A semi-parametric Spatially Informed Hawkes process (SPIN-Hawkes).

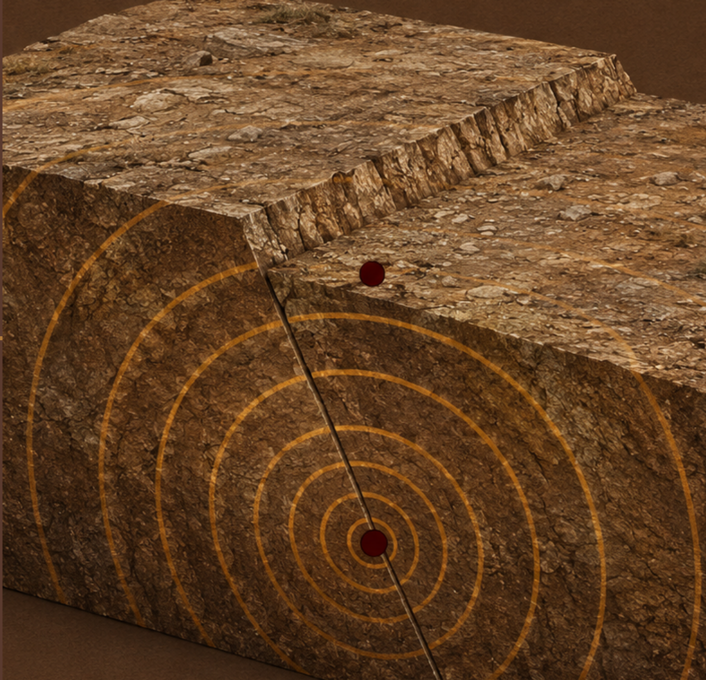
An application to french probabilistic seismic hazard analysis

Ibrahim SEYDI

Sophie DONNET · Merlin KELLER
Joseph MURÉ · Julien STOEHR

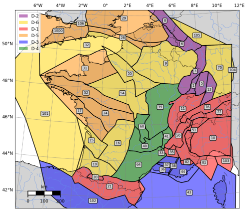
19/06/2026

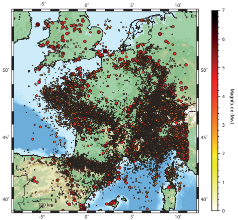
MIA PARIS-SACLAY SOLSTIS



Industrial context, motivations and limitations

This work contributes to the development of learning methods for the **selection and validation of seismic source models** used in EDF's seismic risk assessment.





***(Left)** Seismotectonic model of France proposed by EDF. **(Right)** French earthquake catalog for PSHA.*

Limitations of current models :

- **Zoning models :** too many zones $\Rightarrow$ highly uncertain parameters (data scarcity)
- **Zoneless models :** better data fit, but ignore physical/geological constraints

$\Rightarrow$ **Need for a "hybrid" model integrating both data and geological expertise, while properly accounting for uncertainties**

Seismic data and marked Hawkes process

Full seismic catalog main/after-shocks

$$\mathcal{D} = \{(t_i, x_i, y_i, m_i)\}_{i=1}^{N}, \qquad (x_i, y_i) \in \mathcal{X} \subset \mathbb{R}^2,\ t_i \in (0, T],\ m_i \in \mathbb{R}_+.$$

Conditional intensity

$$\lambda^\star(t, x, y, m \mid \mathcal{H}_t) \;=\; p_M(m) \;\times\; \lambda(t, x, y \mid \mathcal{H}_t),$$

$$\lambda(t, x, y \mid \mathcal{H}_t) \;=\; \underbrace{\mu(x, y)}_{\text{background}} + \underbrace{\sum_{j\,:\,t_j<t} \phi\big(t - t_j,\ x - x_j,\ y - y_j,\ m_j\big)}_{\text{self-exciting triggering}}.$$

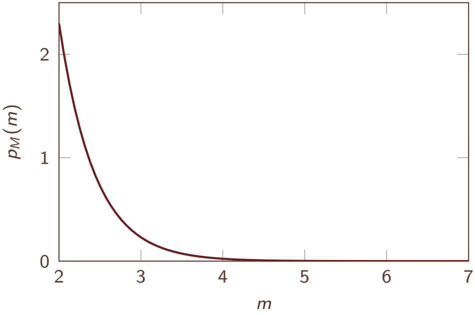
- $\mu(x, y)$: spatially varying rate of spontaneous background events
- ϕ : self-exciting contribution of each past event to future seismicity

Model — Magnitude distribution p_M

Gutenberg–Richter law (truncated exponential) :

$$p_M(m) = \frac{\beta\, e^{-\beta(m-m_c)}}{1 - e^{-\beta(m_{\max}-m_c)}} \mathbb{1}_{\{m_c \leq m \leq m_{\max}\}} \;, \qquad \beta = \ln(10)\, b$$

