Stochastic Weather Generators: an overview

Denis Allard

Biostatistics and Spatial Processes (BioSP), MIA division, INRA
Avignon, France

Statistical and mathematical tools for the study of climate extremes,
Cargèse, 9th-13th November 2015

Definition

Adapted from (Yiou, 2014)

SWG are tools that generate random series of meteorological variables such as precipitation, temperature, wind speed, etc., with statistics similar to those of recorded data:

- Mean, variance, quantiles
- Skewness, extremes
- Covariance (dependence) between variables
- Temporal dependence / coherence (persistence)
- Spatial dependence / coherence

- Calibrated on recorded series
- Computational efficiency $\Rightarrow$ long series and/or large number of realizations

SWGs are not climate models

From (Ailliot et al., 2015)

- Driven by data
  - GCMs are driven by physics
  - SWGs use statistical/algorithmic approaches on recorded data
- Focus on local scale
  - GCMs are global numerical models, on very large grids
  - SWGs focus on small spatial scales: one or few sites over a limited region
- Computation speed
  - Inclusiveness of GCMs imply costly computations, hence very few runs
  - SWGs are cheap to compute $\Rightarrow$ long series and/or large number of realizations

SWGs are complementary to GCMs, focused on local weather patterns and fast reproduction

For what purpose?

Used in impact studies

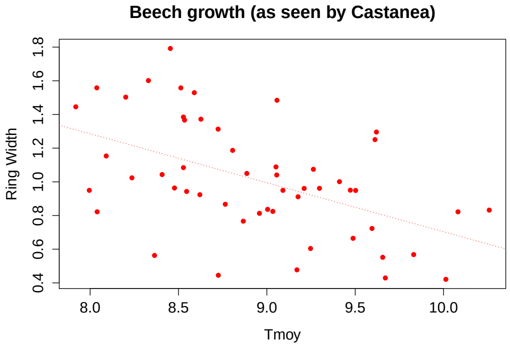
Outputs of SWGs are used as inputs in process-based models, e.g. energy demand models, crop models, hydrological models, insurance models, ...

- Assessing complex, non linear, responses to climate in agro-ecological systems
- Explore unmeasured climates
- Explore plant / ecosystem models as functions of climate variability
- Optimal decision under uncertainty: simulate up to year $t + k$, optimize decision
- Disaggregating (downscaling) meteorological variables from GCM outputs

Example: Beech forest on the Ventoux Massif

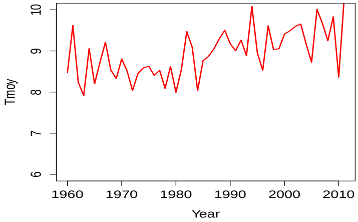
- Beech is one the major tree species in France
- Southern limit of its range is on the Ventoux Massif
- Sensitive to drought
- Growth expected to decrease due to climate change

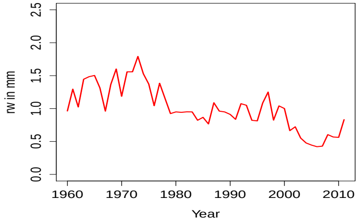
Example: Beech forest on the Ventoux Massif
Can we asses the effect of Climate Change on Ring Width?



- Ring width vs. annual mean Temperature, as given by Castanea (Le Dantec, 2000; Davi, 2000)
- Castanea is an eco-physiological model, calculating photosynthesis and transpiration on an hourly basis, and C and water balance on a daily basis, for a stand

Example: Beech forest on the Ventoux Massif
Can we asses the effect of Climate Change on Ring Width?





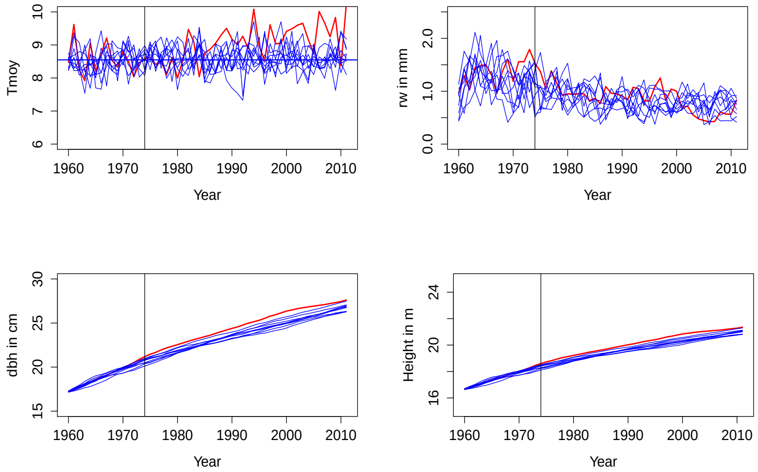
Can we separate the effect of age and the effect of climate change?

Use WACSgen !

- Estimate parameters on 1960-1974 data
- Generate 10 simulated series 1960-2012 from those
- Compare simulated series to measured one in Castanea

INRA SCIENCE & IMPACT département mia Biostatistique BIO/Π & Processus Spatiaux

Beech forest on the Ventoux Massif
Can we asses the effect of Climate Change on Ring Width?



