Space-time stochastic rainfall modelling at sub-kilometer scale

Denis Allard, with Lionel Benoit and Grégoire Mariethoz (U. Lausanne), and with Denis Marcotte (Polytechnique Montréal)

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

DESIRES – atelier spatiotempmeteo

Rennes, 28–30 novembre 2018

Outline

1. Rainfall modelling with a latent, truncated, Gaussian field
   - Motivations
   - Model
   - Estimation / simulation

   Benoit, L., Allard, D., & Mariethoz, G. (2018). Stochastic Rainfall Modelling at Sub Kilometer Scale. *Water Resources Research.*

2. Using GMRF / SPDE representation to accelerate computations
   - Very short introduction to GMRF / SPDE
   - Illustrate how it can be used on a geoscience example

   Marcotte, D., & Allard, D. (2018). Gibbs sampling on large lattice with GMRF. *Computers & Geosciences*, 111, 190-199.

Outline

1. Rainfall modelling with a latent, truncated, Gaussian field
   - Motivations
   - Model
   - Estimation / simulation

   Benoit, L., Allard, D., & Mariethoz, G. (2018). Stochastic Rainfall Modelling at Sub Kilometer Scale. *Water Resources Research.*

2. Using GMRF / SPDE representation to accelerate computations
   - Very short introduction to GMRF / SPDE
   - Illustrate how it can be used on a geoscience example

   Marcotte, D., & Allard, D. (2018). Gibbs sampling on large lattice with GMRF. *Computers & Geosciences*, 111, 190-199.

First part

Rainfall modelling with a latent, truncated, Gaussian field

Motivation

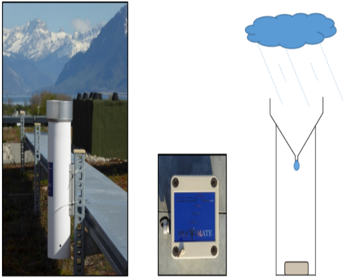
Main Objective

Analyzing and modeling precipitation **within** a radar pixel i.e. at **very fine scale**

- Fine scale modeling of precipitations: $\sim$ min; $\sim 10 - 1000$ m
- Accounting for large amount of 0 values
- Space-time dependence structure, including transport (advection)
- Large data-set
- Bayesian analysis for (conditional) simulation of rain events with uncertainties on parameters

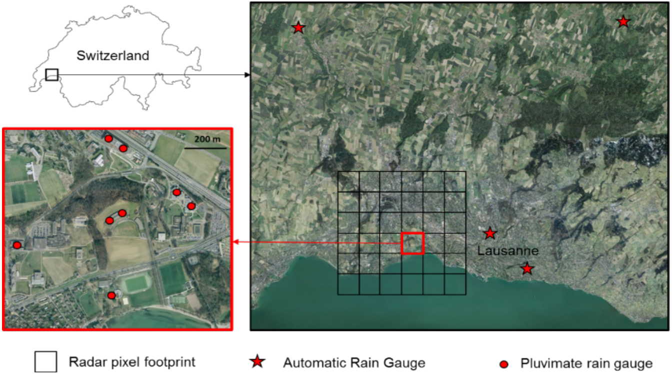
New rain gauges

- New sorts of rain gauges, called pluvimates
- Drop counting rain gauges
  $\sim$ 0.01 mm resolution
- Have been calibrated with lab experiments
- Integration time is 30s

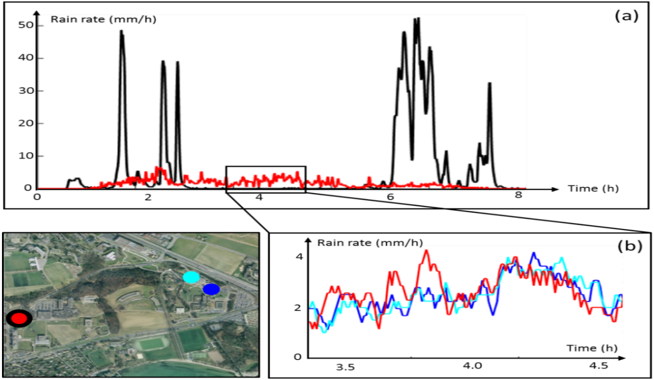


(a) (b) (c)