- $b \approx 1$ (seismology), so $\beta \approx 2.3$
- Independent of (t, x, y)
- Same for background and triggered events
- $\int_{m_c}^{m_{\max}} p_M(m)\, dm = 1 \Rightarrow$ separable from spatio-temporal part



Model — Background intensity μ

The background intensity is modeled as a **Spatially Structured sigmoidal Gaussian Cox process (SSGC)** :

$$\mu(x,y) = \tilde{\mu}(x,y)\,\sigma\big(f(x,y)\big) = \frac{\tilde{\mu}(x,y)}{1+e^{-f(x,y)}}$$

Baseline $\tilde{\mu}$: macroscopic structure

Piecewise constant over J sub-regions $\mathcal{X}_j$:

$$\tilde{\mu}(x,y) = \exp\left(\sum_{j=1}^{J} \varepsilon_j\, u_j(x,y)\right) = e^{\varepsilon_j}$$

for $(x,y) \in \mathcal{X}_j$, where u_j is the indicator of $\mathcal{X}_j$.

Latent GP f : residual fluctuations

Zero-mean GP with squared-exponential kernel :

$$f \mid v,\ell \sim \mathcal{GP}\Big(0,\ k(\cdot,\cdot \mid v,\ell)\Big)$$

$$k = v\exp\left(-\frac{\|s-s'\|^2}{2\ell^2}\right)$$

v : marginal variance, ℓ : length-scale.

$\tilde{\mu}$ captures the prior macroscopic spatial structure ; f models local residual fluctuations around this baseline.

Model — Triggering kernel ϕ

The excitation kernel ϕ decomposes into three factors :

$$\phi(t - t_j, x - x_j, y - y_j, m_j) = \phi_m(m_j)\, \phi_t(t - t_j)\, \phi_s(x - x_j, y - y_j \mid m_j)$$

Productivity

$$\phi_m(m_j) = A\, e^{\alpha(m_j - m_c)}$$

Exponential scaling with parent magnitude

Modified Omori–Utsu

$$\phi_t(\Delta t) = \frac{(p-1)\, c^{p-1}}{(\Delta t + c)^p}$$

Temporal aftershock decay ($c > 0$ and $p > 1$)

Spatial dispersion

$$\phi_s = \frac{q-1}{\pi R(m_j)} \left[1 + \frac{r^2}{R(m_j)}\right]^{-q}$$

$R(m_j) = d\, e^{\gamma(m_j - m_c)}$, $d > 0$, $\gamma > 0$, $q > 1$

- ϕ_t : normalized density on $\mathbb{R}_+$
- ϕ_s : normalized density on $\mathbb{R}^2$; radius scales with magnitude
- Total aftershock rate : $\iint \phi\, dx\, dy\, dt = A\, e^{\alpha(m_j - m_c)}$

Towards Bayesian inference — Structured Gaussian prior on zonal baselines

The zonal coefficients $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_J)^\top$ are chosen to be **not independent**. They are endowed with a centered Gaussian prior that encodes spatial proximity :

$$\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}_J, \boldsymbol{\Sigma}_\varepsilon), \qquad (\boldsymbol{\Sigma}_\varepsilon)_{jj'} = \delta_0 \exp\left(-\frac{\|c_j - c_{j'}\|^2}{2\delta_1^2}\right)$$

where $c_j = (c_x^j, c_y^j)$ are the **centroids** of region $\mathcal{X}_j$.

- $\delta_0 > 0$: marginal variance
- $\delta_1 > 0$: correlation length-scale

This regularizes the baseline by inducing correlations between adjacent zones, tempering the strength of the prior belief

Towards Bayesian inference — Priors on remaining parameters

Weakly informative Gamma priors enforce positivity :

Parameter	Constraint	Prior	Rationale
A, α, c, d, γ	> 0	$\mathrm{Gamma}(1,1) = \mathrm{Exp}(1)$	diffuse, mean 1
p	> 1	$p - 1 \sim \mathrm{Gamma}(1,1)$	shifted support
q	> 1	$q - 1 \sim \mathrm{Gamma}(1,1)$	shifted support
β	> 0	$\mathrm{Gamma}(2,1)$	mode ≈ 1, mean 2

GP kernel hyperparameters

(v, ℓ) calibrated prior to sampling via a heuristic based on kriging the following predictions of the latent GP :

$$z_i = \frac{2\,N\,|\mathcal{X}_j|}{N_j}\,\hat{p}(x_i, y_i) - 2, \qquad (x_i, y_i) \in \mathcal{X}_j, \tag{1}$$

using a linearization of sigmoïd fonction, and then can be **held fixed** throughout inference.

