

Inférence INLA* pour modèles spatiotemporels

* *Integrated Nested Laplace Approximations*

Thomas Opitz
BioSP, INRA Avignon

Journée RESSTE
RESeau Statistiques pour données Spatio-TEmporelles

Paris, 11/05/2015

Objectifs pour cet exposé

- ▶ “*INLA*, c’est quoi ? ”
 - ▶ modèles à **processus Gauss–Markov latents**
 - ▶ **approximation analytique** des lois a posteriori univariées
- ▶ champs **Gauss–Markov en espace continu**
 - ▶ champs Gauss–Markov solutions d’EDPs stochastiques
 - ▶ triangulation et éléments finis pour des solutions approximatives
- ▶ **approches spatiotemporelles avec INLA**

Spatial vs. spatiotemporel

- ▶ temps \times espace $\mathbb{R}^2 \neq \mathbb{R}^3$
 \rightsquigarrow **flèche temporelle**
- ▶ **tendances temporelles**
(saisonnalités, dérives, épidémies, changement climatique, ...)
- ▶ typiquement **pas de copies i.i.d.** dans le cadre spatiotemporel
- ▶ la **dimension** peut être **très grande** dans le cadre spatiotemporel

INLA

Champs Gauss–Markov

Approche EDPS

Exemple d'application

Conclusion et perspectives

Modèles hiérarchiques à processus gaussiens latents

- ▶ lois marginales **gaussiennes** souvent **pas réalistes**
 - ↪ données nonnégatives/discrètes/zéro-inflatéés/...
 - ⚠ structures de dépendance souples et pilotables pour l'inférence ???
- ▶ structures **hiérarchiques**
 - ▶ champs gaussiens latents pour modéliser la **dynamique** du processus
 - ▶ reproduire les **processus physiques**
 - ▶ erreurs de mesure

lci :

$$\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$$

hyperparamètres

$$\mathbf{x} \mid \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \mathbf{Q}(\boldsymbol{\theta})^{-1})$$

champ gaussien latent

$$\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta} \sim \prod_i \pi(y_i \mid \eta_i(\mathbf{x}), \boldsymbol{\theta})$$

observations

INLA en bref

[Rue et al., 2009]

- ▶ modèle de **régression additif** : $\mathbb{E}Y = g(\eta)$
- ▶ le **prédicteur** η est un processus **gaussien**
 - ▶ peut être complexe : hiérarchique, spatiotemporel, ...
 - ▶ processus **latent** \rightsquigarrow cadre **bayésien**
 - ▶ η est de **grande dimension**
 \rightsquigarrow utiliser des processus **Gauss-Markov** ($Q = C^{-1}$ est creuse)
- ▶ **approximations analytiques** (Laplace) dans les densités a posteriori
- ▶ un petit nombre de hyperparamètres nongaussiens peut être estimés (dispersion, variance, etc.)
- ▶ **algorithmes numériques efficaces** (GMRFLib, R-INLA)

Construction de modèles

données y_i **sachant** **prédicteur** η_i i.i.d. \sim **famille exponentielle**

- ▶ **trois briques de base :**
 - ▶ **vraisemblance** des données et **lien** $\mathbb{E}Y_i \mid \eta_i = g(\eta_i)$
 - ▶ **modèle latent** gaussien pour η_i
 - ▶ **lois a priori** des hyperparamètres
- ▶ vraisemblances disponibles en R-INLA :
gaussien, gamma, loi extrême, survie (Weibull, expontiel, Cox, ...),
Poisson/binomiale (+Zéro-Inflation), t, ...
- ▶ modèles latents disponibles en R-INLA :
i.i.d., linéaire, marches aléatoires, autorégressif, modèles de Besag spatiaux
(Gauss–Markov), champ gaussien Matérn (**approche EDPS**), modèle
mixte, “user-defined”
+ emboîtement (additif)
+contraintes $Ax = e$

Quelques exemples d'application

- ▶ modélisation **spatiotemporelle** de la concentration de particules fines [Cameletti et al., 2013]
- ▶ processus de **Cox log-gaussiens** : en écologie [Illian et al., 2012] ; pour les feux de forêts [Serra et al., 2014]
- ▶ épidémiologie [Schrödle and Held, 2011, Schrödle et al., 2012]
- ▶ régression quantile [Yue and Rue, 2011]
- ▶ modèles GLMM [Fong et al., 2010]
- ▶ survie [Martino et al., 2011]
- ▶ ...

