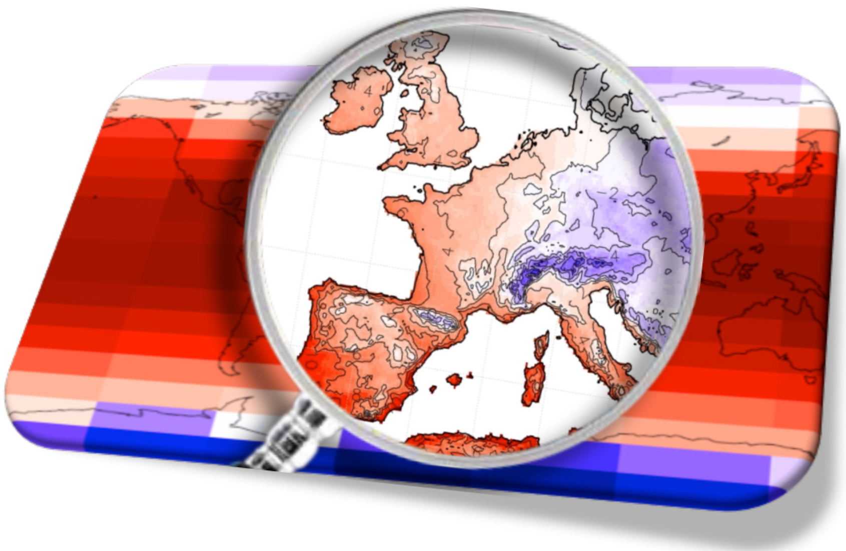
LABORATOIRE DES SCIENCES DU CLIMAT & DE L'ENVIRONNEMENT

Modélisations statistiques à différentes échelles climatiques et environnementales

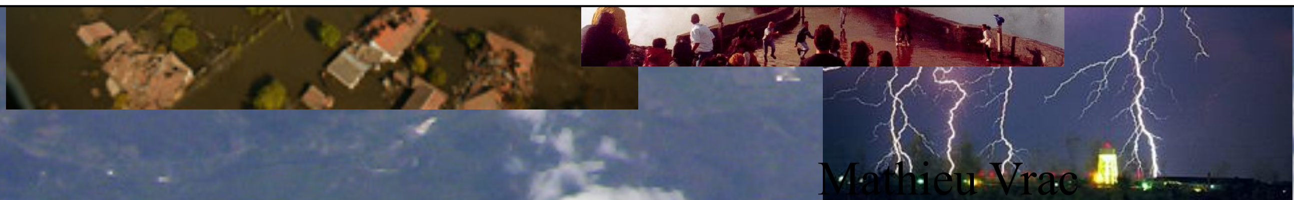


Mathieu Vrac

Soutenance d'Habilitation à Diriger des Recherches - UVSQ
30 mars 2012

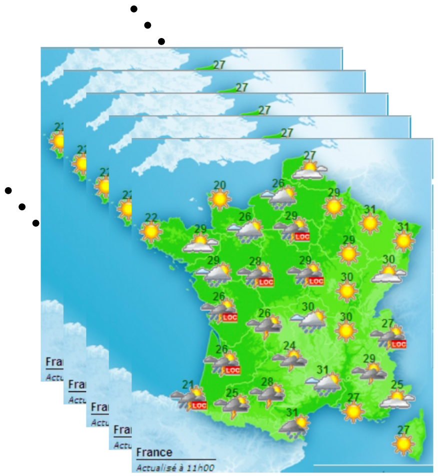
Motivations

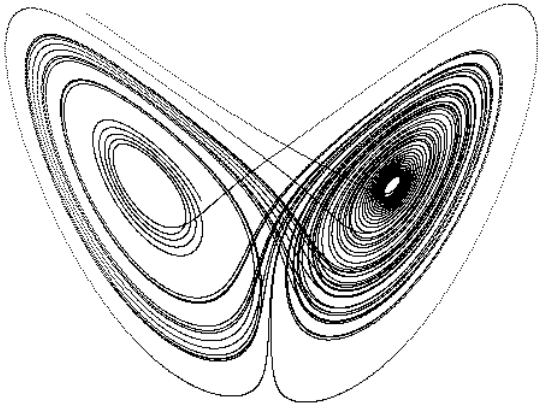
- **~30%** of the world economic activities are affected by the meteorological conditions (source: IPCC AR4)
- IPCC scenarios of climate change and GCMs have:
  - A **coarse spatial resolution** (~250km) **!!**
    - ⇒ Needs for *downscaling*
  - **Uncertainties/spatial biases/errors** (many causes)
    - ⇒ Large-scale *weather regimes*
  - Different types of "**extreme events**" to model
    - ⇒ *Occurrence – persistence – magnitude*
- **My approach**: *Statistical climatology* via **pdf modelling**



Meteorology ≠ Climate

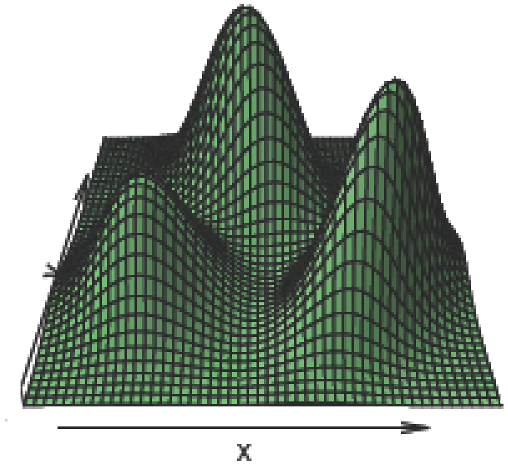
- <u>Time</u>: ~1 week vs. 100 years





- <u>Dynamics</u>: 1 trajectory vs. the “attractor”

- <u>Statistics</u>:
  1 realization vs. its **random variable**



Main thread of this talk:

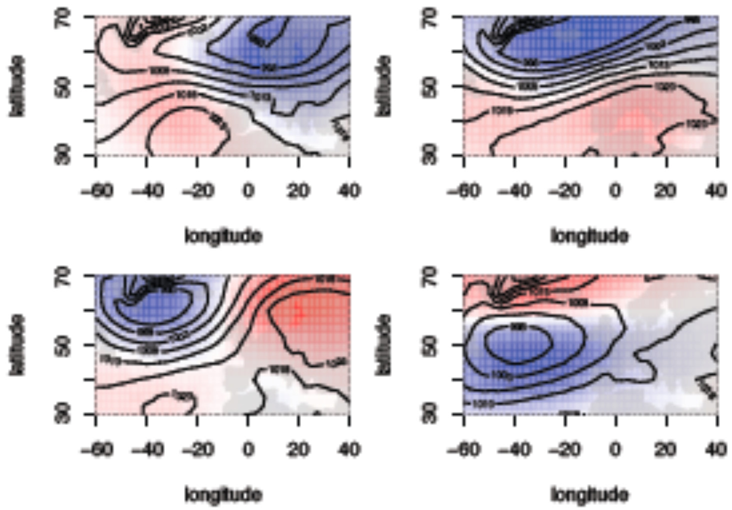
What we need is the **pdf** describing the climate variables

Outline

- Large-scale Weather Regimes
  - WRs through mixture modelling via the EM algorithm
  - "Distance" measures between pdfs associated with regimes
  - Seasonal regimes and evolutions
- Statistical Downscaling (a.k.a. regionalization)
  - What's that? How can we do?
  - Stochastic weather generators => Focus on pdf
- Extreme Events Modelling
  - Main (univariate) notions
  - DS of extreme
- And now, what? (i.e., Perspectives)

Outline

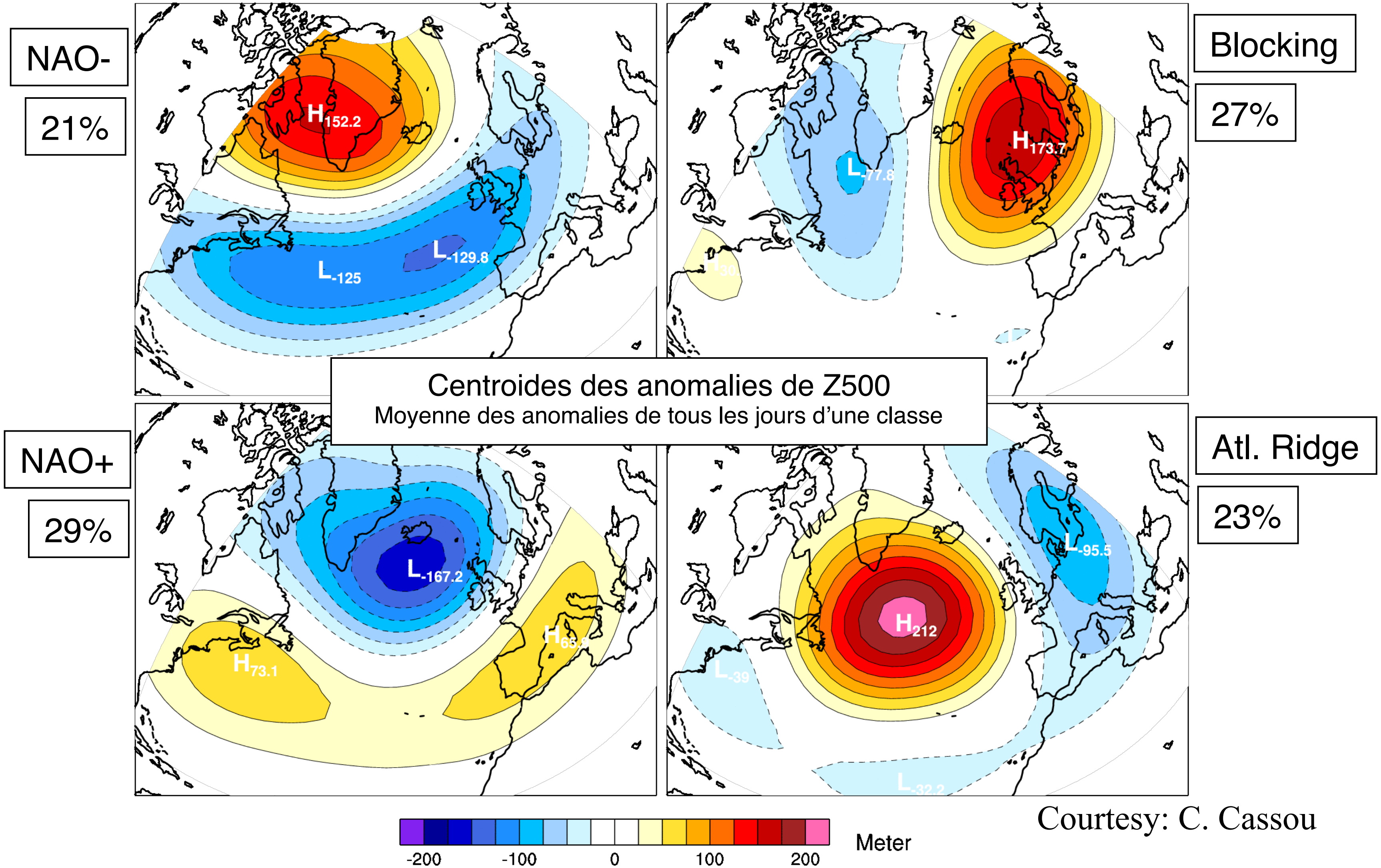
- Large-scale Weather Regimes
  - WRs through mixture modelling via the EM algorithm
  - "Distance" measures between pdfs associated with regimes
  - Seasonal regimes and evolutions



- Statistical Downscaling (a.k.a. reg
  - What's that? How can we do?
  - Stochastic weather generators => Fc
- Extreme Events Modelling
  - Main (univariate) notions
  - DS of extreme
- And now, what? (i.e., Perspectives)

Large-Scale Weather Regimes (WR)

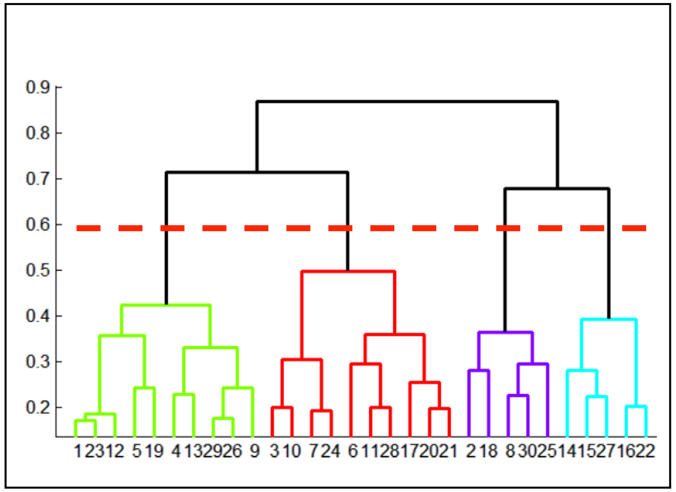
- **What?**
  - Elementary "pattern" of the large-scale extra-tropical circulation
  - Recurrent / persistent (several days) / transitions



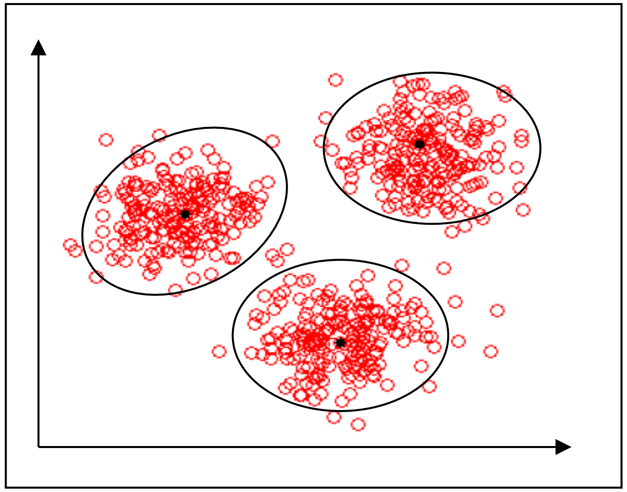
Courtesy: C. Cassou

Large-Scale Weather Regimes (WR)

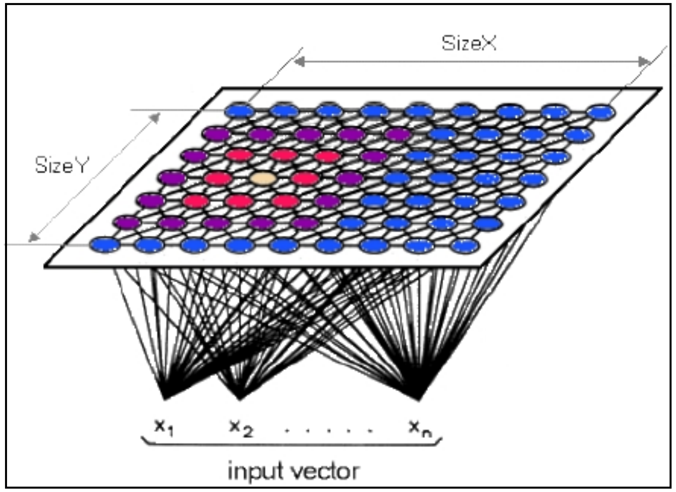
- What?
  - Elementary "pattern" of the large-scale extra-tropical circulation
  - Recurrent / persistent (several days) / transitions
- Why?
  - Describe the main variability of the climate
    ⇒ Ease the physical interpretation (discretization)
  - Evaluation tool: How do GCMs simulate the main circulation properties?
- How?
  - *Subjective* WRs (e.g., Lamb, 1972, UK; Hess & Brezowsky, 1976, EU)
  - *Objective* WRs: **Mathematical clustering methods**



Hierarchies (e.g., Ward, 1963)



k-means (Diday, 1977)



SOM (Kohonen, 1982)

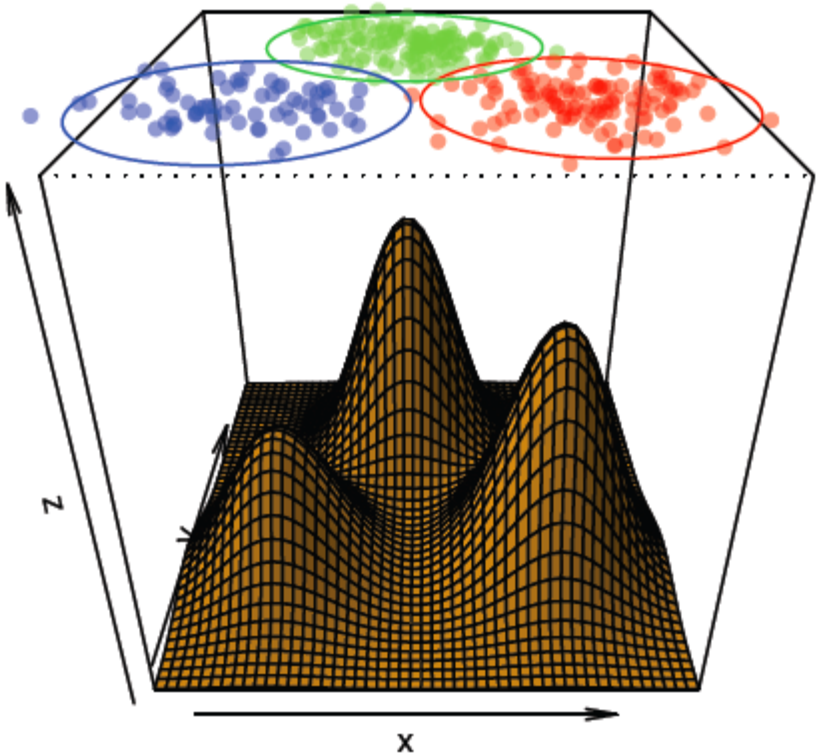
WRs via the Mixture Model approach

- Probability density function (pdf) -based approach
  - Any density can be approximated by a mixture of Gaussian densities
- Looks for pdf *f(x)* as a weighted sum of parametric pdf's:

$$f(x) = \sum_{k=1}^{K} \underbrace{\pi_k}_{\text{Weight (or mixture ratio)}} \underbrace{f_k(x|\alpha_k)}_{\text{Parametric (Gaussian) pdf}} \qquad \text{with } 0 < \pi_k < 1 \text{ and } \sum_{k=1}^{K} \pi_k = 1$$