Towards Bayesian inference — Posterior & Likelihood

The posterior is : $p(\boldsymbol{\theta} \mid \mathcal{D}) \propto L_N(\boldsymbol{\theta})\, p(\varepsilon)\, p(f \mid v, \ell)\, p(v, \ell)\, p(\beta)\, p(\boldsymbol{\theta}_\phi)$. The likelihood of $\mathcal{D}$ under the Hawkes process model factorizes as :

$$L_N(\boldsymbol{\theta}) = \underbrace{\prod_{i=1}^{N} p_M(m_i)}_{\text{magnitudes}} \times \underbrace{\left[\prod_{i=1}^{N} \lambda(t_i, x_i, y_i \mid \mathcal{H}_{t_i})\right] \exp\left(-\int_0^T \int_{\mathcal{X}} \lambda \, dx \, dy \, dt\right)}_{\text{spatio-temporal}}$$

Towards Bayesian inference — Posterior & Likelihood

The posterior is : $p(\boldsymbol{\theta} \mid \mathcal{D}) \propto L_N(\boldsymbol{\theta})\, p(\varepsilon)\, p(f \mid v, \ell)\, p(v, \ell)\, p(\beta)\, p(\boldsymbol{\theta}_\phi)$. The likelihood of $\mathcal{D}$ under the Hawkes process model factorizes as :

$$L_N(\boldsymbol{\theta}) = \underbrace{\prod_{i=1}^{N} p_M(m_i)}_{\text{magnitudes}} \times \underbrace{\left[\prod_{i=1}^{N} \lambda(t_i, x_i, y_i \mid \mathcal{H}_{t_i})\right] \exp\left(-\int_0^T \int_{\mathcal{X}} \lambda \, dx\, dy\, dt\right)}_{\text{spatio-temporal}}$$

The integrated intensity splits into background and triggering (recall $\lambda = \mu + \sum \phi$) :

Triggering integral

The triggering term is tractable by Fubini :

$$\int_0^T \int_{\mathcal{X}} \sum_{j: t_j < t} \phi \, dx\, dy\, dt = \sum_{j=1}^{N} A\, e^{\alpha(m_j - m_c)}\, T_j\, S_j$$

$$T_j = 1 - \left(\frac{c}{T - t_j + c}\right)^{p-1}, \qquad S_j \approx 1$$

Towards Bayesian inference — Posterior & Likelihood

The posterior is : $p(\boldsymbol{\theta} \mid \mathcal{D}) \propto L_N(\boldsymbol{\theta})\, p(\varepsilon)\, p(f \mid v, \ell)\, p(v, \ell)\, p(\beta)\, p(\boldsymbol{\theta}_\phi)$. The likelihood of $\mathcal{D}$ under the Hawkes process model factorizes as :

$$L_N(\boldsymbol{\theta}) = \underbrace{\prod_{i=1}^{N} p_M(m_i)}_{\text{magnitudes}} \times \underbrace{\left[\prod_{i=1}^{N} \lambda(t_i, x_i, y_i \mid \mathcal{H}_{t_i})\right] \exp\left(-\int_0^T \int_{\mathcal{X}} \lambda \, dx\, dy\, dt\right)}_{\text{spatio-temporal}}$$

The integrated intensity splits into background and triggering (recall $\lambda = \mu + \sum \phi$) :

Triggering integral

The triggering term is tractable by Fubini :

$$\int_0^T \int_{\mathcal{X}} \sum_{j:t_j<t} \phi \, dx\, dy\, dt = \sum_{j=1}^{N} A\, e^{\alpha(m_j - m_c)}\, T_j\, S_j$$

$$T_j = 1 - \left(\frac{c}{T - t_j + c}\right)^{p-1}, \qquad S_j \approx 1$$

Background integral

$$\int_0^T \int_{\mathcal{X}} \mu \, dx\, dy\, dt = T \sum_{j=1}^{J} e^{\varepsilon_j} \int_{\mathcal{X}_j} \sigma\big(f(x, y)\big)\, dx\, dy$$

No closed form when f is a GP
$\Longrightarrow$ *source of intractability*

Towards Bayesian inference — Data augmentation strategy

A **two-stage** augmentation strategy to :

1. Decouple background (μ) from triggering (ϕ)
2. Render the background component conjugate to the GP prior

Stage 1
Latent branching
$Z = (z_1, \ldots, z_N)$

$\longrightarrow$

Stage 2
Pólya–Gamma ω
$+$ marked PP $\pi_{\widehat{\mathcal{S}}}$

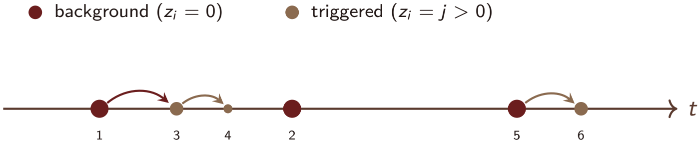
$\longrightarrow$

Result
Augmented likelihood
quadratic in f

Data augmentation — Stage 1 : Latent branching structure

Introduce latent branching variables :

$$Z = (z_1, \dots, z_N), \qquad z_i = \begin{cases} 0 & \text{if event } i \text{ is a background event,} \\ j > 0 & \text{if event } i \text{ was triggered by event } j. \end{cases}$$