Densité jointe a posteriori

Notons

- ▶ $\mathbf{x} = (x_1, \dots, x_{m_x})$ le **vecteur des variables gaussiennes** déterminant $\boldsymbol{\eta}$
- ▶ $\boldsymbol{\theta} = (\boldsymbol{\theta}_y, \boldsymbol{\theta}_x)$ le vecteur des **hyperparamètres** du modèle

Alors

- ▶ $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}(\boldsymbol{\theta}_x)^{-1})$
- ▶ **formule de Bayes** \Rightarrow densité a posteriori

$$\pi(\mathbf{x}, \boldsymbol{\theta} \mid \mathbf{y}) \propto \pi(\boldsymbol{\theta}) \times |\mathbf{Q}(\boldsymbol{\theta}_x)|^{0.5} \exp \left(-0.5 \mathbf{x}' \mathbf{Q}(\boldsymbol{\theta}_x) \mathbf{x} + \sum_i \log \pi(y_i \mid \eta_i, \boldsymbol{\theta}_y) \right)$$

- ▶ typiquement,
 m_x est grand (10^2 à 10^6),
 m_θ petit (≤ 10)
- ▶ m_θ doit être petit pour INLA

Densités marginales a posteriori

$$\pi(x_i | y) = \int \pi(x, \theta | y) d\mathbf{x}_{-i} d\theta = \int \pi(x_i | \theta, y) \pi(\theta | y) d\theta$$

$$\pi(\theta_j | y) = \int \pi(x, \theta | y) d\mathbf{x} d\theta_{-j} = \int \pi(\theta | y) d\theta_{-j}$$

L'étude de ces deux densités est cruciale :

- ▶ θ_j : souvent des **paramètres interprétables**, importants
- ▶ x_i : **cartes spatiales** ou **spatiotemporelles** de la moyenne, de la variance, de quantiles, etc. du processus latent

⚠ pour calculer ces densités, il faut **intégrer par rapport à x**

↪ **impossible** en grande dimension !

Approximation de Laplace pour intégrer par rapport à x

Soit x_0 le maximum global unique d'une fonction g . Alors

$$\begin{aligned} & \int_{-\infty}^{\infty} \exp(ng(x)) dx \\ & \stackrel{n \gg 0}{\approx} \int_{-\infty}^{\infty} \exp(ng(x_0) + 0.5n(x - x_0)'H(g)(x_0)(x - x_0)) dx \\ & = \left(\frac{2\pi}{n}\right)^{d/2} |H(g)(x_0)|^{-1/2} \exp(ng(x_0)) \end{aligned}$$

- ▶ déterminer x_0 et $H(g)(x_0)$ par un algorithme de **Newton–Raphson**
- ▶ dans le contexte statistique, n peut être le nombre de copies i.i.d.
- ▶ INLA : souvent **pas de copies i.i.d.**, mais bonne approximation quand même

Integrated Nested Laplace Approximations

[Rue and Martino, 2007, Rue et al., 2009]

- ▶ **“Integrated”** : schéma d’intégration numérique par rapport à θ
- ▶ **“Laplace”** : pour intégrer par rapport à x
 - ▶ approcher $\pi(x \mid \theta, y)$ par une densité gaussienne $\tilde{\pi}_G(x \mid y, \theta)$
 - ▶ Laplace + intégration numérique : $\tilde{\pi}(\theta_j \mid y) = \int \tilde{\pi}(\theta \mid y) d\theta_{-j}$
 \rightsquigarrow bonne approximation $\tilde{\pi}(\theta_j \mid y)$ pour les modèles courants
- ▶ **“Nested”** : appliquer encore Laplace pour intégrer par rapport à x_{-i} dans $\pi(x_i \mid y)$
 + d’autres approximations et astuces numériques $\rightsquigarrow \tilde{\pi}(x_i \mid \theta, y)$