Data 1/Lausanne

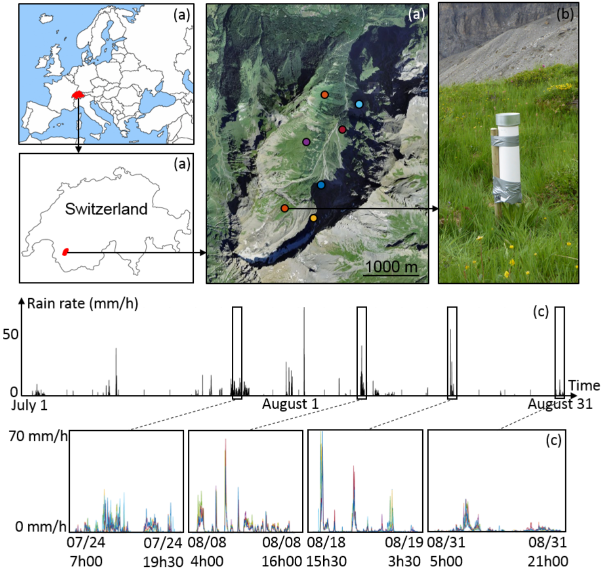


Data 1/Lausanne



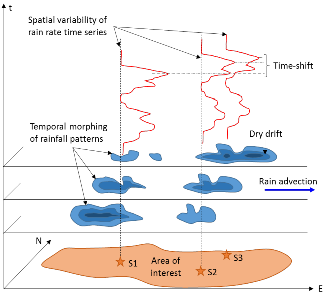
Space-time rain rate fluctuations as observed by a dense network of Pluvimates. (a) Global view of a stratiform (red, 3-4 January 2016) and a convective (black, 24-25 October 2016) rain events. (b) Zoom on a two hours period for the convective event, and comparison between co-located (blue curves) and distant (red vs blue curves) measurement locations.

Data 2/ Alpine catchment



Conceptual model

- Fine space-time scales $\Longrightarrow$ stationarity
- Evolving shapes along time $\Longrightarrow$ space-time correlation
- Dry-drift $\Longrightarrow$ correlation between intensity and presence of rainfall
- Advection of clouds $\Longrightarrow$ transport term
- Diffusivity $\Longrightarrow$ non-separable ST covariance function



Hierarchical model

Latent Gaussian field

Allcroft and Glasbey (2003), Allard and Bourotte (2014), Baxevani and Lennartsson (2015), etc...

- Spatio-temporal coordinates $(\mathbf{s}, t) \in (D \times T)$
- Precipitations $R_m(\mathbf{s}, t)$ arise from a latent, standardized, stationary Gaussian random field $Y(\mathbf{s}, t)$ with

$$Cov(Y(\mathbf{s}, t), Y(\mathbf{s}', t')) = \rho(\mathbf{s}' - \mathbf{s}, t' - t; \boldsymbol{\eta}),$$

  where $\rho(\cdot, \cdot, \boldsymbol{\eta})$ is a spatio-temporal covariance function, with parameters $\boldsymbol{\eta}$
- Marginal transformation with truncation:

$$R_m(\mathbf{s}_i, t_i) = \psi(Y(\mathbf{s}_i, t_i) + \epsilon_i) = \left(\frac{Y(\mathbf{s}_i, t_i) + \epsilon_i - a_0}{a_1} \right)^{1/a_2}, \quad \text{if } Y(\mathbf{s}_i, t_i) \geq a_0$$

  with $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$ i.i.d and $\boldsymbol{\alpha} = (a_0, a_1, a_2)^t$.
- $R_m(\mathbf{s}_i, t_i) = 0$ if $Y(\mathbf{s}_i, t_i) < a_0$.

