A stochastic weather generator based on new spatio-temporal cross-covariance function

Marc Bourotte[1], Denis Allard[1] & Emilio Porcu[2]

[1] INRA Avignon - Unité BioSP
[2] University Federico Santa Maria, Chile

19 September 2014 : SWGen Avignon

Région Provence Alpes Côte d'Azur

Motivations

- Use of Gaussian Random Field (latent or not) to model climatic variables
- Multivariate space-time data

$\Rightarrow$ **Need to model a multivariate spatio-temporal second order structure**

Separability

$M \otimes S \otimes T$: For all $(\mathbf{h}, u)$ in $\mathbb{R}^d \times \mathbb{R}$ and all $i, j = 1, ..., p$

$$C_{ij}(\mathbf{h}, u) = \mathbb{C}ov\left(Z_i(s,t), Z_j(s+h, t+u)\right) = \rho_{ij}.C_S(\mathbf{h}).C_T(u)$$

Nonseparable models :

- $M \otimes (ST)$: $C_{ij}(\mathbf{h}, u) = \rho_{ij}.C(\mathbf{h}, u) \rightarrow$ *Gneiting (2002)*
- $(MS) \otimes T$: $C_{ij}(\mathbf{h}, u) = C_{ij}(\mathbf{h}).C_T(u) \rightarrow$ *Gneiting, Kleiber & Schlather (2010) and Apanasovich, Genton & Sun (2012)*

New nonseparable model for multivariate spatio-temporal random fields (MST)

Motivations

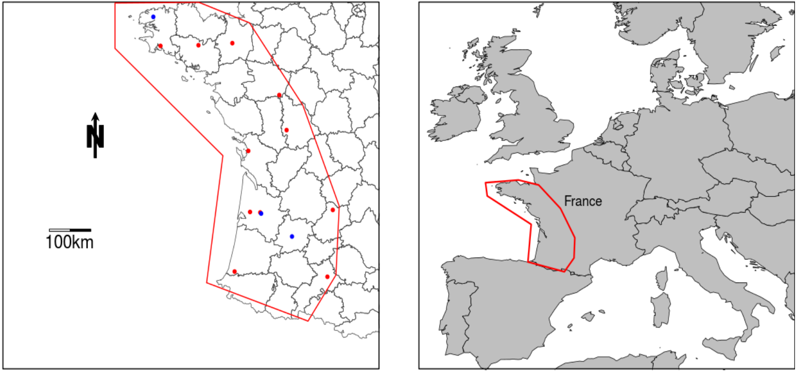


Figure : *Location of the 14 weather stations over western France. Red points are used for estimation, blue points for validation (INRA Climatik portal).*

Outline

Outline

Main result (1/2)

Proposition 1 *(Matérn model)*

Let $\psi(t), t \geq 0$, be a positive function with a completely monotone deriative. The multivariate spatio-temporal model given below

$$C_{ij}(\mathbf{h},u) := \frac{\sigma_i \sigma_j}{\psi(|u|^2)^{d/2}} \rho_{ij} M\left(\frac{\mathbf{h}}{\psi(|u|^2)^{1/2}} | \nu_{ij}, r_{ij}\right), \quad (\mathbf{h},u) \in \mathbb{R}^d \times \mathbb{R},$$

provides a valid second-order structure if

- $\nu_{ij} = (\nu_i + \nu_j)/2, \ \forall i,j = 1,\dots,p;$
- $r_{ij} = \{(r_i^2 + r_j^2)/2\}^{1/2}, \ \forall i,j = 1,\dots,p$ and $r_i > 0, \ \forall i = 1,\dots,p;$
- $\rho_{ij} = \beta_{ij} \frac{\Gamma\{(\nu_i+\nu_j)/2\}}{\Gamma(\nu_i)^{1/2}\Gamma(\nu_j)^{1/2}} \frac{r_i^{\nu_i} r_j^{\nu_j}}{\{(r_i^2+r_j^2)/2\}^{(\nu_i+\nu_j)/2}}, \ \forall i,j = 1,\dots,p;$
- $[\beta_{ij}]_{i,j=1}^p$ is a correlation matrix.