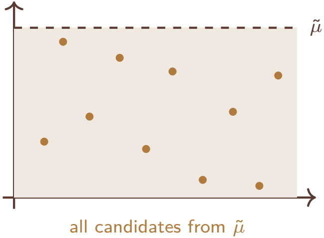
This **factorizes the augmented likelihood** into :

$$L_N = \underbrace{L_{\text{background}}(\varepsilon, f)}_{\text{depends on } \mu} \times \underbrace{L_{\text{self-excited}}(\boldsymbol{\theta}_\phi)}_{\text{depends on } \phi} \times p(Z)$$

Data augmentation — Stage 2 : Marked PP & PG augmentations

Classical thinning (simulation)

Start from $\tilde{\mu}$, **reject** to get $\tilde{\mu}\,\sigma(f)$



Step 1 : draw all candidates uniformly under $\tilde{\mu}$

Step 2 : the curve $\tilde{\mu}\,\sigma(f)$ sets the acceptance boundary

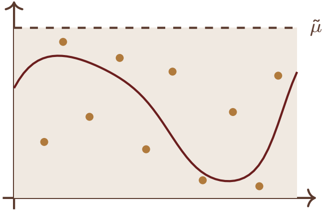
Step 3 : accept with prob. $\sigma(f)$, reject otherwise

Result : Inhomogeneous PP $(\tilde{\mu} \cdot \sigma(f))$

Data augmentation — Stage 2 : Marked PP & PG augmentations

Classical thinning (simulation)

Start from $\tilde{\mu}$, **reject** to get $\tilde{\mu}\,\sigma(f)$



curve $\tilde{\mu}\,\sigma(f)$ separates the two regions

Step 1 : draw all candidates uniformly under $\tilde{\mu} \cdot (\mathcal{X})$

Step 2 : the curve $\tilde{\mu}\,\sigma(f)$ sets the acceptance boundary

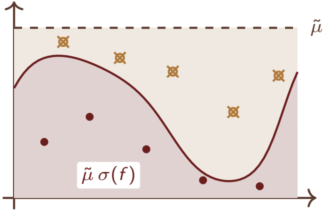
Step 3 : accept with prob. $\sigma(f)$, reject otherwise

Result : Inhomogeneous PP $(\tilde{\mu} \cdot \sigma(f))$

Data augmentation — Stage 2 : Marked PP & PG augmentations

Classical thinning (simulation)

Start from $\tilde{\mu}$, **reject** to get $\tilde{\mu}\,\sigma(f)$



keep •, discard ×

Step 1 : draw all candidates uniformly under $\tilde{\mu} \cdot (\mathcal{X})$

Step 2 : the curve $\tilde{\mu}\,\sigma(f)$ sets the acceptance boundary

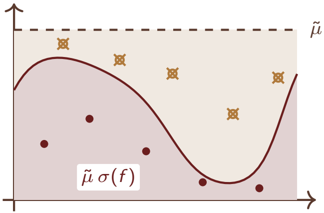
Step 3 : accept with prob. $\sigma(f)$, reject otherwise

Result : Inhomogeneous PP $(\tilde{\mu} \cdot \sigma(f))$

Data augmentation — Stage 2 : Marked PP & PG augmentations

Classical thinning (simulation)

Start from $\tilde{\mu}$, **reject** to get $\tilde{\mu}\,\sigma(f)$



keep •, discard ×

Step 1 : draw all candidates uniformly under $\tilde{\mu} \cdot (\mathcal{X})$

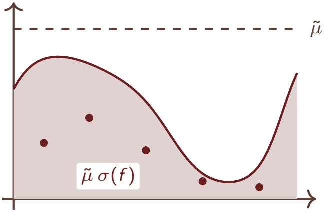
Step 2 : the curve $\tilde{\mu}\,\sigma(f)$ sets the acceptance boundary

Step 3 : accept with prob. $\sigma(f)$, reject otherwise

Result : Inhomogeneous PP $(\tilde{\mu} \cdot \sigma(f))$

Marked Poisson process augmentation

Start from data, **add back** latent events



• observed data only

Start : only observed background events • are known

Identify : the thinned region $\tilde{\mu}\,\sigma(-f)$ above the curve

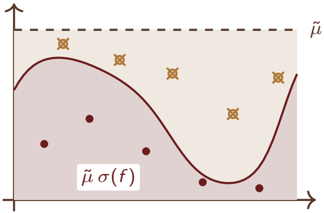
Augment : sample latent • from $\tilde{\mu}\,\sigma(-f)$

Result : Union $\longrightarrow$ homogeneous PP$(\tilde{\mu})$ + mark P–G $\Rightarrow$ Tractable integral & conjugacy GP