$Y(\mathbf{s}, t)$ is hidden, since not observed when $R_m(\mathbf{s}, t) = 0$.

Spatio-temporal covariance model

Non separable spatio-temporal covariance function, with transport

Gneiting (2002), Gneiting et al. (2007), Ailliot et al. (2011), etc.
Bourotte et al. (2016) in the multivariate case.

- Advection is modeled by a single vector $\mathbf{V}$ with

$$\rho(||\mathbf{s}' - \mathbf{s}|| - \mathbf{V}(t' - t), |t' - t|) = \rho_L(||\mathbf{s}' - \mathbf{s}||, |t' - t|). \quad (1)$$

- Non separable model for ρ_L, with

$$\rho_L(\mathbf{h}, u) = \frac{1}{(u/d)^{2\delta} + 1} \exp\left\{ \frac{-(|||\mathbf{h}||/c)^{2\gamma}}{\{(u/d)^{2\delta} + 1\}^{\beta\gamma}} \right\} \quad (\mathbf{h}, u) \in \mathbb{R}^2 \times \mathbb{R}, \quad (2)$$

with $\boldsymbol{\eta} = (c, d, \beta, \gamma, \delta, S_V, \theta_V, \sigma_\epsilon^2)$.

Separability

Definition

A covariance function is separable if

$$C_{ST}(\mathbf{h}, u) \propto C_S(\mathbf{h}) C_T(u)$$

i.e.

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

Is the covariance function of:

$$Y(\mathbf{s}, t) = Y_S(\mathbf{s}).Y_T(t); \qquad Z_S(\mathbf{s}) \perp Z_T(t)$$

Main property

Equivalent to

$$Z(\mathbf{s}, t) \perp Z(\mathbf{s}', t') \mid Z(\mathbf{s}, t')$$

$\forall \mathbf{s}, \mathbf{s}' \in D; \quad \forall t, t' \in T$

Separable models

Advantages

- Easy to understand
- Simplifies coding and memory storage
- Matrix computation are accelerated

But

- No complex interaction between space and time
- Overly simplistic for most applications on climate/weather variables

Need for non separable models

Here, we use the popular/widely used / flexible Gneiting class

Parameter $\beta \in [0, 1]$ controls separability

Likelihood

Let $R_m(\mathbf{s}_1, t_1), \ldots, R_m(\mathbf{s}_n, t_n)$ be the rainfall measurements.

- For $i = 1, \ldots, n$

$$Z(\mathbf{s}_i, t_i) = \begin{cases} \psi^{-1}(R_m(\mathbf{s}_i, t_i)) = a_0 + a_1 R_m(\mathbf{s}_i, t_i)^{a_2} & \text{when } R_m(\mathbf{s}_i, t_i) > 0 \\ Z(\mathbf{s}_i, t_i) \leq a_0 & \text{otherwise} \end{cases}$$

- Let $I = \{i : R_m(\mathbf{s}_i, t_i) > 0\}$ and $I^0 = \{i : R_m(\mathbf{s}_i, t_i) = 0\}$ and let $\mathbf{R} = (\mathbf{R}_{I^0}, \mathbf{R}_I)$.

Likelihood

$$\ell(\mathbf{R}; \boldsymbol{\theta}) = \ell(\mathbf{R}_I; \boldsymbol{\theta}).\mathbb{P}(\mathbf{R}_{I^0} \leq \mathbf{0} \mid \mathbf{R}_I; \boldsymbol{\theta}),$$

where $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\eta})$.