INLA vs. MCMC

- ▶ **MCMC** : méthode “par défaut”, approximation par **simulation**
 - ++ OpenBUGS, JAGS, Stan, ...
 - + simulation de réalisations selon la loi a posteriori $\pi(\mathbf{x}, \boldsymbol{\theta} \mid \mathbf{y})$
 - + étude des lois a posteriori multivariées facile
 - lenteur
 - problèmes de convergence et instabilités, e.g. mise à jour de $\boldsymbol{\theta}$ difficile
- ▶ **INLA** : approximation **analytique**
 - +++ **rapidité**
 - ++ inférence plus générique que MCMC, moins de calibrations à faire
 - ++ R-INLA (<http://www.R-INLA.org/>)
 - ▶ tutoriels, exemples, forum de discussion
 - ▶ documentation, publications
 - seulement les lois a posteriori univariées sont faciles à étudier
 - contraintes par rapport à la structure des modèles

Modèles spatiotemporels et inférence INLA

En principe, tout prédicteur gaussien $\eta(s, t)$ est utilisable avec INLA ;
voir [Cressie and Wikle, 2011] pour une revue de modèles spatiotemporels.

Cependant, le vecteur gaussien latent \mathbf{x} peut être de très grande dimension.

⇒ Il faut **utiliser un modèle Gauss–Markov** numériquement convenable.

Quelques références INLA :

[Blangiardo et al., 2013, Lindgren and Rue, 2013, Blangiardo and Cameletti, 2015].

INLA

Champs Gauss–Markov

Approche EDPS

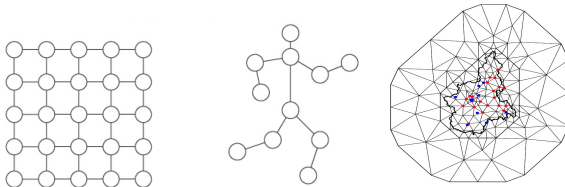
Exemple d'application

Conclusion et perspectives

Le principe d'un champ Gauss–Markov [Rue and Held, 2005]

x champ Gauss–Markov

- ▶ $x_i \mid \mathbf{x}_{-i} \stackrel{d}{=} x_i \mid \mathbf{x}_{\text{voisinage}(i)}$ avec un “petit” voisinage
 \Rightarrow indépendance conditionnelle de x_i et x_j si $j \notin \text{voisinage}(i)$
- ▶ matrice de précision : $q_{ij} = 0$ si $j \notin \text{voisinage}(i)$
 \Rightarrow matrice de précision creuse \mathbf{Q} avec $\#\{q_{ij} \neq 0\} = \mathcal{O}(m_x)$
- ▶ $x_i = \sum_j q_{ij}/q_{ii} \times x_j = \sum_{j \in \text{voisinage}(i)} q_{ij}/q_{ii} \times x_j$



Complexité des calculs

Modèles de grande dimension par rapport à x

⇒ calculs avec $Q \in \mathbb{R}^{m_x \times m_x}$ deviennent lourds

(Q^{-1} , $|Q|$, Qz , $z'Qz$, simulation, ...)

Réduction de la complexité :

	Q pleine	Q creuse
Qz	m_x^2	m_x
Q^{-1}	m_x^3	$m_x^{3/2}$

Modèles Gauss–Markov spatiaux

- exploiter $x_i = \sum_{j \in \text{voisinage}(i)} q_{ij} / q_{ii} \times x_j$

Modèles Gauss–Markov spatiotemporels

- trouver une **approximation Gauss–Markov** pour une covariance spatiotemporelle

 ...

- ▶ **modèles séparables** à composantes Gauss–Markov :
 - ▶ propriété de Gauss–Markov est conservée
 - ▶ produits de Kronecker pour les matrices de covariance/précision :

$$\mathbf{C}_{st} = \{C((s_1, t_1), (s_2, t_2))\} = \{C(t_1, t_2) \times C(s_1, s_2)\} = \mathbf{C}_t \otimes \mathbf{C}_s$$

$$\Rightarrow \mathbf{Q}_{st} = \mathbf{Q}_t \otimes \mathbf{Q}_s$$
 - ▶ exemple : **modèles autorégressifs**
- ▶ **modèles non séparables** : à venir avec l'approche EDPS ...