Data augmentation — Stage 2 : Marked PP & PG augmentations

Classical thinning (simulation)

Start from $\tilde{\mu}$, **reject** to get $\tilde{\mu}\,\sigma(f)$



keep ●, discard ×

Step 1 : draw all candidates uniformly under $\tilde{\mu} \cdot (\mathcal{X})$

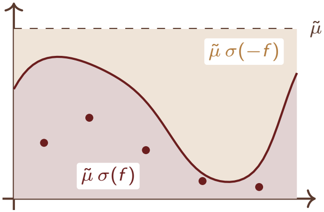
Step 2 : the curve $\tilde{\mu}\,\sigma(f)$ sets the acceptance boundary

Step 3 : accept with prob. $\sigma(f)$, reject otherwise

Result : Inhomogeneous PP $(\tilde{\mu} \cdot \sigma(f))$

Marked Poisson process augmentation

Start from data, **add back** latent events



thinned region : $\tilde{\mu}\,\sigma(-f)$

Start : only observed background events ● are known

Identify : the thinned region $\tilde{\mu}\,\sigma(-f)$ above the curve

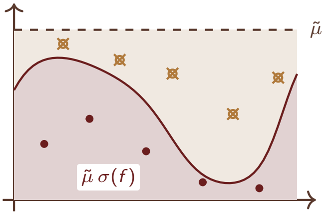
Augment : sample latent ● from $\tilde{\mu}\,\sigma(-f)$

Result : Union $\longrightarrow$ homogeneous PP$(\tilde{\mu})$ + mark P–G $\Rightarrow$ Tractable integral & conjugacy GP

Data augmentation — Stage 2 : Marked PP & PG augmentations

Classical thinning (simulation)

Start from $\tilde{\mu}$, **reject** to get $\tilde{\mu}\,\sigma(f)$



keep •, discard ×

Step 1 : draw all candidates uniformly under $\tilde{\mu} \cdot (\mathcal{X})$

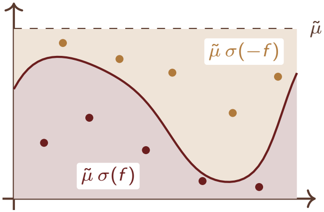
Step 2 : the curve $\tilde{\mu}\,\sigma(f)$ sets the acceptance boundary

Step 3 : accept with prob. $\sigma(f)$, reject otherwise

Result : Inhomogeneous PP $(\tilde{\mu} \cdot \sigma(f))$

Marked Poisson process augmentation

Start from data, **add back** latent events



• ∪ • = PP($\tilde{\mu}$)

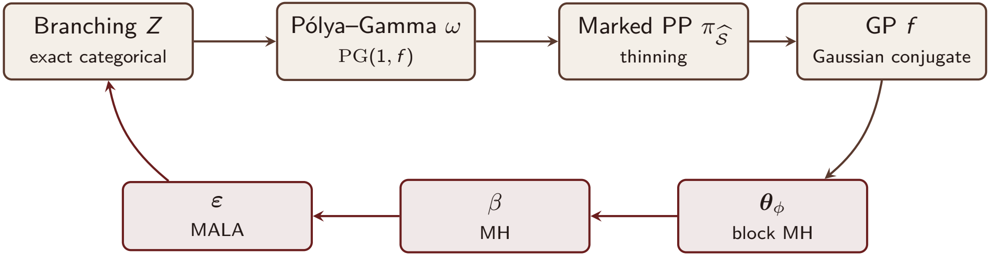
Start : only observed background events • are known

Identify : the thinned region $\tilde{\mu}\,\sigma(-f)$ above the curve

Augment : sample latent • from $\tilde{\mu}\,\sigma(-f)$

Result : Union $\longrightarrow$ homogeneous PP($\tilde{\mu}$) + mark P–G $\Rightarrow$ Tractable integral & conjugacy GP

Bayesian Inference — Gibbs sampling

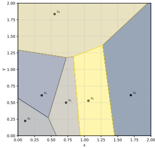


- **Brown :** Conjugate updates (closed-form or exact sampling)
- **Red :** Non-conjugate (MALA or Metropolis–Hastings)

Exp. 1 : SSGC vs SGCP vs KDE - Profile 1, Setting A

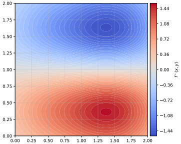
Three models compared : **SSGC** (our model, structured $\tilde{\mu}$ + prior $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon})$) / **SGCP** ($\propto$ Donner and Opper (2018), $J=1$, constant $\tilde{\mu}$) / **KDE** (with Silverman bandwidth)

