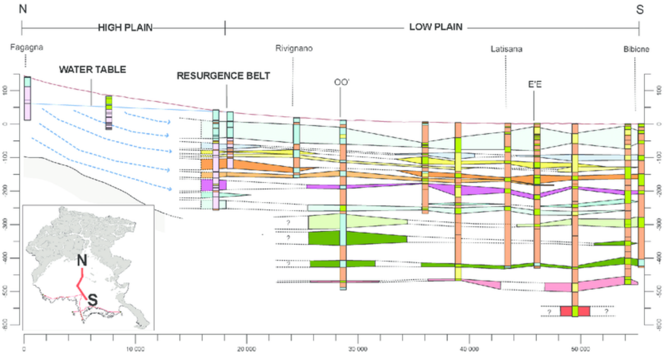
A latent Gaussian approach for modeling geological sequences

Denis Allard[1], Carlo Gaetan[2], Paolo Fabbri[3]

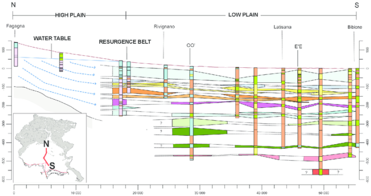
[1] Biostatistics and Spatial Processes (BioSP), MIA division, INRAE, Avignon, France
[2] DAIS, Università Ca' Foscari, Venezia, Italia
[3] Dipartimento di Geoscienze, Università degli Studi di Padova, Italia

Statistiques à Rochebrune, 2024

Motivating case study: the Venice aquifer



Motivating case study: the Venice aquifer



Characteristics:

- Succession of facies
- Vertical wells
- Data along wells = sequence of facies and depth between different facies
- Only part of the sequence is visible along wells

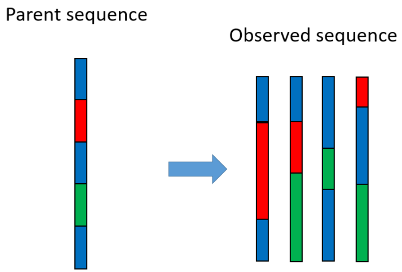
Our aim:

- Building a stochastic model for this spatial process
- Addressing the lateral continuity issue
- Doing inference based on well data
- Performing stochastic simulations, conditional on well observations

Geological sequence

Parent sequence

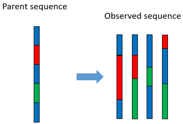
There always exists at least one sequence of facies of minimal length, compatible with all observed sequences. We call it the parent sequence

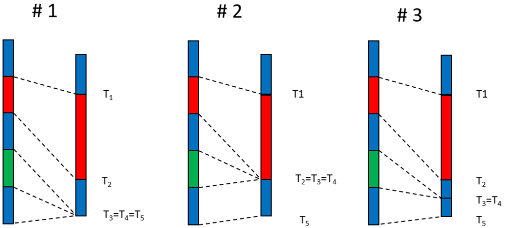


Corollary

Many observed sequences can arise from a single parent sequence

Geological sequence





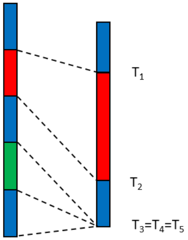
- Different scenarios are compatible with this observed sequence
- How do we assign a thickness to each layer of the parent sequence, based on the observed sequences?