INLA

Champs Gauss–Markov

Approche EDPS

Exemple d'application

Conclusion et perspectives

Champs Gauss-Markov solution d'EDP stochastiques

[Lindgren et al., 2011, Krainski and Lindgren, 2013]

EDP stochastique (fractionnaire)

$$(\kappa(s) - \Delta)^{\alpha/2} \tau(s) x(s) = W(s), \quad \alpha = \nu + 0.5d, \quad \nu \geq 0$$

où

- ▶ W bruit blanc gaussien,
- ▶ $\kappa(s)$ portée (locale),
- ▶ $\tau(s)$ variabilité (locale),
- ▶ $\alpha \in \{0, 1, 2, \dots\} \rightsquigarrow$ solution $x = \{x(s)\}$ est Gauss-Markov
- ▶ si κ, τ constant \rightsquigarrow champ de **Matérn stationnaire** (paramètre de régularité ν)
- ▶ solution approximative par formulation variationnelle : méthode des **éléments finis**

$$x^*(s) = \sum_{k=1}^n \psi_k(s) \omega_k$$

- ▶ triangulation de l'espace
- ▶ ψ_k linéaires par morceaux (=triangles)
- ▶ poids gaussiens ω_k

Solution approximative par les éléments finis

- ▶ pour n_{tri} points de discrétisation, faire une **triangulation** de l'espace
- ▶ différentes méthodes (selon α) pour déterminer les poids ω_k

↪ solution \mathbf{x}^* :

matrice de précision creuse, définie en espace continu, convergent vers \mathbf{x}

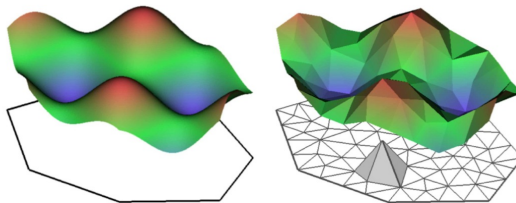


Fig. 3 Left panel: example of a spatial random field (left) given by $X(\mathbf{s}) = \cos(s_1) + \sin(s_2)$, where $\mathbf{s} = \{s_1, s_2\}$. Right panel: corresponding finite element representation of the spatial random field $X(\mathbf{s})$ according to Eq.(9).

Approche EDPS : ce qu'on gagne ...

- ▶ une pléiade de **modèles Gauss–Markov souples en espace continu** :
 - ▶ Matérn stationnaire
 - ▶ sur des variétés (sphères, ...)
 - ▶ autres dépendances nonstationnaires (périodicités, portée locale $\kappa(s)$, covariables dans la portée et la variance, ...)
 - ▶ champs multivariés [Hu et al., 2013]
- ▶ triangulation et éléments finis \rightsquigarrow **discrétisation adaptative** de l'espace

INLA

Champs Gauss–Markov

Approche EDPS

Exemple d'application

Conclusion et perspectives

Un modèle autorégressif spatiotemporel

[Cameletti et al., 2013, Cosandey-Godin et al., 2014]

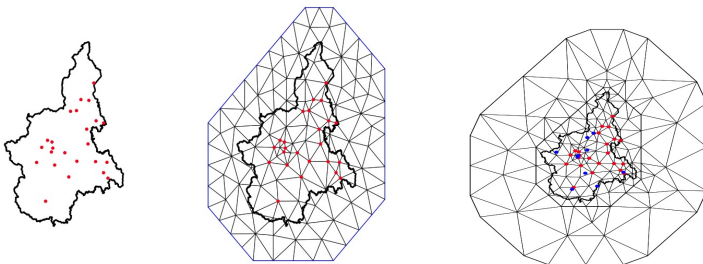
$$\begin{cases} \eta(s, t) &= \mathbf{z}(s, t)\boldsymbol{\beta} + \xi(s, t) + \varepsilon(s, t) \\ \xi(s, t) &= a\xi(s, t-1) + \omega_t(s) \end{cases}$$

