Une introduction aux modèles bayésiens hiérarchiques
– avec une application aux feux de forêts en France –

Thomas Opitz

BioSP, INRAE, Avignon

École TDMA 2023

Grenoble
8 juin 2023

Wildfires in France..

Accueil > Actu > Environnement > Incendies

Méga-feu en Gironde et dans les Landes: 1.000 pompiers mobilisés dont un très grièvement blessé et 10.000 évacués face à un incendie hors de contrôle

Le méga-feu de Gironde mobilise 1.000 pompiers. / Sécurité civile

… and elsewhere

Superficie brûlée en 2023

Provinces les plus touchées par les feux de forêt
En date du 6 juin 2023

Alberta	1 192 489 ha
Saskatchewan	976 439 ha
Territoires du Nord-Ouest	383 277 ha
Colombie-Britannique	344 767 ha
Québec	327 521 ha
Ontario	31 853 ha
Nouvelle-Écosse	26 655 ha

Source: Ressources naturelles Canada, Saskatchewan Public Safety Agency, SOPFEU

RADIO-CANADA

Wildfires

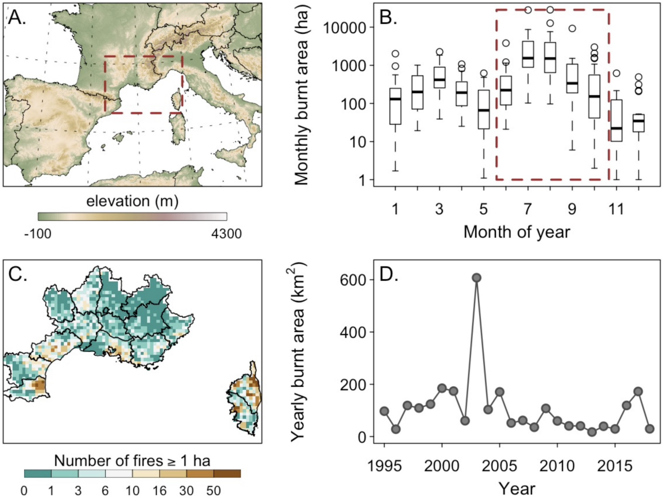
- **Wildfires** represent a major environmental risk worldwide with strong impacts on societies, ecosystems and climate change.
- Many **statistical and machine-learning approaches** have been deployed to better understand and predict wildfires.
- There are two important aspects about wildfires:
  - **Occurrence** (Where? When? How many?)
  - **Size** (How extreme?)
- **Extreme fires** (such as those exceeding a high size threshold) contribute very strongly to total burnt areas and must be modeled with great care.

Explanatory variables for wildfire risk

- **Weather and climate conditions** $\Rightarrow$ **Vulnerability** to wildfires
  - Complex interactions among temperature, precipitation, humidity, wind
  - Various weather-based fire-danger indices have been developed, such as the Fire-Weather Index (FWI)
- Forests and other vegetation cover $\Rightarrow$ **Exposure**
- **Human-related factors:** human behavior, and wildfire management and prevention, are highly important but often difficult to assess quantitatively

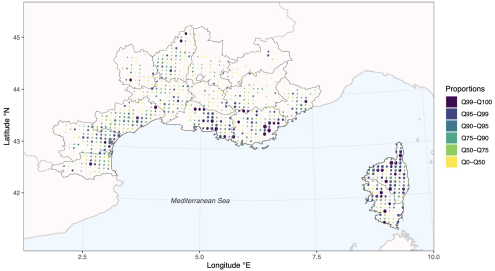
Prométhée database (since 1970s): for each wildfire,

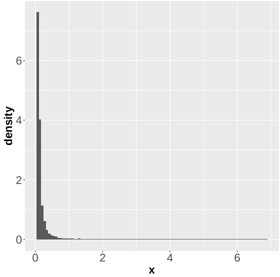
- **position** (2km precision) and date of ignition, and
- **burnt area** ($\approx$ 10000 wildfires larger than 1ha since 1995)



“1% of fires do 99% of the damage”

Left: **Aggregated burnt areas** per pixel (regular 8km grid)
Right: Histogram of **individual wildfire sizes** in km (diameters $= \frac{2}{\pi}\sqrt{\text{surface}}$)





Complex data...