Set-up and notations

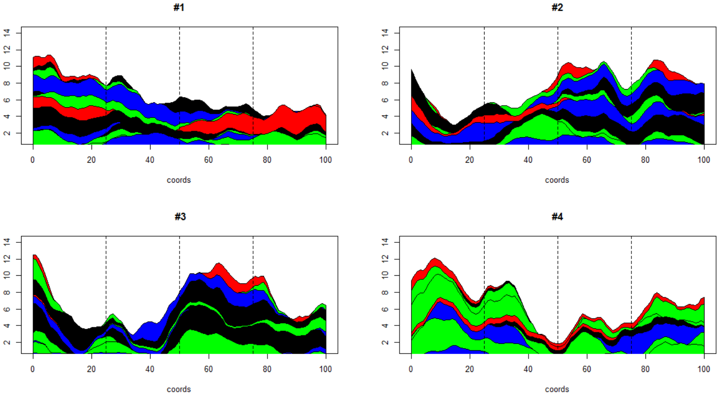
- n wells at locations $s_i \in D$
- Observed depths $\mathbf{T}_i^o$ and thicknesses $\mathbf{Z}_i^o$ at well i
- Complete depths $\mathbf{T}_i$ and thicknesses $\mathbf{Z}_i$ at well i
- Complete depths can include successive identical colors
- Complete thicknesses include 0s
- Independent truncated Gaussian (Tobit) random fields for each thickness

$$
\begin{aligned}
Z_j(s) &= \mu_j(W_j(s) - \tau_j)^{\beta_j} \quad \text{if} \quad W_j(s) > \tau_j \\
&= 0 \quad \text{otherwise}
\end{aligned}
$$

- $W_j(s)$ is a standard GRF with (stationary) covariance $C(\cdot; a_j)$
- $T_j(s) = T_{j-1}(s) + Z_j(s)$



Four simulations



Likelihood

Simplifying assumptions:

- Random fields $W_j(s)$ are independent
- Parameters of layers in same facies are identical:

$$\theta = (\mu_1, \ldots, \mu_K, \beta_1, \ldots, \beta_K, \tau_1, \ldots, \tau_K, a_1, \ldots, a_K)$$

- For each layer j, wells are re-ordered s.t. $\{Z_{j,i} > 0\} \Leftrightarrow \{i \leq n_j\}$

Conditional on the complete sequence of thicknesses, the likelihood of thickness j is

$$L_j(\theta; Z_{j,1}, \ldots, Z_{j,n}) = \phi_{n_j}(W_{j,1}, \ldots, W_{j,n_j}; \mathbf{0}, \boldsymbol{\Sigma}_j)\, \Phi_{n_j}(\tau_j, \ldots, \tau_j; \mathbf{m}_j, \mathbf{V_j})\, \frac{1}{\mu_j \beta_j} \left(\frac{Z_{j,i}}{\mu_j}\right)^{1-1/\beta_j}$$

where $\mathbf{m}_j$ and $\mathbf{V_j}$ are conditional expectations and variance, as provided by usual Kriging

But:

- Complete sequences of thicknesses are not known (incomplete data)
- Impossible to integrate out for large dataset, because of huge number of possibilities

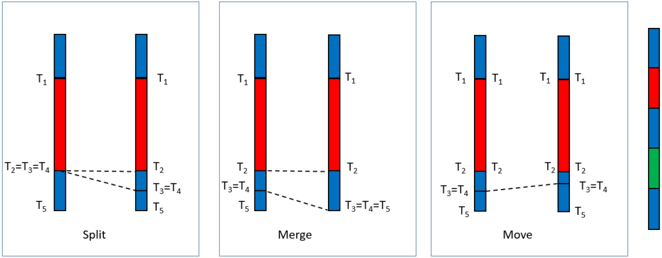
$\hookrightarrow$ Use MCMC and Bayesian setting

Metropolis-Hastings MCMC

- Iterative updates of parameters and complete sequence of thicknesses
- Special care for exploring all possible configurations ($\mathbf{T}, \mathbf{Z}$) compatible with the parent sequence and the data: split, merge and move
- Uninformative flat priors on parameters $\Phi(\tau_k)$ and $\beta_k \in [1/4, 4]$
- PC priors (Fulgstad, 2019) for (σ_k, a_k), with $\sigma_k^2 \propto \mu_k^2$

$$\pi(\sigma, a) = \lambda_a a^{-2} \exp(-\lambda_a / a) \lambda_\sigma \exp(-\lambda_\sigma \sigma), \quad (1)$$