Likelihood

Log-likelihood

$$
\begin{aligned}
L(\mathbf{R};\boldsymbol{\theta}) &= L(\mathbf{R}_I;\boldsymbol{\theta}) + \log \mathbb{P}(\mathbf{R}_{I^0} \leq \mathbf{0} \mid \mathbf{R}_I;\boldsymbol{\theta}). \\
&= -0.5 \log|\boldsymbol{\Sigma}_{I,I}| - 0.5\mathbf{Z}_I^t\boldsymbol{\Sigma}_{I,I}^{-1}\mathbf{Z}_I - N_I \log(2\pi) \\
&\quad + \log \Phi_{N_{I^0}}(\mathbf{a}_0; \boldsymbol{\Sigma}_{I^0,I}\boldsymbol{\Sigma}_{I,I}^{-1}, \boldsymbol{\Sigma}_{I^0,I^0} + \boldsymbol{\Sigma}_{I_0,I}\boldsymbol{\Sigma}_{I,I}^{-1}\boldsymbol{\Sigma}_{I,I^0})
\end{aligned}
$$

with

$$
\boldsymbol{\Sigma}_{I,I}[i,j] = \rho_L(\mathbf{s}_j - \mathbf{s}_i - \mathbf{V}(t_j - t_i), t_j - t_i) \quad \text{for } i \neq j \in I, \quad \text{and } \boldsymbol{\Sigma}_{I,I}[i,i] = 1 + \sigma_\epsilon^2
$$

- Computing $\boldsymbol{\Sigma}_{I,I}^{-1}$ and $|\boldsymbol{\Sigma}_{I,I}|$: necessitate $\mathcal{O}(N_I^3)$ operations
- Computing $\Phi_{N_{I^0}}$, which is a N_{I^0} dimensional normal probabilities, see e.g. Genz (1992, 2004, 2009, etc.) and R package `mvtnorm`.
- Limited to moderate size datasets, ~ 1000 i.e. $\simeq$ 2h with 8 pluvimates

If larger datasets: block likelihood (BL) or alternative strategies

Estimation methods

Block Likelihood

- The original dataset is divided into a series of N_τ blocks, B_p of τ consecutive measurements at all sites
- For $p = 1, \ldots, N_\tau$, let

$$\begin{aligned} I_p &= \{i \in B_p : R_m(\mathbf{s}_i, t_i) > 0\} \\ I_p^0 &= \{i \in B_p : R_m(\mathbf{s}_i, t_i) = 0\} \end{aligned}$$

- The Blockwise log-likelihood, $L_{BL}(\mathbf{R}; \boldsymbol{\theta})$ is

$$L_{BL}(\mathbf{R}; \boldsymbol{\theta}) = \sum_{p=1}^{N_\tau} \Big[-0.5 \left(|\boldsymbol{\Sigma}_{I_p,I_p}| + \mathbf{Z}_{I_p}^t \boldsymbol{\Sigma}_{I_p,I_p}^{-1} \mathbf{Z}_{I_p} + N_{I_p} \log(2\pi)\right) + \log \Phi(a_0; \boldsymbol{\Sigma}_{I_p^0,I_p} \boldsymbol{\Sigma}_{I_p,I_p}^{-1}, \boldsymbol{\Sigma}_{I_p^0,I_p^0} - \boldsymbol{\Sigma}_{I_p^0,I_p} \boldsymbol{\Sigma}_{I_p,I_p}^{-1} \boldsymbol{\Sigma}_{I_p,I_p^0})\Big].$$

What size of blocks ?

Bayesian setting

The full model is

$$\begin{aligned}
R_m(\mathbf{s}_i, t_i) &= \psi(Y(\mathbf{s}_i, t_i)) \\
Y(\cdot, \cdot) &\sim \mathcal{G}(\rho) + \epsilon; \qquad \rho: \ \mathbf{V} \text{ and } \rho_L(\mathbf{h}, u) \sim \text{Gneiting class} \\
\epsilon &\sim \text{i.i.d. } \mathcal{N}(0, \sigma_\epsilon^2) \\
\sigma_\epsilon^2 &\sim \pi_\epsilon \\
\boldsymbol{\eta} &\sim \pi_{\boldsymbol{\eta}} \\
\boldsymbol{\alpha} &\sim \pi_{\boldsymbol{\alpha}} \\
\mathbf{V} &\sim \pi_{\mathbf{V}}
\end{aligned}$$