- Each parametric pdf f_k is associated to a weather regime

Example with K=3, and
f_k = 2-d Gaussian distribution



The "EM" algorithm

- Successive iterations (i) of the E and M steps
  - Expectation (E) step: for each k=1,…, K (WRs), and j=1,…,n (data), computation of

$$\tau_k^i(x_j) = \frac{\pi_k^i f_k(x_j | \alpha_j^i)}{\sum_{k=1}^{K} \pi_k^i f_k(x_j | \alpha_j^i)}$$

Posterior proba that x_j belongs to WR k at iteration i

  - Maximization (M) step: for each k=1,…, K (WRs), computation of

$$\pi_k^{i+1} = \frac{1}{n} \sum_{j=1}^{n} \tau_k^i(x_j)$$ (= Max. Likelihood estimator of the ratios)

and solve likelihood equations

$$\sum_{j=1}^{n} \tau_k^i(x_j) \frac{\partial Log(f_k(x_j | \alpha_k^{i+1}))}{\partial \alpha_{m,j}} = 0$$

✓ Dempster et al. (1977)

✓ McLachlan and Peel (2000)

The "EM" algorithm

- Successive iterations (i) of the E and M steps
  - Expectation (E) step: for each k=1,…, K (WRs), and j=1,…,n (data), computation of

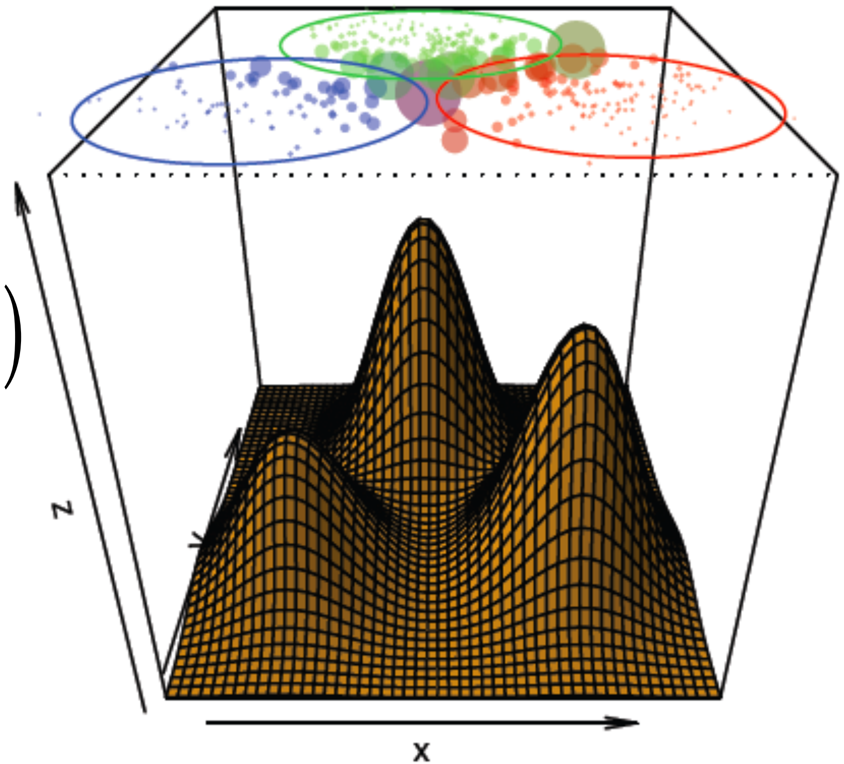
$$\tau_k^i(x_j) = \frac{\pi_k^i f_k(x_j | \alpha_j^i)}{\sum_{k=1}^{K} \pi_k^i f_k(x_j | \alpha_j^i)}$$

Posterior proba that x_j belongs to WR k at iteration i

Uncertainty:

$$u_k(x_j) = 1 - \Pr(x_j \in WR_k | x_j)$$

$$= 1 - \tau_k(x_j)$$



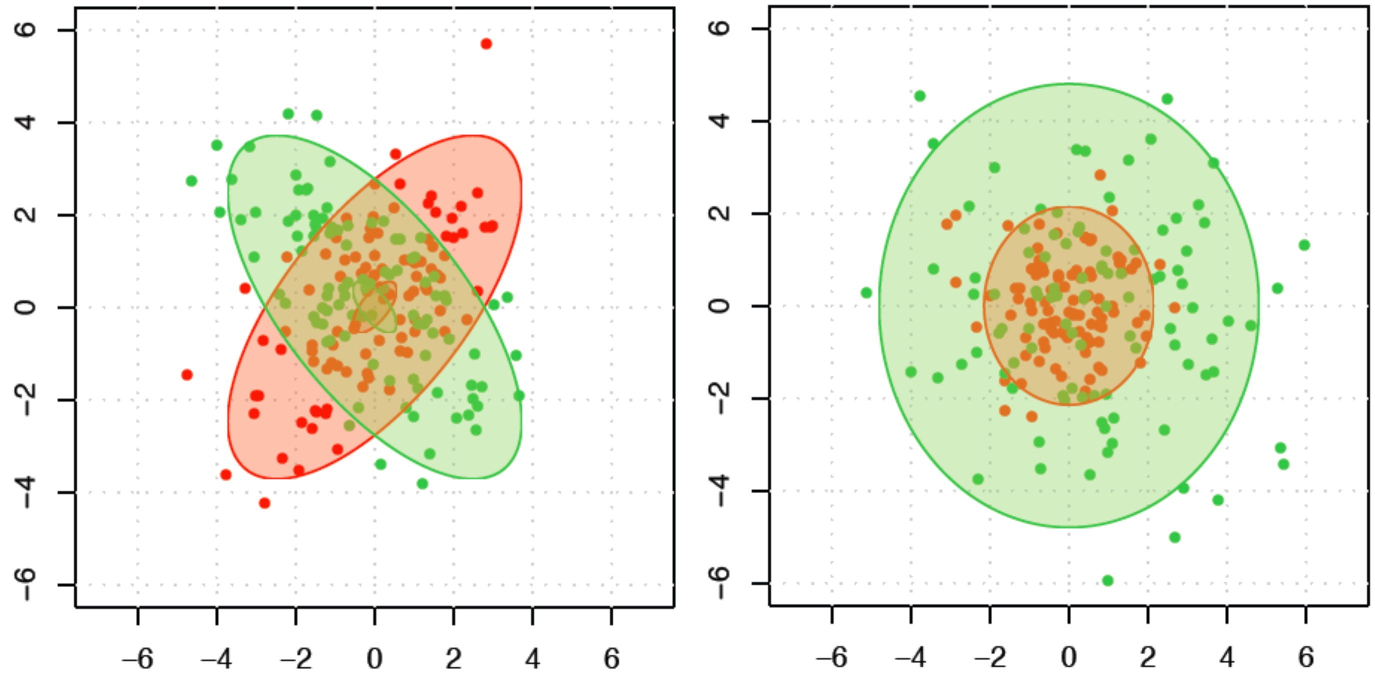
✓ Dempster et al. (1977)
✓ McLachlan and Peel (2000)

pdf-based metrics for comparing WRs

(Rust, Vrac, Lengaigne, Sultan, 2010, JClim)

pdf-based metrics for comparing WRs

- Classically, comparisons through "mean states"
  - Visually
  - Euclidean distance
  - Pattern correlation
  - Are mean states representative?
  - Variance (spread) / Shape?



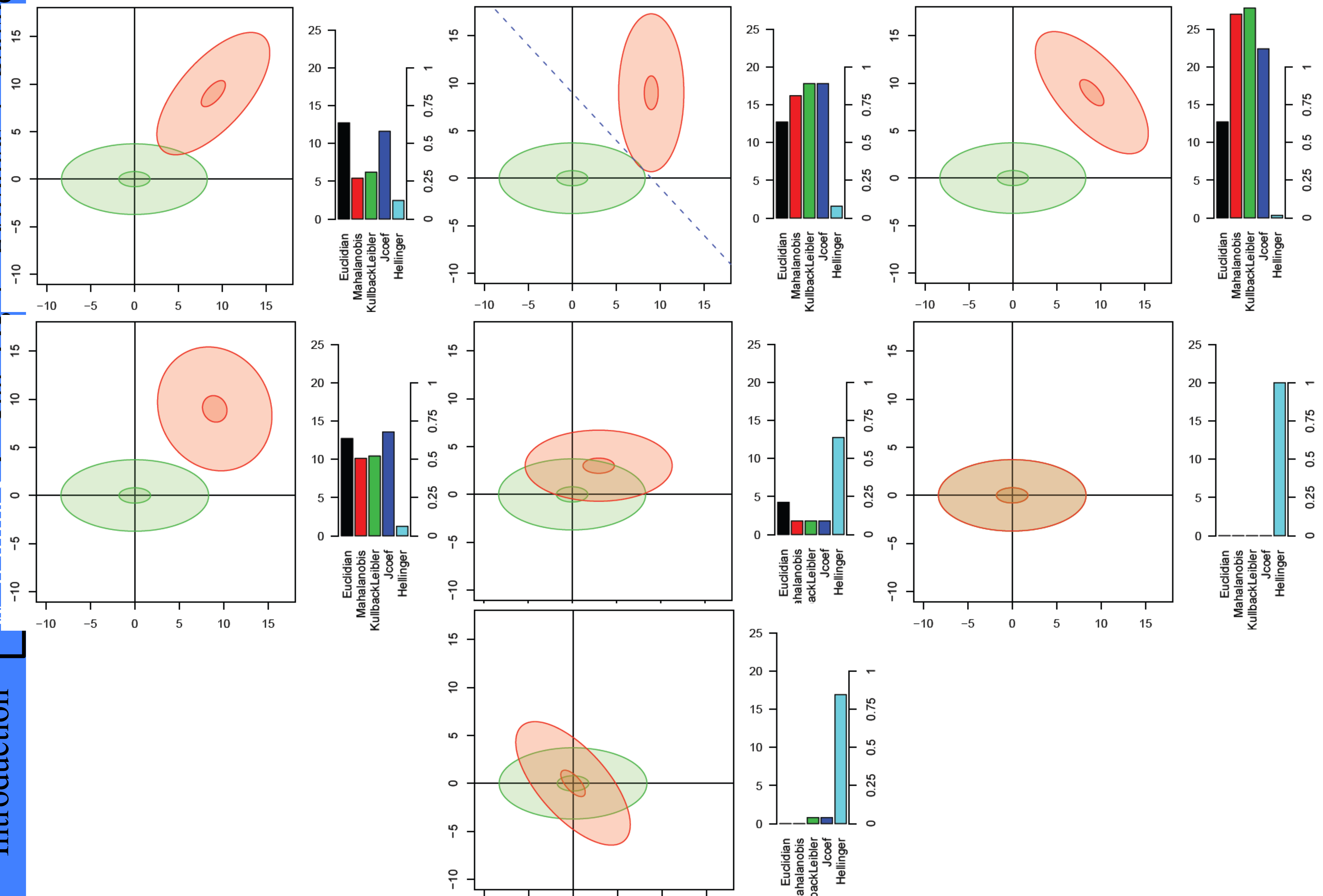
Same means,
different shapes and variances

pdf-based metrics for comparing WRs

- Classically, comparisons through "mean states"
  - Visually
  - Euclidean distance
  - Pattern correlation
  - Are mean states representative?
  - Variance (spread) / Shape?

- Distance measures between pdfs
  - Euclidean distance $d^2_{Eucl}(P,Q) = \left\| \mu_p - \mu_q \right\|^2 = (\mu_p - \mu_q)^T I (\mu_p - \mu_q)$
  - Mahalanobis dist. $d^2_{Mah}(P,Q) = \left\| \mu_p - \mu_q \right\|^2_{\Sigma_p^{-1}} = (\mu_p - \mu_q)^T \Sigma_p^{-1} (\mu_p - \mu_q)$
  - Kullback-Leibler discrimination (KL, or I-coef) $d_{KL}(P,Q) = I(P|Q) = \int_R \log\left(\frac{q(x)}{p(x)}\right) q(x)dx$
  - J-coefficient $d_J(P,Q) = \big(I(P|Q) + I(Q|P)\big)/2$
  - Hellinger coefficient $d_H^{(s)}(P,Q) = \int_R q(x)^s p(x)^{(s-1)} dx, \quad d_H \in [0,1]$

pdf-based metrics for comparing WRs

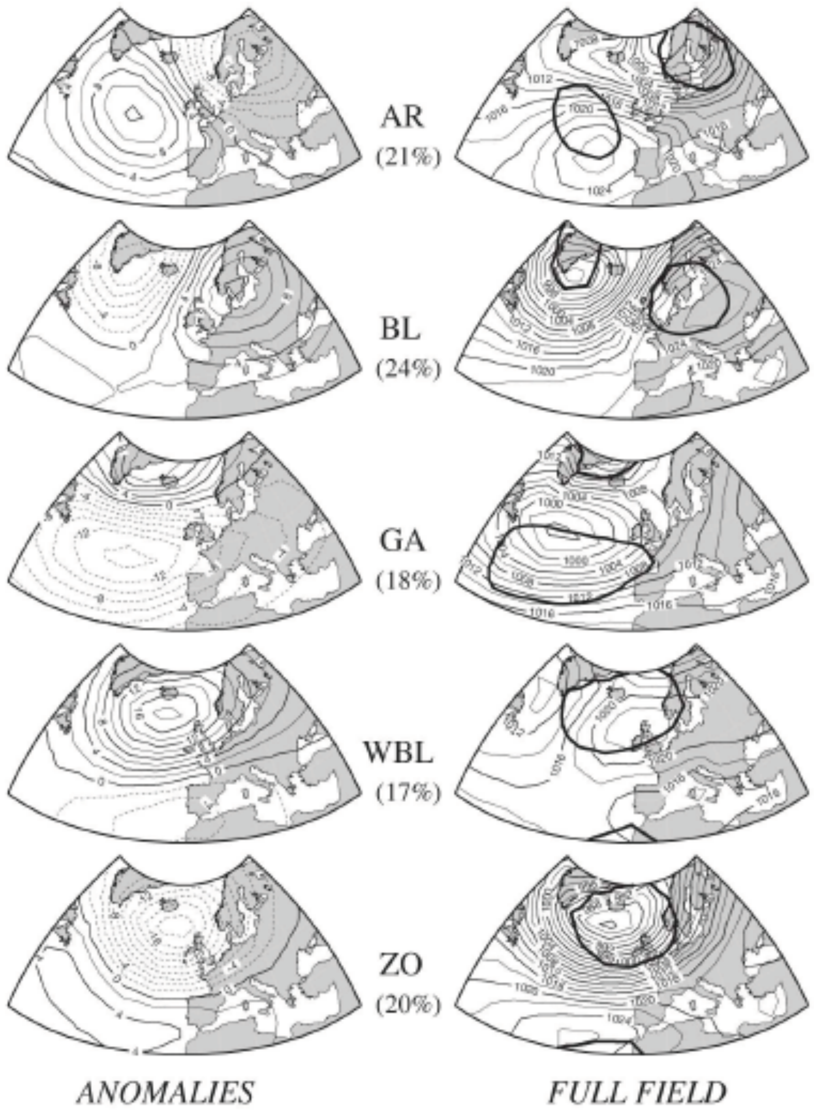


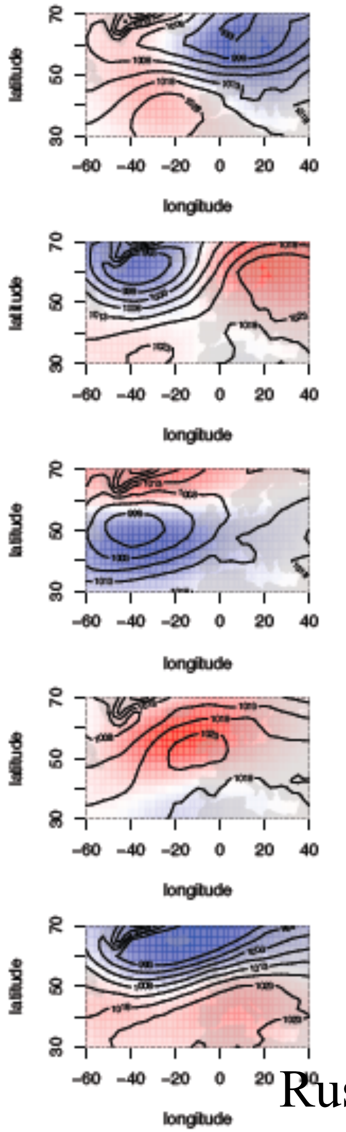
5 "spherical" WRs

- Plaut & Simonnet (2001): k-means on North Atlantic daily NCEP SLP winter (NDJFM) anomalies => **5 spherical WRs**

⇒ EM with K=5 with spherical covariance matrix (i.e., $\Sigma_k = \mathrm{diag}(\sigma_k)$)

Plaut & Simonnet (2001)