Typology

Model based (parametric)

- (?) appropriateness/robustness of the statistical model
- $(-)$ temporal and spatial coherence difficult to capture
- $(+)$ can create non recorded situations
- $(+)$ can simulate more extreme conditions than those observed
- $(+)$ identification through a set of parameters $\Rightarrow$ sensitivity analysis

Resampling / analogs (non parametric)

- $(+)$ compatibility between climatic variables is guaranteed
- $(+)$ statistical features are reproduced by construction
- $(+)$ temporal and spatial coherence also
- $(-)$ cannot create unobserved meteorological situations
- $(-)$ Implicit assumption: the most extreme observation has been observed

Typology of parametric SWGs

Parametric SWGs

- A statistical model + simulation method, translated into a computer code.
- There are three main types of parametric SWGs (Ailliot et al., 2015)

- ARMA models, based on Box-Jenkins methodology, also called Richardson-Type SWGs

$$\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t-1} + \mathbf{B}\epsilon$$

  Implies exponential distribution for wet/dry spells
- Point process models. Only for precipitations, in particular for storms (Onof et al., 2000; Kaczmarska, Isham and Onof, 2014)
- Hierarchical models:
  1. At the lower level, a finite variable, $S(t) \in \{1, \ldots, K\}$, corresponding to weather "types", "states" or "regimes"
  2. At the upper level, statistical model for

$$\mathbf{Y}_t \mid \mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots, S_t, S_{t-1}, S_{t-2}, \ldots$$

Brief history

- 60's Early work in Hydrology
- 1977 Markov Chain for W/D days; independent Gamma pdf for Rain (Katz, 1977)
- 1981 Addition of T_n, T_x, R with time dependence (Richardson, 1981)
- 1991 Introduction of the concept of weather states using HMM (Zucchini and Guttorp, 1991)
- 90's First models to link weather states to external large scale variables
- > 95 Explosion of improvements
- > 10 Used as a method for downscaling GCM outputs at the local scale

Weather Generators are characterized by their features:

1. Single-site / Multi-site
2. Weather type models
3. Model for the variables, conditionally to the weather type

To follow, focus on

1. Hierarchical parametric models
2. Resampling / analog algorithms
3. SWGs for downscaling
4. Trends and challenges
5. WACSgen (Flecher et al., 2010) and `R` package `WACS`

I. Hierarchical parametric models

Weather states

A disrete variable $S_t \in 1, \ldots, K$, which models weather "types", "states" or "regimes".

- S_t is observed when external variables, such as descriptors of large scale synoptic climatological patterns, are available
- S_t is latent / hidden when it is estimated from the data

Versatile approach for building SWGs, in particular for multi-site SWGs and/or when using SWGs for downscaling GCM outputs.

Most general framework

$$\begin{array}{ccccccccc} \cdots & \rightarrow & S_{t-1} & \rightarrow & S_t & \rightarrow & S_{t+1} & \rightarrow & \cdots \\ & \nearrow & \downarrow & \nearrow & \downarrow & \nearrow & \downarrow & \nearrow & \\ \cdots & \rightarrow & \mathbf{Y}_{t-1} & \rightarrow & \mathbf{Y}_t & \rightarrow & \mathbf{Y}_{t+1} & \rightarrow & \cdots \end{array}$$

Some arrows might be absent in simpler models

I. Hierarchical parametric models
I.1 Single site parametric SWGs

Katz (1977)

$$\begin{array}{ccccccccc} \cdots & \to & \text{dry} & \to & \text{wet} & \to & \text{wet} & \to & \cdots \\ & & \downarrow & & \downarrow & & \downarrow & & \\ \cdots & & 0 & & Y_t & & Y_{t+1} & & \cdots \end{array}$$

- $S_t \in \{\text{wet}, \text{dry}\}$
- Homogenous Markov chain for S_t
- Y_t is Gamma distributed, conditionally independent in time

I. Hierarchical parametric models
I.1 Single site parametric SWGs

Richardson (1981)

$$\begin{array}{ccccccccc} \cdots & \rightarrow & \text{dry} & \rightarrow & \text{wet} & \rightarrow & \text{wet} & \rightarrow & \cdots \\ & & \downarrow & & \downarrow & & \downarrow & & \\ \cdots & \rightarrow & \mathbf{Y}_{t-1} & \rightarrow & \mathbf{Y}_t & \rightarrow & \mathbf{Y}_{t+1} & \rightarrow & \cdots \end{array}$$

- $S_t \in \{\text{wet}, \text{dry}\}$, with homogenous Markov chain
- Precip as before. Non precip. variables

$$\tilde{\mathbf{Y}}_t = \mathbf{A}_{S_t}\tilde{\mathbf{Y}}_{t-1} + \mathbf{B}_{S_t}\boldsymbol{\epsilon}_{S_t}, \quad \boldsymbol{\epsilon}_{S_t} \sim (0, \mathbf{R}_{S_t})$$

$$\mathbf{Y}_t = \tilde{\mathbf{Y}}_t \boldsymbol{\Sigma}_{t,S_t}^{1/2} + \boldsymbol{\mu}_{t,S_t}$$

- Parameters depend on S_t being wet/dry, and on position in the year
- $\boldsymbol{\mu}_{t,S_t}$ and $\boldsymbol{\Sigma}_{t,S_t}^{1/2}$ vary according to an annual cycle, using e.g. cosine functions

Beyond Richardson (1981). Improvements:

- weather state modeling
- weather variables modeling

I. Hierarchical parametric models
I.1 Single site parametric SWGs

Improving the weather state modeling: External weather states

- clustering algorithm on large scale atmospheric variables (Bogardi et al., 1993; Wilson et al., 1992; Hay et al., 1991; Garavaglia et al., 2010), e.g. using *k*-means (Cattiaux et al., 2010; Garavaglia et al., 2010; Guanche et al., 2013, e.g.), mixture models (Vrac et al., 2007), simulated annealing optimization (Bárdossy, 2010; Haberlandt et al., 2014).
- allows to investigate the impact of large scale changes on the weather type distribution (Hughes and Guttorp, 1994; Haberlandt et al., 2014; Wilks, 2012)
- has been adapted to non-stationarity (climate change) in Jones et al. (2011)

Can be useful for downscaling

Not always relevant / optimal to capture the stochastic properties of meteorological variables of interest

I. Hierarchical parametric models
I.1 Single site parametric SWGs

Improving the weather state modeling: Latent / Hidden weather states

- Hidden Markov Models (HMMs) used to model the succession of weather states citepZucchini91.
- Non-homogeneous HHMs: transition probabilities allowed to depend on time and covariates via a link function (Katz and Parlange, 1995; Furrer and Katz, 2007; Ailliot and Monbet, 2012).

$$\begin{array}{ccccccccc}
\cdots & \rightarrow & S_{t-1} & \rightarrow & S_t & \rightarrow & S_{t+1} & \rightarrow & \cdots \\
 & \nearrow & \downarrow & \nearrow & \downarrow & \nearrow & \downarrow & \nearrow & \\
\cdots & \rightarrow & \mathbf{Y}_{t-1} & \rightarrow & \mathbf{Y}_t & \rightarrow & \mathbf{Y}_{t+1} & \rightarrow & \cdots
\end{array}$$

I. Hierarchical parametric models
I.1 Single site parametric SWGs

Improving the weather state modeling: Latent / Hidden weather states

- Large scale variables, ENSO, NAO may also be introduced in the transition probabilities. Improves the description of the inter-annual climate variability and offers a way to link WGs to global climate models (e.g. Hughes and Guttorp, 1994; Hughes et al., 1999; Bellone et al., 2000; Qian et al., 2002; Robertson et al., 2004; Vrac et al., 2007; Zheng and Katz, 2008; Kim et al., 2012).
- With large number of weather states, we evade from exponential sojourn time for dry/wet conditions (Flecher et al., 2010)
- Semi-Markov models with sojourn durations in the regimes modeled by parametric (Racsko et al., 1991; Wilby et al., 1998) or semi-empirical distributions (Semenov et al., 1998). Inference in this setting becomes difficult (Sansom and Thomson, 2001; Bulla et al., 2010).

I. Hierarchical parametric models
I.1 Single site parametric SWGs

Improving the modeling for weather variables: Precipitation

- Evade from Gamma using mixtures of Gamma (Kenabatho et al., 2012), powered exponential of a truncated Gaussian distribution (Allard and Bourotte, 2014), semi-parametric distributions (Lennartsson et al., 2008)
- Use of distributions specifically designed to model extreme values, e.g. generalized Pareto distribution (Lennartsson et al., 2008), e.g. dynamic mixture of the Gamma and GPD (Vrac et al., 2007)
- Time dependence can be modeled using autoregressive process (Hutchinson, 1995; Flecher et al., 2010), parametric auto-correlation function (Allard and Bourotte, 2014) or Gaussian copula (Lennartsson et al., 2008).

I. Hierarchical parametric models
I.1 Single site parametric SWGs

Improving the modeling for weather variables: Other variables

- Generally modeled by a multivariate autoregressive model (Parlange and Katz, 2000)

$$\tilde{\mathbf{Y}}_t = \mathbf{A}_{S_t}\tilde{\mathbf{Y}}_{t-1} + \mathbf{B}_{S_t}\epsilon_{S_t}, \quad \epsilon_{S_t} \sim (0, \mathbf{R}_{S_t})$$

$$\mathbf{Y}_t = \tilde{\mathbf{Y}}_t \mathbf{\Sigma}_{t,S_t}^{1/2} + \boldsymbol{\mu}_{t,S_t}$$

- When autoregressive parameters depend on weather types, the marginal distributions are not be Gaussian anymore
- In Flecher et al. (2010), use of multivariate closed skew-normal distribution, allowing for flexible skewness (more details when presenting WACS)
- For the simultaneous modeling of wind speed and wind direction, Ailliot et al. (2014) used Markov-switching autoregressive processes.

I. Hierarchical parametric models
I.2 Multi-site parametric SWGs

Weather states

Main challenge

Defining spatial models for categorical variables, while keeping a small number of parameters. Still an open problem in spatial statistics.

- Easy way: define a constant weather state over the region; back to the previous case. Ok on small regions.
- Tying together multiple single-site chains by drawing correlated random number in the Markov Chains, as in (Wilks, 1998), may lead to inconsistencies. Alternative inference schemes proposed in Thompson et al. (2007)
- A very interesting option is to censor a Gaussian random fields (Allard and Bourotte, 2014; Kleiber et al., 2012; Baxevani and Lennatsson, 2015).
- The threshold for censoring can also depend on covariates (Qian et al., 2002).

I. Hierarchical parametric models
I.2 Multi-site parametric SWGs

Precipitation

- In many models (Zucchini and Guttorp, 1991; Hughes and Guttorp, 1994; Bellone et al., 2000; Robertson et al., 2004), precipitation amounts are conditionally independent in space and time. Not satisfactory
- Transforms of Gaussian random fields are used to model rainfall occurence in Allard and Bourotte (2014); Kleiber et al. (2012); Baxevani and Lennatsson (2015)
- Kleiber et al. (2012) developped a multisite extension of the chain-dependent model where rainfall amount at each site was modeled by Gamma distributions with shape and scale parameters varying according to latent Gaussian fields.