$$M(\mathbf{h}|\nu,r) = \frac{2^{1-\nu}}{\Gamma(\nu)} (r\|\mathbf{h}\|)^\nu \mathscr{K}_\nu (r\|\mathbf{h}\|), \quad \mathbf{h} \in \mathbb{R}^d$$

Main result (2/2)

Proposition 2 *(Cauchy model)*

Let $\psi(t), t \geq 0$, be a positive function with a completely monotone deriative. The multivariate spatio-temporal model given below

$$C_{ij}(\mathbf{h},u) := \frac{\sigma_i\sigma_j}{\psi(|u|^2)^{d/2}}\rho_{ij}\mathscr{C}\left(\frac{\mathbf{h}}{\psi(|u|^2)^{1/2}}|\nu_{ij},r_{ij},\alpha\right), \quad (\mathbf{h},u) \in \mathbb{R}^d \times \mathbb{R},$$

provides a valid second-order structure if

- $\nu_{ij} = (\nu_i + \nu_j)/2, \ \forall i,j = 1,\ldots,p;$
- $r_{ij} = \{(r_i^2 + r_j^2)/2\}^{1/2}, \ \forall i,j = 1,\ldots,p$ and $r_i > 0, \ \forall i = 1,\ldots,p;$
- $\rho_{ij} = \beta_{ij}\frac{\Gamma\{(\nu_i+\nu_j)/2\}}{\Gamma(\nu_i)^{1/2}\Gamma(\nu_j)^{1/2}}\frac{r_i^{\nu_i}r_j^{\nu_j}}{\{(r_i^2+r_j^2)/2\}^{(\nu_i+\nu_j)/2}}, \ \forall i,j = 1,\ldots,p;$
- $[\beta_{ij}]_{i,j=1}^p$ is a correlation matrix.

$$\mathscr{C}(\mathbf{h}|\nu,r,\alpha) = \left(1 + \frac{\|\mathbf{h}\|^\alpha}{r^2}\right)^{-\nu}, \quad \mathbf{h} \in \mathbb{R}^d, 0 < \alpha \leq 2$$

Simulated random field

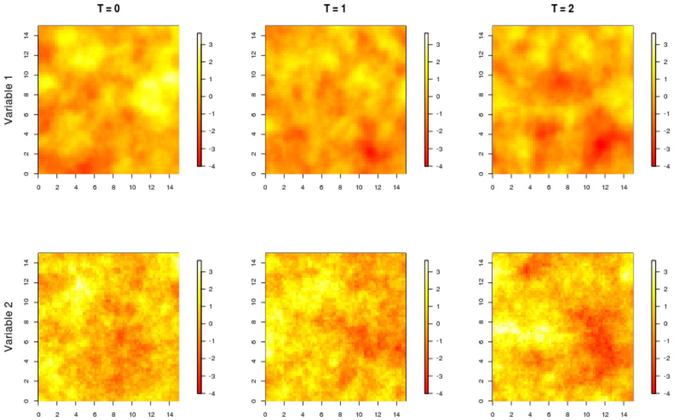


Figure : *Bivariate spatio-temporal Gaussian random field with Matérn model ($\beta_{12} = 0.5$). Top row : smoother field ($\nu_1 = 1.5$ and $r_1 = 1$) & Bottom row : rougher field ($\nu_2 = 0.5$ and $r_2 = 0.5$).*

Estimation (1/2)

Weighted Pairwise (Log-)Likelihood (Varin et al. (2011), Bevilacqua et al. (2012))

Let $(Y_k, Y_{k'}) = \big(Y_i(\mathbf{s}_j, t), Y_{i'}(\mathbf{s}_{j'}, t')\big)$

$$wpl(\theta) = \sum_{k=1}^{n-1} \sum_{k'=k+1}^{n} w_{kk'} \left\{ -\log(2\pi) - \frac{1}{2}\log|\Sigma_{kk'}| - \frac{1}{2}(Y_k, Y_{k'})\Sigma_{kk'}^{-1}(Y_k, Y_{k'})^T \right\}$$

where

$$w_{kk'} = \begin{cases} 1 & \text{if} \quad \mathbf{h} = \|\mathbf{s}_j - \mathbf{s}_{j'}\| \leqslant \lim_S \;\text{ and }\; u = |t - t'| \leqslant \lim_T \\ 0 & \text{otherwise} \end{cases}$$

$$\Sigma_{kk'} = \begin{pmatrix} \sigma_i^2 & C_{ii'}(\mathbf{h}, u) \\ C_{ii'}(\mathbf{h}, u) & \sigma_{i'}^2 \end{pmatrix}$$