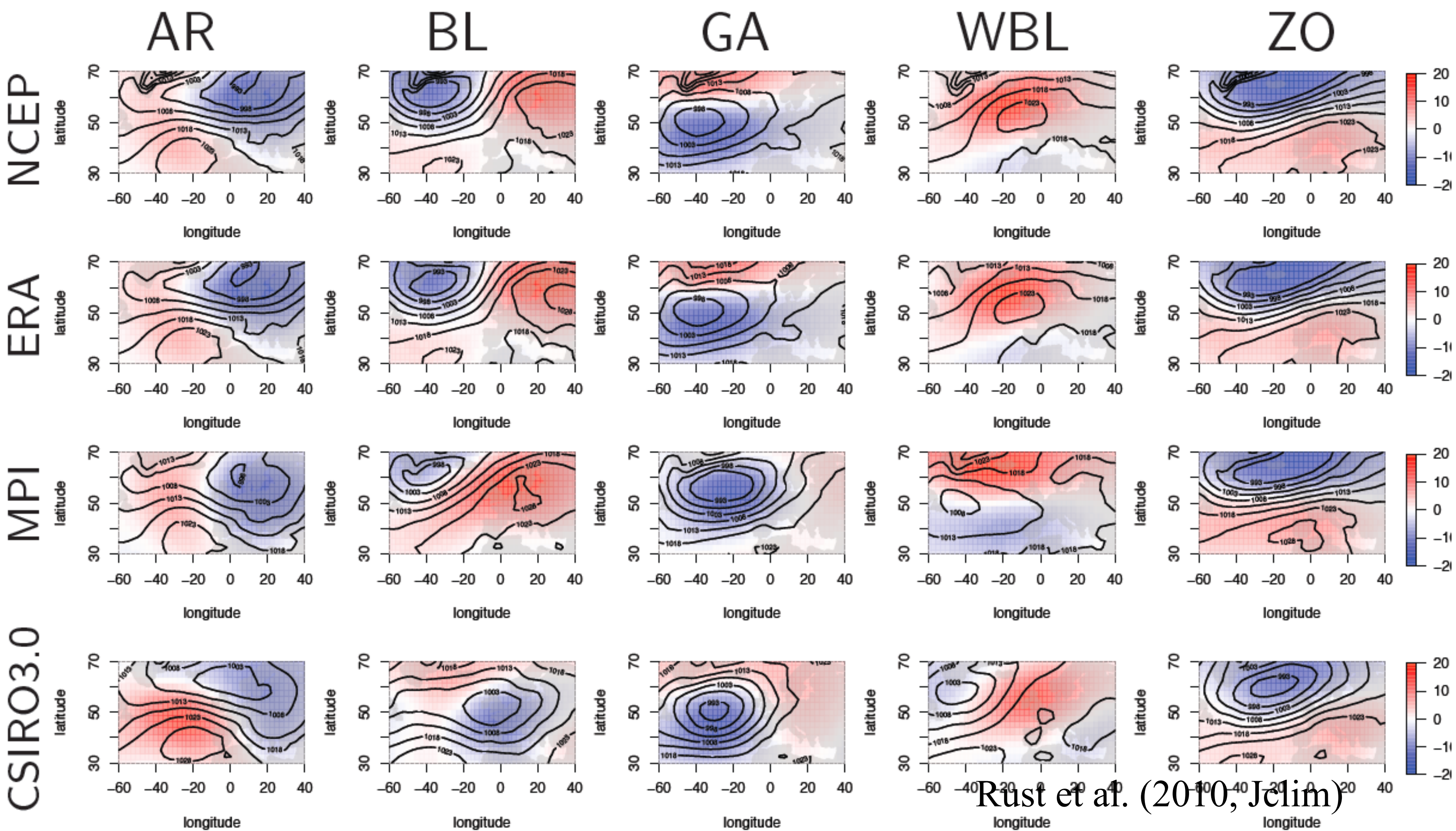


Rust et al. (2010, Jclim)

5 "spherical" WRs

- Plaut & Simonnet (2001): k-means on North Atlantic daily NCEP SLP winter (NDJFM) anomalies => **5 spherical WRs**

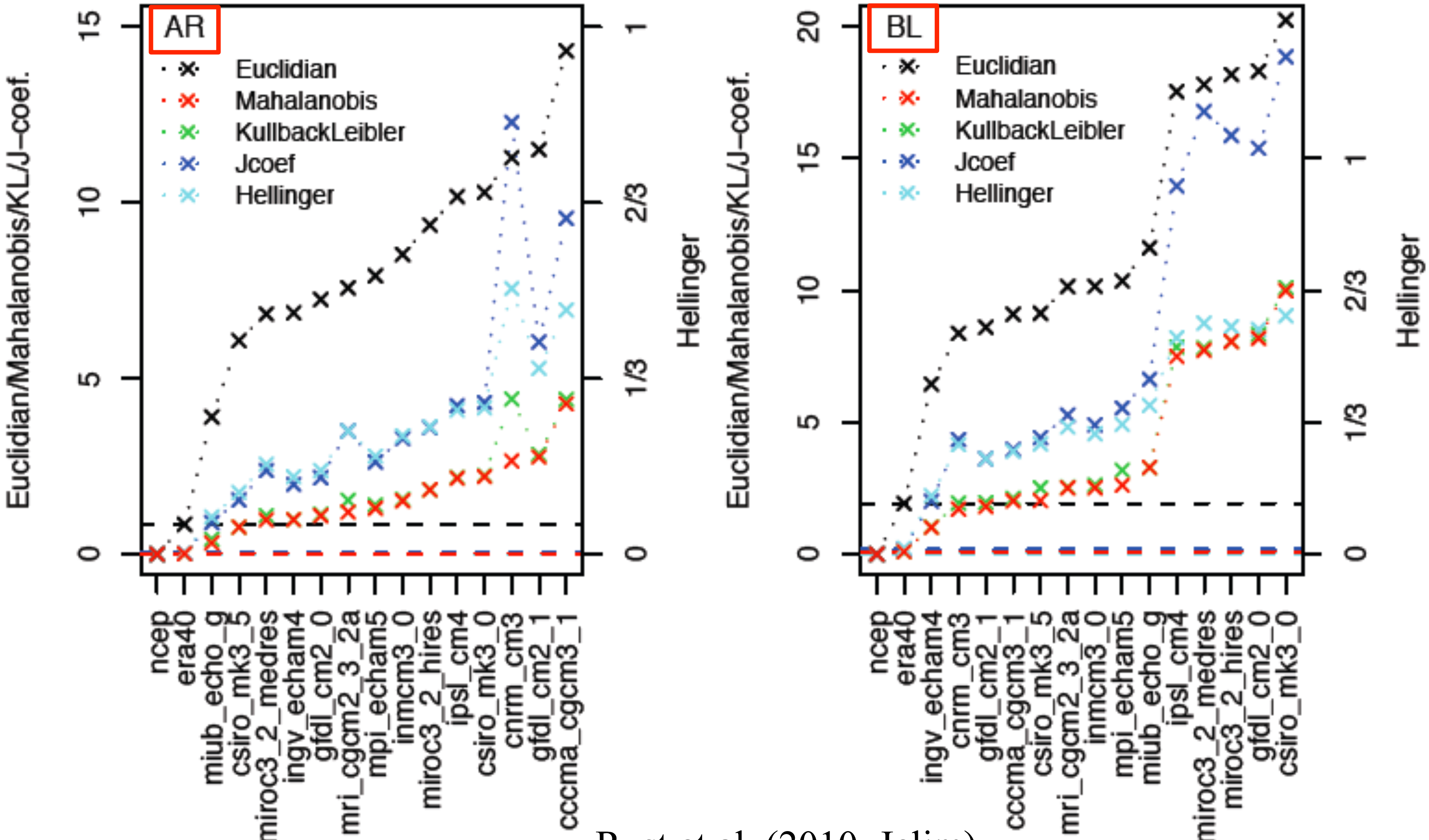
⇒ EM with K=5 with spherical covariance matrix (i.e., $\Sigma_k = \mathrm{diag}(\sigma_k)$)



Rust et al. (2010, Jclim)

5 “spherical” WRs

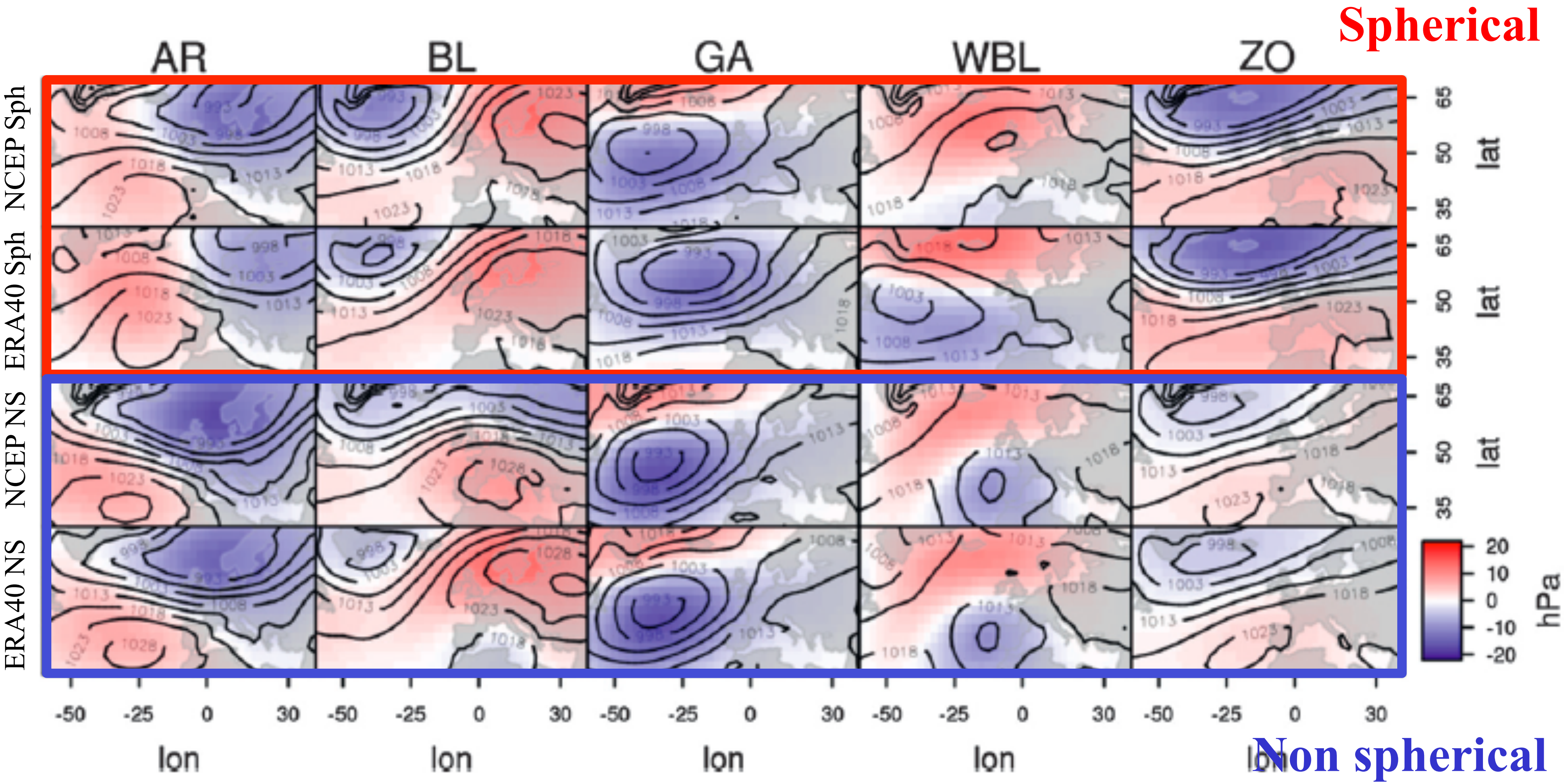
- Quantifying differences to NCEP/NCAR WRs
  - ✓ Association is made by minimizing the Euclidian distances



Rust et al. (2010, Jclim)

5 “NON-spherical” WRs

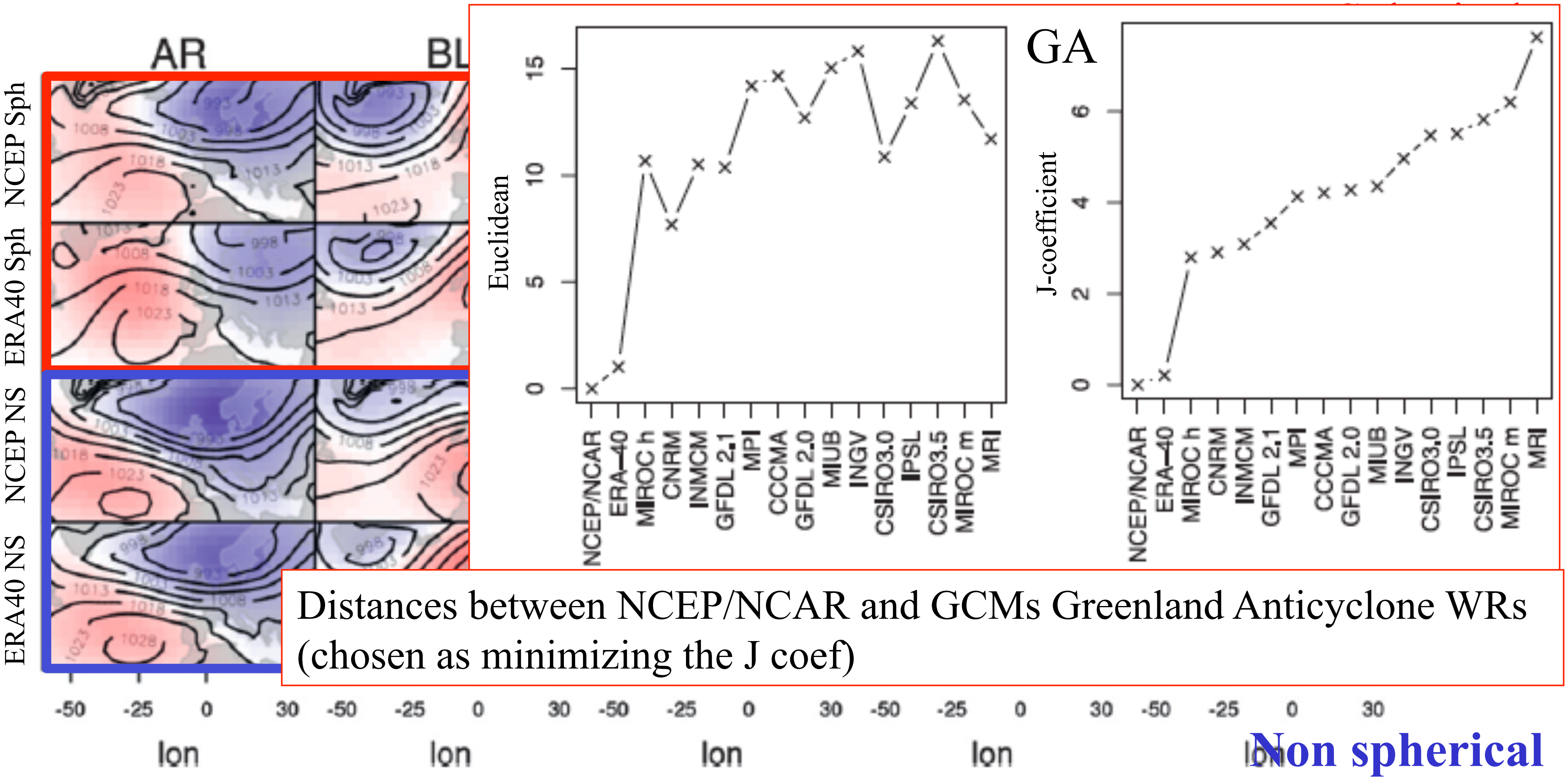
- “Optimal WRs are **not spherical** !” BIC says



Rust et al. (2010, Jclim)

5 "NON-spherical" WRs

- "Optimal WRs are **not spherical** !" BIC says
  - ⇒ Euclidean distance is not appropriate anymore !
  - ⇒ K-means is not suited ! Mixtures are more flexible !



