A multivariate multi-site Weather Generator based on a Flexible Class of non-separable Cross-Covariance Functions

Denis Allard, with **M. Bourotte**

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

Workshop on Stochastic Weather Generators,
Vannes, 17–20 May, 2016

Context

- Provide a multivariate, multi-site SWG for the Mediterranean region of France
- Modeling of the space-time dependence structure, not only within each variable, but also between the variables
- This requires models allowing for different range and smoothness parameters for each variable
- With special care for rain

Context

SITES :

1 Gruissan (11)
2 Millau (12)
3 Salon De Provence (13)
4 Montélimar (26)
5 St Marcel Les Valence (26)
6 Bellegarde (30)
7 Fourques (30)
8 St Gilles (30)
9 Auzeville-Tolosane (31)
10 Marseillan Plage (34)
11 Mauguio (34)
12 Montpellier (34)
13 Roujan (34)
14 Alenya (66)
15 Perpignan (66)
16 Lavaur (81)
17 Avignon (84)
18 Carpentras (84)

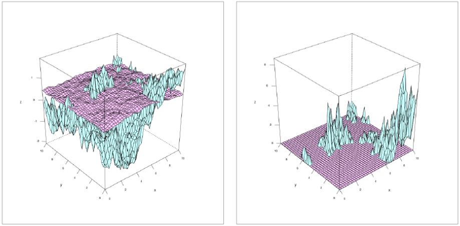
- 18 weather stations, two not exactly Mediterranean
- Variables: Rain, Tn, Tx, W speed, Total R

Pre-processing

1. Box-Cox transforms: $\Rightarrow$ Log-transform of Wind speed; no transformation of other variables
2. Latent Gaussian transform for rain

$$Z(\mathbf{s}, t) = \psi(Y(\mathbf{s}, t)) = z_m + b\left(\exp\{a(Y(\mathbf{s}, t) - y_0)^c - 1\}\right) \text{ if } Y(\mathbf{s}, t) > y_0$$

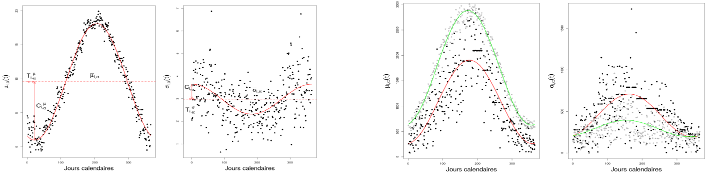
and $Z(\mathbf{s}, t) = 0$ otherwise, $Y(\mathbf{s}, t)$ being a latent Gaussian Random Field.



Occurrence and rainfall are modeled with a single Random Field

Pre-processing

1. Box-Cox transforms: $\Rightarrow$ Log-transform of Wind speed; no transformation of other variables
2. Latent Gaussian transform for rain
3. Remove seasonal cycles for other variables



We now consider a p random vector $(Y_1(\mathbf{s}, t), \ldots, Y_p(\mathbf{s}, t))$.

What kind of spatio-temporal models for these residuals?

Set-up and notations

- p-dimensional multivariate random field $\mathbf{Y}(\mathbf{x}) = \{Y_1(\mathbf{x}), \dots, Y_p(\mathbf{x})\}^\top$
- $\mathbf{x} = (\mathbf{s}, t) \in D \times T \subset \mathbb{R}^{d+1}$, $d \geq 1$, with $\mathbf{s} \in D \subset \mathbb{R}^d$ and $t \in T \subset \mathbb{R}$
- Stationarity:

$$\text{Cov}\left\{Z_i(\mathbf{s}, t), Z_j(\mathbf{s} + \mathbf{h}, t + u)\right\} = C_{ij}(\mathbf{h}, u)$$

$$\mathbf{C}(\mathbf{k}) = \left[C_{ij}(\mathbf{h}, u)\right]_{1 \leq i,j \leq p}, \quad \mathbf{k} = (\mathbf{h}, u) \in D \times T.$$

Goal

Elaborate valid, flexible parametric classes of space-time matrix-valued covariance functions for $\mathbf{Z}(\mathbf{x})$, hence

$$\boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda} \geq 0,$$

for any $n \in \mathbb{N}$, for any finite set of points $(\mathbf{s}_1, t_1), \dots, (\mathbf{s}_n, t_n)$ and for any vector $\boldsymbol{\lambda} \in \mathbb{R}^{np}$