Voronoi partition $(10, 1, 2, 10, 8, 2)$



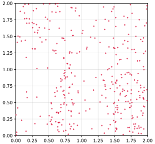
&

Latent field $f_A^\star$



$\Longrightarrow$

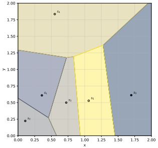
Dataset generated



Exp. 1 : SSGC vs SGCP vs KDE - Profile 1, Setting A

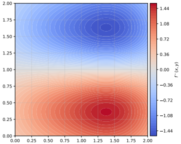
Three models compared : **SSGC** (our model, structured $\tilde{\mu}$ + prior $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$) / **SGCP** ($\propto$ Donner and Opper (2018), $J = 1$, constant $\tilde{\mu}$) / **KDE** (with Silverman bandwidth)

Voronoi partition
$(10, 1, 2, 10, 8, 2)$



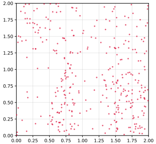
&

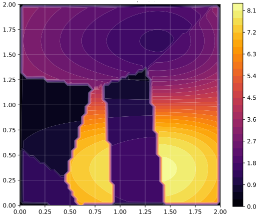
Latent field $f_A^\star$



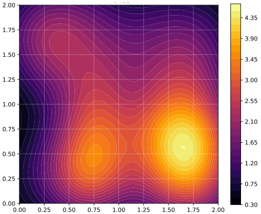
$\Longrightarrow$

Dataset generated

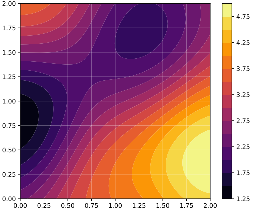




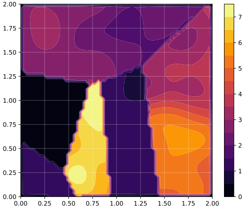
True $\mu^\star$



KDE



SGCP

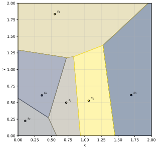


SSGC

Exp. 1 : SSGC vs SGCP vs KDE - Profile 1, Setting B

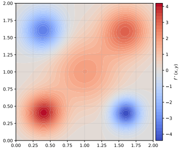
Three models compared : **SSGC** (our model, structured $\tilde{\mu}$ + prior $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$) / **SGCP** ($\propto$ Donner and Opper (2018), $J=1$, constant $\tilde{\mu}$) / **KDE** (with Silverman bandwidth)

Voronoi partition $(10, 1, 2, 10, 8, 2)$



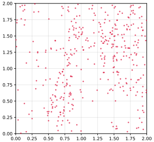
&

Latent field $f_B^\star$



$\Longrightarrow$

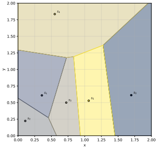
Dataset generated



Exp. 1 : SSGC vs SGCP vs KDE - Profile 1, Setting B

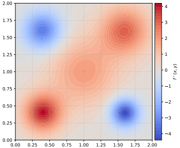
Three models compared : **SSGC** (our model, structured $\tilde{\mu}$ + prior $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon})$) / **SGCP** ($\propto$ Donner and Opper (2018), $J=1$, constant $\tilde{\mu}$) / **KDE** (with Silverman bandwidth)

Voronoi partition $(10, 1, 2, 10, 8, 2)$



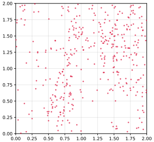
&

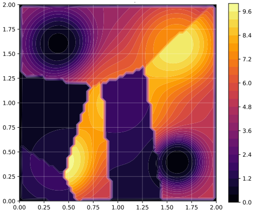
Latent field $f_B^\star$



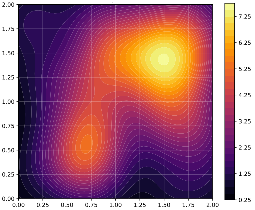
$\Longrightarrow$

Dataset generated

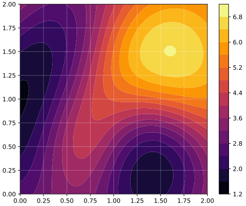




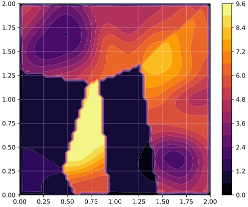
True $\mu^\star$



KDE



SGCP



SSGC

Exp 1 : Quantitative summary

Reconstruction metrics — SSGC vs SGCP vs KDE

		RMSE			MAE			$\overline{\text{CRPS}}$		
Profile	**Setting**	SSGC	SGCP	KDE	SSGC	SGCP	KDE	SSGC	SGCP	KDE
P1	A	**0.943**	2.036	2.067	**0.608**	1.581	1.528	**0.596**	1.577	n/a
	B	**0.914**	2.155	2.040	**0.722**	1.732	1.672	**0.698**	1.726	n/a

*n/a : CRPS not defined for KDE ; **Bold** : best result per row per metric*

- **SSGC** performs better while allowing to take acount uncertainties
- Better modelisation can have **significant consequences** in risk assessment

