# Comment prédire et cartographier des risques environnementaux spatio-temporels: l'approche bayésienne avec INLA*

* Integrated Nested Laplace Approximation

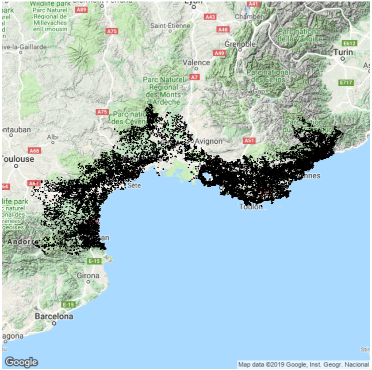
Thomas Opitz, BioSP, INRA Avignon

AG du département MIA, INRA
Jouy-en-Josas
22 Mai 2019

# Wildfires as an example of environmental risks

$\approx 23,000$ occurrences 1995-2018 (Prométhée database `www.promethee.com`)

- important **human, economical and ecological losses** in the Mediterranean basin
- **complex interplay of factors such as human activity, land cover, vegetation cycles and weather** contributing to wildfire occurrences
- **climate change** may further aggravate occurrence intensity and wildfire activity
- **(marked) point pattern data :** space-time ignition points and burnt surfaces



# Typical modeling goals and challenges

**Goals :**

- **space-time mapping** of susceptibility/risk and specific risk factors
- identify **relevant risk factors**, and assess their contribution
- statistical inferences and **probabilistic uncertainty assessment**

**Challenges :**

- data observed at **different spatial and temporal scales**
  *e.g., land cover raster, irregularly spaced weather stations, DFCI grid*
  - spatial and temporal scales of covariate influence on the response ?
- **nonlinear trends** with respect to covariates and to time and space, unobserved / unavailable covariates $\Rightarrow$ need for **random effect modeling**
- highly "nongaussian" data (point patterns, counts, threshold exceedances)
- **"moderately big" data** ($10^4$-$10^7$) : many observation points or control points

**Numerous applications** : **(agro-)ecolocy/epidemiology**, **atmospheric processes**,...

- discrete variables : landslide risk, epidemiology, ...
- continuous variables : soil, air pollution, weather/climate

# About 10 years ago...INLA was born

Journal of the Royal Statistical Society
SERIES B Statistical Methodology

*J. R. Statist. Soc.* B (2009)
**71**, *Part* 2, *pp.* 319–392

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

Håvard Rue and Sara Martino
*Norwegian University for Science and Technology, Trondheim, Norway*

and Nicolas Chopin
*Centre de Recherche en Economie et Statistique and Ecole Nationale de la Statistique et de l'Administration Economique, Paris, France*

**Bayesian computing** remains challenging even though MCMC is available :
good theoretical properties of **MCMC**, but in practice **often too slow**.

**INLA is an alternative, usually (much) faster solution for latent Gaussian models :**

- in practice, most Bayesian models are in this class
- **conditional independence** of observations $\boldsymbol{y}$ with respect to latent Gaussian $\boldsymbol{x}$ :
  $\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta} \sim \prod_i \pi(y_i \mid \eta_i, \theta)$ with **linear predictor** $\eta = Ax$ and **hyperparameters** $\theta$
- $\dim(\boldsymbol{x})$ typically large ($10^2$ to $10^5$), $\dim(\boldsymbol{\theta})$ typically small (0 to 15)
- implementation : `R-INLA` (`www.r-inla.org/`)

# Regression modeling with R-INLA in a nutshell

**Regression models** with **generalized additive structure** :
**observed response variable** is explained through **covariates and random effects**

$$y \sim F_\eta \text{ with } \eta = \underbrace{\underbrace{\beta_0 + \beta_1 \times \mathrm{covar.1} + \ldots + \beta_m \times \mathrm{covar.m}}_{\textbf{fixed effects}} + \textbf{random effects}}_{\textbf{linear predictor}}$$