Other variables

Simultaneous use of

- linear or generalized linear models for trends and standard deviations in spatial context
- standardized Gaussian random fields for spatialy correlated residuals Kleiber et al. (2012); Bourotte et al. (2015)

II. Resampling / analog methods

Some good reasons to evade from parametric models

- Modeling temporal / spatial coherence is a difficult task: model selection issue
- Lack of temporal / spatial coherence in many SWGs, in particular in a multivariate setting
- Modeling the relationship to circulation patterns adds a layer of complexity
- Analog methods are of algorithmic nature: they resample the patterns contained inside the data, without inferring a statistical model

I will briefly present two examples of such algorithms: Direct Sampling technique and AnaWEGE.

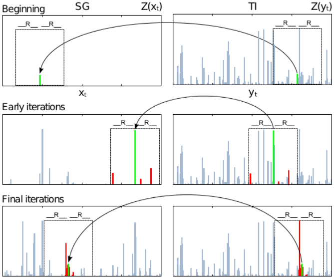
II. Resampling / analog methods
Direct sampling

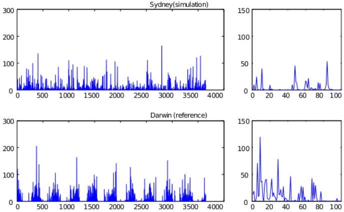
Direct Sampling (Oriani et al., 2014) has been proposed to simulate precipitation.

Denote S = simulation; D = data. For $i = 1, \ldots, n$:

1. Go to a random time t_i
2. Retrieve the "data-event" from Z_S within a neighborhood of radius R around t_i, with at most N values
3. A random time t is visited; $Z_D(t)$ and the corresponding "data-event" at t are retrieved
4. A distance $d(\cdot,\cdot)$ is computed between the two "data-events"
5. If the distance is below a threshold T set $Z_S(t_i) = Z_D(t)$. Otherwise, go to 1.
6. If a fraction F has been scanned and all distances are above T, the datum $Z_D(t)$ minimizing the distance is kept.

II. Resampling / analog methods
Direct sampling





From Oriani et al. (2014)

II. Resampling / analog methods
Direct sampling

Features

- $+$ Tailored secondary variables can be added to compute the distance $d(\cdot,\cdot)$ (e.g. seasonal effects, moving averages, ...)
- $-$ Parameters need to be carefully tuned: R, N, T, F, distance $d(\cdot,\cdot)$. No general methodology
- $+$ Once well tuned, good performances in reproducing statistics of interest (wet spells, dry spells, moving averages, ...)
- $+$ Short patches of verbatim copies: $\geq$ 6 days is very rare
- $-$ Cannot generate new extreme events
- $+$ But, able to generate new aggregated events (reshuffling)
- ▶ Has been extended to multivariate setting (OK) and spatial setting (to be improved)

Needs parametric models to extend to non recorded situations: new extreme values, climate change, spatio-temporal setting, ...

II. Resampling / analog methods
AnaWEGE

AnaWEGE (Yiou, 2014) is a random weather generator based on circulation analogues (computed on Z = SLP on a $2.5^\circ \times 2.5^\circ$ grid), in the North-Atlantic Region, between, say, 1/1/1948 and 31/12/2012.

Set radius, $R = 30$; and # analogs, $K = 20$. For each day j:

1. Find the set of K days, in an interval $R(j)$, of smallest RMS distances

$$d(j,j')^2 = \sum_x (Z(x,j) - Z(x,j'))^2, \quad |j - j'| \leq R$$

   We thus have K sorted RMS distances, $d_1 \leq \cdots \leq d_K$.

2. Compute the corresponding rank correlation coefficients, $c_1, \ldots, c_K$.

II. Resampling / analog methods
AnaWEGE

Static generator (for generating ensembles of seasons, e.g. 90 days)

Copy a season of a random year
For each day in the season

1. $Z_S(x, j)$ remains identical with probability $p_0 = \beta\alpha_1$
2. $Z_S(x, j)$ is replaced with $Z_D(x, j_k)$, with probability $p_k = \beta(1 + c_k)$.

Note:

- α_1 controls the time persistence
- β is a normalizing constant, i.e. $\sum_{k=0}^{K} p_k = 1$
- Each day of a given trajectory is replaced independently $\Rightarrow$ no possibility to create very different trajectory.

II. Resampling / analog methods
AnaWEGE

Dynamic generator (for generating very long series, e.g. several years)

Draw a random day for $Z_S(x, 1)$
For each day j

1. Day $(j+1)$ has K analogues at days $j_1, \ldots, j_K$, with correlations $(c_1, \ldots, c_K)$
2. $Z_S(x, j+1)$ remains identical with probability $p_0 \beta \alpha_1$
3. $Z_S(x, j+1)$ is replaced with $Z_D(x, j_k)$, with probability

$$p_k = \beta(1 + c_k) \exp\{-\alpha_2 |j - j_k|\}$$

Note:

- α_1 controls the time persistence (e.g. $\alpha_1 = 0.5$)
- α_2 controls the weight given to calendar proximity (e.g. $\alpha_2 = 0.4$)