Split, merge, move

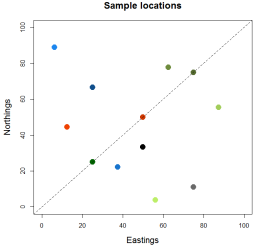


Theorem

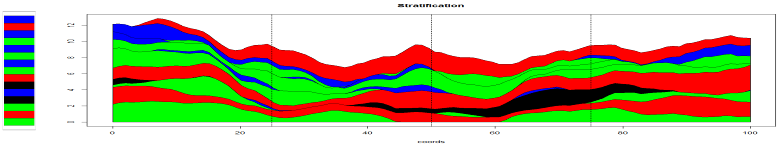
Thanks to Split, Merge and Move, all configurations can be explored. The Markov Chain is ergodic.

Simulation: Matérn(a; $\nu = 3/2$)

12 wells

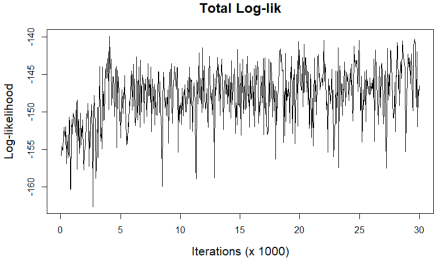


	black	red	blue	green
p	0.3	0.8	0.3	0.8
μ, β	1	1	1	1
a	20	20	10	10
$E[T]$	1.8	1.1	1.8	1.1.
$\hat{p}$	0.58	0.83	0.28	0.80
$\bar{T}$	0.31	1.31	0.62	1.25

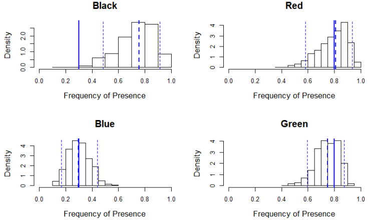


Random Walk Metropolis-Hastings MCMC

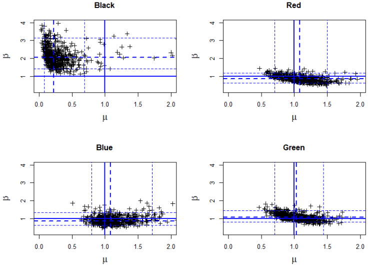
- 5000 iterations for burn-in
- 30000 iterations for sampling
- Sample every 20 iterations
- Proposals with uniform RWs; priors as described above
- Acceptance ratio $\simeq 0.5$

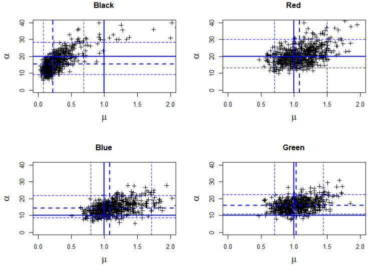


Parameter estimation

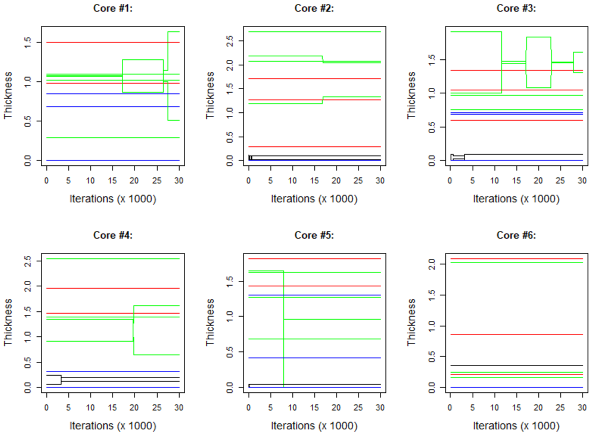


	black	red	blue	green
p	0.3	0.8	0.3	0.8
μ, β	1	1	1	1
a	20	20	10	10
$E[T]$	1.8	1.1	1.8	1.1.
$\hat{p}$	0.58	0.83	0.28	0.80
$\bar{T}$	0.31	1.31	0.62	1.25