- All parameters are independent
- Gibbs for censored $Y(\cdot, \cdot)$ within Metropolis Hastings for $\boldsymbol{\theta}$
- Uniform vague priors
- Random Walk proposals
- 5000 iterations for burn-in; 10000 iterations; 100 samples

Data

Synthetic data

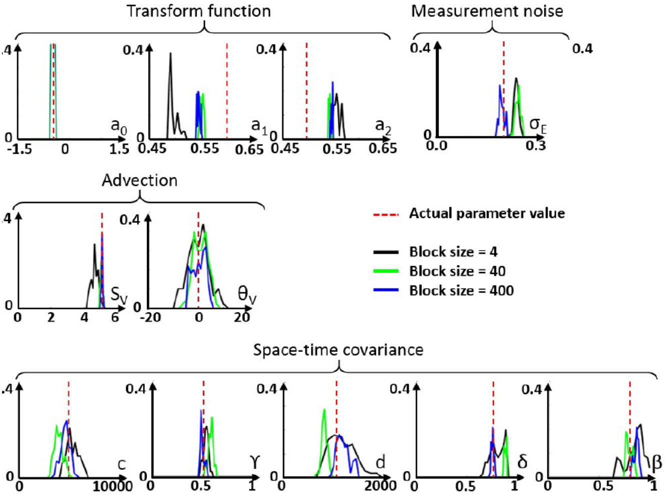
Situation mimicking real data

- 400 60s-measurements. $\Rightarrow$ 6h40m
- 9 pluvimates on 1000m $\times$ 1000m regular grid
- ρ_L: $c = 5000m,\ \gamma = 0.5,\ d = 1000s,\ \delta = 0.8,\ \beta = 0.8,\ \sigma_\epsilon = 0.2$
- **V**: $Sv = 5m/s,\ \theta_v = 0$
- Transform: $a_0 = -0.5,\ a_1 = 0.6,\ a_2 = 0.5$
- 100 blocks of size 4; 4 blocks of size 100; 1 block of size 400

CPU

- Matlab; 20 cores machine for matrix inversion (Intel Xéon CPU E5-2699 v4, 2.2GHz)
- From 45' (BL4) to 550' (BL100) and 1300' (BL400)

Synthetic data – Posteriors



De we need sophisticated models ?

Experiment: we model from the complete model, and estimate according to

- Model A: No advection & separable covariance function [Kleiber et al, 2012; Baxevani and Lennartsson, 2015].

$$\rho_L(\mathbf{h}, u) = \frac{1}{(u/d)^{2\delta} + 1} \exp\left\{ \frac{-||\mathbf{h}||}{c} \right\} \quad (\mathbf{h}, u) \in \mathbb{R}^2 \times \mathbb{R},$$

  with $\delta = 1/2$

- Model B: Advection $\mathbf{V}$, separable covariance function [Lepioufle et al., 2012; Leblois and Creutin, 2013; Bárdossy and Pegram, 2016]).

$$\rho_L(\mathbf{h}, u) = \frac{1}{(u/d) + 1} \exp\left\{ \frac{-||\mathbf{h}|| + \mathbf{V}u}{c} \right\} \quad (\mathbf{h}, u) \in \mathbb{R}^2 \times \mathbb{R},$$