II. Resampling / analog methods
AnaWEGE

Features

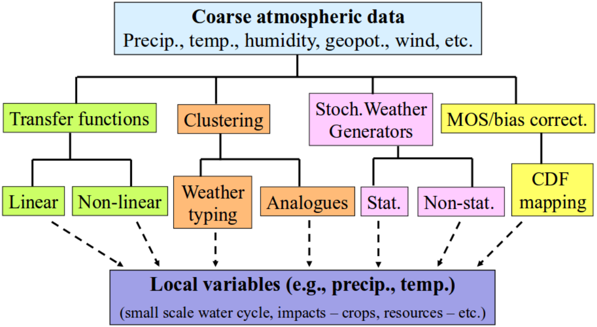
- + Can easily be adapted to simulate other variables, including in a multivariate setting: select the analogues on Z and copy variables $\mathbf{Y}$
- + Spatial and temporal coherence are verified by construction
- – Parameters need to be carecully tuned: R, K, α_1, α_2,
- + Once well tuned, good performances in reproducing statistics of interest
- + Short patches of verbatim copies in time (exponential decay)
- – Exact copies in space
- – Cannot generate new extreme events
- + But, able to generate new aggregated events in time (reshuffling)

III. SWGs for downscaling
Motivations

- IPCC scenarios of climate change have a coarse spatial resolution (250 km)
- Not adapted to ecological, social, economic scales of impact studies at local scale
- Downscaling: to derive regional or local meteorological variables from GCM or reanalysis outputs ($\rightarrow$ 1–5 km)
- Dynamical downscaling (RCMs): GCM outputs drive regional models determining atmosphere dynamics. Expensive resources and computing
- Statistical downscaling: based on statistical relationships between large- and local-scale variables: inexpensive, fast, provides uncertainty quantification

More during Mathieu's talk, Wednesday morning.

III. SWGs for downscaling
Main statistical approaches



From Vrac (2014)

III. SWGs for downscaling

Inhomogeneous SWGs
Include large scale information (from GCM ouptuts, reanalysis data)

$$Y_t \sim f(\cdot \mid \theta(\mathbf{X}_t)) \times P(O_t \mid O_{t-1}, \mathbf{X}_t)$$

- Y_t is wind or precipitation
- $\mathbf{X}_t$ contains GCM features (Pryor et al., 2005) or data (Furrer and Katz, 2007)
- O_t is occurrence (for rain)
- The parameters $\theta(\mathbf{X}_t)$ depend on large scale info, e.g. by use Vector Generalized Linear Models (Wong et al., 2014) or Neural Network Conditional Mixture Models (Carreau and Vrac, 2011)

If N stations, precipitations are conditionally independent.

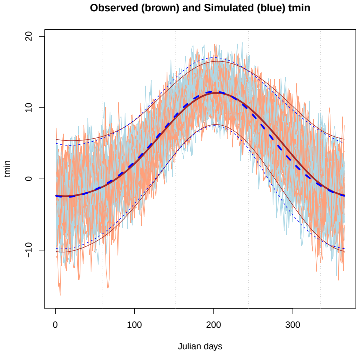
III. SWGs for downscaling
Discussion

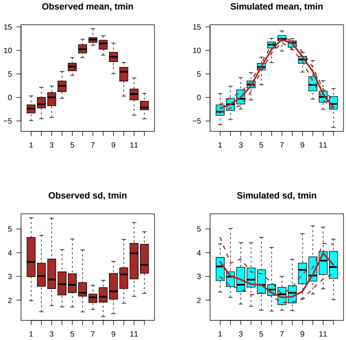
SWGs for downscaling (from Vrac, 2014)

- Many possible models
- Choice of predictors is a major issue
- To reach the very local scale, applying SWGs to CGM outputs may be sometimes better than applying SWGs to RCM outputs
- RCMs and SWGs for downscaling are not in conflict. They provide complementary approaches
- There is no universally better SWG: use ensembles, whenever possible

IV. Trends and challenges

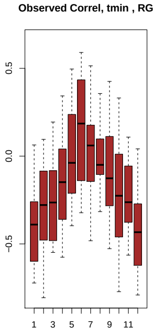
- SWGs must be validated against data. Validation statistics must be chosen in relation to the problem at hand

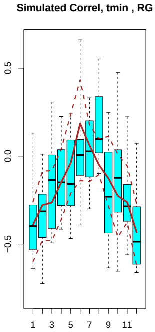


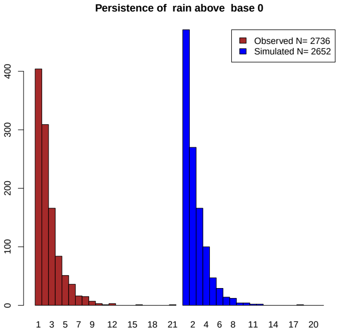


IV. Trends and challenges

- SWGs must be validated against data. Validation statistics must be chosen in relation to the problem at hand







IV. Trends and challenges

- SWGs must be validated against data. Validation statistics must be chosen in relation to the problem at hand
- Many SWGs were presented. How can we compare them? Not on the same variables, spatial scale, context, etc... See European VALUE network (http///www.value-cost.eu/).
- Still a lot to do for multi-site, multivariate SWGs: flexible multivariate space-time models (Bourotte et al., 2015), to account for spatio-temporal coherence and spatio-temporal motion
- In particular, SWGs with a space-time modeling of weather states are required
- Stationarity is often an implicit assumption (for resampling techniques and parametric models). Non stationarity is required in a climate change context.
- Resampling techniques need parametric models to: simulate values outside recorded bounds, account for climate change, simulations at un-recorded stations, etc...
  Can we come up with SWGs taking advantage of the best of both techniques?