Estimation (2/2)

Parameters to estimate

Parameters	Number	Initial Values
σ_i	p	empirical standard deviation
β_{ij}	$\frac{p(p-1)}{2}$	empirical correlation
ν_i, r	$p+1$	$C_{ii}(\mathbf{h},0) = \sigma_i^2 M(\mathbf{h} \mid \nu_i, r)$
a, α, β	3	$C_{ii}(\mathbf{0},u) = \sigma_i^2 (au^{2\alpha}+1)^{-\frac{\beta \boldsymbol{d}}{2}}$

Simulation study (1/2)

Parameters	True	Min	Q_1	Median	Mean	Q_3	Max
σ_1	600	428	537	582	676	647	4173
σ_2	3	1.69	2.54	2.87	3.19	3.26	19.55
σ_3	12	8.63	11.02	11.97	15.80	13.18	78.37
β_{12}	0.45	-0.62	0.24	0.44	0.42	0.66	1
β_{13}	-0.6	-1	-0.75	-0.63	-0.63	-0.53	0.18
β_{23}	-0.5	-1	-0.66	-0.47	-0.47	-0.29	0.52
ν_1	0.5	0.05	0.44	0.51	0.53	0.61	2.14
ν_2	0.9	0.27	0.84	0.92	0.95	1.01	2.30
ν_3	0.4	0.01	0.35	0.42	0.43	0.50	1.02
r_1	0.0030	0.001	0.0018	0.0031	0.0035	0.0044	0.0181
r_2	0.0019	0.001	0.0016	0.0021	0.0023	0.0029	0.0142
r_3	0.0028	0.001	0.0015	0.0027	0.0032	0.0045	0.0185
a	0.5	0.03	0.40	0.53	0.62	0.73	2.59
α	0.5	0.00	0.41	0.53	0.54	0.66	1.00
β	0.8	0.00	0.16	0.80	0.60	1.00	1.00

Table : *Summary statistics of estimates of the trivariate spatio-temporal Matérn model over 500 replicates for the spatio-temporal boundary* $\mathbf{d} = (300km, 3days)$.

Simulation study (2/2)

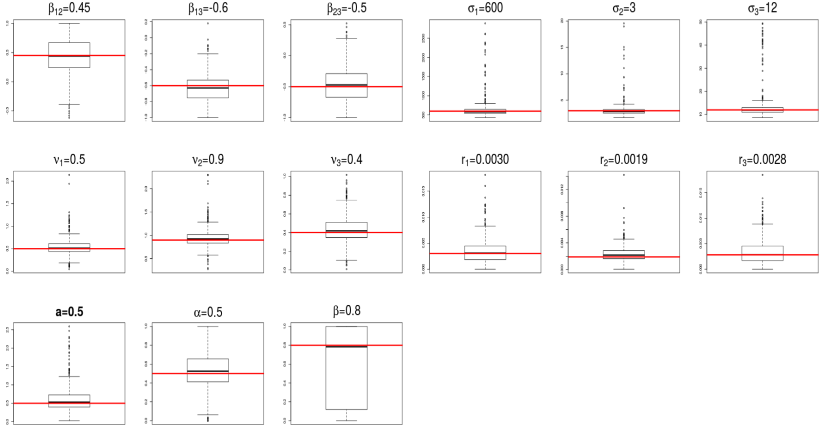


Figure : *Boxplots of parameters inference of the trivariate parsimonious model over 500 replicates for the spatio-temporal boundary* $\mathbf{d} = (300km, 3days)$.

Outline

Presentation of data

Variable 1 : Solar Radiation ; Variable 2 : Mean Temperature ; Variable 3 : Humidity
Daily recorded since 2003 to 2013.
Focus on April month for stationarity assumption.