Distances between NCEP/NCAR and GCMs Greenland Anticyclone WRs (chosen as minimizing the J coef)

Non spherical

Rust et al. (2010, Jclim)

Seasonal WRs via mixture models

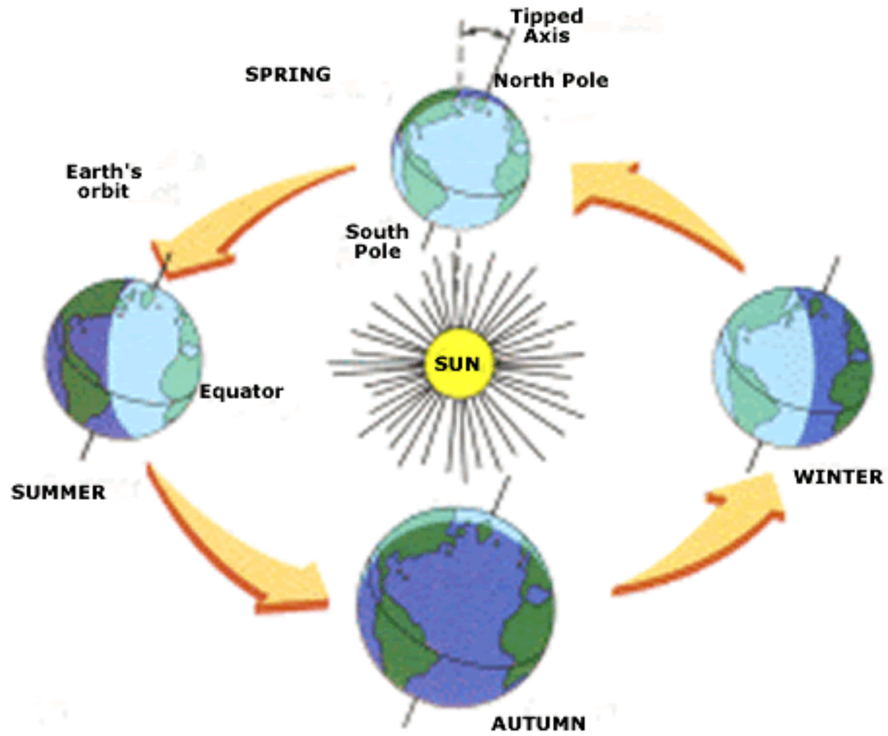
Stage de fin d'étude de P. Vaittinada-Ayar

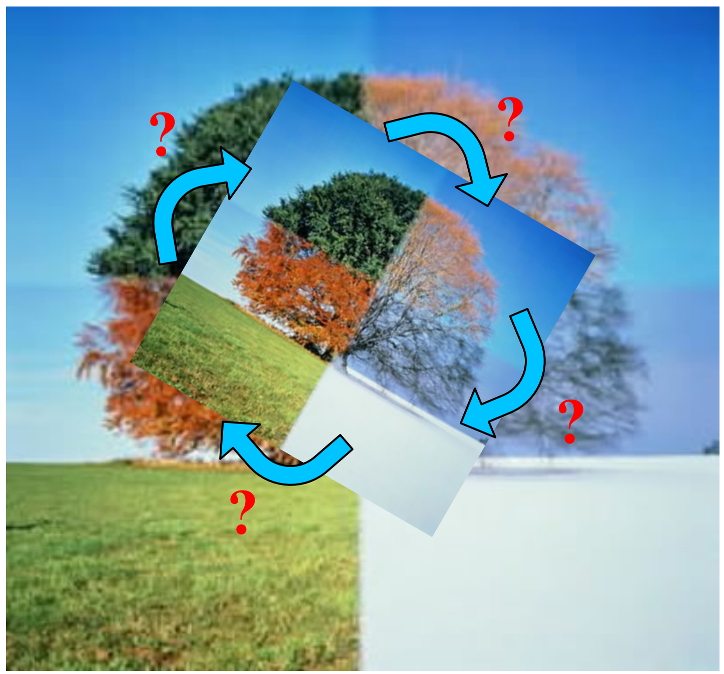
(Vrac, Yiou & Vaittinada-Ayar, 2012, submitted)

Seasonal WRs via mixture models

(Il n'y a plus de saisons – ma brave dame (?))

- Changes of *seasonality* (e.g., Barnston and Livezey, 1987; Jacobeit et al., 2002; Zveryaev, 2006; Trenberth et al., 2007, etc.):
  - May impact: storm activities (Ulbrich et al., 2009), tourism (Amelung et al., 2007), energy, vegetation, health, etc.
  - Never done through WR modelling





Seasonal WRs via mixture models

(Il n'y a plus de saisons – ma brave dame (?))

- Changes of *seasonality* (e.g., Barnston and Livezey, 1987; Jacobeit et al., 2002; Zveryaev, 2006; Trenberth et al., 2007, etc.):
  - May impact: storm activities (Ulbrich et al., 2009), tourism (Amelung et al., 2007), energy, vegetation, health, etc.
  - Never done through WR modelling
- Classical clustering: per season and on anomalies (e.g., Michelangeli et al., 1995, Yiou and Nogaj, 2004, Cassou, 2008, etc.)
  - Here: characterize the seasonality of regimes all along the year

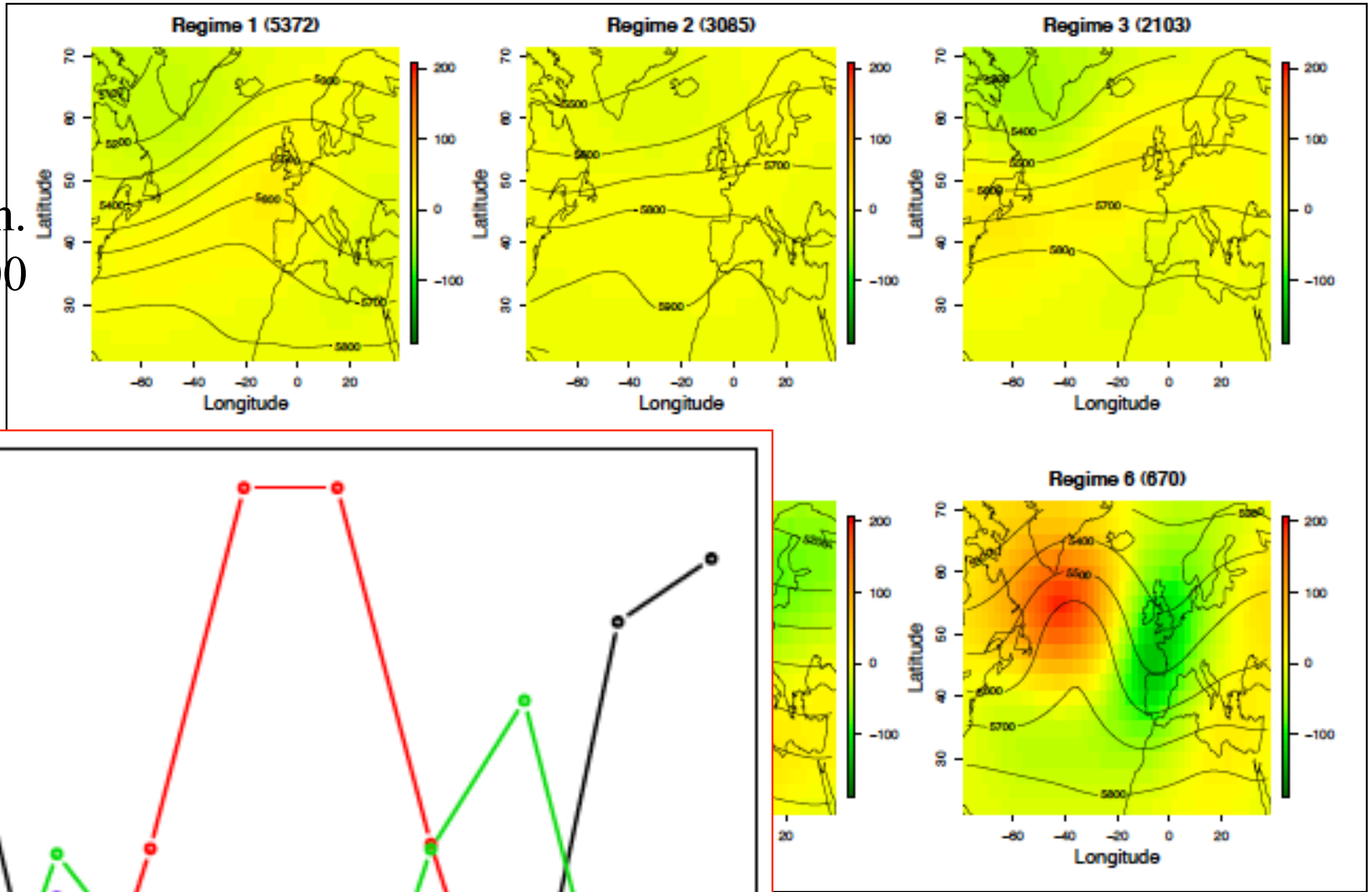
    $\Rightarrow$ **All seasons** altogether + **NOT deseasonalized** data
- Data: NCEP/NCAR daily Z500 (1975-2009) over North Atlantic
  - EM on weighted PCA (95% of variance => 11 PCs)

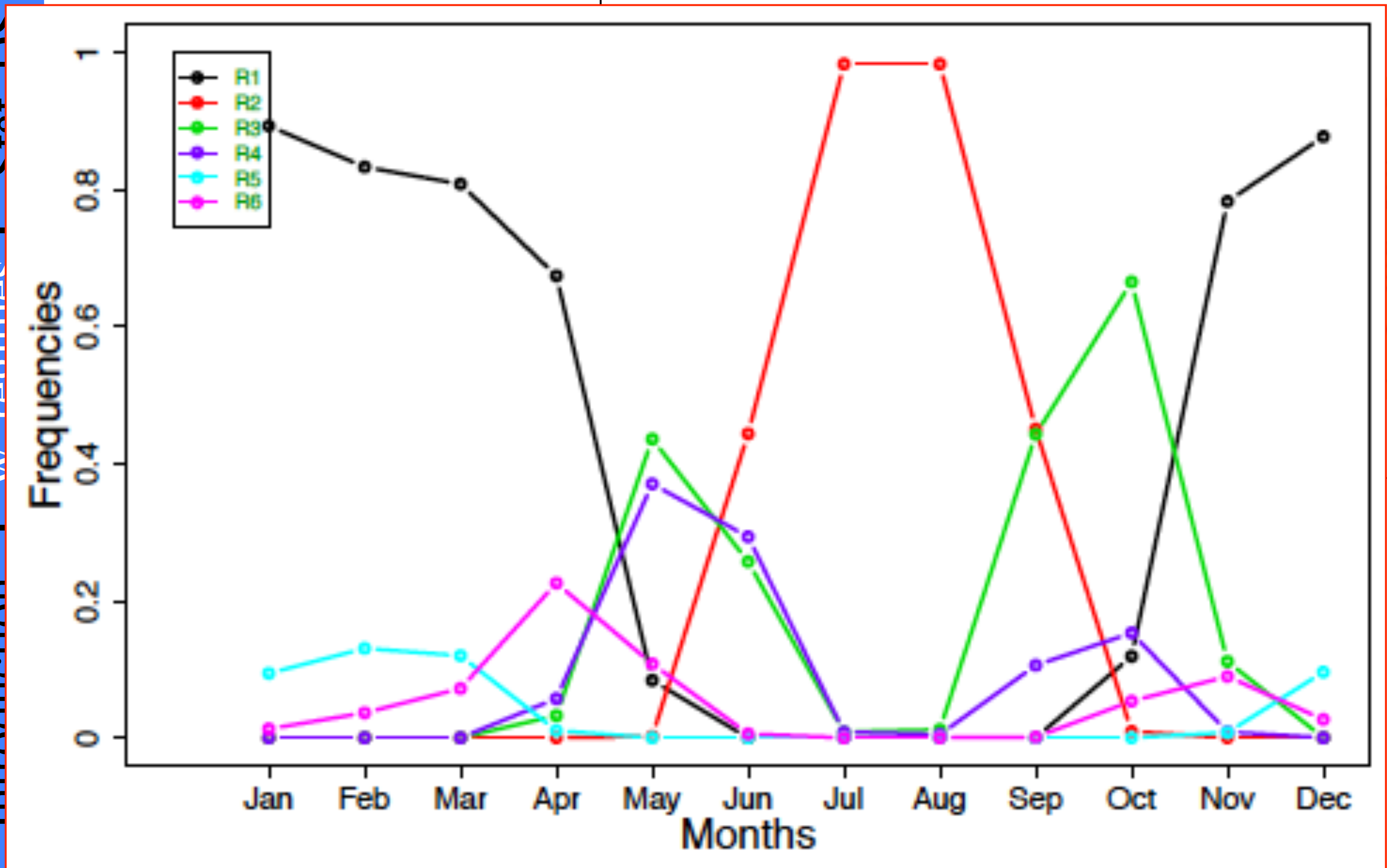
Vrac et al. (2012)

Seasonal WRs: Seasonal cycles

BIC => K=6 WRs
(ellipsoidal cov.)

Colors = mean Z500 anom.
Contours = mean raw Z500



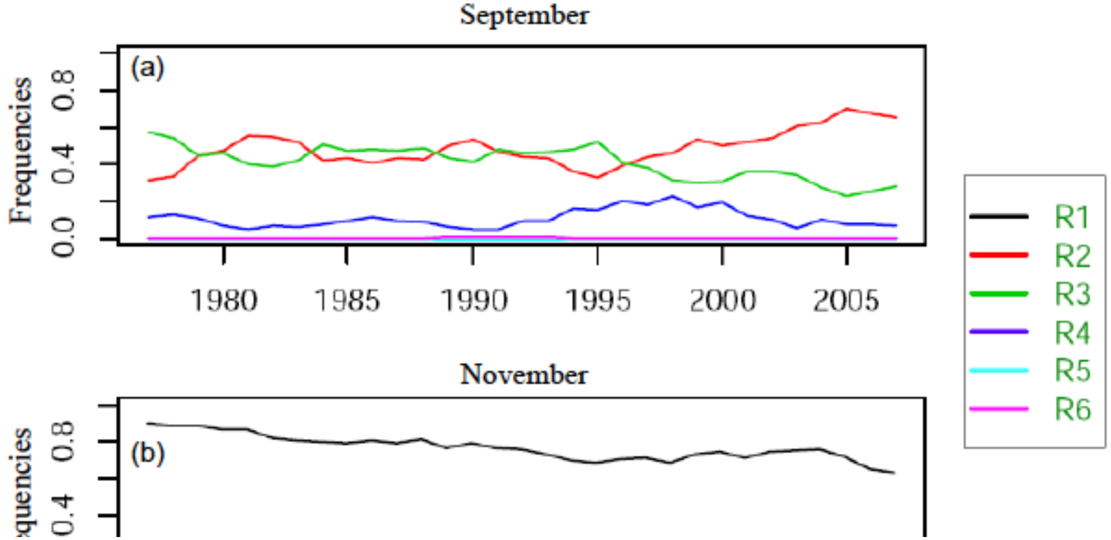


Seasonal cycles of the occurrence monthly frequencies

Vrac et al. (2012)

Seasonal WRs: Frequencies evolutions

Temporal evolution (1977-2007, 5-year moving average) of the monthly frequencies:
(a) September; (b) November



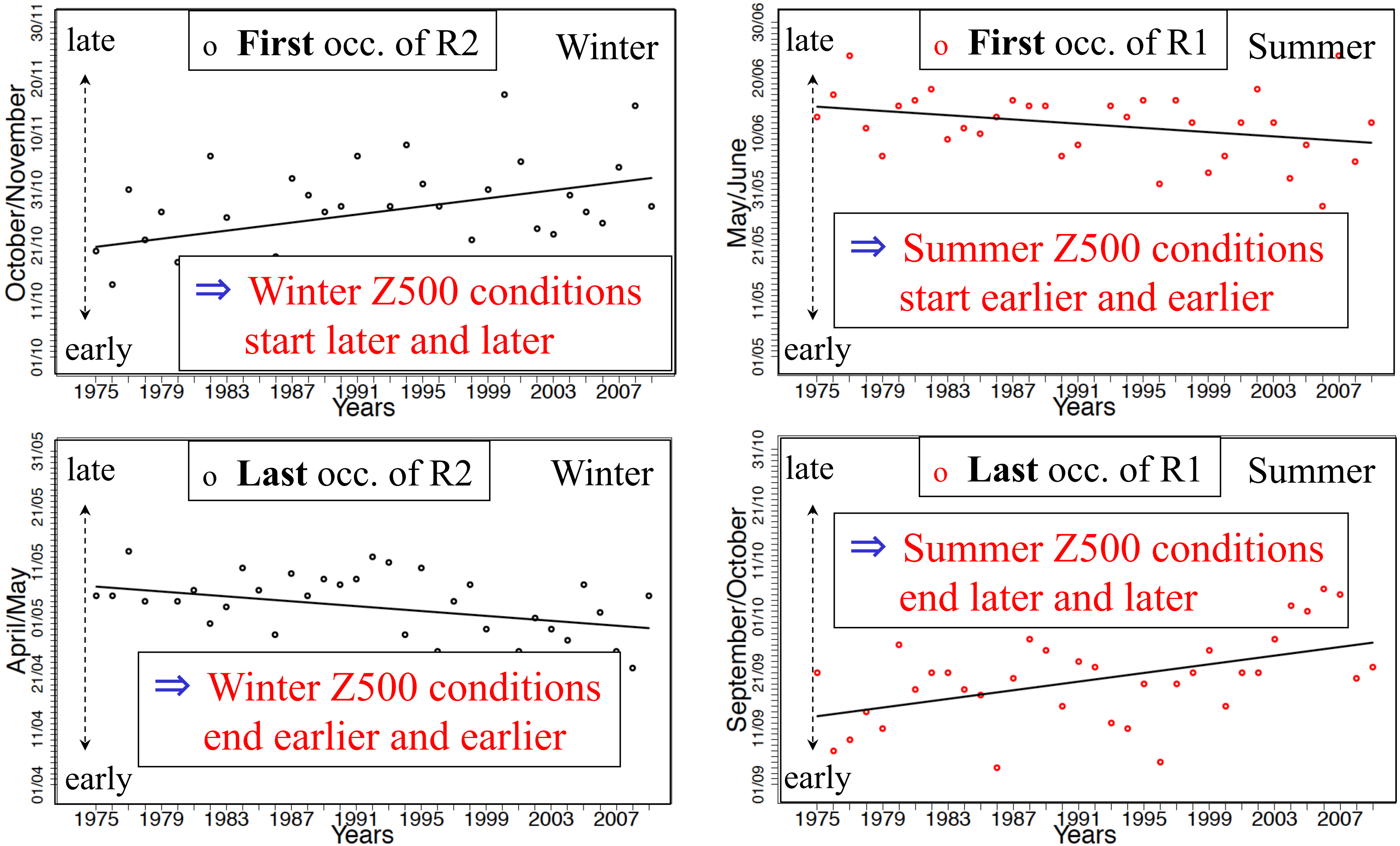
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
WR 1	0.004	0.008									-0.006	
WR 2					3.10^{-4}	0.005	7.10^{-4}	7.10^{-4}	0.007	0.002		
WR 3				0.002		-0.006	-6.10^{-4}	-7.10^{-4}	-0.008	-0.004	0.004	
WR 4				0.003	0.005	0.005				0.003		
WR 5	-0.003	-0.009									8.10^{-4}	-0.002
WR 6				-0.004	-0.004	-0.004					0.004	0.001

Slopes (signif. at 95%) of the frequencies' trends per WR (1975-2009)

- Summer-like conditions are getting more frequent for May – October
- Winter-like frequencies increase in January and February
- Winter-like conditions are fewer and fewer in November

Vrac et al. (2012)

Seasonal WRs: Beginning & End dates



Year-to-year evolution of the first (top row) and last (bottom row) dates of observation (circles) for winter and summer regimes. Evolution of the frequencies for the 2-month periods of beginning and end of the seasons is indicated with triangles.