$\Rightarrow$ amenable to **inference with INLA :**

- **Gaussian process priors** for fixed and random random effects
- **analytical Laplace approximations** and **numerical integration schemes** to **estimate posterior distributions**
- **Gauss–Markov structures** ensure sparse precision matrices
- **SPDE approach** [Lindgren et al., 2011] : Gauss–Markov approximations for spatial random effects $W(s)$ with flexible Matérn covariance function

**Posterior distributions** (means, standard deviations, credible intervals,...) of

- hyperparameters (precisions, spatial/temporal dependence range,...)
- latent variables $\boldsymbol{x}$ (regression coefficients $\beta_j$, spline functions, spatial fields, ...)
- predictions of the response variable

are obtained using **Bayes' formula**.

**Example :** posterior mean of a regression coefficient $\beta_j$ is

$$\mathbb{E}(\beta_j \mid \boldsymbol{y}) = \int \beta_j \int \int \pi(\beta_j, \boldsymbol{x}, \boldsymbol{\theta} \mid \boldsymbol{y}) \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{\theta} \, \mathrm{d}\beta_j$$

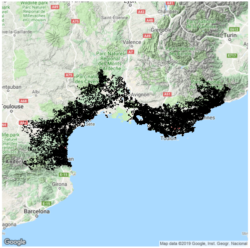
- $\pi(\beta_j, \boldsymbol{x}, \boldsymbol{\theta} \mid \boldsymbol{y}) \propto \pi(\beta_j, \boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})$
- $\boldsymbol{x}$ is the vector of latent Gaussian variables with $\beta_j$ removed

**Integrated Nested Laplace Approximation :** analytical approximations

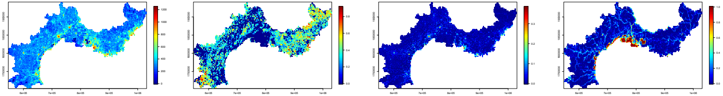
- **Laplace approximation** for integration with respect to Gaussian densities $\mathrm{d}\boldsymbol{x}$
- **multivariate integration schemes** for small number of hyperparameters $\mathrm{d}\boldsymbol{\theta}$

## Wildfire occurrence data

- ≈ $23,000$ occurrences 1995-2018
- daily wildfire counts over DFCI grid ($\approx (2km)^2$)
- we build a model with monthly resolution

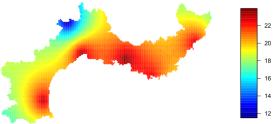



- land cover data (e.g., road length, coniferous trees, buildings, water)

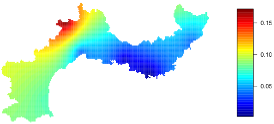


- weather station data, interpolated by space-time kriging

Temperature (average °C, July 2014)




Precipitation (July 2014)



# Building a log-Gaussian Cox process model

**Point process model** : $N(U) \mid \Lambda(s,t) \sim \mathrm{Pois}(\Lambda(s,t))$ for a space-time unit $U$ at $(s,t)$

**Log-Gaussian intensity :** $\log \Lambda(s,t)$ has additive structure

$$\beta_0 + \beta^{\mathrm{time}}\tilde{t} + \sum_{j=1}^{30} \beta_j^{\mathrm{land}} z_j^{\mathrm{land}}(s) + \sum_{j=1}^{3} \beta_j^{\mathrm{clim}} \hat{z}_j^{\mathrm{clim}}(s,t) + f(\mathrm{month}(t)) + W(s,a(t))$$

- land : land cover, roads, vegetation types (IGN databases)
  $\Rightarrow$ preprocessing towards **average/sd/co-occurrence values for DFCI pixels**
- linear time trend : $\beta^{\mathrm{time}}$ is difference “end − beginning” of study period
- $f(\mathrm{month}(t))$ : seasonal effect with monthly resolution
- clim : **gridded monthly anomalies** of temperature and precipitation
- **spatio-temporal random effect** at yearly resolution with temporal autoregression,