Advertisement

Workshops on Stochastic Weather Generators

- 2016 Vannes, 17–21 May 2016,
  http://lebesgue.fr/fr/content/sem2016-climate
- 2014 Avignon, 17–19 Sept. 2014,
  http://informatique-mia.inra.fr/swg2014/accueil
- 2012 Roscoff, 29 May – 1st June 2012:
  http://pagesperso.univ-brest.fr/~ailliot/SWGEN_workshop

Advertisement

Workshops on Stochastic Weather Generators

Advertisement

Recent review papers on SWGs

- Ailliot, P., Allard, D., Monbet, V., and Naveau, P. (2015). Stochastic weather generators: an overview of weather type models. *Journal de la Société Fran caise de Statistique*, 156(1).
- Wilks, D. (2012). Stochastic weather generators for climate-change downscaling, part ii: multivariable and spatially coherent multisite downscaling. *Wiley Interdisciplinary Reviews: Climate Change*, 3(3):267–278.

Exercise session with WACSgen, Tuesday and Wednesday afternoon

V. WACSgen

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
- But, site specific

Flecher, C., Naveau, P., Allard, D., and Brisson, N. (2010). A stochastic daily weather generator for skewefbd data. *Water Resources Research*, 46:W07519.

General architecture

1. Transforming Precip. and removing the trend $\Leftarrow$ work on vector of residuals $\mathbf{Y}_t$
2. Find "weather types", S_t, and transition matrix / season or month
3. Model multivariate/temporal distribution of residual for each "weather type" /season

General architecture

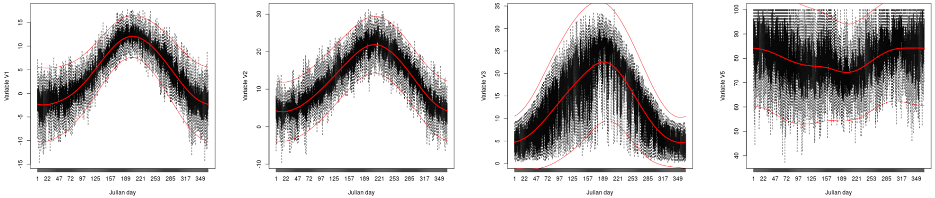
1. Transforming Precip. and removing the trend $\Leftarrow$ work on vector of residuals $\mathbf{Y}_t$
2. Find "weather types", S_t, and transition matrix / season or month
3. Model multivariate/temporal distribution of residual for each "weather type" /season

General architecture

1. Transforming Precip. and removing the trend $\Leftarrow$ work on vector of residuals $\mathbf{Y}_t$
2. Find "weather types", S_t, and transition matrix / season or month
3. Model multivariate/temporal distribution of residual for each "weather type" /season

1. Removing the trend

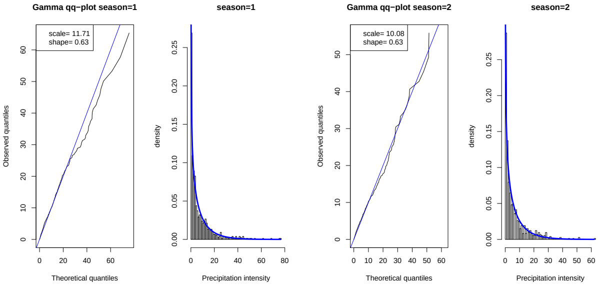
a/ For each variable, build standardized residuals using smoothed means and standard deviations



Create 4 seasons: MAM, JJA, SON, DJF and work on residuals, independently for each season

1. Modeling rainfall

b/ Apply Gamma transform on P for each season



2. Finding 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

$\neq$ Hidden Markov Models !

3. Estimating the Multivariate density

For each season simultaneous residuals follow a Complete Skew-Normal distribution $CSN_{n,m}(\mu, \Sigma, D, \nu, \Delta)$:

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

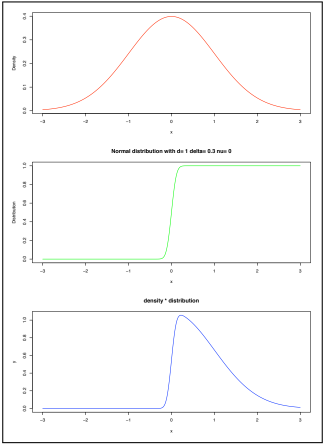
If $D = 0$: $N_n(\mu, \Sigma)$

Good mathematical properties: closed under

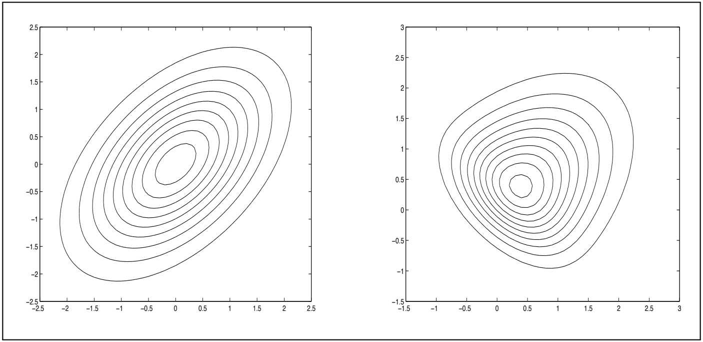
1. Linear transformation
2. Marginalization
3. Conditioning
4. Very simple simulation using algorithm for Gaussian vectors