State-of-the-art

Full separability

$$\mathbf{C}(\mathbf{k}) = \mathbf{A}\,\rho_S(\mathbf{h})\,\rho_T(u)$$

Then,

$$\boldsymbol{\Sigma} = \mathbf{A} \otimes \mathbf{C}_S \otimes \mathbf{C}_T,$$

Space-time separability

$$\rho_{ST}(\mathbf{k}) = \rho_S(\mathbf{h})\,\rho_T(u)$$

$$\Longleftrightarrow$$

$$\mathrm{Cov}\left\{Z(\mathbf{s},t), Z(\mathbf{s}',t')\right\} \propto \mathrm{Cov}\left\{Z(\mathbf{s},t), Z(\mathbf{s}',t)\right\} \mathrm{Cov}\left\{Z(\mathbf{s}',t), Z(\mathbf{s}',t')\right\},$$

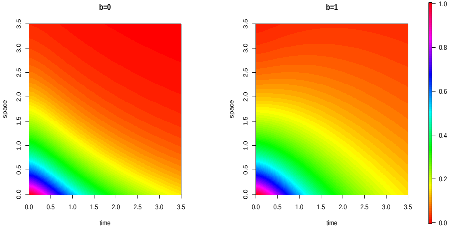
- With separability, sophisticated interactions between space and time cannot be captured
- In a multivariate framework, separability between variables and space-time variations implies that all variables are characterized by the same space-time correlation function

Non separable space-time covariances

Gneiting class: (Gneiting, 2002; Gneiting et al., 2007)

$$\mathcal{G}(\mathbf{h}, u) = \frac{\sigma^2}{\psi(u^2)^{d/2}} \, \varphi\left(\frac{\|\mathbf{h}\|^2}{\psi(u^2)}\right), \quad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R},$$

where $\varphi : [0, \infty) \to \mathbb{R}$ is completely monotone on the positive real line, with $\varphi(0) < \infty$, ψ is variogram $+c > 0$.



Spatial covariance: Multivariate Matérn family

Matérn

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

where $\nu > 0$ is a smoothness parameter, $r > 0$ is a scale parameter.

Gneiting et al. (2010); Apanasovich et al. (2012)

$$C_{ij}(\mathbf{h}) = \sigma_i \sigma_j \rho_{ij} \mathcal{M}(\mathbf{h}; r_{ij}, \nu_{ij}), \quad \mathbf{h} \in \mathbb{R}^d,$$

- Each covariance function $C_{ii}(\mathbf{h})$, has a different smoothness parameter
- Allows some cross correlations between the variables
- Restrictions on the parameters $\{r_{ij}, \nu_{ij}, \rho_{ij}\}_{i,j=1,\dots,p}$ to ensure the validity of the matrix-valued covariance function $\mathbf{C}(\mathbf{h}) = \left[C_{ij}(\mathbf{h})\right]_{i,j=1}^{p}$.

Extension to spatio-temporal data

Combine space-time non separability of Gneiting class with non-separable multivariate Matérn.

Gneiting-Matérn

Theorem (Bourotte, Allard and Porcu, *Spatial Statistics*, 2016)

The multivariate Gneiting-Matérn space-time model $\mathbf{C}^{\mathcal{M}} = \left[C_{ij}^{\mathcal{M}}(\cdot,\cdot)\right]_{i,j=1}^{p}$, with

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

is a valid matrix-valued covariance function if, for all $i,j = 1,\dots,p$,

$$\begin{aligned} r_{ij} &= \{(r_i^2 + r_j^2)/2\}^{1/2}, \\ \nu_{ij} &= (\nu_i + \nu_j)/2, \\ \rho_{ij} &= \beta_{ij}\frac{\Gamma(\nu_{ij})}{\Gamma(\nu_i)^{1/2}\Gamma(\nu_j)^{1/2}}\frac{r_i^{\nu_i} r_j^{\nu_j}}{r_{ij}^{2\nu_{ij}}}, \end{aligned}$$

with $r_i, \nu_i > 0$, $\boldsymbol{\beta} = \left[\beta_{ij}\right]_{i,j=1}^{p}$ a correlation matrix and $\psi(t)$, $t \geq 0$, a positive function with a completely monotone derivative.