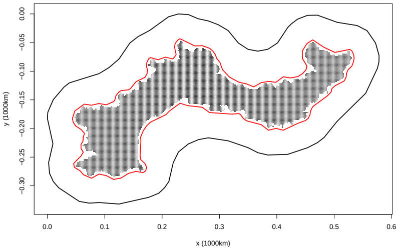
$$W(s,a(t)) = \rho W(s,a(t)-1) + \sqrt{1-\rho^2}\varepsilon_{a(t)}(s), \quad \rho \in (-1,1)$$

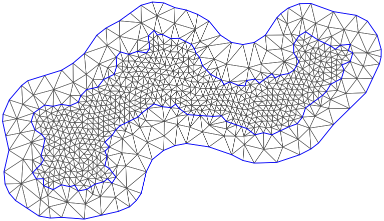
with innovation fields $\varepsilon_{a(t)}(s)$ obeying the Matérn covariance

# Spatial discretisation of the model

**SPDE approach : triangulation of study region** (mesh) + extension to avoid boundary effects $\Rightarrow \approx 900$ nodes to define Markov approximation of Matérn field

⚠ In INLA-SPDE models, the spatial resolution of the latent Gaussian model can be chosen independently of the observation process.



# Computational aspects

- dataset and latent model are **high-dimensional** :

  $\approx$ 2.7 million observations, $\approx 20,000$ latent Gaussian variables
  $\Rightarrow$ sparse precision matrix of dimension $20,000 \times 20,000$

- some stability issues have arisen in estimating the model with INLA

  $\Rightarrow$ solution : **year/pixel-stratified subsampling scheme** for 0 counts

- **Penalized Complexity priors** control departure of model components from simpler baselines and stabilize the estimation

- estimation runtime of model is several hours

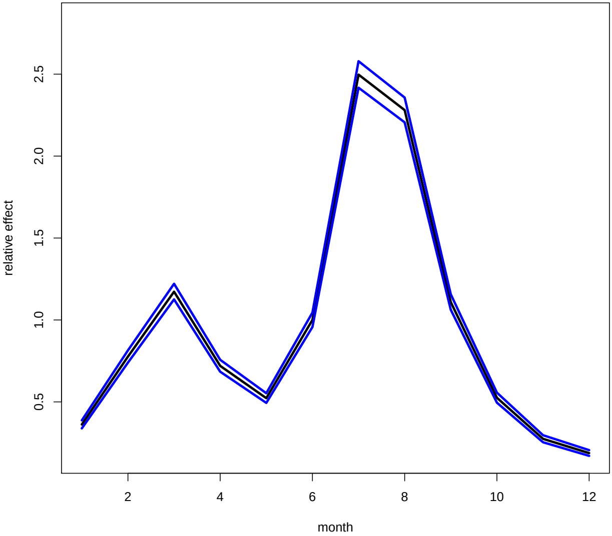
## Covariate (fixed) effects

Some results of estimated $\beta$ coefficients, ordered by decreasing significance :

| covariate | estimate | CI |
|---|---|---|
| altitude (average) | -1.48 | [-1.64,-1.33] |
| temperature anomaly | 0.09 | [0.08,0.1] |
| precipitation (square root) | -3.15 | [-3.66,-2.65] |
| road length (average) | 2.45 | [2,2.91] |
| water (average coverage) | -1 | [-1.21,-0.8] |
| ... | ... | ... |
| building cover (average) | -5.21 | [-6.71,-3.7] |
| road length (standard deviation) | -1.87 | [-2.45,-1.29] |
| forest cover (standard deviation) | 0.77 | [0.49,1.04] |
| ... | ... | ... |
| building cover (standard deviation) | 2.71 | [1.38,4.04] |
| forest cover+building cover | 4.53 | [2.27,6.79] |
| ... | ... | ... |
| forest cover+paths | -2.54 | [-4.06,-1.02] |
| ... | ... | ... |
| time | -0.48 | [-0.91,-0.05] |
| ... | ... | ... |

