A weather generator for simulating multivariate climatic series

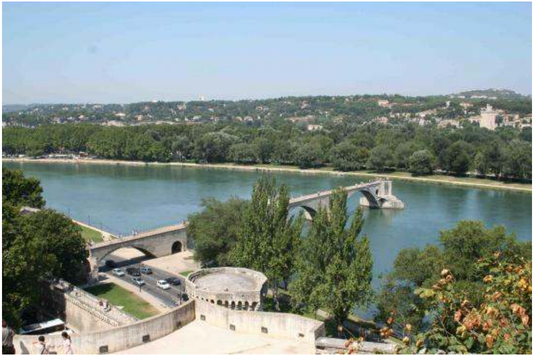
Denis Allard, with
Nadine Brisson (AgroClim, INRA), Cédric Flecher (MetNext) and
Philippe Naveau (LSCE, CNRS)

Biostatistics and Spatial Processes (BioSP)
INRA, Avignon

Modélisation pour l'environnement et expérimentation numérique,
21 Juin 2011, AgroParisTech, Paris

Outline

What is a Weather generator (WG)?



Sur le Pont d'Avignon

What is a Weather generator (WG)?

December 2003

What is a Weather generator (WG)?

March 2010

What is a Weather generator (WG)?

December 2003

Summer

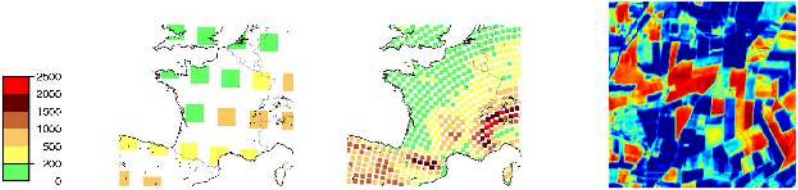
March 2010

- Stochastic model for generating series of daily climatic variables (here, P, Tn, Tx, R, W)
- Calibrated on recorded series

For what purpose?

- Impact studies when climate is involved: e.g. crop models for agriculture
- Explore unmeasured climates
- Disaggregating (downscaling) climatic variables, in particular for GCM outputs (not yet done)

Disaggregation of climatic variables



- Resolution of GCM outputs is at best $\sim$ at a 50 km scale
- But impact studies need climatic variables at the appropriate spatial resulation and time step
- Some plant models need hourly rainfall, which are rarely measured

Outline

Some existing weather generators (WG)

Data-base oriented (non parametric)

- Pros: compatibility between climatic variables is guaranteed
- statistical features are reproduced
- Cons: cannot create unobserved meteorological situations (impact studies !)

Model based (parametric)

- Pros: can create non recorded situations
- Cons: existing WG are limited two a wet/dry classification; non flexible classes of densities

WACS-Gen

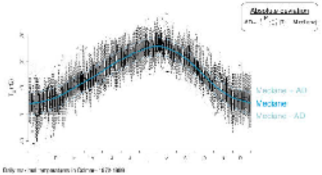
Weather-state Approach Conditionally Skewed- generator

- Parametric, model-based approach
- Accounts for seasonality and inter-annual trend
- Several dry and wet states
- Mixture of multivariate skew-normal densities
- With temporal correlation

PhD Thesis of Cédric Flécher, (co-supervision with Philippe Naveau, LSCE-CNRS, and Nadine Brisson, AgroClim, INRA)

Removing the trend

- For P, Tn, Tx, R, W: build standardized residuals using smoothed median and absolute standard deviations



- Create 4 seasons: MAM, JJA, SON, DJF
- Apply Gamma transform on P for each season

Now work on residuals, for each season independently

Weather States

For each season

- Model-based clustering (Mclust, Fraley & Raftery, 2002) for dry and wet days
  - estimate # states (using BIC)
  - provides a soft classification of days

  Weather states as Markov Chain with transition matrix estimated from soft classification

Multivariate density

For each seaon and each weather state, residuals ($k = 4$ or 5) are distributed according to Complete Skew-Normal distribution

$$f_k(y) = 2^{-k}\phi_k(y;\mu,\Sigma)\Phi_k(S\Sigma^{-1/2}(y-\mu);0,I_k - S^2)$$

- μ vector of location parameter
- Σ is a matrix of co-variation
- S is a diagonal matrix of skewness