- Model C: Complete model, as above

Unconditional simulation

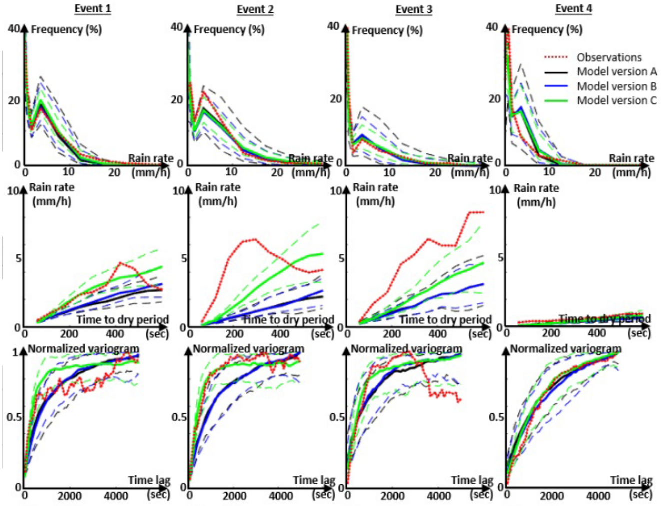


Figure: Reproduction of rainfall statistics by unconditional simulation. Rows correspond to the summary statistics: Row 1: histogram, Row 2: Dry Drift, Row 3: temporal variogram. The solid lines are the medians of realizations while the dashed lines are Q10 and Q90.

Unconditional simulation

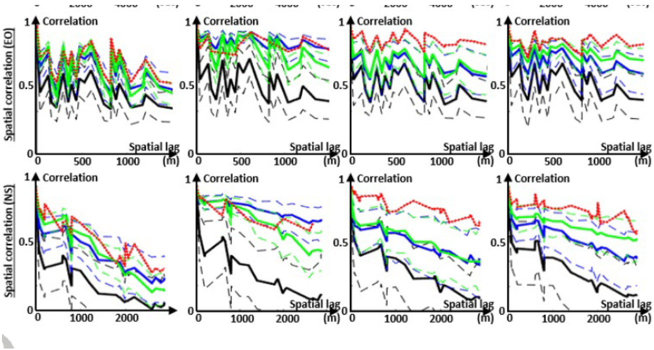


Figure: Reproduction of rainfall statistics by unconditional simulation. Rows correspond to spatial correlogram, along East-West and North-South direction respectively. The solid lines are the medians of realizations while the dashed lines are Q10 and Q90.

De we need sophisticated models ?

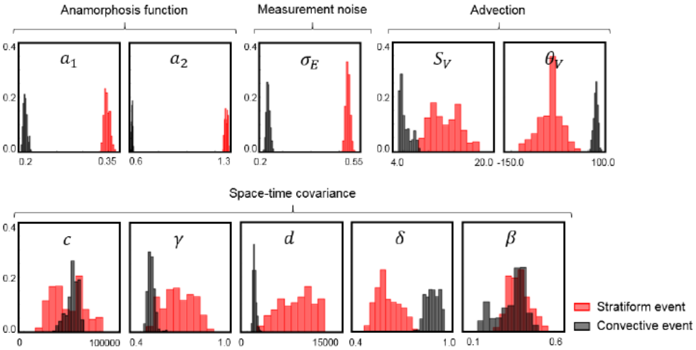
YES !

- C better than B better than A
- Advection is particularly important for spatial variogram
- In addition, non separability is important for dry drift

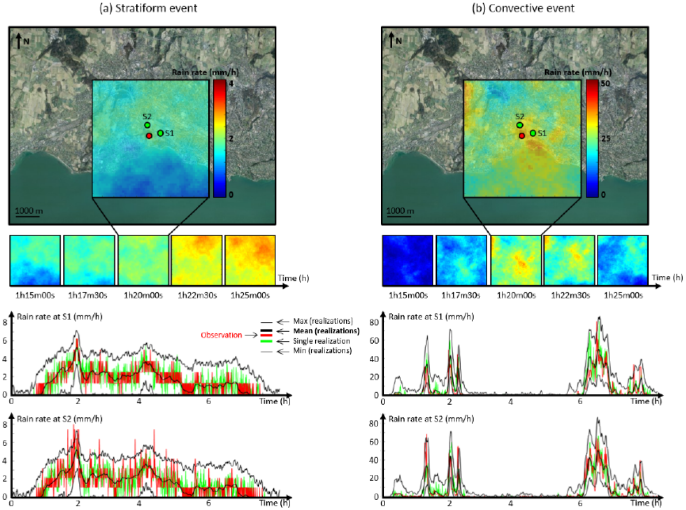
BUT ...