# Estimated seasonal effect

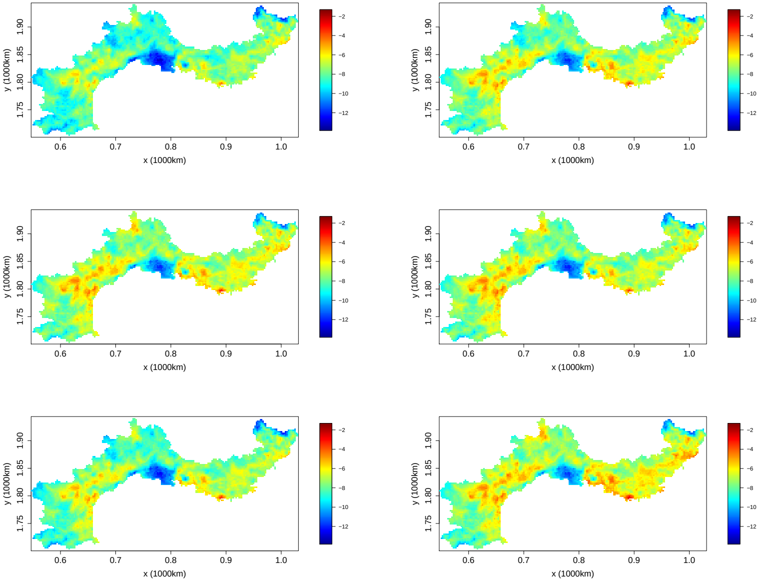
Relative fire occurrence intensity and credible intervals (= 1 for end of June)




Nonlinearity !

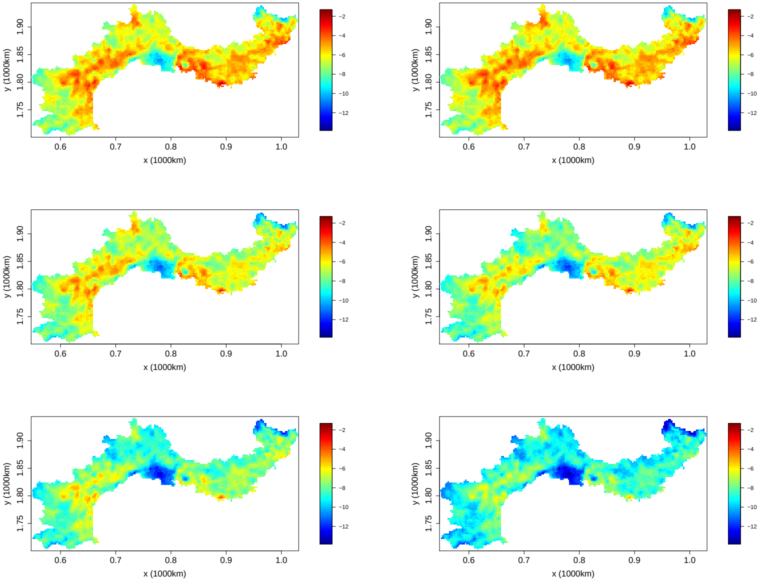
# Space-time mapping of occurrence log-intensity

Months January to June 2017 (top row : 1,2 ; middle row : 3,4 ; bottom row : 5,6)



# Space-time mapping of occurrence log-intensity

Months July to December 2017 (top row : 7,8 ; middle row : 9,10 ; bottom row : 11,12)



# Wrap-up

- **space-time regression and prediction** formulated as **Bayesian hierarchical model** for INLA

- two tools at the core of INLA :
  - **latent Gauss-Markov random fields**
  - **Laplace approximation**

- **SPDE approach** combines Gauss-Markov representation and physical interpretation for spatial modeling

- tuning INLA for very high-dimensional datasets and latent models remains challenging...

