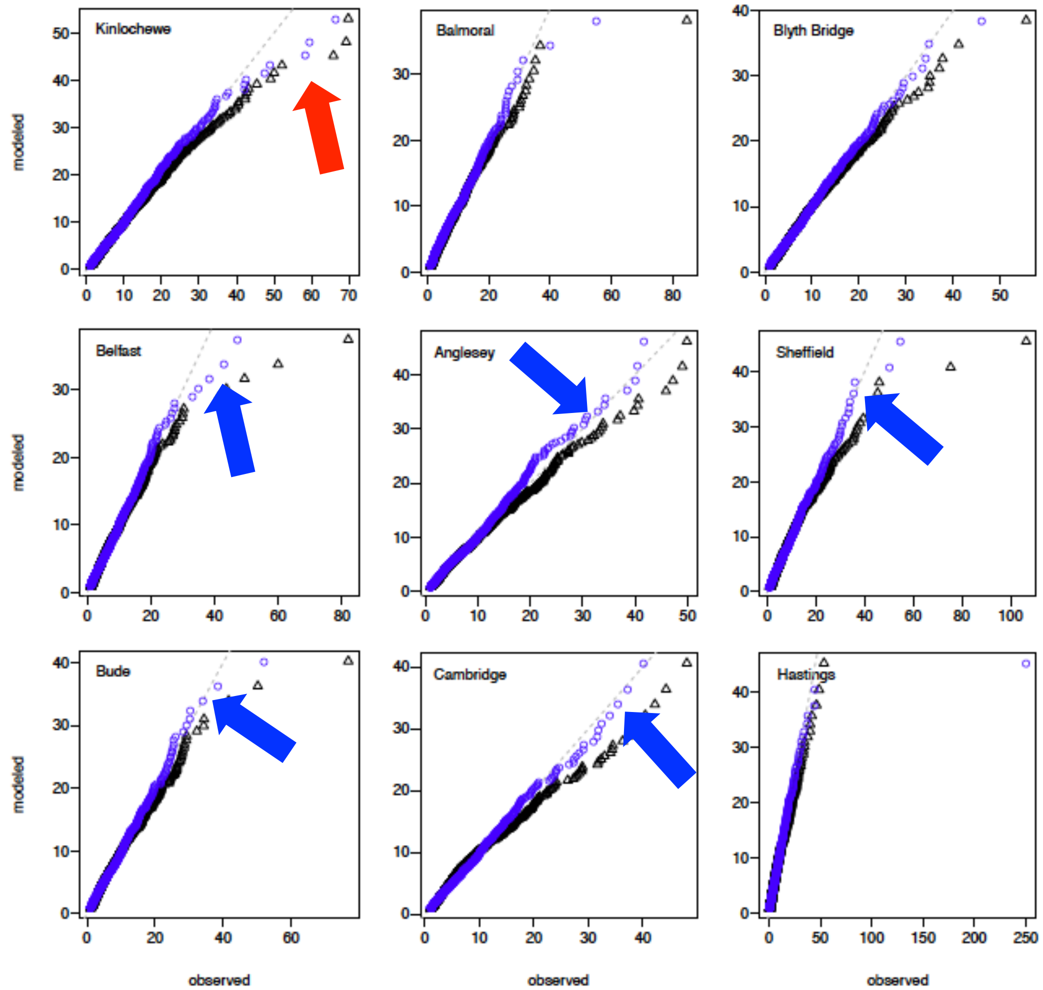
(Values have been standardized to stationary gamma distribution fitted to observed wet day intensities)

Black triangles: VGLM-Γ.