Outline

The Closed Skew-Normal (CSN) distribution

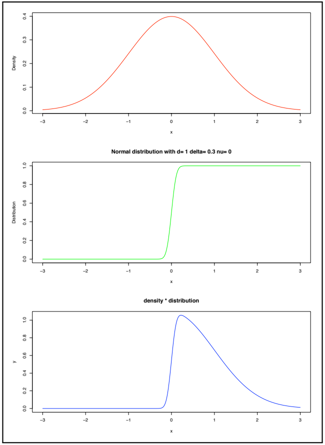
CSN pdf $CSN_{n,m}(\mu, \Sigma, D, \nu, \Delta)$

$$f(y) = \frac{1}{\Phi_m(0; \nu, \Delta + D^t \Sigma D)} \phi_n(y; \mu, \Sigma) \Phi_m(D^t(y - \mu); \nu, \Delta)$$

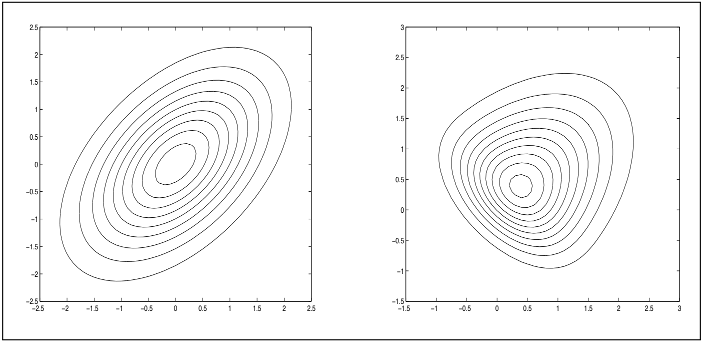
- If $D = 0$: $N_n(\mu, \Sigma)$
- If $m = 1$: skew-normal distribution [Azzalini, 2005; Arellano–Valle, and Azzalini, 2008; Dominguez-Molina and gonzales-Farias, 2003]

Example

$$m = n = 1; \mu = 0, \sigma^2 = 1, d = 1, \nu = 0.3, \Delta = 0.3$$



Gaussian and CSN bivariate density



Some properties of CSN distributions

Linearity

$$A \times CSN_{n,m}(\mu, \Sigma, D, \nu, \Delta) \sim CSN_{r,m}(A\mu, \Sigma_A, D_A, \nu, \Delta_A)$$

where

$$\Sigma_A = A\Sigma A^T, \quad D_A = D\Sigma A^T \Sigma_A^{-1}, \quad \Delta_A = \Delta + D\Sigma D^T - D_A \Sigma_A D_A^T$$

Sum (particular case)

$$N(\mu, \Sigma) + CSN_{n,m}(\psi, \Omega, D, \nu, \Delta) \sim CSN_{n,m}(\psi + \mu, \Omega + \Sigma, D, \nu, \Delta)$$

Some properties of CSN distributions

Linearity

$$A \times CSN_{n,m}(\mu, \Sigma, D, \nu, \Delta) \sim CSN_{r,m}(A\mu, \Sigma_A, D_A, \nu, \Delta_A)$$

where

$$\Sigma_A = A\Sigma A^T, \quad D_A = D\Sigma A^T \Sigma_A^{-1}, \quad \Delta_A = \Delta + D\Sigma D^T - D_A \Sigma_A D_A^T$$

Sum (particular case)

$$N(\mu, \Sigma) + CSN_{n,m}(\psi, \Omega, D, \nu, \Delta) \sim CSN_{n,m}(\psi + \mu, \Omega + \Sigma, D, \nu, \Delta)$$

Some properties of CSN distributions

Conditioning

Consider $Y = (Y_1, Y_2) \sim CSN_{n,m}(\mu, \Sigma, D, \nu, \Delta)$.
Then, $Y_2 | Y_1 = y_1$ is

$$CSN(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(y_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}, D_2, \nu - D_1 y_1, \Delta)$$

Moment generating function

$$M(t) = \frac{\Phi_m(D^t\Sigma t; \nu, \Delta + D\Sigma D^T)}{\Phi_m(0; \nu, \Delta + D\Sigma D^T)} \exp\{\mu^T t + \frac{1}{2}(t^T \Sigma t)\}$$