A few (last?) words on weather regimes

- Simple & efficient way to describe/model main atmospheric variability
- Many studies but still underexploited in GCMs evaluation context
  - Statistical MM brings flexibility, quantitative eval. of uncertainty

And now, what? (some perspectives)

- Heat-waves & cold-spells (modelling & GCM evaluations):
  - Frequency: Occurrence/Persistence
  - Intensity: Spatial structure & variability
- Operational context / climate indices
  - Start/end of upcoming seasons (from seasonal forecasts)
  - Heat-wave early warning (from seasonal/regular forecasts)

Yes, we can (use it) !

- **R packages** developed for <u>clustering and clusters analyses</u>:
  - `GaussDiff` (Rust et al., 2010)
    - ✓ EM clustering
    - ✓ Distances between pdf's
  - `SeasonalRegimes`, in progress (Vrac et al., 2012)
    - ✓ Seasonal cycles of frequencies
    - ✓ Trends significances
    - ✓ Projections for operational context
  - `CCMtools` (Vrac and Yiou, 2010)
    - ✓ Different (original) clustering methods with local-scale info
    - ✓ Evaluations of WRs
    - ✓ Criteria of WRs quality in downscaling applications

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

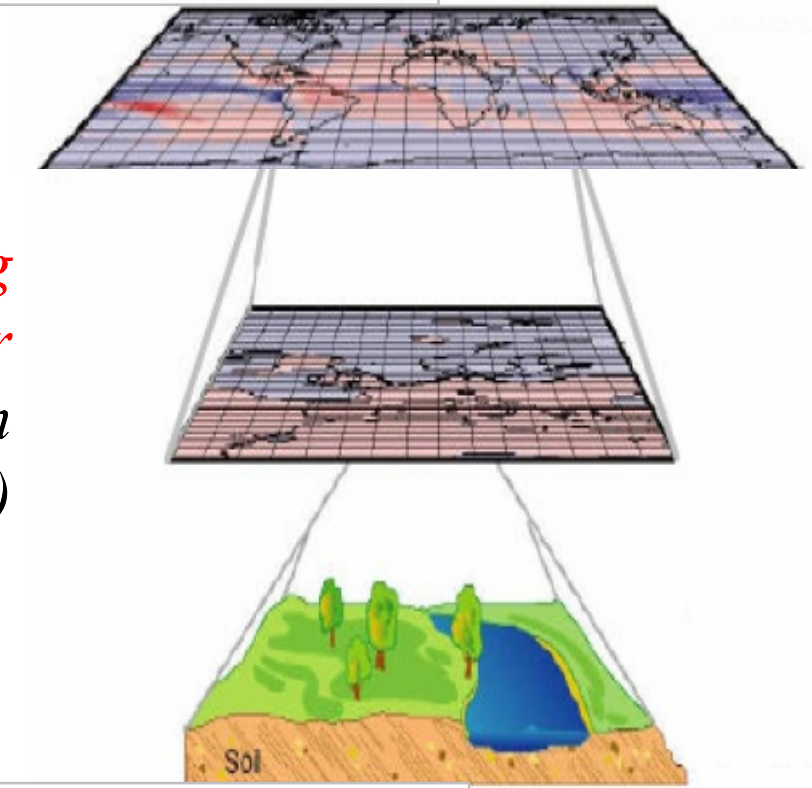
Outline

- Large-scale Weather Regimes
  - WRs through mixture modelling via the EM algorithm
  - "Distance" measures between pdfs associated with regimes
  - Seasonal regimes and evolutions
- Statistical Downscaling (a.k.a. regionalization)
  - What's that? How can we do?
  - Stochastic weather generators => Focus on pdf
- Extreme Events Modelling
  - Main (univariate) notions
  - DS of extreme
- Conclusions and Perspectives

What is downscaling ???

≈ 250 km

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



<u>Definition</u>:

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:

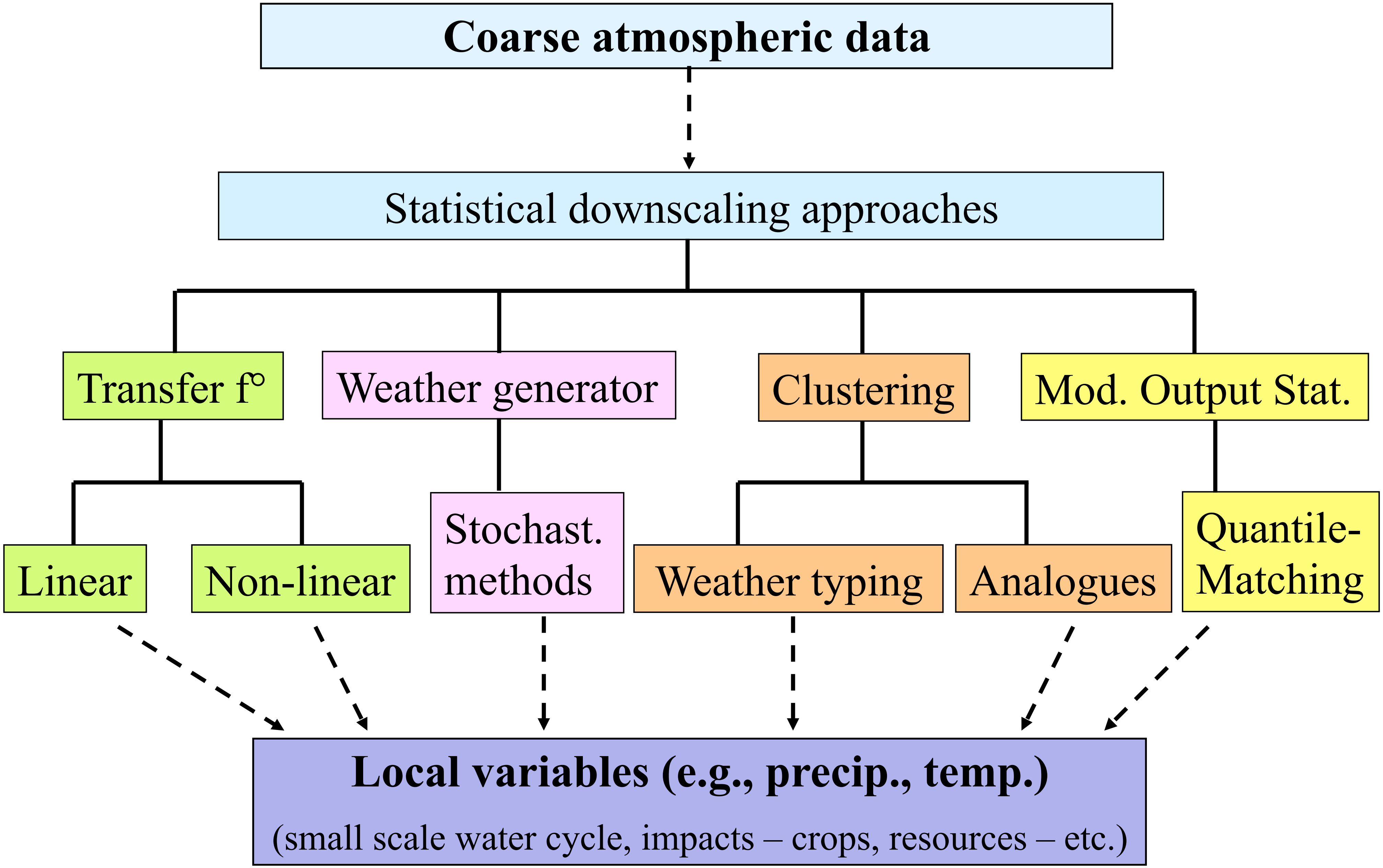
- Based on statistical relationships between large- and local-scale variables
- Low costs and rapid simulations applicable to any spatial resolution
- Uncertainties (results, propagation, etc)

Region, city, fields, station

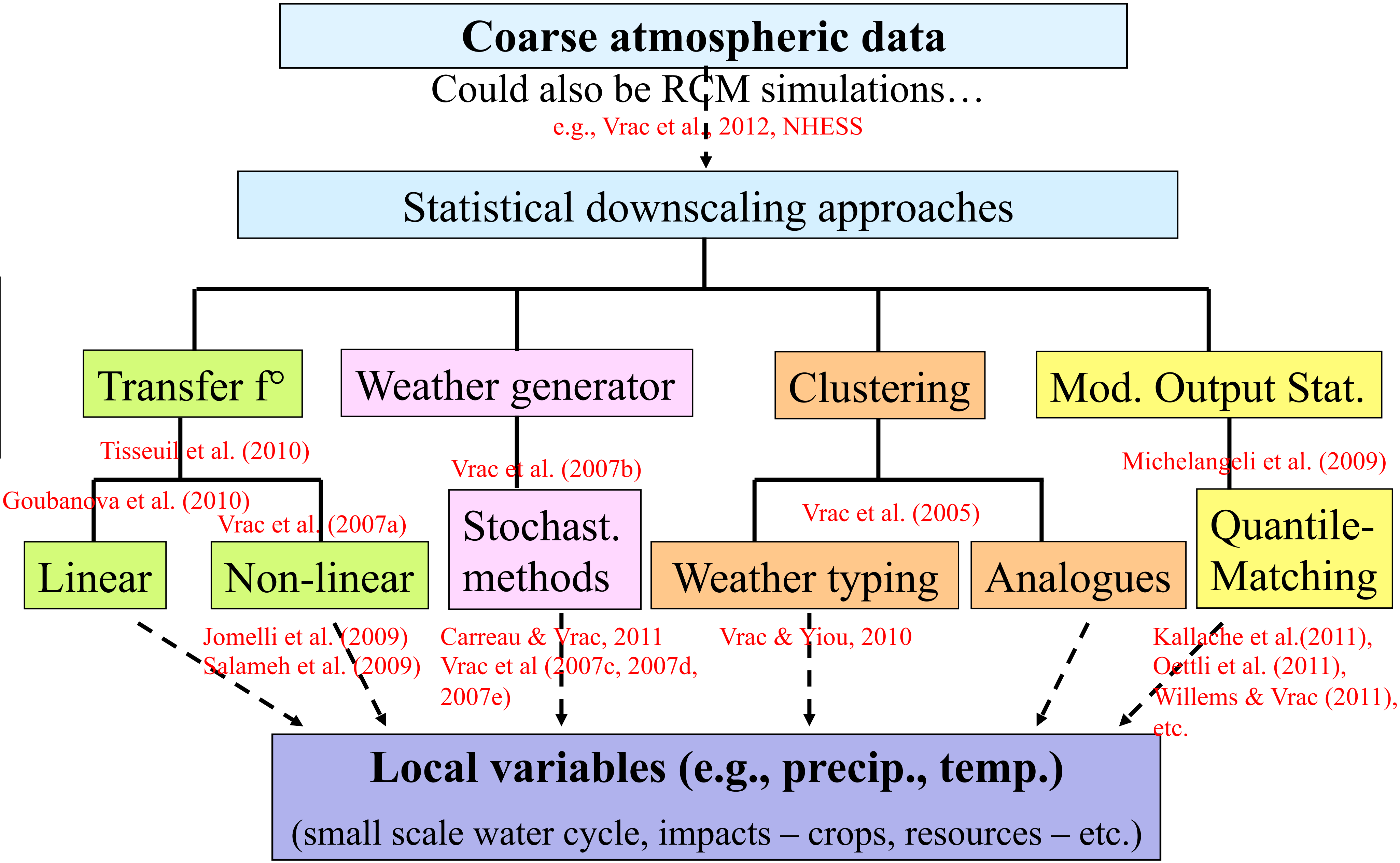
Local variables (e.g., precip., temp.)

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

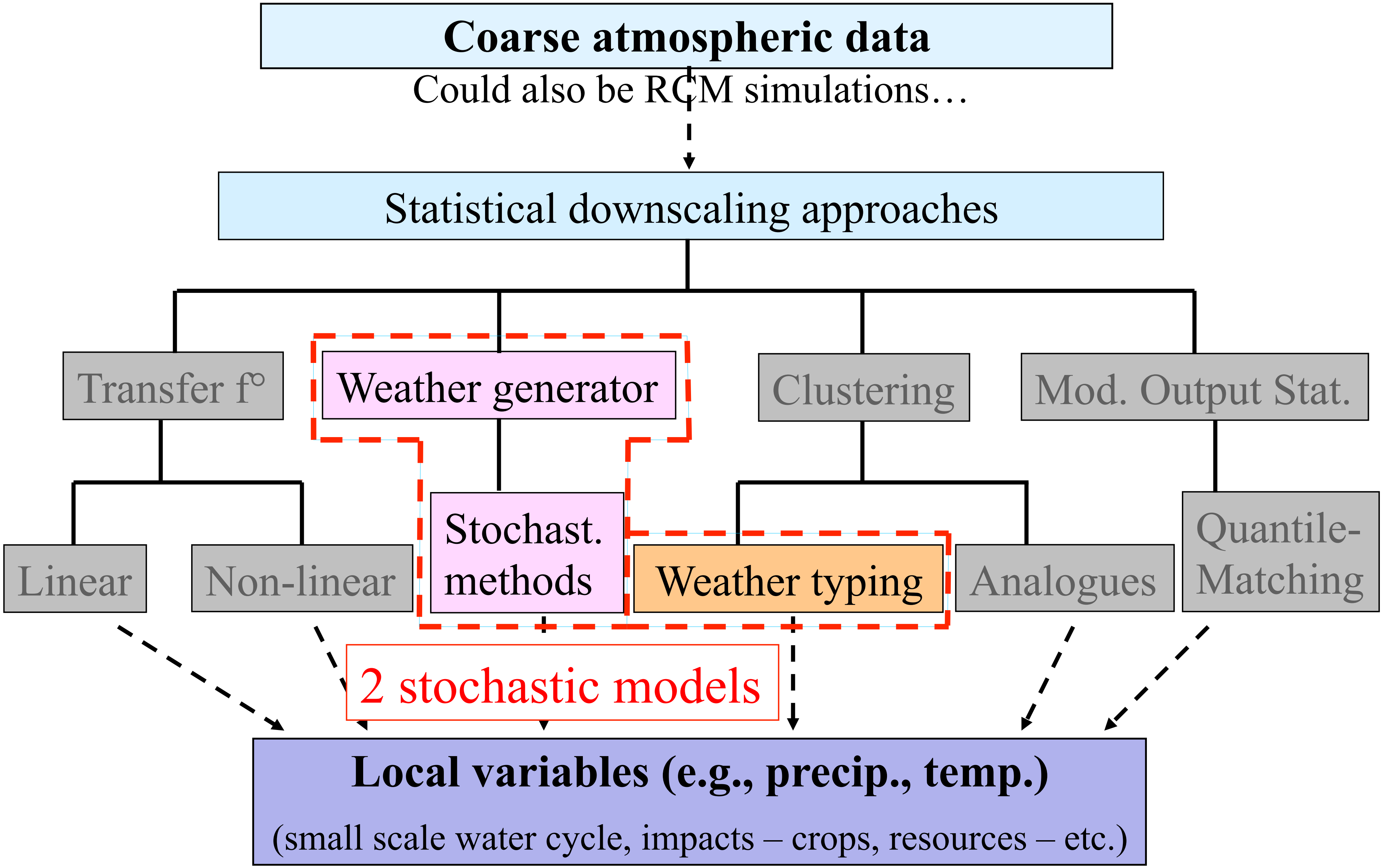
Main statistical approaches



Main statistical approaches



Main statistical approaches



*Non-homogeneous Stochastic Weather Generators for **precipitation downscaling***

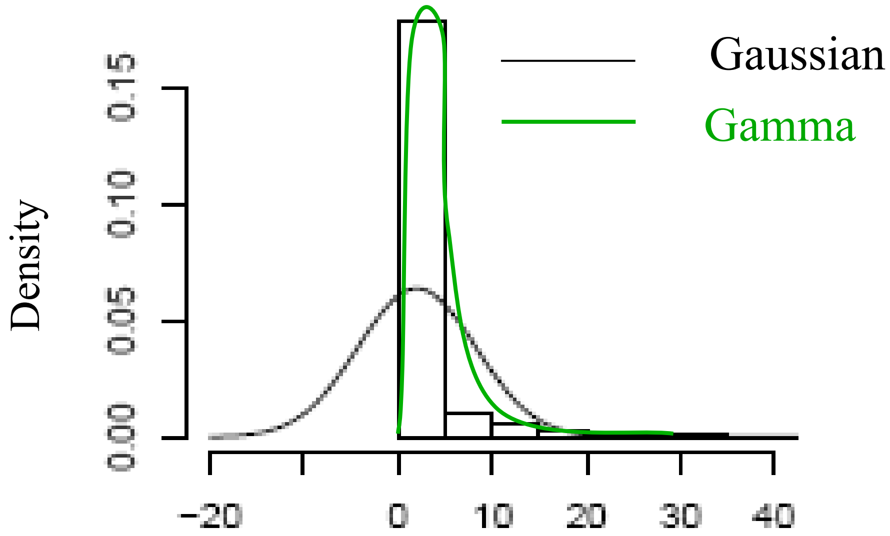
Vrac, Stein, Hayhoe, 2007

Carreau and Vrac, 2011

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)

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

"Non-homogeneous" WGs

- 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)

- Our approach:

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

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

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

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

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

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

NSWT & NN-CMM

Non-homogeneous Stochastic Weather Typing — Neural Network – Conditional Mixture Model

Precipitation pdf (at one station)

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

Parameters are fonctions of (atmospheric, etc.) predictors

NSWT:

$WR_{t-1}(X_{t-1}, WR_{t-2})$ ← X_{t-1} . . .