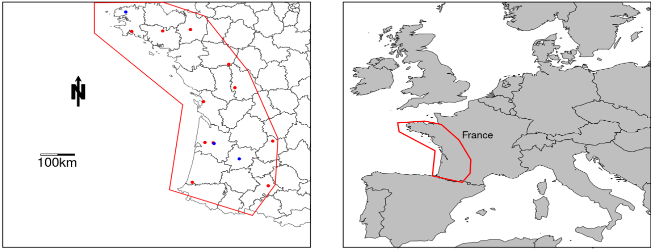


Figure : *Location of the 14 weather stations over western France. Red points are used for estimation, blue points for validation (INRA Climatik portal).*

Model framework (1/2)

A classical framework for climatic variables modelling
$\mathbf{Y}(\mathbf{s},t) = \{Y_1(\mathbf{s},t),\dots,Y_p(\mathbf{s},t)\}^\top$

$$Y_i(\mathbf{s},t) = \mu_i(\mathbf{s}) + Z_i(\mathbf{s},t), \quad i = 1,\dots,p.$$

1. Trend
   For each variable : quadratic trend of longitude and latitude
2. Covariance Matrix

$$C_{ij}(\mathbf{h},u) = \frac{\sigma_i\sigma_j\beta_{ij}}{a|u|^{2\alpha}+1} \frac{\Gamma\{(\nu_i+\nu_j)/2\}}{\Gamma(\nu_i)^{1/2}\Gamma(\nu_j)^{1/2}} \frac{r_i^{\nu_i} r_j^{\nu_j}}{\{(r_i^2+r_j^2)/2\}^{(\nu_i+\nu_j)/2}}$$
$$\times \; M\left(\frac{\mathbf{h}}{(a|u|^{2\alpha}+1)^{\beta/2}} \Big| \frac{\nu_i+\nu_j}{2}, \sqrt{\frac{r_i^2+r_j^2}{2}}\right).$$

Model framework (2/2)

Different models

1. The separability parameter $\beta = 0$ and all the components have the same smoothness and scale parameters;
2. The separability parameter $0 < \beta \leq 1$ and all the components have the same smoothness and scale parameters;
3. The separability parameter $\beta = 0$ and each component has its own smoothness and scale parameters;
4. The separability parameter $0 < \beta \leq 1$ and each component has its own smoothness and scale parameters.

Inference results

	σ_1	σ_2	σ_3	β_{12}	β_{13}	β_{23}	v_1	v_2	v_3
Model 1	621.13	3.12	11.90	0.48	-0.61	-0.34	0.58	0.58	0.58
Model 2	592.66	3.12	11.96	0.48	-0.56	-0.35	0.46	0.46	0.46
Model 3	605.41	3.16	11.84	0.57	-0.64	-0.36	0.47	0.79	0.35
Model 4	605.41	3.14	11.83	0.56	-0.64	-0.39	0.61	0.77	0.38
	$1/r_1$	$1/r_2$	$1/r_3$	a	α	β	-LogLik	Nb	
Model 1	435	435	435	0.52	0.56	0	5006082	10	
Model 2	588	588	588	0.48	0.69	0.85	5005745	11	
Model 3	357	526	434	0.48	0.52	0	5003501	14	
Model 4	286	556	400	0.46	0.59	0.53	5003358	15	

Table : *Maximum pairwise likelihood estimates of parameters for the 4 models with* $\mathbf{d} = (300km, 3days)$. *Units are Joules per square centimeter for solar radiation, degree Celsius for temperature, percentage for humidity and kilometer for distances.*

Fitting (1/3)

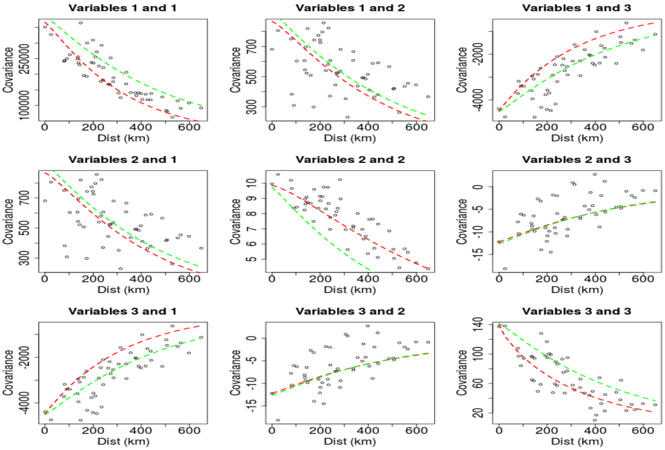


Figure : *Empirical and cross-covariance functions at lag time $u = 0$ with fitted models : fully separable model (model1 ; broken green line) and fully nonseparable model (model4 ; broken red line).*