- Dry-drift is still under-represented

Lausanne – Estimation

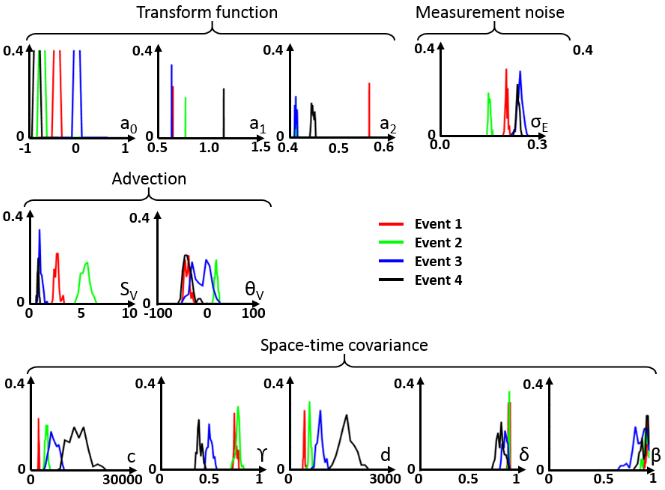


Lausanne – Simulations



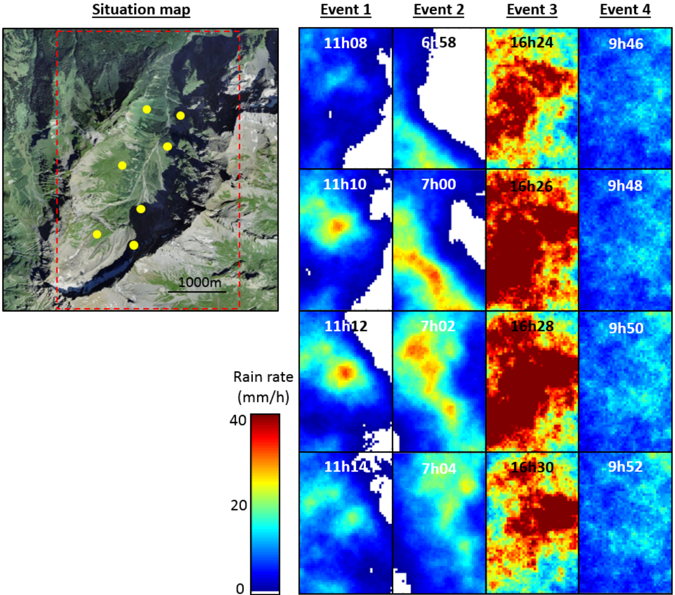
Alpine catchment – Estimation

1 & 3: shower events; 2 & 3: convective events. 4: stratiform event



Alpine catchment – Simulations

1 & 3: shower events; 2 & 3: convective events. 4: stratiform event



Challenge: censored latent Gaussian process for large datasets

Full likelihood / conditional simulation

- Either compute $\Phi_{N_{I0}}$
- Or simulate conditional truncated Gaussian with M-H step

Computer intensive if dataset is large.

Composite likelihood

- products of smaller likelihoods Varin et al. (2011) $\Rightarrow$ easy to compute, e.g. pairs or smaller blocks
- Pairwise Likelihood (PL) is efficient and unbiased, see e.g. Bevilacqua and Gaetan (2015), Bourotte et al. (2016)
- But PL is under dispersed $\Rightarrow$ cannot be used directly in a Bayesian setting
- Adjustments and re-calibration proposed in Ribatet et al. (2012) and Stoehr and Friel (2015).