Thicknesses in wells



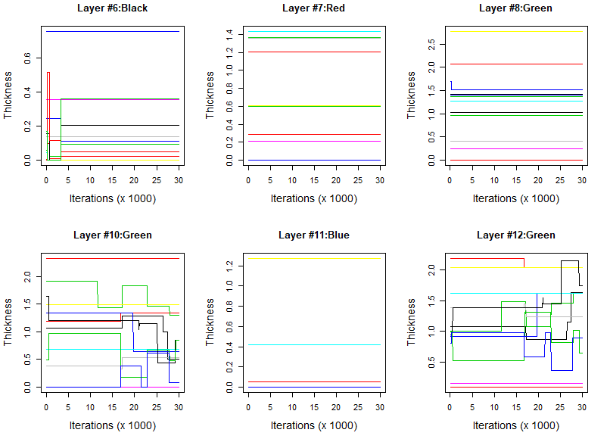
Thickness of different layers in wells ♯1 to ♯6 as a function of iterations. Layers are represented according to the color of their category

INRAE

Biostatistique BIO/Π & Processus Spatiaux

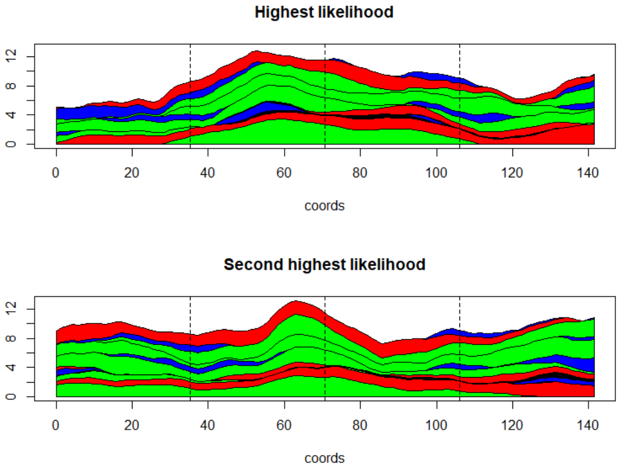
Università Ca'Foscari Venezia

Thicknesses of layers

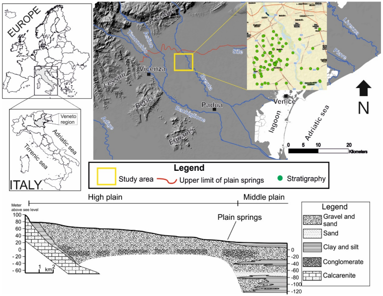


Thickness of layers ♯6 to ♯12 as a function of iterations. Each borehole is represented with a different color

Conditional simulations



Case study in the Veneto plaine

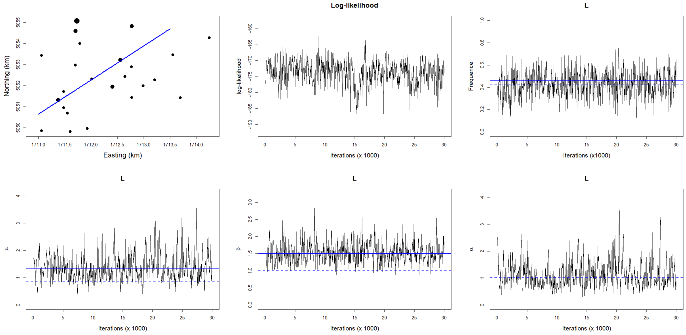


- 24 wells
- Parent sequence: L−S−G−L−A−G
- Matérn regularity: $\nu = 1/2$

	L	S	G	A	Overall
Number of records	22	18	12	3	55
Proportion of presence, $p_j(0)$	0.46	0.75	0.25	0.13	0.38
Average thickness (in m), $\bar{T}_j$	0.73	2.25	3.89	1.10	1.94
Initial value, $\tau_j(0)$	0.10	-0.67	0.67	1.15	–
initial value, $\mu_j(0)$	0.96	2.06	6.52	2.21	–