## Collaborations and projects around INLA

**INLA+SPDE workshop** (Avignon, 11/2018) $\approx$ 40 participants

- funded by RESSTE network and GdR EcoStat
- materials online : `informatique-mia.inra.fr/resste/SPDE`

More tutorials will come...

**Ongoing projects :**

- landslide susceptibility mapping in space and time
  *with Luigi Lombardo (U Twente), Haakon Bakka, Raphael Huser (KAUST)*
- wildfire risk (occurrences and burnt area) under climate projections
  *with colleagues of URFM, INRA Avignon (PhD project of Hélène Fargeon)*
- soil variables in France / UK : inference on temporal trends, space-time mapping
  *"Pari scientifique" EA, with colleagues of Infosol, INRA Orléans, and Ben Marchant (British Geological Survey)*
- space-time dynamics of *varroa destructor*, a bee parasite
  *with André Kretzschmar (BioSP), Nicolas Desassis (MinesParisTech)*
- fight against Asian hornets : is trapping young nest-founding hornets efficient ?
  *with colleagues from ITSAP Avignon and BioSP*
- wolf attacks on sheep herds : which impact of wolf kills and protective measures ?
  *with Oksana Grente, Olivier Gimenez (CEFE, Montpellier)...*
- pesticide treatments in simulated agricultural landscapes
  *part of Patrizia Zamberletti's PhD project at BioSP*
- conditional space-time extremes
  *with Emma Simpson, Jenny Wadsworth (Lancaster University)*

# Outlook

- **network structure :** INLA is useful for estimating models on fixed networks, but what about **learning the network** at the same time ?

  $\rightsquigarrow$ develop **"INLA-within-MCMC" algorithms** :

  - MCMC for modifying the network structure
  - INLA for observations conditional to network

- **statistical learning for "big" space-time point pattern data**

  $\rightsquigarrow$ use machine learning techniques appropriate for regression modeling

- **beyond INLA** : statistical learning for **"big" space-time data** with **strongly "nonlinear"/multiscale space-time effects** ?

  - use **Bayesian probabilistic tools for mechanistical-statistical modeling as foundation**,
  - then **integrate machine learning tools for handling big data** ?

# References


Fargeon, H., Martin-StPaul, N., Pimont, F., De Cáceres, M., Ruffault, J., Opitz, T., Allard, D., and Dupuy, J.-L. (2018).
Assessing the increase in wildfire occurrence with climate change and the uncertainties associated with this projection.
In *8th International conference on forest fire research*.

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.

Lombardo, L., Opitz, T., and Huser, R. (2018a).
Numerical recipies for landslide spatial prediction by using R-INLA : A step-by-step tutorial.
In Pourghasemi, H. R. and Gokceoglu, C., editors, *Submitted for Spatial Modeling in GIS and R for Earth and Environmental Sciences*. Elsevier.

Lombardo, L., Opitz, T., and Huser, R. (2018b).
Point process-based modeling of multiple debris flow landslides using INLA : an application to the 2009 Messina disaster.
*Stochastic environmental research and risk assessment*, 32(7) :2179–2198.

Opitz, T. (2017).
Latent Gaussian modeling and INLA : A review with focus on space-time applications.
*Journal of the French Statistical Society (Special Issue on Space-Time Statistics)*, 158(3).

Opitz, T., Bonneu, F., and Gabriel, E. (2019).
Point-process based modeling of space-time structures of forest fire occurrences in Mediterranean France.
*Submitted*.

Opitz, T., Huser, R., Bakka, H., and Rue, H. (2018).
INLA goes extreme : Bayesian tail regression for the estimation of high spatio-temporal quantiles.
*Extremes*, 21(3) :441–462.

Rue, H., Martino, S., and Chopin, N. (2009).
Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations.
*Journal of the royal statistical society (Series B)*, 71(2) :319–392.