- ▶ covariables observées $\mathbf{z}(s, t)$, coefficients linéaires $\boldsymbol{\beta}$
- ▶ processus autorégressif $\xi(s, t)$
avec innovations $\omega_t(s)$ spatialement corrélées (e.g. approche EDPS)
- ▶ erreurs de mesure $\varepsilon(s, t)$
- ▶ $\eta(s, t)$ peut être latent ou directement observé ; $\xi(s, t)$ est latent

Application du modèle autorégressif spatiotemporel : Particules fines

[Cameletti et al., 2013]

- ▶ données journalières observées $y(s_i, t_j) = \eta(s_i, t_j)$: particules fines PM10, Oct 2005 – Mar 2006, Piemont, Italie
- ▶ covariables : longitude, latitude, vitesse du vent, précipitation, température, "mixing height", émissions, altitude
- ▶ innovations : modèle EDPS
- ▶ hyperparamètres : σ_ε , σ_ω , coefficient AR $|a| < 1$, portée ρ



Quelques résultats

Parameter	Quantiles				
	Mean	St.Dev	0.025	0.5	0.975
σ_ε^2	0.0326	0.0014	0.0300	0.0325	0.0353
σ_w^2	1.2762	0.240	0.9300	1.2302	1.8603
ρ	275	16.8	244	275	310
a	0.9601	0.0081	0.9453	0.9598	0.9759

Covariate	Quantiles				
	Mean	St.Dev.	0.025	0.5	0.975
Intercept	3.69	0.45	2.79	3.69	4.57
<i>A</i>	-0.20	0.05	-0.29	-0.20	-0.10
<i>UTMX</i>	-0.16	0.16	-0.48	-0.16	0.16
<i>UTMY</i>	-0.18	0.15	-0.48	-0.18	0.11
<i>WS</i>	-0.06	0.01	-0.08	-0.06	-0.04
<i>TEMP</i>	-0.12	0.04	-0.19	-0.12	-0.05
<i>HMIX</i>	-0.02	0.01	-0.05	-0.02	0.002
<i>PREC</i>	-0.05	0.01	-0.07	-0.05	-0.04
<i>EMI</i>	0.04	0.01	0.01	0.04	0.07

- ▶ R-INLA :
résumés des lois a posteriori
- ▶ $\rho = \sqrt{8\nu}/\kappa$ portée effective
- ▶ $\sigma_w^2 = 1/(4\pi\kappa^2\tau^2)$
variance du champ spatial
- ▶ σ_ε^2 variance résiduelle
- ▶ forte corrélation temporelle et spatiale

Lois marginales a posteriori pour le 30/01/2006

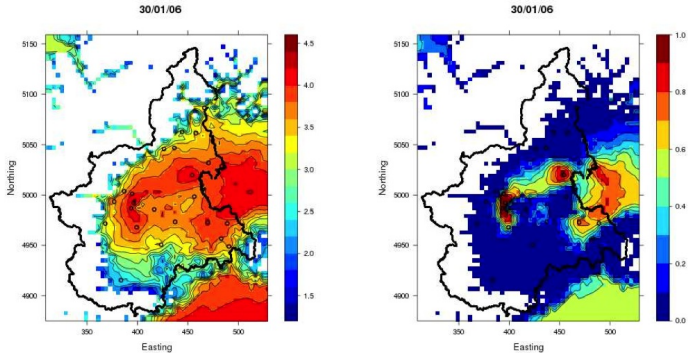


Fig. 5 Map of the PM₁₀ posterior mean on the logarithmic scale (left) and exceedance probability for 50 $\mu\text{g}/\text{m}^3$ (right) for January 30th, 2006. Only locations with an altitude below 1 km are shown.

INLA

Champs Gauss–Markov

Approche EDPS

Exemple d'application

Conclusion et perspectives

Perspectives

Certaines extensions sont conceptuellement simples, mais leur implémentation dans R-INLA sera complexe.