VGLM-mixture model: improving the extremes

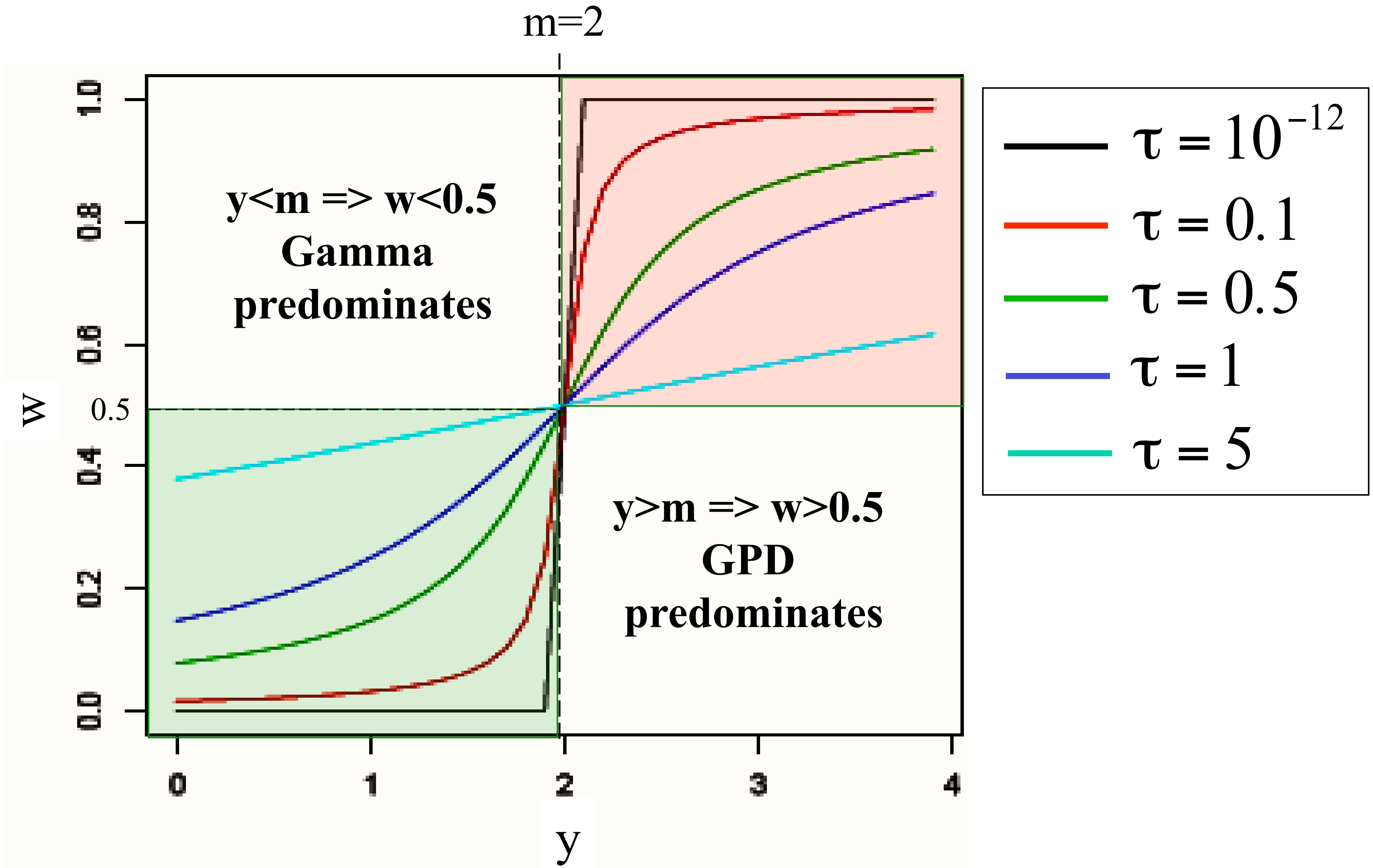
From Wong et al. (2014)





QQ plots of wet day intensities (mm/day) for nine example gages in Summer (JJA).
(Values have been standardized to stationary gamma distribution fitted to observed wet day intensities)
Black triangles: VGLM-Γ. **Blue circles: VGLM-mixture model.**

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



AIC (DJF)

Rain gage	Stationary mixture	VGLM gamma	VGLM mixture
Kinlochewe	15650	**14855**	14905
Balmoral	7786	7704	**7701**
Blyth Bridge	8129	**8040**	8054
Belfast	7963	**7792**	7803
Anglesey	9024	**8861**	8875
Sheffield	7718	**7638**	7640
Bude	8958	**8865**	8889
Cambridge	4845	**4821**	4825
Hastings	6712	**6605**	6626

AIC (JJA)

Station	Mixture	VGLM gamma	VGLM mixture
Kinlochewe (Src. Id. 66)	10391	**10167**	10170
Balmoral (Src. Id. 148)	**5691**	5754	5691
Blyth Bridge (Src. Id. 274)	7084	**7036**	7038
Belfast (Src. Id. 16374)	6211	6201	**6163**
Anglesey (Src. Id. 11463)	6493	6428	**6428**
Sheffield (Src. Id. 525)	5589	5581	**5549**
Bude (Src. Id. 1418)	6084	6044	**6016**
Cambridge (Src. Id. 454)	4871	4902	**4847**
Hastings(Src. Id. 818)	4700	4700	**4673**