X_t —NMM→ $WR_t(X_t, WR_{t-1})$

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

$\psi_0(WR_t, st)$

$WR_{t+1}(X_{t+1}, WR_t)$ ← X_{t+1} . . .

NN-CMM:

X_{t-1} —ANN→ . . .

X_t —ANN→ $\alpha(\mathbf{X}_t, st)$, $\psi_0(\mathbf{X}_t, st)$

X_{t+1} —ANN→ . . .

Same philosophy, different implementations

Vrac et al. (2007, Clim.Res.)

Carreau & Vrac (2011, WRR)

The modelling part of **NSWT**

- Prior weather states $(WR_t=k)_{k=1,\ldots,K}$: "*circulation*" or "*precipitation*"
- Fit a Non-homogeneous Markov Model to characterize transition proba.:

$$\mathrm{P}\left(WR_t = l \middle| \mathbf{WR}_1^{t-1}, \mathbf{X}_1^T\right) = \mathrm{P}\left(WR_t = l \middle| WR_{t-1} = k, \mathbf{X}_t\right)$$
$$\propto \gamma_{kl} \exp\left(-\frac{1}{2}(\mathbf{X}_t - \mu_{kl})\Sigma^{-1}(\mathbf{X}_t - \mu_{kl})'\right)$$

- Precipitation probability density function (N stations):

$$\phi_{\mathbf{Y}|WR_t=k,\mathbf{X}_t}(\mathbf{y}) = \prod_{i=1}^{N}\left[\phi\left(y_i;\psi_i\left(k,\mathbf{X}_t\right)\right)\right]$$
$$= \prod_{i=1}^{N}\left[\left(1-\alpha_{ki}\left(\mathbf{X}_t\right)\right)\delta_0\left(y_t^i\right) + \left(\alpha_{ki}\left(\mathbf{X}_t\right)\phi_0\left(y_t^i;\psi_{0,ki}\right)\right)\right]$$

with $\alpha_{ki}\left(\mathbf{X}_t\right) = \dfrac{\exp\left(\mathbf{X}_t{}'\lambda_{ki}\right)}{1+\exp\left(\mathbf{X}_t{}'\lambda_{ki}\right)}$ and ϕ_0 = Gamma pdf

The modelling part of **NN-CMM**

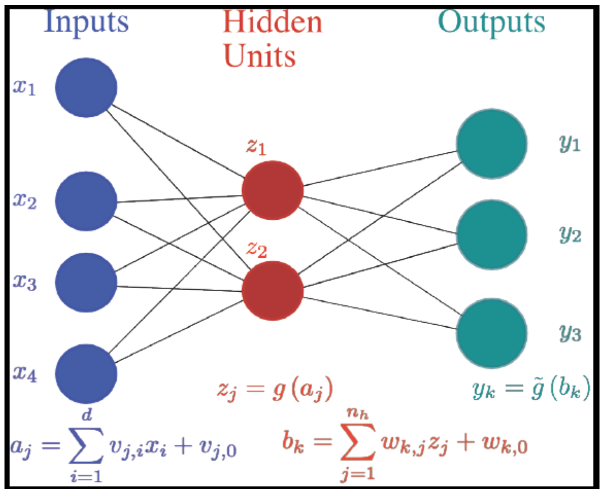
- Precipitation probability density function (N Stations):

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

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

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

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



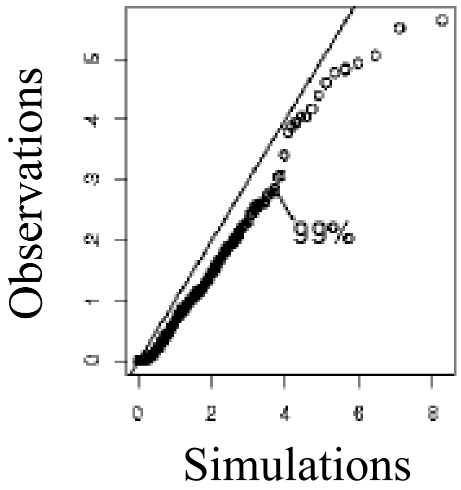
$f = \begin{cases} \text{Gaussian} \\ \text{or} \\ \textbf{Log-Normal} \\ \text{or} \\ \text{Hybrid Pareto} \end{cases}$

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

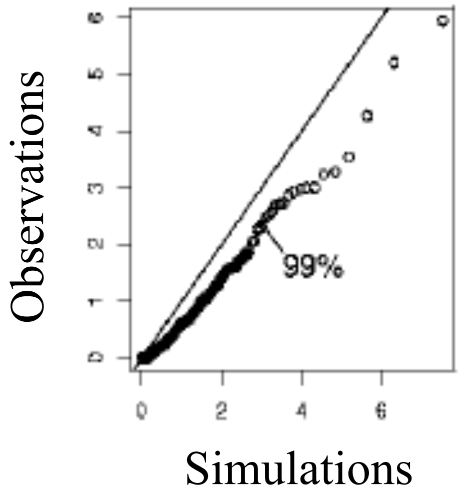
NSWT evaluation: Quantile-Quantile (QQ) plot