MCMC outputs

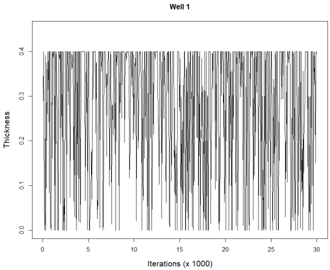
Category L

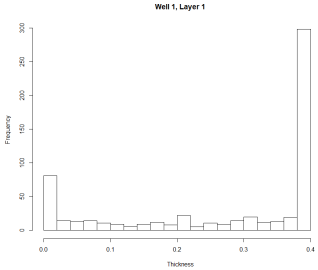


Location of the 24 boreholes analyzed in the Veneto dataset (top left); diameter is proportional to the number of thicknesses recorded (from 2 to 4); thick blue line: cross-section for conditional simulation. Then, from top to bottom and from left to right: total likelihood, *p*, μ, β and α as a function of iterations for category $\mathtt{L}$. Continuous lines: posterior medians. Dashed lines: initial values.

Thickness in Well ♯ 1

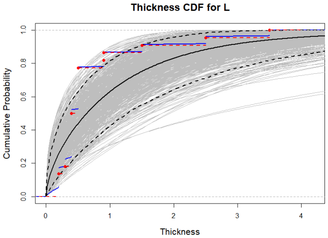
Oserved sequence is $\mathtt{L} - \mathtt{A} - \mathtt{G}$

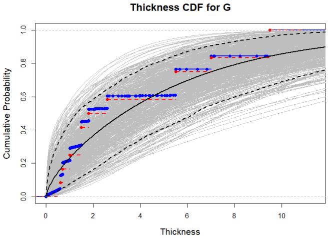




Thickness of the first layer $\mathtt{L}$ in borehole # 1. Left: as a function of iterations. Right: posterior histogram.

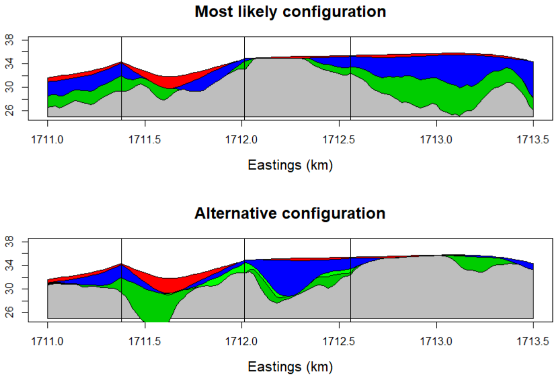
Posterior distributions of thickness





Thickness Cumulative Distributions (TCD). In gray: MCMC samples of the posterior theoretical TCD; Black continuous curve: pointwise posterior median TCD; Black dashed curves: pointwise posterior 0.05 and 0.95 posterior quantiles. Red dashed curve: TCD of the original data; Blue curve: TCD of the MCMC samples.

Conditional simulations



$L_1 = -162.5$; $L_2 = -171.3$ Notice that there are two different layers for G in the bottom cross-section.

To conclude

To sum up:

- Flexible model for deposition processes in geosciences
- Bayesian setting offers fair estimates of the parameters
- and coherent conditional simulations

And also:

- Can also be viewed as categorical function data modeling with spatial dependence
- In geology, depth can be converted to time through sedimentation → spatio-temporal model for categorical variables
- Together with a Markov Chain model on the parent sequence, offers a framework in many other situations: health, social mobility, ...

Allard, D., Fabbri, P., & Gaetan, C. (2021). Modeling and simulating depositional sequences using latent Gaussian random fields. *Mathematical Geosciences,* 53, 469-497.

The real conclusion

THANK YOU VERY MUCH

INRAE

Biostatistique BIO/Π & Processus Spatiaux

Università Ca'Foscari Venezia