Example

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



Gaussian and CSN bivariate density



CSN for WACS-gen

To simplify the model, we set some simplifying assumptions, thus defining

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

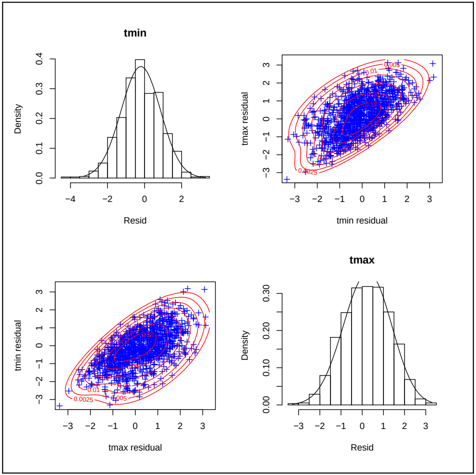
- $m = n = \#variables$
- $\nu = 0$
- $D = \Sigma^{-\frac{1}{2}}\mathbf{S}$
- $\Delta = I_k - \mathbf{S}^2$
- $\mathbf{S} = \text{diag}(\delta_1, \ldots, \delta_K)^t$.

Then, for each season, simultaneous residuals at day d and $d+1$ follow:

$$\begin{pmatrix} \mathbf{Y}_d \\ \mathbf{Y}_{d+1} \end{pmatrix} \sim \text{CSN}^* \left(\begin{pmatrix} \boldsymbol{\mu}_d \\ \boldsymbol{\mu}_{d+1} \end{pmatrix}, \begin{pmatrix} \Sigma_d & \Sigma_d^{1/2}\mathbf{R}\Sigma_{d+1}^{1/2} \\ \Sigma_{d+1}^{1/2}\mathbf{R}\Sigma_d^{1/2} & \Sigma_{d+1} \end{pmatrix}, \begin{pmatrix} \mathbf{S}_d \\ \mathbf{S}_{d+1} \end{pmatrix} \right), \tag{1}$$

where **R** is the vector of temporal correlation.

CSN for WACS-gen



Estimation and Simulation workflow

Estimation

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

  }

Simulation

- Simulate the Markov Chain WS(t)
- For $d = 1, \ldots, T$ {
  Conditionally on WS, simulate $\mathbf{Y}_d \sim$ CSN given $\mathbf{Y}_{d-1}$
- Add seasonal and interannual trend

Bibliography

Ailliot, P., Allard, D., Monbet, V., and Naveau, P. (2015). Stochastic weather generators: an overview of weather type models. *Journal de la Société Fran caise de Statistique*, 156(1).

Ailliot, P., Bessac, J., Monbet, V., and Pène, F. (2014). Non-homogeneous hidden Markov-switching models for wind time series. *Journal of Statistical Planning and Inference*.

Ailliot, P. and Monbet, V. (2012). Markov-switching autoregressive models for wind time series. *Environmental Modelling and Software*, 30:92–101.

Allard, D. and Bourotte, M. (2014). Disaggregating daily precipitations into hourly values with a transformed censored latent Gaussian process. *preprint*.

Bárdossy, A. (2010). Atmospheric circulation pattern classification for south-west germany using hydrological variables. *Physics and Chemistry of the Earth, Parts A/B/C*, 35(9):498–506.

Baxevani, A. and Lennatsson, J. (2015). A spatiotemporal precipitation generator based on a censored latent gaussian field. *Water Resources Research*.

Bellone, E., Hughes, J., and Guttorp, P. (2000). A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts. *Climate Research*, 15:1–12.

Bogardi, I., Matyasovsky, I., Bardossy, A., and Duckstein, L. (1993). Application of a space-time stochastic model for daily precipitation using atmospheric circulation patterns. *Journal of Geophysical Research*, 98(D9):16653–16667.

Bourotte, M., Allard, D., and Porcu, E. (2015). A flexible class of non-separable cross-covariance functions for multivariate space-time data. *arXiv preprint arXiv:1510.07840*.

Bulla, J., Bulla, I., and Nenadig, O. (2010). hsmm - an R package for analyzing hidden semi-Markov models. *Computational Statistics and Data Analysis*, 54(3):611–619.

Carreau, J. and Vrac, M. (2011). Stochastic downscaling of precipitation with neural network conditional mixture models. *Water Resources Research*, 47(10).

Cattiaux, J., Vautard, R., Cassou, C., Yiou, P., Masson-Delmotte, V., and Codron, F. (2010). Winter 2010 in Europe: a cold extreme in a warming climate. *Geophysical Research Letters*, 37(20).

Flecher, C., Naveau, P., Allard, D., and Brisson, N. (2010). A stochastic daily weather generator for skewefbd data. *Water Resources Research*, 46:W07519.

Furrer, E. M. and Katz, R. W. (2007). Generalized linear modeling approach to stochastic weather generators. *Climate Research*, 34:129–144.

Garavaglia, F., Gailhard, J., Paquet, E., Lang, M., Garçon, R., Bernardara, P., et al. (2010). Introducing a rainfall compound distribution model based on weather patterns sub-sampling. *Hydrology and Earth System Sciences Discussions*, 14.

Guanche, Y., Mínguez, R., and Méndez, F. (2013). Climate-based Monte Carlo simulation of trivariate sea states. *Coastal Engineering*, 80:107–121.

Haberlandt, U., Belli, A., and Bárdossy, A. (2014). Statistical downscaling of precipitation using a stochastic rainfall model conditioned on circulation patterns–an evaluation of assumptions. *International Journal of Climatology*.