Exp 2 : Robustness under prior misspecification

Scenario	True partition	Inference partition
M1 : Superfluous	$J^\star = 1$ (homogeneous)	$J = 6$
M2 : Wrong partition	$J^\star = 6$	$J = 5$
M3 : Missing zones	$J^\star = 6$	$J = 1$ (homogeneous)

Reconstruction metrics — oracle vs misspecified (Settings A & B)

		Setting A			Setting B		
Scen.	**Model**	RMSE	MAE	CRPS	RMSE	MAE	CRPS
M1	Oracle (SGCP, $J=1$)	**0.147**	**0.109**	**0.106**	**0.734**	**0.591**	**0.588**
	SSGC superfluous ($J=6$)	0.882	0.597	0.541	1.243	0.983	0.897
M2	Oracle (SSGC, $J=6$)	**0.943**	**0.616**	**0.604**	**0.917**	**0.721**	**0.697**
	Wrong partition ($J=5$)	1.727	1.288	1.234	1.980	1.597	1.506
M3	Oracle (SSGC, $J=6$)	**0.952**	**0.620**	**0.607**	**0.949**	**0.751**	**0.728**
	Missing zones ($J=1$)	2.038	1.582	1.578	2.160	1.739	1.734

- **Hierarchy** : oracle $\gg$ wrong partition $\gg$ missing zones $>$ superfluous zones

Take-Home messages & Perspectives

The SPIN-Hawkes / SSGC model

- Spatially informative background via structured baseline $\tilde{\mu}(x, y)$ with centroid-based spatial correlations
- Zero-mean residual GP isolating local fluctuations
- Self-excited part modelled parametrically
- Two-stage augmentation (branching $+$ marked PP/Pólya–Gamma) $\Rightarrow$ efficient Gibbs sampler
- Promising results from tests on simulated data and on a part of french catalog

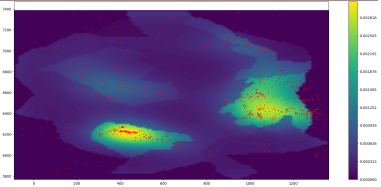
Take-Home messages & Perspectives

The SPIN-Hawkes / SSGC model

- Spatially informative background via structured baseline $\tilde{\mu}(x, y)$ with centroid-based spatial correlations
- Zero-mean residual GP isolating local fluctuations
- Self-excited part modelled parametrically
- Two-stage augmentation (branching + marked PP/Pólya–Gamma) $\Rightarrow$ efficient Gibbs sampler
- Promising results from tests on simulated data and on a part of french catalog

Limitation & Perspectives

- **GP scalability** : $\mathcal{O}(N_f^3)$ cost
- Sparse GP needed for large datasets
- Development of a **variational inference**
- Validation on full **real seismological catalogs**

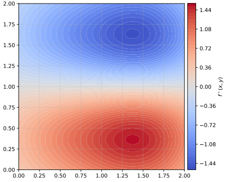


Appendix — Experimental setup

Domain : $\mathcal{X} = [0,2]^2$, Voronoi partitions

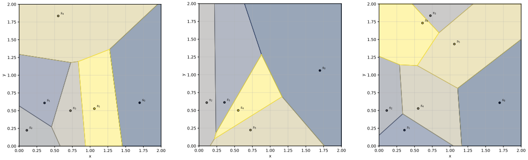
Two latent fields

- **A** : smooth, low amplitude, range $[-1.5, 1.5]$
- **B** : sharp peaks with sigmoid saturation, range $[-3.5, 4.0]$



Three zonal contrast profiles

Profile	J	$\mu^\star_{\max}/\mu^\star_{\min}$	**Values**
P1	6	10	$(10, 1, 2, 10, 8, 2)$
P2	5	2	$(3.5, 2, 4, 3, 2.5)$
P3	7	20	$(20, 1, 1, 1, 20, 1, 1)$



Appendix — Evaluation metrics of experiments

Point-based metrics

$$\mathrm{RMSE} = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(\hat{\mu}_m - \mu_m^\star)^2}$$

$$\mathrm{MAE} = \frac{1}{M}\sum_{m=1}^{M}\left|\hat{\mu}_m - \mu_m^\star\right|$$

Distribution-based metric

$$\overline{\mathrm{CRPS}} = \frac{1}{M}\sum_{m=1}^{M}\mathrm{CRPS}(F_m, \mu_m^\star)$$

$$\mathrm{CRPS}(F, y) = \int_{-\infty}^{+\infty}\left(F(z) - \mathbf{1}\{z \geq y\}\right)^2 dz$$