Based on **circulation** patterns

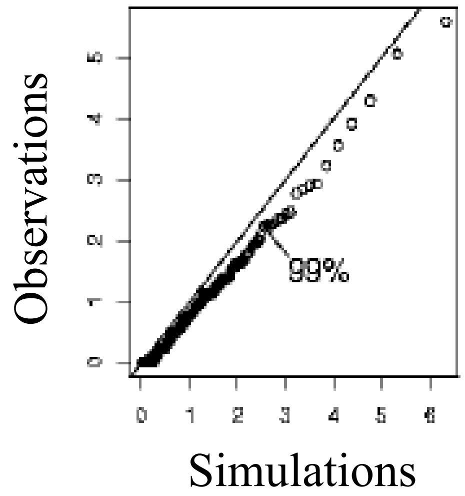
QQplot Charleston



QQplot Galva

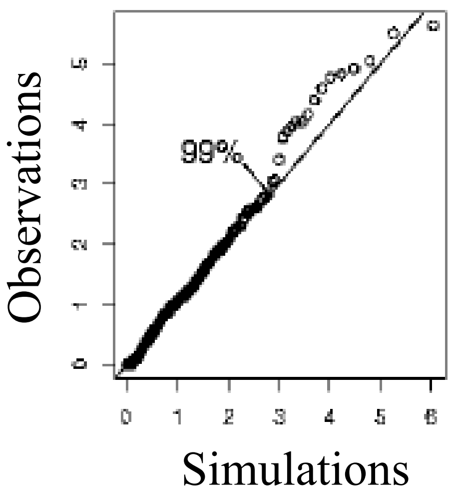


QQplot Walnut

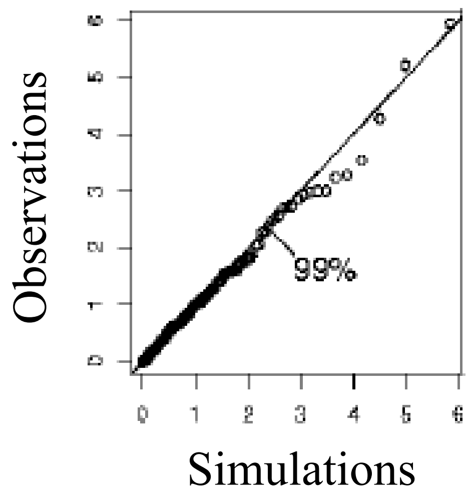


Based on **precipitation** patterns

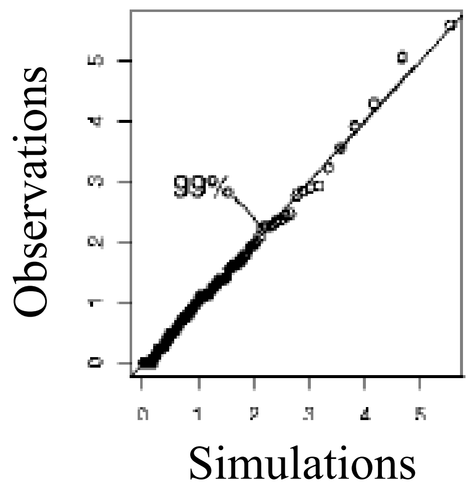
QQplot Charleston



QQplot Galva



QQplot Walnut

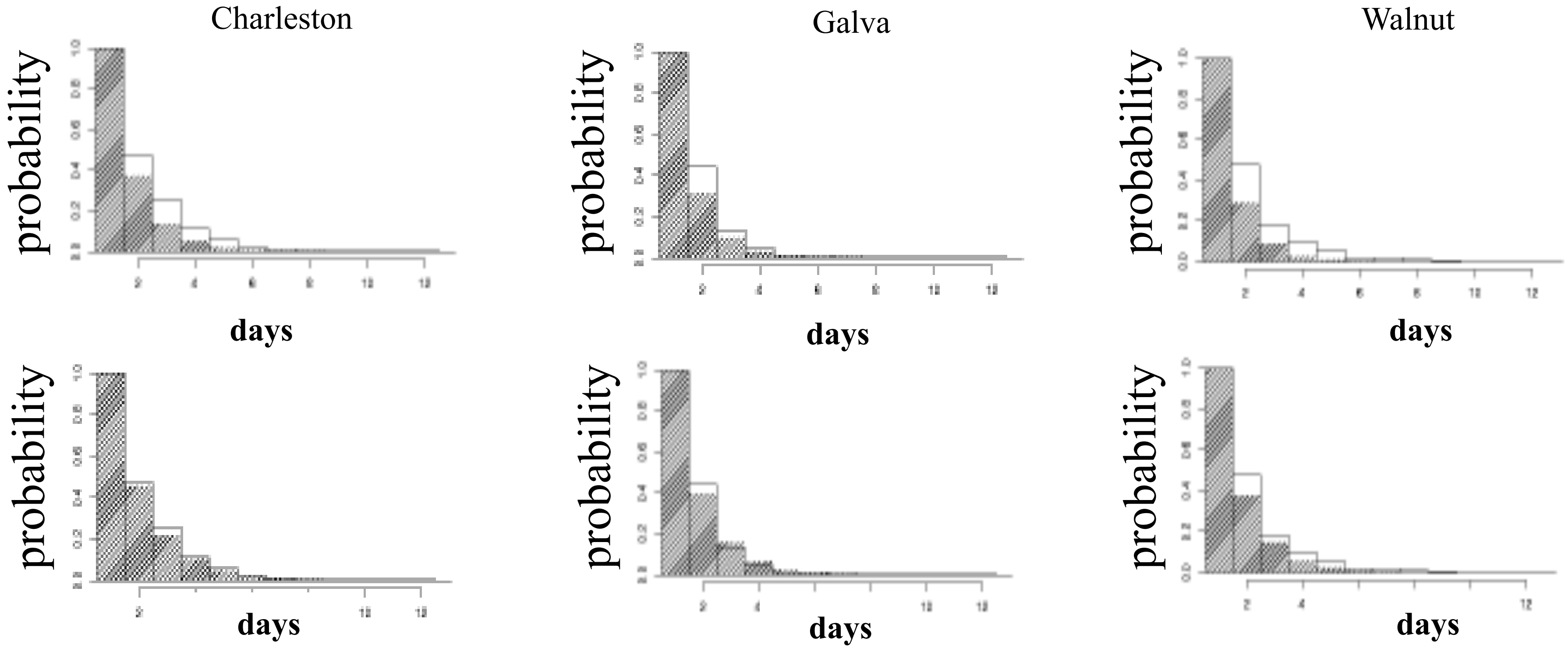


From

NSWT: Wet spells probabilities

Circulation patterns

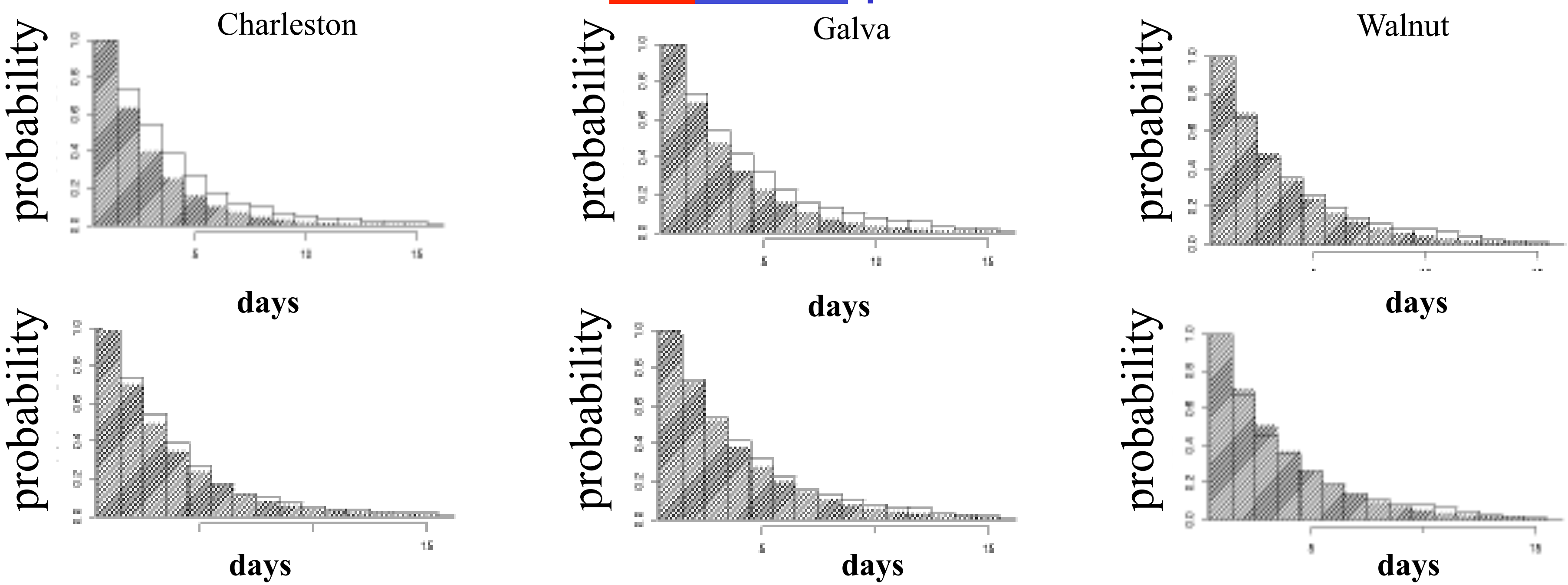
Precipitation patterns



NSWT: Dry spells probabilities

Circulation patterns

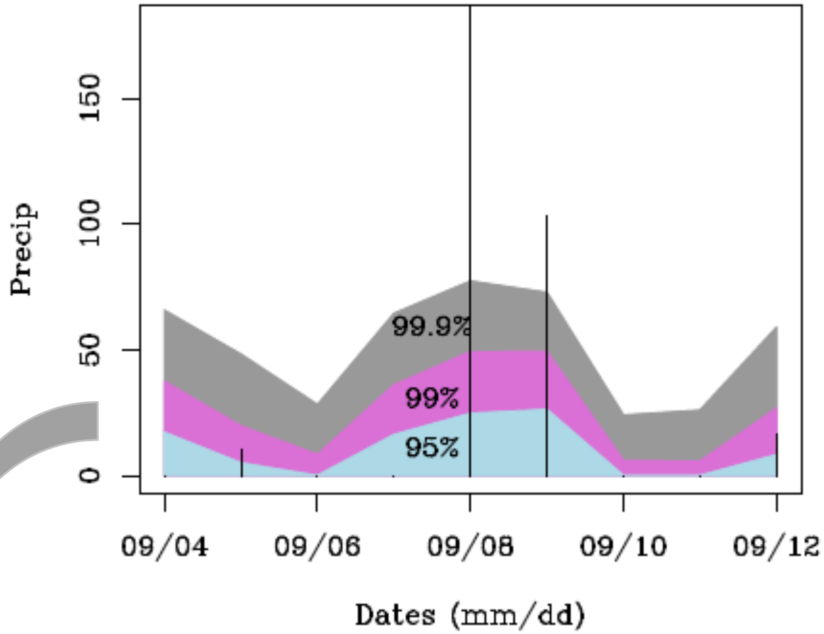
Precipitation patterns



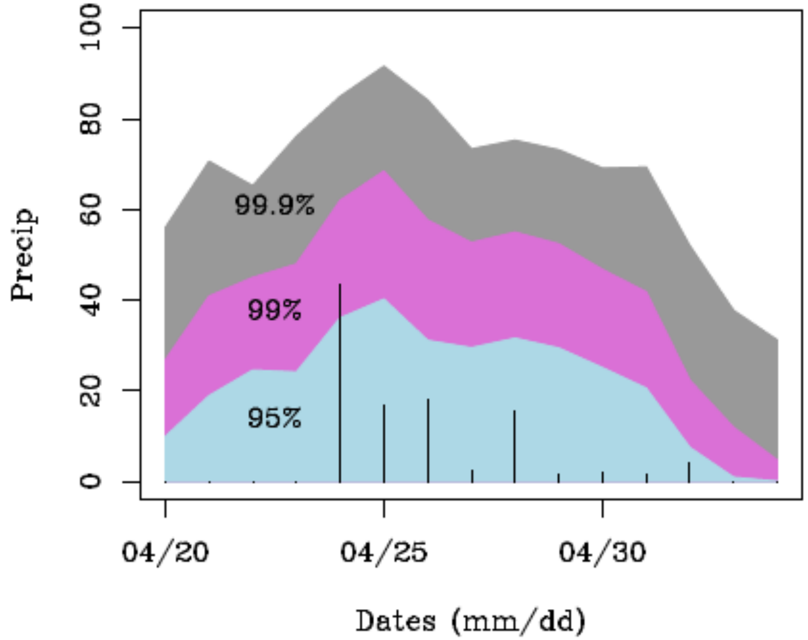
NN-CMM-2L evaluation: Daily pdfs

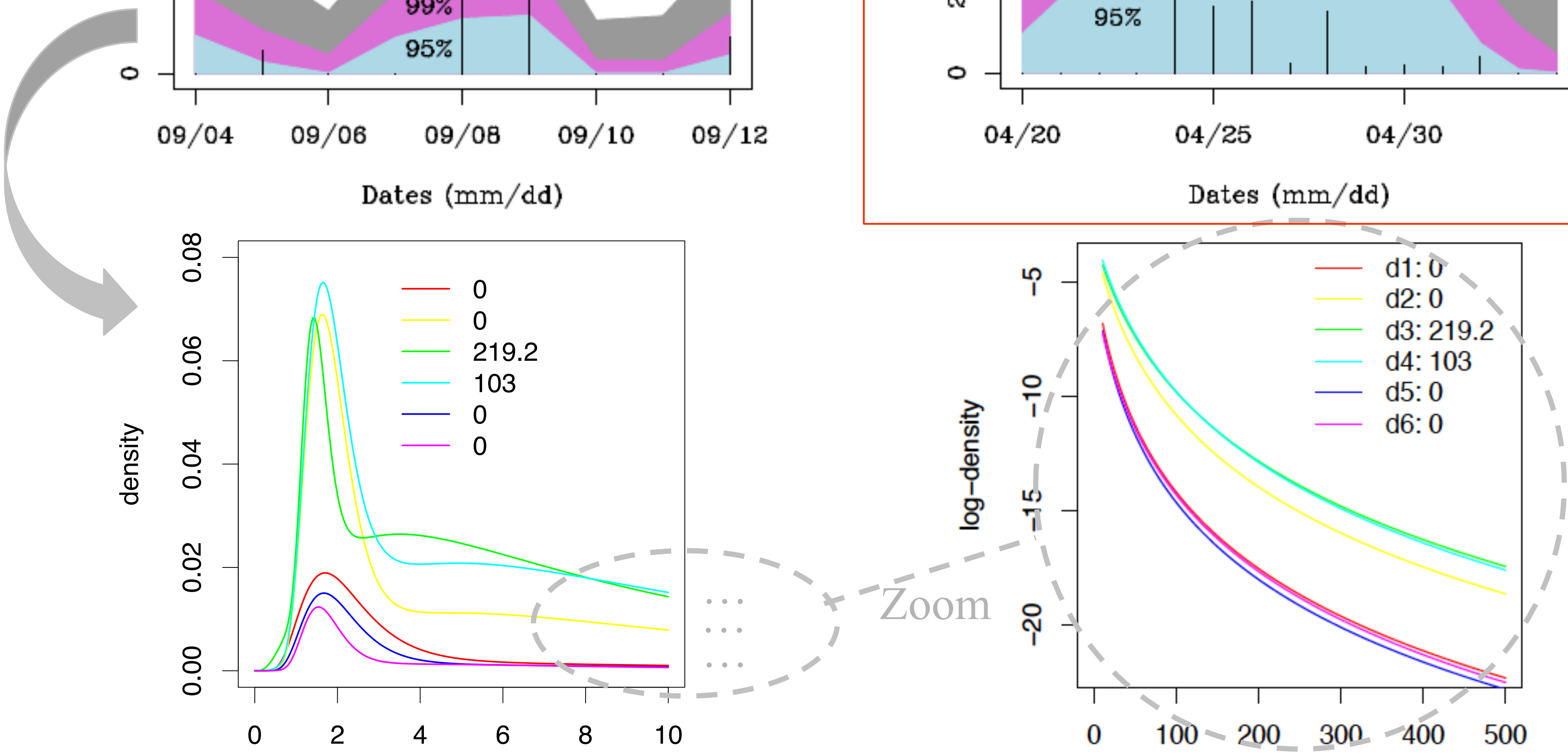
Illustration on the Orange station

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



Spell with the **longest** wet spell

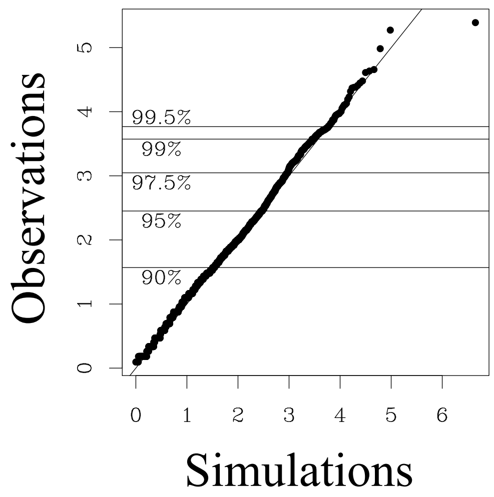




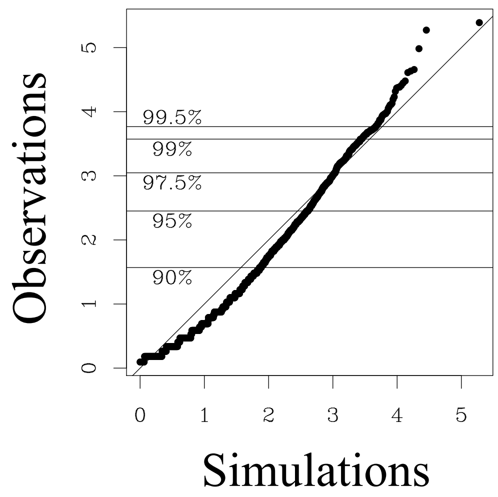
NN-CMM-2L vs. (NN-Cond'l) Gamma

Illustration on the Orange station

QQ-plot (log) CMM-2L



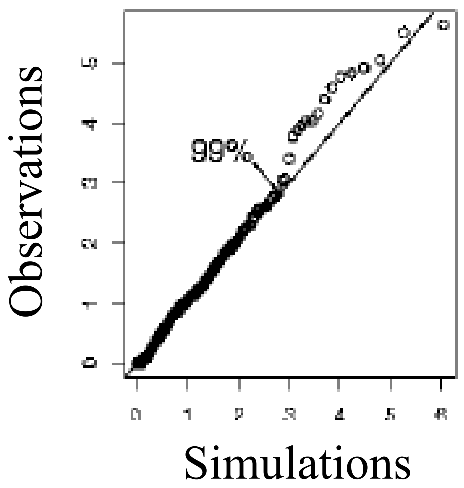
QQ-plot (log) Ber-Gamma



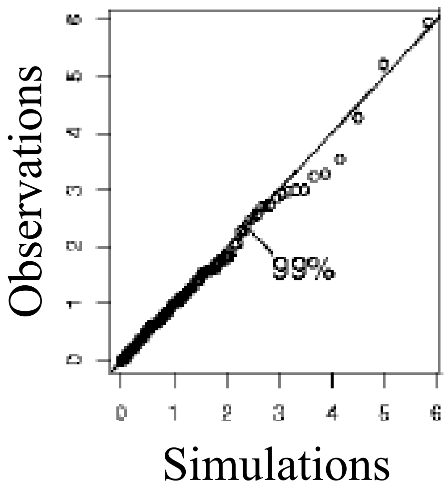
Williams (1998)

NSWT w/ Gamma: good but... not always enough!

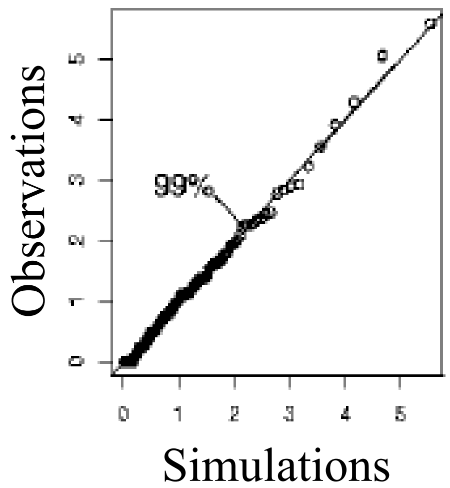
QQplot Charleston



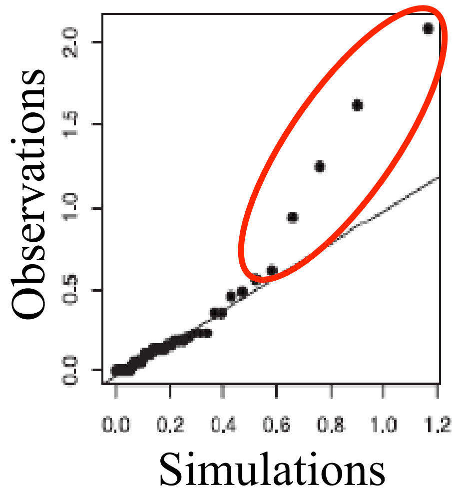
QQplot Galva



QQplot Walnut



QQplot Quincy

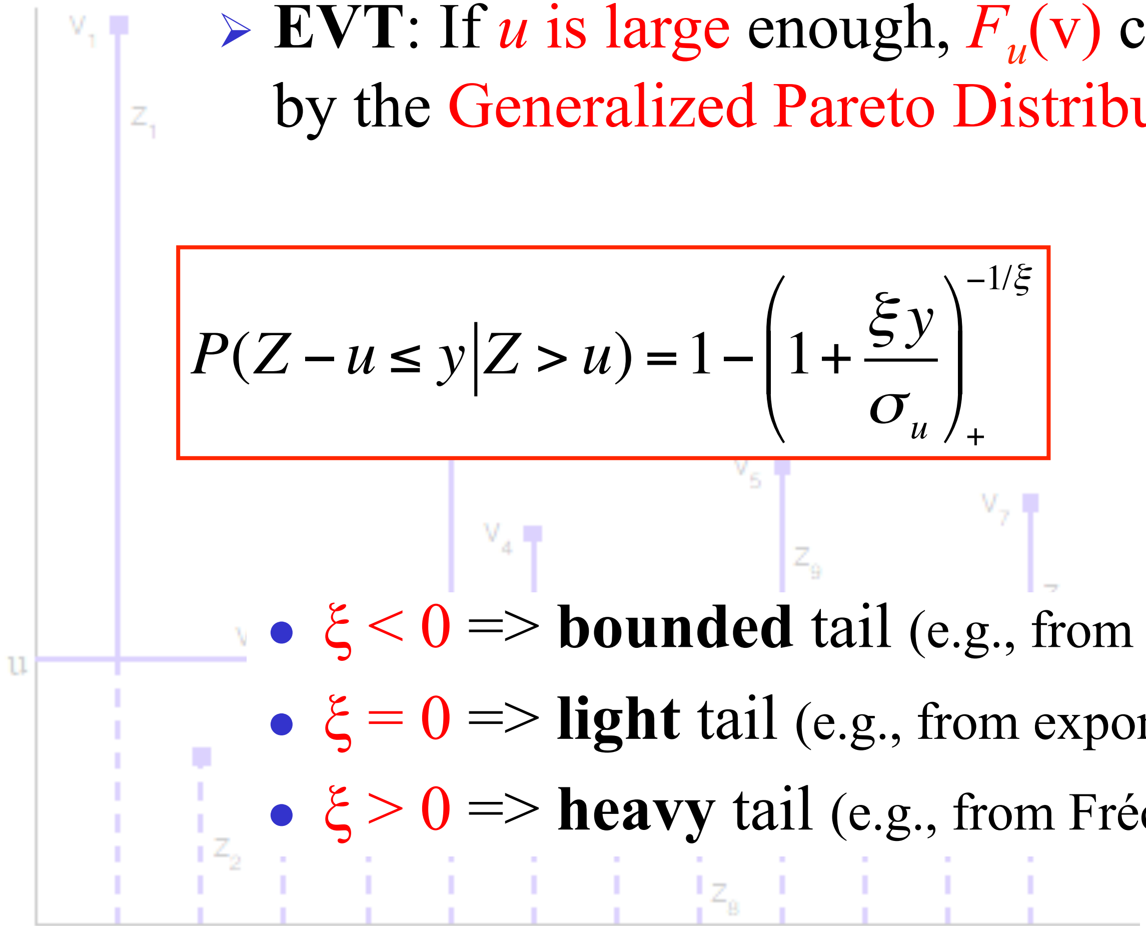


Outline

- Large-scale Weather Regimes
  - WRs through mixture modelling via the EM algorithm
  - "Distance" measures between pdfs associated with regimes
  - Seasonal regimes and evolutions
- Statistical Downscaling (a.k.a. regionalization)
  - What's that? How can we do?
  - Stochastic weather generators => Fo
- **Extreme Events Modelling**
  - Main (univariate) notions
  - DS of extreme
- And now, what? (i.e., Perspectives)

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 $Z\text{-}u$, 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)



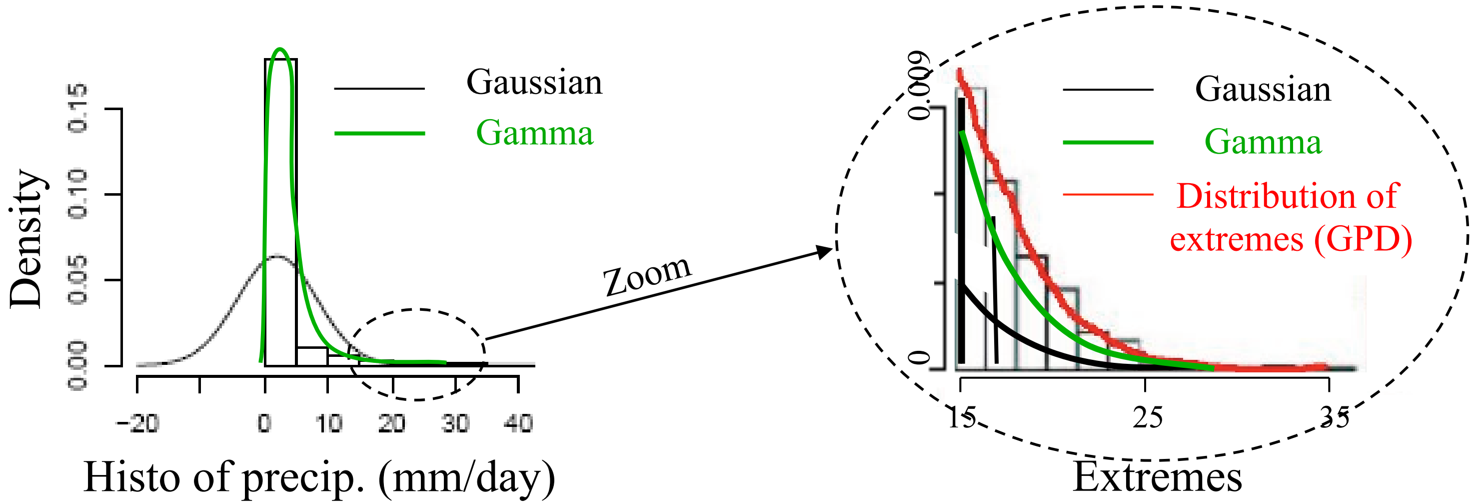
$$\boxed{P(Z-u \le y \mid Z>u) = 1-\left(1+\frac{\xi y}{\sigma_u}\right)_+^{-1/\xi}}$$

- ✓ 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)

Modelling the **whole** precipitation distribution?