Hay, L., McCabe, G., Wolock, D., and Ayers, M. (1991). Simulation of precipitation by weather-type analysis. *Water Resources Research*, 27:493–501.

Hughes, J. and Guttorp, P. (1994). A class of stochastic models for relating synoptic atmospheric patterns to local hydrologic phenomenon. *Water Resources Research*, 30:1535–1546.

Hughes, J., Guttorp, P., and Charles, S. (1999). A non-homogeneous hidden Markov model for precipitation occurrence. *Applied Statistics*, 48(1):15–30.

Hutchinson, M. (1995). Stochastic space-time weather models from ground-based data. *Agricultural and Forest Meteorology*, 73(3):237–264.

Jones, P., Harpham, C., Goodess, C., and Kilsby, C. (2011). Perturbing a weather generator using change factors derived from regional climate model simulations. *Nonlinear Processes in Geophysics*, 18(4):503–511.

Katz, R. (1977). Precipitation as a chain-dependant process. *Journal of Applied Meteorology*, 16:671–676.

Katz, R. and Parlange, M. (1995). Generalizations of chain-dependent processes: Application to hourly precipitation. *Water Resources Research*, 31(5):1331–1341.

Kenabatho, P., McIntyre, N., Chandler, R., and Wheater, H. (2012). Stochastic simulation of rainfall in the semi-arid limpopo basin, botswana. *International Journal of Climatology*, (32):1113–1127.

Kim, Y., Katz, R. W., Rajagopalan, B., Podestá, G. P., and Furrer, E. M. (2012). Reducing overdispersion in stochastic weather generators using a generalized linear modeling approach. *Climate research*, 53:13–24.

Kleiber, W., Katz, R., and Rajagopalan, B. (2012). Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes. *Water Ressources Research*, 48:W01523.

Lennartsson, J., Baxevani, A., and Chen, D. (2008). Modelling precipitation in Sweden using multiple step Markov chains and a composite model. *Journal of Hydrology*, 363(1):42–59.

Oriani, F., Straubhaar, J., Renard, P., and Mariethoz, G. (2014). Simulation of rainfall time series from different climatic regions using the direct sampling technique. *Hydrology and Earth System Sciences*, 18(8):3015–3031.

Parlange, M. and Katz, R. (2000). An extended version of the Richardson model for simulating daily weather variables. *Journal of Applied Meteorology*, 39:610–622.

Pryor, S., Schoof, J. T., and Barthelmie, R. (2005). Empirical downscaling of wind speed probability distributions. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 110(D19).

Qian, B., Corte-Real, J., and Xu, H. (2002). Multisite stochastic weather models for impact studies. *International Journal of climatology*, 22(11):1377–1397.

Racsko, P., Szeidl, L., and Semenov, M. (1991). A serial approach to local stochastic weather models. *Ecological Modelling*, 57:27?41.

Richardson, C. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, 17(1):182–190.

Robertson, A., Kirshner, S., and Smyth, P. (2004). Downscaling of daily rainfall occurrence over northeast Brazil using a hidden Markov model. *Journal of climate*, 17(22):4407–4424.

Sansom, J. and Thomson, P. (2001). Fitting hidden semi-Markov models to breakpoint rainfall data. *Journal of Applied Probability*, 38:142–157.

Semenov, A., Brooks, R., Barrow, E., and Richardson, C. (1998). Comparison of the wgen and lars-wg stochastic weather generators for diverse climates. *Climate Research*, 10:95–107.

Thompson, C., Thomson, P., and Zheng, X. (2007). Fitting a multisite rainfall model to New Zealand data. *Journal of Hydrology*, 340:25–39.

Vrac, M., Stein, M., and Hayhoe, K. (2007). Statistical downscaling of precipitation through non homogeneous stochastic weather typing. *Climate Research*, 34:169–184.

Wilby, R., Wigley, T., Conway, D., Jones, P., Hewitson, B., Main, J., and Wilks, D. (1998). Statistical downscaling of general circulation model output: a comparison of methods. *Water Resources Research*, 34:2995–3008.

Wilks, D. (1998). Multisite generalization of a daily stochastic precipitation generation model. *Journal of Hydrology*, 210(1):178–191.

Wilks, D. (2012). Stochastic weather generators for climate-change downscaling, part ii: multivariable and spatially coherent multisite downscaling. *Wiley Interdisciplinary Reviews: Climate Change*, 3(3):267–278.

Wilson, L., Lettenmaier, D., and Skyllingstad, E. (1992). A hierarchial stochastic model of large-scale atmospheric circulation patterns and multiple station daily precipitation. *Journal of Geophysical Research*, 97(D3):2791–2809.

Wong, G., Maraun, D., Vrac, M., Widmann, M., Eden, J. M., and Kent, T. (2014). Stochastic model output statistics for bias correcting and downscaling precipitation including extremes. *Journal of Climate*, 27(18):6940–6959.

Yiou, P. (2014). Anawege: a weather generator based on analogues of atmospheric circulation. *Geoscientific Model Development*, 7(2):531–543.

Zheng, X. and Katz, R. (2008). Mixture model of generalized chain-dependent processes and its application to simulation of interannual variability of daily rainfall. *Journal of Hydrology*, 349(1):191–199.

Zucchini, W. and Guttorp, P. (1991). A hidden Markov model for space-time precipitation. *Water Resources Research*, 27:1917–1923.