Specific model used in the rest of this talk

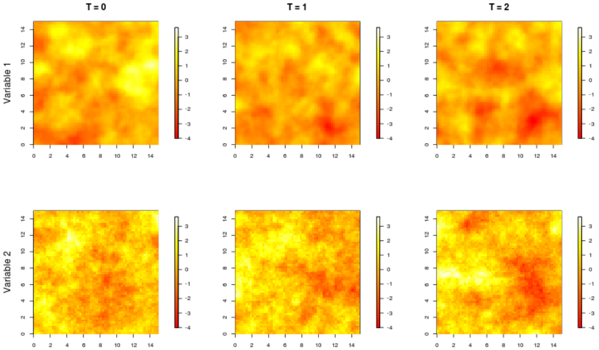
Gneiting-Matérn

$$\psi(x) = (\alpha x^a + 1)^b, \quad x \geq 0,$$

with $\alpha > 0, 0 < a \leq 1, 0 \leq b \leq 1$.

$$C_{ij}(\mathbf{h}, u) = \frac{\sigma_i \sigma_j}{(\alpha|u|^{2a} + 1)^{\tau}} \rho_{ij} \mathcal{M}\left(\frac{\mathbf{h}}{(\alpha|u|^{2a} + 1)^{b/2}}; r_{ij}, \nu_{ij},\right), \quad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R},$$

Simulation



Bivariate Gneiting-Matérn space-time Gaussian random field at 3 consecutive instants. Top row: smoother component ($\nu_1 = 1.5$ and $r_1 = 1$); bottom row: rougher component ($\nu_2 = 0.5$ and $r_2 = 0.5$). The co-located correlation parameter ρ_{12} is set to 0.5.

Estimation of the parameters

Problem

- maximum full likelihood (FL) requires $\mathcal{O}\left((np)^3\right)$ operations (determinant and inverse of $np \times np$ covariance matrix)
- There are $(p+2)(p+3)/2$ parameters to estimate, i.e. 15 parameters when $p = 3$

Our solution: use composite likelihood

- products of smaller likelihoods
- defined on certain small subsets of data, such as marginal or conditional events
- easy to compute

See Varin et al. (2011) for an overview

Pairwise Composite Likelihood

- Pairwise marginal Gaussian log-likelihoods computed on pairs of data (Bevilacqua and Gaetan, 2015)

$$\ell(i, j, \mathbf{s}_\alpha, \mathbf{s}_\beta, t_\alpha, t_\beta; \theta).$$

- The Weighted Pairwise log-Likelihood (WPL) is thus (Bevilacqua and Gaetan, 2015)

$$\text{wpl}(\theta) = \sum_{(i,j,\alpha,\beta)} \ell(i, j, \mathbf{s}_\alpha, \mathbf{s}_\beta, t_\alpha, t_\beta; \theta) w_{\alpha\beta}, \quad \theta \in \Theta,$$

- The computational cost is $\mathcal{O}\left((np)^2\right)$.
- Can be reduced by only considering pairs such that

$$(\|\mathbf{h}\|, u) \leq (d_S, d_T).$$

Previous results (Bourotte et al., 2016)

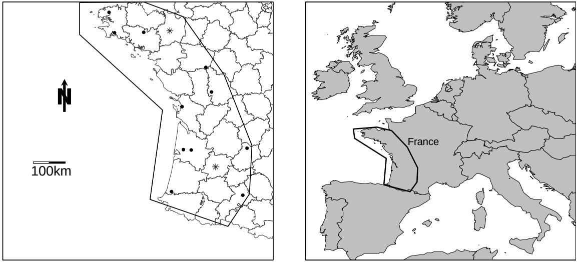
On a setting mimicking the conditions of the data set

- Considering only pairs such that

$$(\|\mathbf{h}\|, u) \leq (d_S, d_T) = (500\text{km}, 2\text{days})$$

  was found sufficient
- Good relative efficiency (rre > 0.65), on most parameters
- Except on β_{ij}, where rre > 0.2
- Better prediction scores (RMSE, RMAE, LogS, CRPS) than simpler models either separable and/or with equal spatial parameters (smoothness, range)

Western France weather dataset (Bourotte et al., 2016



Location of the 13 weather stations over western France. With stars: validation stations