From the MGF one can derive first and second moments

Some properties of CSN distributions

Conditioning

Consider $Y = (Y_1, Y_2) \sim CSN_{n,m}(\mu, \Sigma, D, \nu, \Delta)$.
Then, $Y_2 | Y_1 = y_1$ is

$$CSN(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(y_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}, D_2, \nu - D_1 y_1, \Delta)$$

Moment generating function

$$M(t) = \frac{\Phi_m(D^t \Sigma t; \nu, \Delta + D\Sigma D^T)}{\Phi_m(0; \nu, \Delta + D\Sigma D^T)} \exp\{\mu^T t + \frac{1}{2}(t^T \Sigma t)\}$$

From the MGF one can derive first and second moments

Hierarchical construction

$$\begin{pmatrix} Y \\ X \end{pmatrix} \sim N_{n+m}\left(\begin{pmatrix} 0 \\ \nu \end{pmatrix}, \begin{pmatrix} \Sigma & -D^t\Sigma \\ -\Sigma D & \Delta + D^t\Sigma D \end{pmatrix}\right),$$

Then

$$\mu + (Y|X \leq 0) = CSN_{n,m}(\mu, \Sigma, D, \nu, \Delta)$$

Simulation algorithm

1. simulate a vector $X \sim N_m(\nu, \Delta + D^t\Sigma D)$, conditional on $X \leq 0$
2. simulate a vector Y conditionally on X, according to the bivariate model above
3. return $\mu + Y$

Outline

CSN for WACS-gen

To simplify the model, we set

- $m = n = k$
- $D = \Sigma^{-\frac{1}{2}} S$
- $\Delta = I_k - S^2$
- $S = \text{diag}(\delta_1, \ldots, \delta_K)^t$.

$CSN_k^*(\mu, \Sigma, S)$:

$$f_k(y) = 2^{-k} \phi_k(y; \mu, \Sigma) \Phi_k(S\Sigma^{-1/2}(y - \mu); 0, I_k - S^2)$$

Hence

$$\tilde{X} = \Sigma^{-1/2}(X - \mu) \sim CSN_5^*(0, I_5, S)$$

Estimation of the parameters

Maximum likelihood is known to be difficult and non-robust
Estimation is done by weighted moments

Using as third moment the quatity $E[\Phi_k(Y, 0, I_k)]$ with

$$E[\Phi_k(Y,0,I_k)] = 2^k \Phi_{2k}\left(0; \begin{bmatrix} -mu \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma + I_k & \lambda\Sigma^{1/2} \\ \lambda\Sigma^{1/2} & I_k \end{bmatrix}\right)$$

[Flecher, Allard and Naveau (2009) *Stat. Prob. Letters*]

Estimation workflow

- Remove inter-annual and seasonal trend for Tn, Tx, R, W
- For each season $\{$
    1. Gamma transform P
    2. Mclust classification of weather states (WS)
    3. Estimation of transition matrix
    4. For each WS, estimation of the CSN parameters
    5. Estimation of temporal correlation

  $\}$

Simulation algorithm

- Simulate the Markov Chain WS(t)
- Fort $t = 1, \ldots, T$ $\{$
  1. Conditionnally on WS simulate $\tilde{X}(t) \sim$ CSN given $\tilde{X}(t-1)$
  2. Transform $\tilde{X}(t)$ into $X(t)$
- Add seasonal and interannual trend

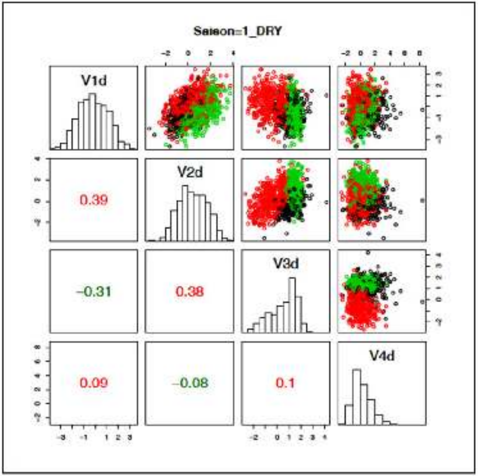
Outline

Illustration

30 year series in Colmar, France

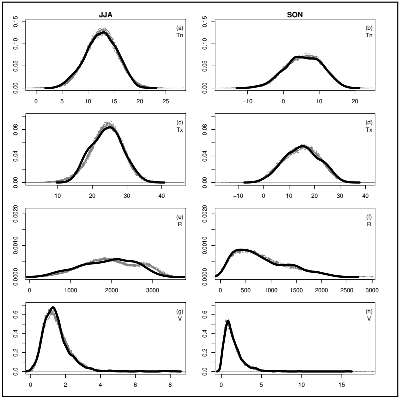
- Well marked seasonal cycle: cold winters ($\bar{T}n = 2^\circ$C in Jan.) and warm summers ($\bar{T}x = 25^\circ$C in Jul.)
- $\approx$ 1/4 wet days
- Average P is 530 mm/year
- 3 wet WS / season (exc. JJA)
- 3 dry WS / season