- ▶ extensions de l'approche INLA
 - ▶ plusieurs prédicteurs η_j pour une variable Y_i
 - ▶ transformations nonlinéaires de \mathbf{x} vers les prédicteurs $\boldsymbol{\eta}$
- ▶ extensions de l'approche EDPS
 - ▶ **covariances nonséparables en espace-temps continu** : à venir (www.R-INLA.org)
 - ▶ gestion de l'anisotropie géométrique dans R-INLA

Conclusion

- ▶ le **principe INLA** :
 - ▶ modèles hiérarchiques
 - ▶ champs Gauss–Markov latents et approche EDPS
 - ▶ approximations analytiques pertinentes
 - ▶ flexibilité pour construire des modèles spatiotemporels utiles et réalistes
- ▶ R-INLA : grande richesse de modèles et méthodes
 - ▶ implémentations très efficaces
 - ▶ prédiction, sélection de modèle (DIC, facteurs de Bayes, ...), ...
 - ▶ www.R-INLA.org : documentation et communauté
 - ▶ **nouvelles fonctionnalités** au fur et à mesure



Blangiardo, M. and Cameletti, M. (2015).
Spatial and Spatio-temporal Bayesian Models with R-INLA.
John Wiley & Sons.



Blangiardo, M., Cameletti, M., Baio, G., and Rue, H. (2013).
Spatial and spatio-temporal models with R-INLA.
Spatial and spatio-temporal epidemiology, 7 :39–55.



Cameletti, M., Lindgren, F., Simpson, D., and Rue, H. (2013).
Spatio-temporal modeling of particulate matter concentration through the SPDE approach.
AStA Advances in Statistical Analysis, 97(2) :109–131.



Cosandey-Godin, A., Krainski, E. T., Worm, B., and Flemming, J. M. (2014).
Applying Bayesian spatiotemporal models to fisheries bycatch in the Canadian Arctic.
Canadian Journal of Fisheries and Aquatic Sciences, 72(999) :1–12.



Cressie, N. and Wikle, C. K. (2011).
Statistics for spatiotemporal data.
John Wiley & Sons.



Fong, Y., Rue, H., and Wakefield, J. (2010).
Bayesian inference for generalized linear mixed models.
Biostatistics, 11(3) :397–412.



Hu, X., Simpson, D., Lindgren, F., and Rue, H. (2013).
Multivariate Gaussian random fields using systems of stochastic partial differential equations.
arXiv preprint arXiv:1307.1379.



Illian, J. B., Sørbye, S. H., Rue, H., et al. (2012).
A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA).
The Annals of Applied Statistics, 6(4) :1499–1530.



Krainski, E. T. and Lindgren, F. (2013).
The R-INLA tutorial : SPDE models.
Technical report.
Warning : work in progress... Suggestions are welcome to elias AT r-inla.org.



Lindgren, F. and Rue, H. (2013).
Bayesian spatial and spatiotemporal modelling with R-INLA.
Journal of Statistical Software.



Lindgren, F., Rue, H., and 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.



Martino, S., Akerkar, R., and Rue, H. (2011).
Approximate Bayesian inference for survival models.
Scandinavian Journal of Statistics, 38(3) :514–528.



Rue, H. and Held, L. (2005).
Gaussian Markov random fields : theory and applications.
CRC Press.



Rue, H. and Martino, S. (2007).
Approximate Bayesian inference for hierarchical Gaussian Markov random field models.



Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations.

Rue, H. and Tjelmeland, H. (2002).

Schrödle, B. and Held, L. (2011).

Schrödle, B., Held, L., and Rue, H. (2012).

Biometrics, 68(3) :736–744.

Díaz-Ávalos, C., and Rue, H. (2014).
Spatio-temporal log-Gaussian Cox processes for modelling
wildfire occurrence : the case of Catalonia, 1994–2008.

Wall, M. M. (2004).

Journal of Statistical Planning and Inference,
121(2) :311–324.

Yue, Y. R. and Rue, H. (2011).

Bayesian inference for additive mixed quantile regression models.

Computational Statistics & Data Analysis, 55(1) :84–96.

Projets, applications, ...

Projets, application, ...

- ▶ modélisation de valeurs extrêmes
~> utiliser la fonctionnalité “survie” de R-INLA :
e.g., lois de Weibull censurées pour les vitesses du vent
- ▶ approche EDPS : définir des champs gaussiens sur des supports fragmentés :
e.g., des paysages agricoles ou des réseaux de linéaires (haies, routes, ...)
- ▶ modéliser des trajectoires d'insectes sur des paysages agricoles ?