Model fitting: spatial

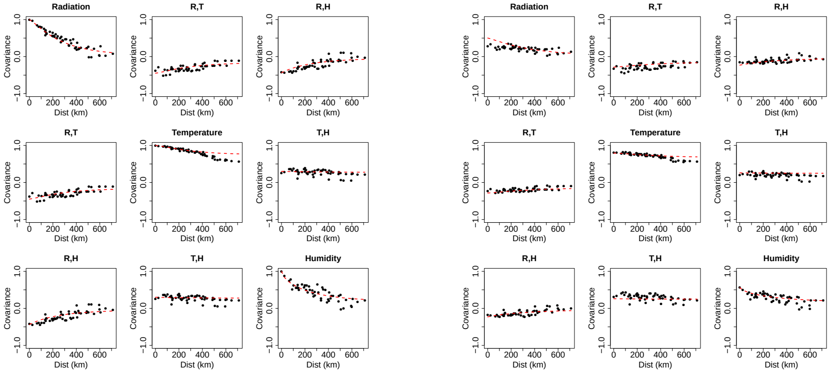


Figure: Spatial direct and cross-covariance functions for R, T and H. Points: empirical values. Solid line: model NS-D with parameters estimated using WPL. Left panel: $u = 0$. Right panel: $u = 1$.

Model fitting: temporal

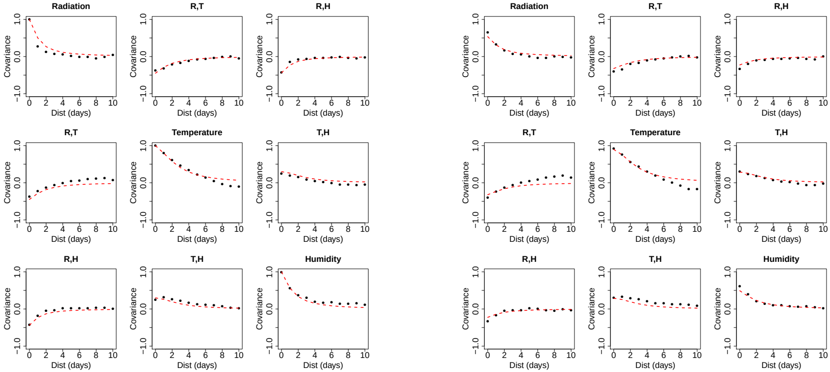


Figure: Temporal direct and cross-covariance functions for R, T and H. Left panel: $\mathbf{h} = 0$. Right panel: $\|\mathbf{h}\| = 200$ km.

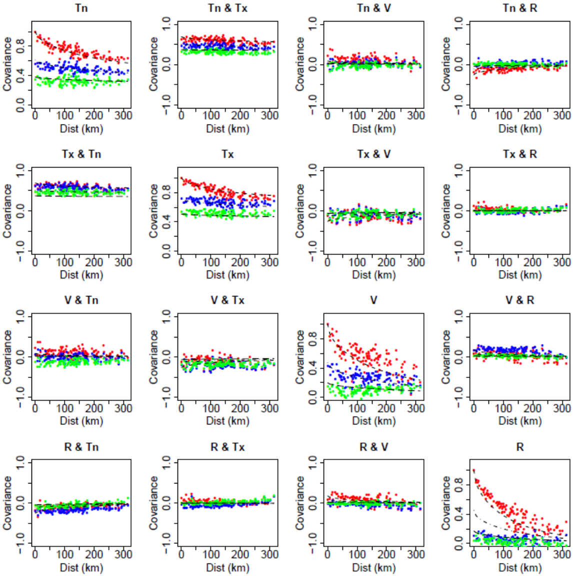
Back to Mediterranean data sets

SITES :

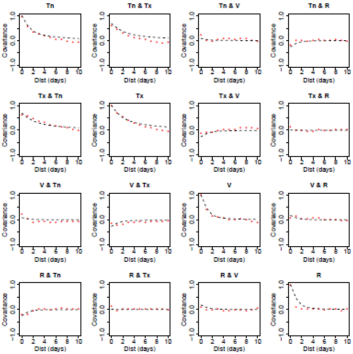
1 Gruissan (11)
2 Millau (12)
3 Salon De Provence (13)
4 Montélimar (26)
5 St Marcel Les Valence (26)
6 Bellegarde (30)
7 Fourques (30)
8 St Gilles (30)
9 Auzeville-Tolosane (31)
10 Marseillan Plage (34)
11 Mauguio (34)
12 Montpellier (34)
13 Roujan (34)
14 Alenya (66)
15 Perpignan (66)
16 Lavaur (81)
17 Avignon (84)
18 Carpentras (84)