Classification of residuals



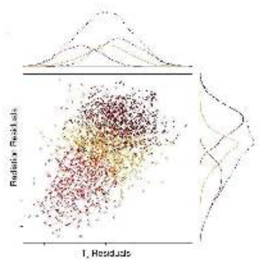
JJA dry days

Marginal densities



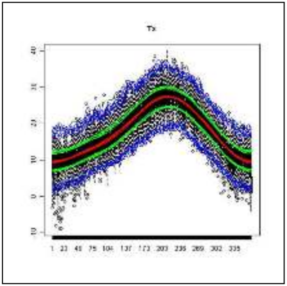
JJA and SON, dry days

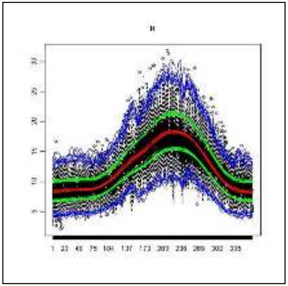
Marginal densities: pair (Tx, R)



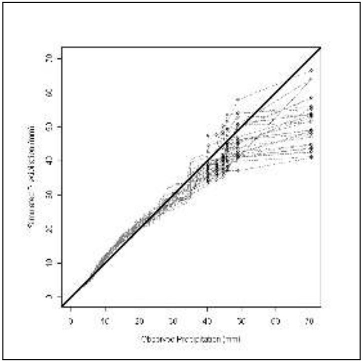
JJA dry days

More general picture: *Tx* and *H*

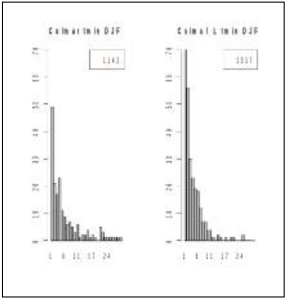


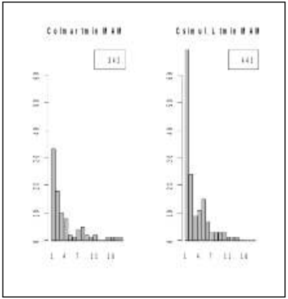


Precipitation

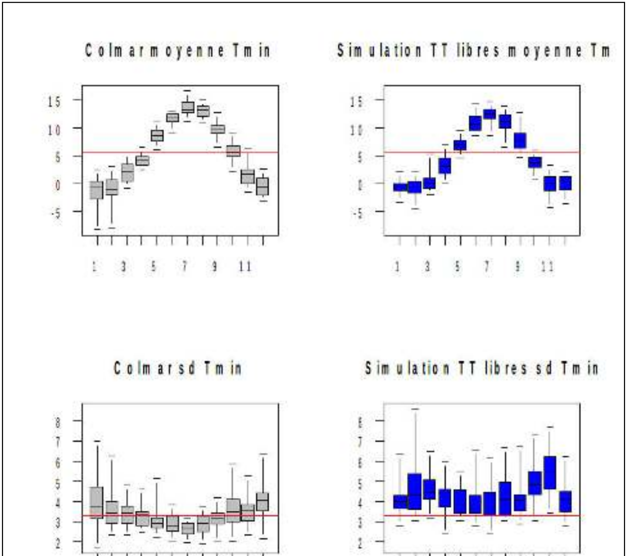


Persistence of $Tn < 0$





Variability



In conclusion

WACS-gen is a flexible weather generator which

- overcomes existing limitations of previous WG
  - beyond dry/wet days
  - skewed densities able to capture natural asymetries of climatic variables
  - good variability of monthly averages
- But still problems
  - non robust estimators for skewness parameter
  - too many free parameters?

[Flecher, Naveau, Allard, Brisson (2010) A stochastic daily weather generator for skewed data, *Water Resources Research*, **46**, W07519]