- *Classical* distributions – for the "main" values
  - "*Gamma*" or "*log-normal*" distributions
  - Model the most current values, i.e. extremes badly represented



- Extreme Value Distributions – for the extremes
  - "*Generalized Extreme Value*" (**GEV**) distribution for maxima
  - "*Generalized Pareto Distribution*" (**GPD**) for peaks over threshold
  - Do NOT characterize precipitation below a given threshold

Merging classical and EV distributions in NSWT

- 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

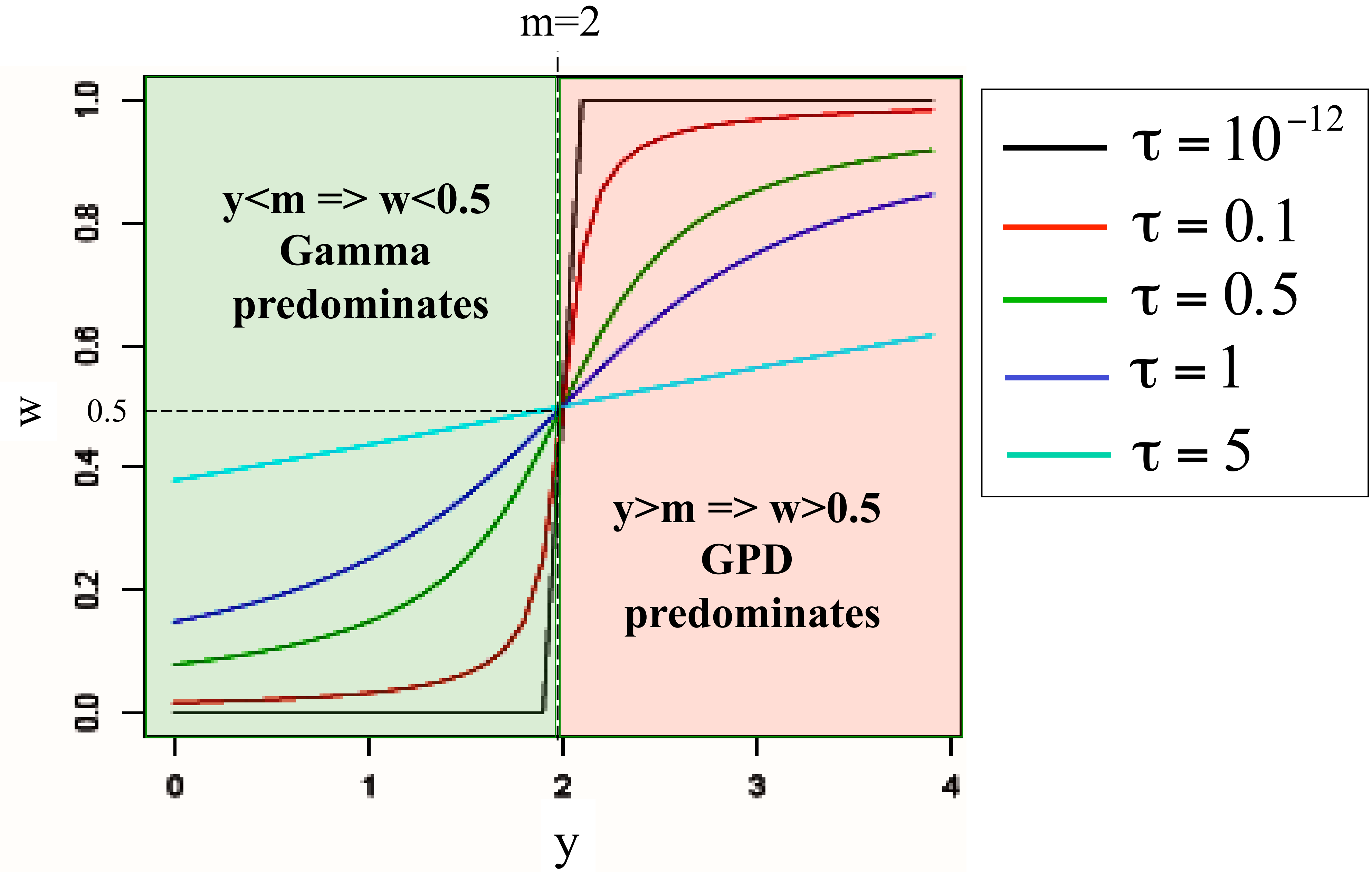
with $$w\left(y \middle| m,\tau\right) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{y-m}{\tau}\right)$$

Value where transition from Γ to GPD

Transition rate

Vrac & Naveau (2007, WRR)

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



Merging classical and EV distributions in NSWT

- Based on Frigessi et al. (2002):

"Dynamic mixture model for unsupervised tail estimation without threshold"

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

6 parameters per station!
⇒ Model selection

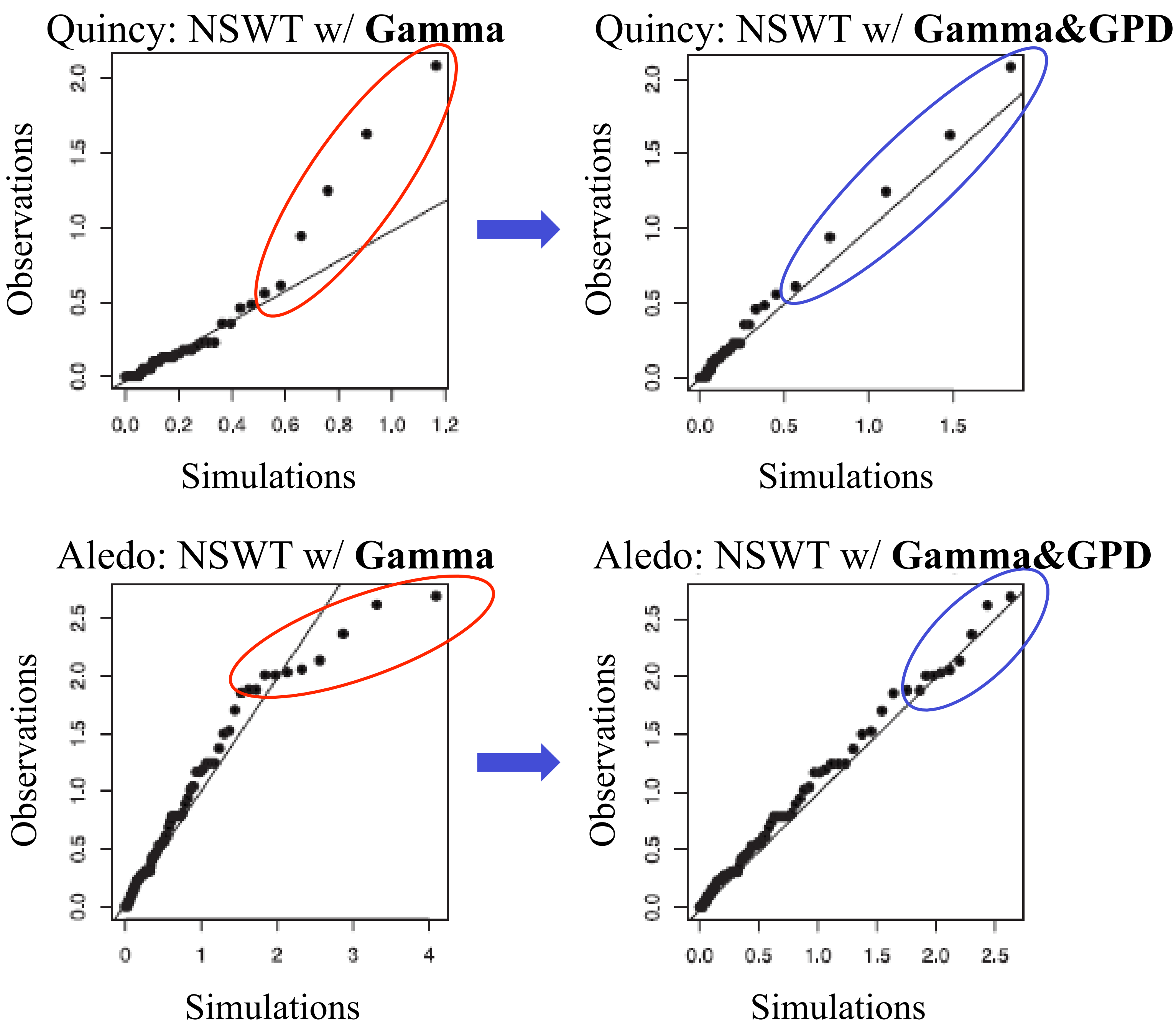
with $$w\left(y\middle|m,\tau\right) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{y-m}{\tau}\right)$$

Value where transition from Γ to GPD

Transition rate

Vrac & Naveau (2007, WRR)

Illustration on two stations



A few (last?) words on Downscaling & Extremes

- **Many** (and many) **applications** of downscaling
  - Past-Present-Future in climate and impacts related studies

  ⇒ **Many** Statistical downscaling **approaches** (still increasing)
  - My favorite ones:
    - ✓ *Stochastic WGs*: cond'l event-wise variability/uncertainty
    - ✓ *Model Output Statistics*: DS of CDFs from CDFs
- **RCMs vs. SDMs**: Not a conflict => complementary approaches
- DS of **extreme** values **vs.** DS of ***indices*** of extremes
- There is not one good SDM for all variables and regions

  ⇒ Different skills according to regions/variables/applications, etc.

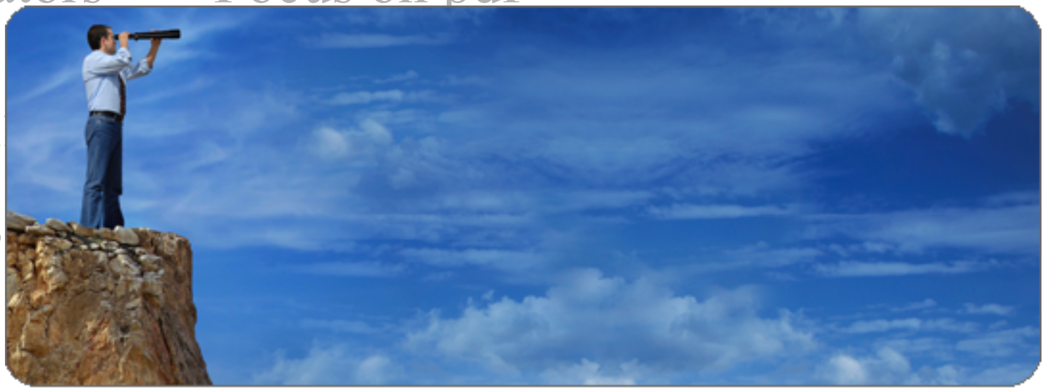
Yes, we can (use it) !

- **R packages** developed for <u>Statistical downscaling & Extremes</u>:
  - `NHMixt` (Vrac & Naveau, 2007)
    - ✓ Statistical mixture model Gamma & GPD
    - ✓ Inclusion of covariates
    - ✓ 2D-extension in progress
  - `CMM`, in progress (Carreau & Vrac et al., 2011)
    - ✓ ANN-Conditional mixture model
    - ✓ Various distributions (Gaussian, Log-N, hybrid Pareto)
  - `CDFt` (Michelangeli et al., 2009, Kallache et al., 2011)
    - ✓ DS of local-scale CDFs from large-scale CDFs
    - ✓ Non-parametric or parametric CDFs (Gaussian, Gamma, GPD)
    - ✓ Inclusion of covariates

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

Outline

- Large-scale Weather Regimes
  - WRs through mixture modelling via the EM algorithm
  - "Distance" measures between pdfs associated with regimes
  - Seasonal regimes and evolutions
- Statistical Downscaling (a.k.a. regionalization)
  - What's that? How can we do?
  - Stochastic weather generators => Focus on pdf
- Extreme Events Modell
  - Main (univariate) notions
  - DS of extreme



- And now, what? (i.e., Perspectives)

Perspectives (DS & extremes)

"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)

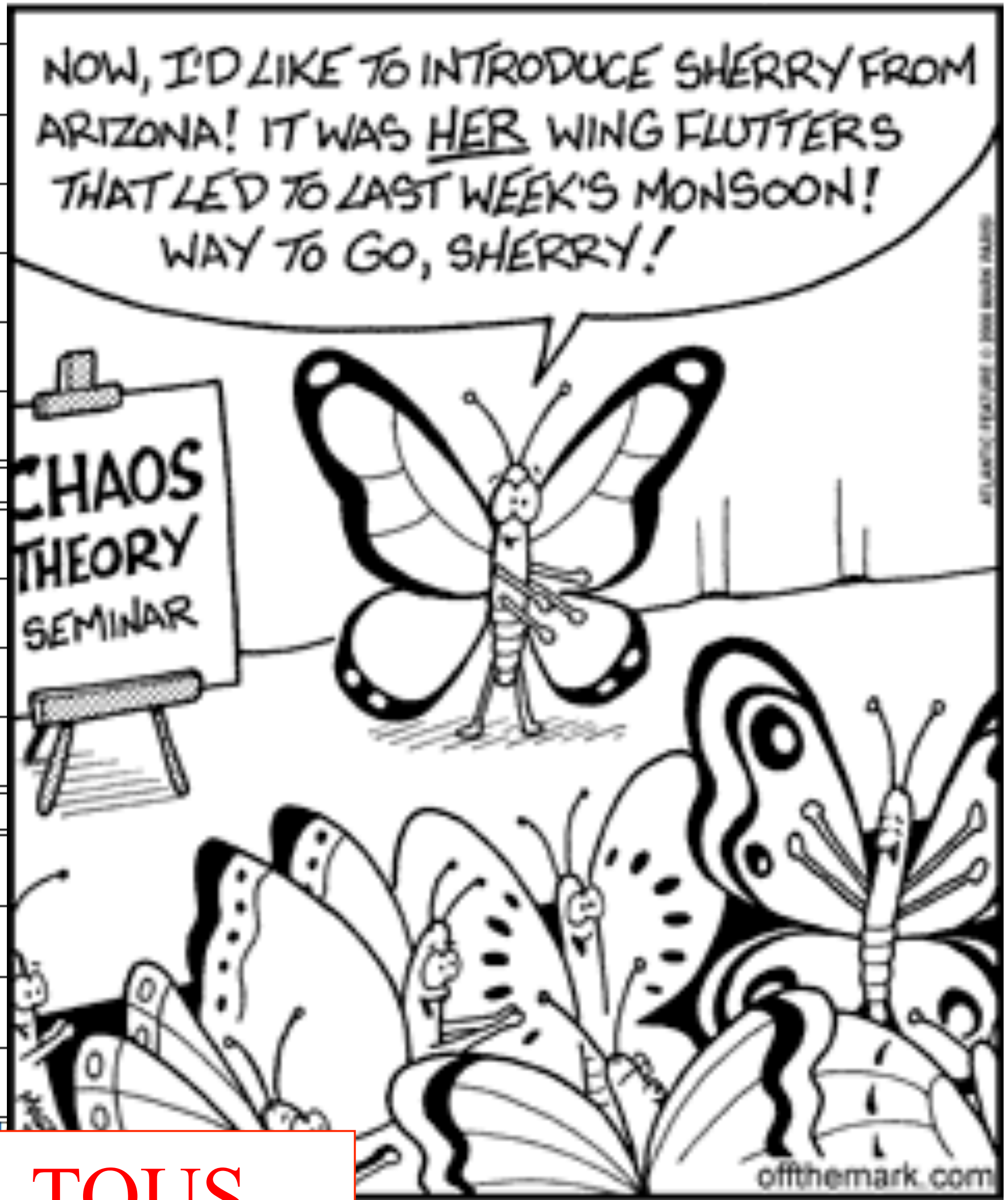
- Only few SDMs are inter-compared:
  - ⇒ **Need more**: between them & with RCMs
    - ✓ Not only "transfer functions" approaches to be tested
      (ANR StaRMIP project (?) inserted in Med-CORDEX)
- Various **evaluations**
  - ⇒ Statistical **indicators of quality** of high-resolution simulations
    - ✓ Spatial & temporal dependences, extremes, inter-var. (COST Value)
  - ⇒ **Uncertainty** modelling (StaRMIP?, TWP3 L-IPSL)
    - ✓ Ensembles of SDMs
    - ✓ "*Model-merging*": (non-) linear, Bayesian, etc., combinations

Perspectives (DS & extremes)

"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 - FP7 RING?)
    - ? MOS ? SWGs?
  - ⇒ Needs for **spatial models**: DS even at locations where no data
    - ? Continuous spatial processes? Inter/Extra-polation of parameters?
  - ➢ Especially for **dependence of extremes** (ANR McSIM)
- R packages: Dissemination / Valorization
  - ✓ A real strength for applications and proper use of developments

Postdocs	Downscaling	Régimes de temps	Extrêmes	Applications ou autres
Thomas Noël (postdoc LSCE/LMD), 2011	Correction de biais			Prec., temp. etc. France (DRIAS)
Maalak Kallache (postdoc LSCE/Climpact), 2009-2010	MNH		[illegible]	[illegible]
Tamara Salameh (postdoc LSCE), 2009-2011	GAM	Attrib./comparaison		
Nicolas Vigaud (postdoc BRGM), 2009-2010	CDF-t			
Armel Martin (postdoc LSCE), 2009-2010	GAM			
Pascal Oettli (postdoc LOCEAN), 2009-2010	CDF-t			
Christophe Lavaysse (postdoc LMD), 2008-2009	CDF-t & RCMs			
Henning Rust (postdoc LSCE), 2008-2009	GLM & MNH	EM & Eval. statistique		
Katerina Goubanova (postdoc LMD), 2008	Fonctions de transfert			
Thésards				
Aurélien Bechler (thésard LSCE-AgroPT) 2011-2014	Modèles spatiaux			
Guillaume Levavasseur (thésard LSCE), 2009-2012	GAM & GLM			
Clément Tisseuil (thésard U.P.S. Toulouse), 2007-2009	ABT, GLM, GAM & ANN	HAC		
Medhi Limam (thésard Dauphine, 2002-2004)		Classification hybride		
Étudiants de Master 1 & 2				
Pradeebane Vaittinada Ayar (Étudiant Master 2), 2010		EM		
Guillaume Levavasseur (Étudiant Master2, 2009)	GAM			
Mohamed Azlim (Étudiant Master 2), 2004		Copules		
Medhi Limam (Étudiant Master2, 2002)		Classification hybride		
Elsa Bernard (Étudiante Master 1), 2011		[illegible]	[illegible]	
Anthony Merlo (Étudiant Master 1), 2010	CDF-t & RCM			
Florian Hechner (Étudiant Magistère 2, 2003)		EM-copules		Simulations numériques



MERCI A TOUS…