MCMC diagnostics

$\hat{R}$,

Effective Sample Size

MALA acceptation rate

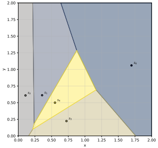
Spatial relative error

$$\Delta_\mu(x_m, y_m) = \frac{|\hat{\mu}_m - \mu_m^\star|}{\mu_m^\star}$$

Appendix — SSGC vs SGCP vs KDE - Profile 2, Setting A

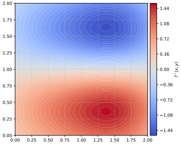
Three models compared : **SSGC** (our model, structured $\tilde{\mu}$ + prior $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$) / **SGCP** ($\propto$ Molkenthin and al. (2022), $J = 1$, constant $\tilde{\mu}$) / **KDE** (with Silverman bandwidth)

Voronoi partition



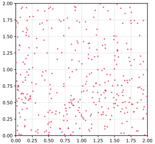
&

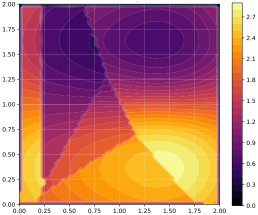
Latent field $f_A^\star$



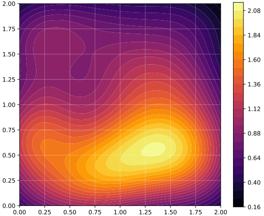
$\Longrightarrow$

Dataset generated

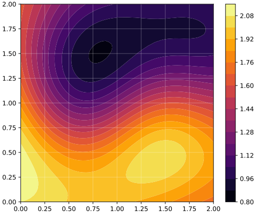




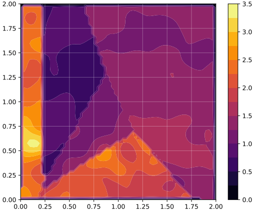
True $\mu^\star$



KDE



SGCP

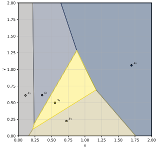


SSGC

Appendix — SSGC vs SGCP vs KDE - Profile 2, Setting B

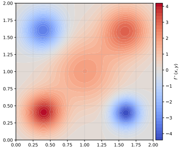
Three models compared : **SSGC** (our model, structured $\tilde{\mu}$ + prior $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon})$) / **SGCP** ($\propto$ Molkenthin and al. (2022), $J=1$, constant $\tilde{\mu}$) / **KDE** (with Silverman bandwidth)

Voronoi partition



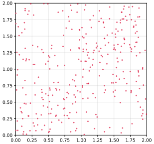
&

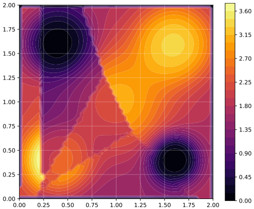
Latent field $f_A^\star$



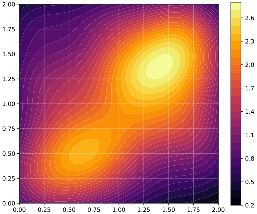
$\Longrightarrow$

Dataset generated

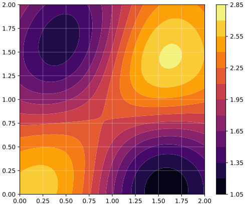




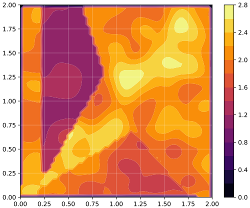
True $\mu^\star$



KDE



SGCP



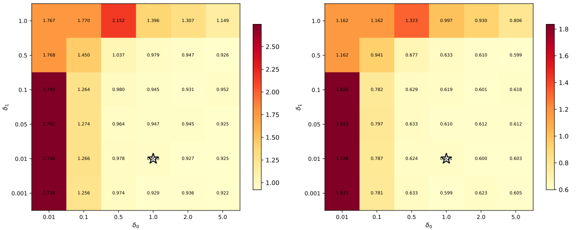
SSGC

Appendix — Sensitivity to (δ_0, δ_1)

Grid : $\delta_0 \in \{0.01, 0.1, 0.5, 1.0, 2.0, 5.0\}$, $\delta_1 \in \{0.001, 0.01, 0.05, 0.1, 0.5, 1.0\}$ (36 configs).

Regime	**Effect**
Small δ_0	Over-shrinkage : $\tilde{\mu} \approx 1$
Large δ_0, small δ_1	Independent zones
Large δ_0, large δ_1	Uniform shrinkage
Moderate (reference $\star$)	Intended regime

- Small δ_0 systematically degrades RMSE/MAE
- Large plateau of good performance for $\delta_0 \in [0.5, 5.0]$, $\delta_1 \leq 0.1$
- $\delta_1 = 1.0$ degrades results : uniform shrinkage incompatible with high-contrast profiles



RMSE (left) & MAE (right) over the (δ_0, δ_1)-grid.
$\star$: reference configuration $(\delta_0, \delta_1) = (1.0, 0.01)$.