Hierarchical nearest-neighbor Gaussian process

Datta, A., Banerjee, S., Finley, A.O., and Gelfand, A.E. 2016. Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. JASA, 111, 800–812.

Second part

Using GMRF / SPDE representation to accelerate computation

- Will be illustrated on spatial data
- Still to be implemented for spatio-temporal data

GMRF – SPDE approach

Sparse precision matrices

Some Matérn spatial covariance matrices induce sparse precision matrices $\mathbf{Q} = \mathbf{C}^{-1}$ Lindgren et al. (2011)

$$C(\mathbf{h}) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(||\mathbf{h}||/r)^{\nu} K_{\nu}(||\mathbf{h}||/r)$$

when $\nu + d/2$ is an integer

- Conditional densities depends only on a few nearest neighbors
- Identical for all pixels, except border effects
- Fast determinant / likelihood computations
- Efficient parallel simulation possible $\Rightarrow$ larger datasets

GMRF – SPDE approach

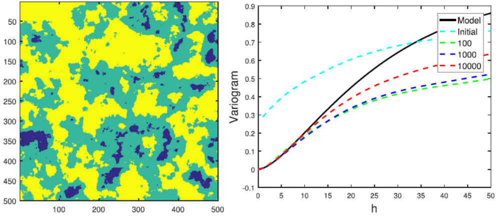
Example: $d = 2, \nu = 1$: Each pixel has only 14 non-zero entries. Top right quadrant reads

$$\sigma^2 \begin{array}{|ccc} 1 & 0 & 0 \\ \text{-}2a & 2 & 0 \\ 4 + a^2 & \text{-}2a & 1 \\ \hline \end{array}$$

with $a = 4 + r^{-2}$.

Application to latent truncated Gaussian fields

Three categories: yellow if $Y(\mathbf{s}) \leq a_1$; green if $a_1 < Y(\mathbf{s}) \leq a_2$; blue if $Y(\mathbf{s}) > a_2$.



Aim: simulate $Y(\mathbf{s})$ conditional on categories (left).

Algorithm

- Initial configuration: at each $\mathbf{s}_i$, generate truncated Gaussian RV, conditional on category
- Iterate until convergence criterion:
  - visit each site $\mathbf{s}_i$, $\quad i = 1, \ldots, N$
    1. Generate Gaussian value $\sim Y(\mathbf{s}_i) \mid \mathbf{s}_{\mathcal{N}_i}$
    2. Accept if category OK, otherwise go to 1.

- Parallel implementation possible on a 14-colour chessboard
- High acceptance rate thanks to truncation
- Convergence gets slower as ν and r increase

Discussion

Done:

- Flexible spatio-temporal model to simulate rainfall events at very fine scale
- Includes transport and non-separability on Lagrangian correlation function
- Bayesian setting allows to simulate events with parameters drawn from the posterior

To do:

- Accelerate simulation using tricks inspired by GMRF/SPDE approach
- Does not model inter-event waiting time
- Catalogue of parameters according to type of events?

Bibliography

Ailliot, P., Baxevani, A., Cuzol, A., Monbet, V., & Raillard, N. (2011). Space?time models for moving fields with an application to significant wave height fields. Environmetrics, 22(3), 354-369.

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*.

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

Benoit, L. and Allard, D. and Mariethoz, G. (2018). Stochastic Rainfall Modelling at Sub Kilometer Scale. *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. (2016). A flexible class of non-separable cross-covariance functions for multivariate space-time data. *Spatial Statistics*, 18, 125–146.

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.

Lindgren, F., Rue, H., Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4), 423-498.

Marcotte, D. and Allard, D. (2018). Gibbs sampling on large lattice with GMRF. *Computers & Geosciences*, 111, 190-199.

Ribatet, M., Cooley, D., Davison, A. C. (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, 813-845.

Stoehr, J., Friel, N. (2015). Calibration of conditional composite likelihood for Bayesian inference on Gibbs random fields. (AISTATS).

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