- 18 weather stations, two not exactly Mediterranean
- Variables: Rain, Tn, Tx, W speed, Total R

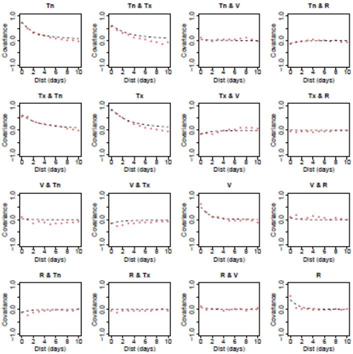
Covariance fitting – spatial domain



Covariance fitting – temporal domain

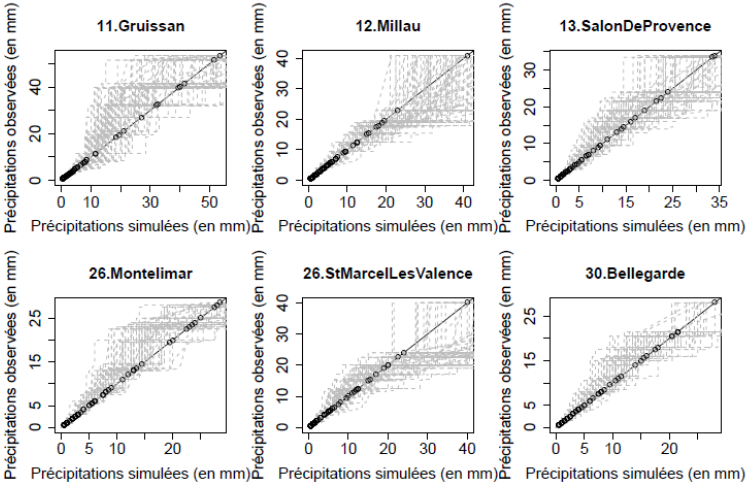


$t = 0$

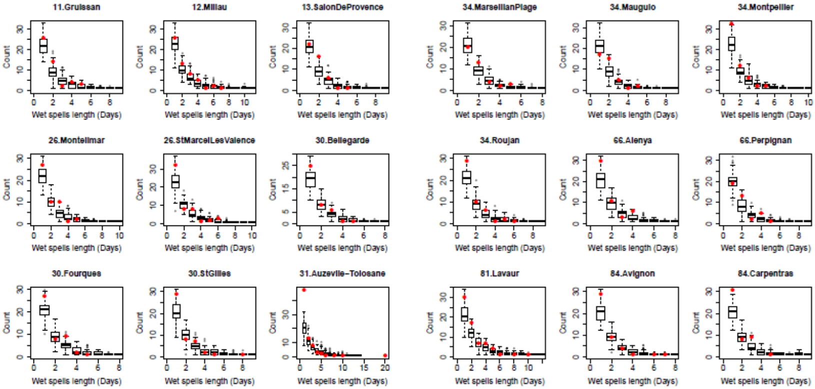


$t = 1$

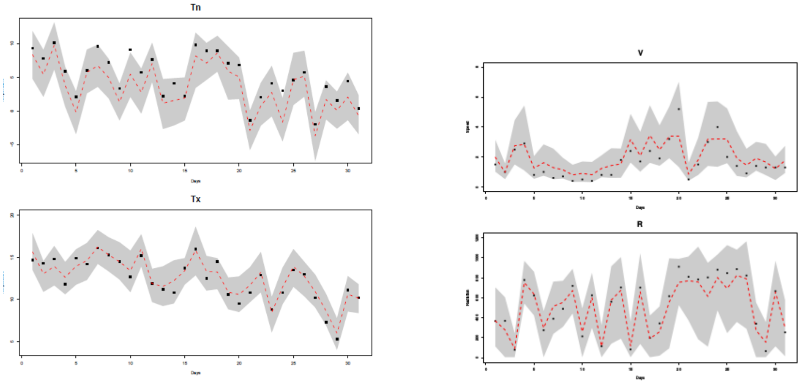
Precipitation



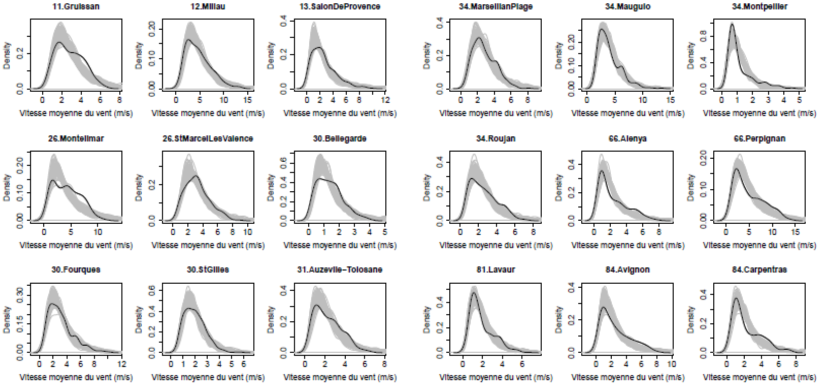
Wet spells


11.Gruissan
12.Millau
13.SalonDeProvence
34.MarseillanPlage
34.Mauguio
34.Montpellier
26.Montelimar
26.StMarcelLesValence
30.Bellegarde
34.Roujan
66.Alenya
66.Perpignan
30.Fourques
30.StGilles
31.Auzeville−Tolosane
81.Lavaur
84.Avignon
84.Carpentras
Count
Wet spells length (Days)

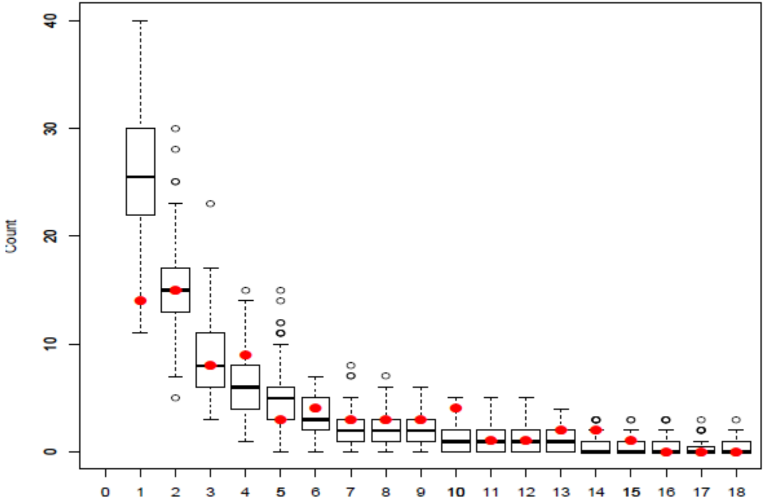
Other variables



Wind densities



Simultaneous Precipitations



Discussion

- New class of model: fills an interesting gap between space-time non separable model (Gneiting's class) and multivariate Matérn
- WPL is an interesting estimation procedure, with mild loss of efficiency, and almost no loss on b
- Very good prediction scores on weather data set
- Precipitations were modeled using truncated Gaussian latent model
- This multisite, multivariate Stochastic Weather Generator provides encouraging results
- Next: introduce wheather types at the regional scale

Bourotte M., Allard, D. and Porcu, E. (On-line) A Flexible Class of Non-separable Cross-Covariance Functions for Multivariate Space-Time Data, *Spatial Statistics*, 10.1016/j.spasta.2016.02.004. ArXiv 1510.07840

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).

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

Apanasovich, T. V. and Genton, M. G. (2010). Cross-covariance functions for multivariate random fields based on latent dimensions. *Biometrika*, 97(1):15–30.

Apanasovich, T. V., Genton, M. G., and Sun, Y. (2012). A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components. *Journal of the American Statistical Association*, 107(497):180–193.

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

Bevilacqua, M. and Gaetan, C. (2015). Comparing composite likelihood methods based on pairs for spatial Gaussian random fields. *Statistics and Computing*, 25(5):877–892.

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*.

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

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, 97(458):590–600.

Gneiting, T., Genton, M., and Guttorp, P. (2007). Geostatistical space-time models, stationarity, separability and full symmetry. In Finkenstädt, B., Held, L., and Isham, V., editors, *Statistical Methods for Spatio-Temporal Systems*, pages 151–175. Chapman & Hall/CRC.

Gneiting, T., Kleiber, W., and Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105(491):1167–1177.

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.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21(1):5–42.

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