Fitting (2/3)

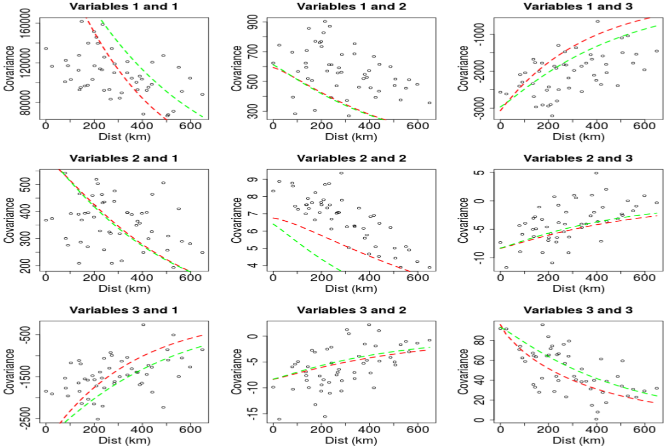


Figure : *Empirical and cross-covariance functions at lag time $u = 1$ with fitted models : fully separable model (model1 ; broken green line) and fully nonseparable model (model4 ; broken red line).*

Fitting (3/3)

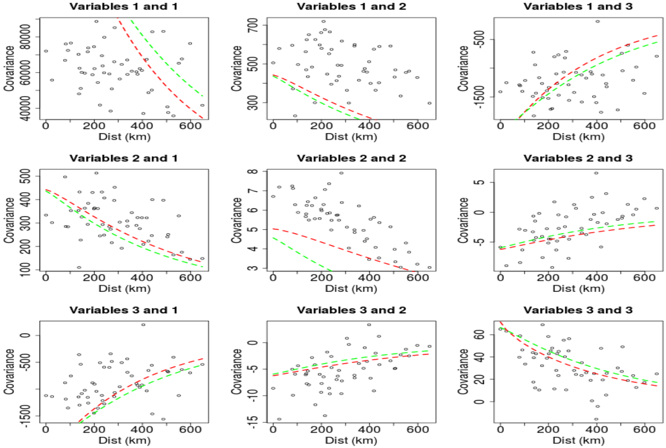


Figure : *Empirical and cross-covariance functions at lag time $u = 2$ with fitted models : fully separable model (model1 ; broken green line) and fully nonseparable model (model4 ; broken red line).*

Validation

- Estimation : red points, April months 2003-2012
- Validation : blue points conditionally to red points, April 2013

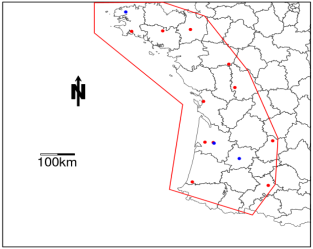


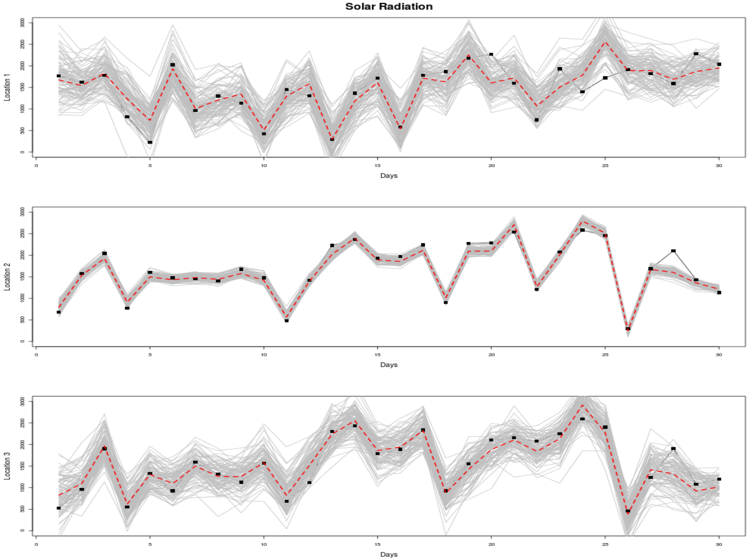
Figure : *Location of the 14 weather stations over western France. Red points are used for estimation, blue points for validation (INRA Climatik portal).*

Validation

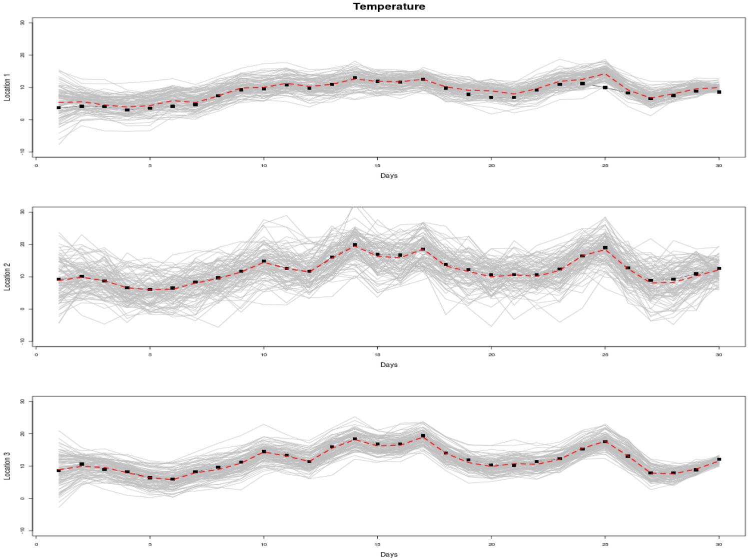
	Solar Radiation		Temperature		Humidity	
	MSPE	MAE	MSPE	MAE	MSPE	MAE
Model 1	38985	133.6	0.77	0.57	25.7	3.88
Model 2	37309	129.3	0.70	0.53	23.7	3.59
Model 3	33314	119.1	0.74	0.58	26.1	3.83
Model 4	35346	125.7	0.64	0.53	20.6	3.40

Table : *Validation results comparing mean squared prediction error (MSPE) and mean absolute error (MAE) for solar radiation, temperature and humidity.*

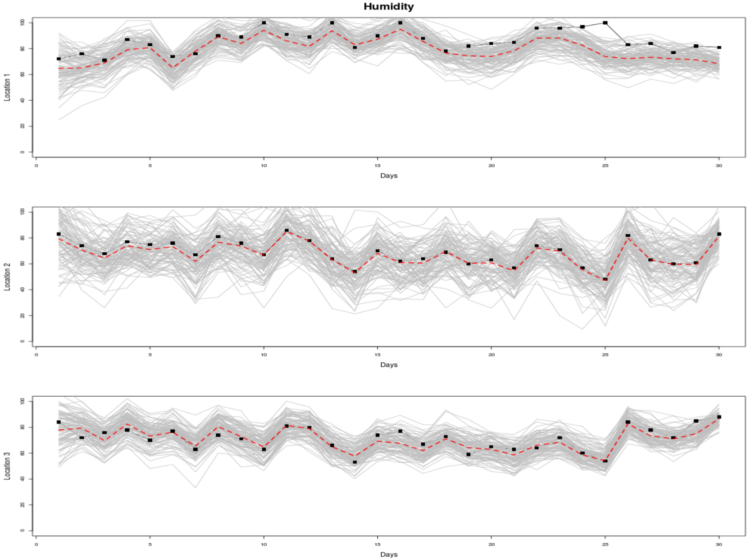
Conditional simulations : Solar Radiation (Joule.cm^{-2})



Conditional simulations : Mean Temperature (°C)



Conditional simulations : Humidity (%)



Outline

1. A new cross-covariance model for multivariate space-time data
2. Application to data
3. Conclusions & Perspectives

Conclusions & Perspectives

- Conclusions
  - Valid theorical model for multivariate data
  - Capiticity to simulate data "look like" observed variables
- Perspectives
  - Asymmetric model ? Non stationary model ?
  - Several variables : wind, min and max temperature, precipitation
  - Investigate the behavior of all variables during heat waves, heavy rain etc