- **Multi-source** data and **multi-scale** data
  (e.g. in-situ, radar, satellite, reanalyses, regional/global climate models)

- **Measurement errors**,
  Preferential sampling (e.g. Citizen Science data)

- **“Small data“**, especially for extreme events

- **“Non-normal“** data: Discrete/Continuous; Asymmetries; Heavy tails

What you see is not what you want to get

Objectif: Transformation des données en connaissances

... and complex processes to model

General setting: Regression modeling
⇒ Explain/predict one or several variables of interest (called **response(s)**) as a function of other variables (called **covariates**/**predictors**/**drivers**/**forcings**)

- Assessment of the influence of **natural and anthropogenic drivers**
- Different **spatial and temporal scales** may be relevant
- **Auto- and cross-correlations** among drivers and responses
- An example of complex processes:

  Risk modeling in the sense of IPCC/UNDRR: for a given "stake",

  $$\textbf{Risk} = \textbf{Hazard}(\text{Drivers}) \times \textbf{Vulnerability}(\text{Drivers}) \times \textbf{Exposure}(\text{Drivers})$$

...require sophisticated models

Desiderata for flexible model classes

- Numerous predictor variables and multiple response variables
- Non-normal probability distributions of the response
- Measurement errors
- Spatial and temporal **autocorrelation**
- **Prior knowledge** about processes

Using models for decision support

- Assess various **sources of uncertainty**:
  - Intrinsic (natural) variability of phenomena
  - Estimation of parameters
  - Model choice
- **Identify drivers** and **quantify how they contribute to the response**
- Make **probabilistic predictions** and long-term projections
- Stochastically generate **new scenarios**

Bayesian setting

We use **prior distributions** for unknown model parameters to be estimated.

Modeling approach

1. Formulate the model for data conditional to parameters.
2. Set prior distributions for parameters by including **prior knowledge**.
3. Infer parameters conditional to data using **Bayes' theorem**.

Bayes' theorem

Prior variables $\boldsymbol{\theta}$.
Data $\boldsymbol{y}$.
Prior probability $\text{Pr}(\boldsymbol{\theta})$.
Likelihood $\text{Pr}(\mathbf{y} \mid \boldsymbol{\theta})$
$\Rightarrow$ Posterior probability

$$\text{Pr}(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{\text{Pr}(\mathbf{y} \mid \boldsymbol{\theta})\text{Pr}(\boldsymbol{\theta})}{\text{Pr}(\mathbf{y})}$$
$$\propto \text{Pr}(\mathbf{y} \mid \boldsymbol{\theta})\text{Pr}(\boldsymbol{\theta})$$

- Mathematically simple.
- Computationally complex, especially if $\boldsymbol{\theta}$ has many components.

⚠ For fixed data $\boldsymbol{y}$, the probability $\text{Pr}(\boldsymbol{y})$ is a parameter-free constant.

Prior probability density $\star$ Likelihood of data $\rightsquigarrow$ Posterior density

With few observations, the prior has strong influence on the posterior model.

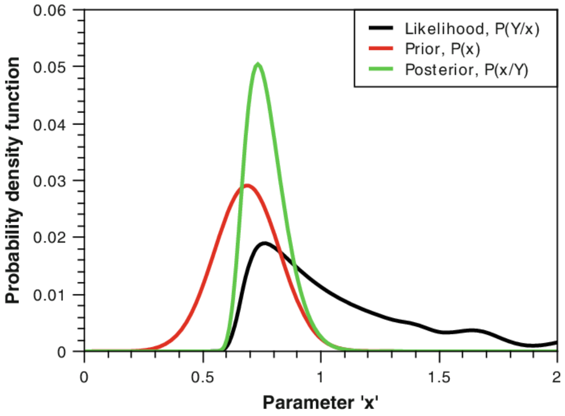


Image credit: Balaji et al. (2013)

The structure of Bayesian hierarchical models

Bayesian hierarchical models flexibly combine **three structured layers**:

$$\begin{array}{c}\text{data}\\ \uparrow \\ \text{latent process}\\ \uparrow \\ \text{hyperparameters}\end{array}$$

- **Observation/data layer**
  defined by the **likelihood model** of observations

  depends on
- **Latent process** capturing trends, dynamics and dependence,

  depends on
- **Hyperparameters** controlling the latent process(es) and likelihood:
  - Signal-to-noise ratio, smoothing ⇒ variance/correlation parameters
  - Range of dependence, *e.g.*, spatial range of correlation function
  - Shape of the response distribution, *e.g.*, gamma or GEV shape

Challenging estimation of numerous parameters
⇒ **Approximate Bayesian inference**

Formal notation of Bayesian hierarchical models

A priori, we suppose that data $\boldsymbol{y}$ have been generated using a latent process that can be represented through a random vector $\boldsymbol{x}$.

Formal notation

Data: $\boldsymbol{y} = (y_1, \ldots, y_n)$
Latent random vector: $\boldsymbol{x} = (x_1, \ldots, x_m)$
Predictor: $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n) = g(\boldsymbol{x})$ with some projection function g

$$y_i \mid \boldsymbol{x}, \boldsymbol{\theta} \overset{ind.}{\sim} \pi(y_i \mid \eta_i, \boldsymbol{\theta}), \qquad \textbf{Likelihood of observations}$$

$$\boldsymbol{x} \mid \boldsymbol{\theta} \sim \pi(\boldsymbol{x} \mid \boldsymbol{\theta}) \qquad \textbf{Latent process}$$

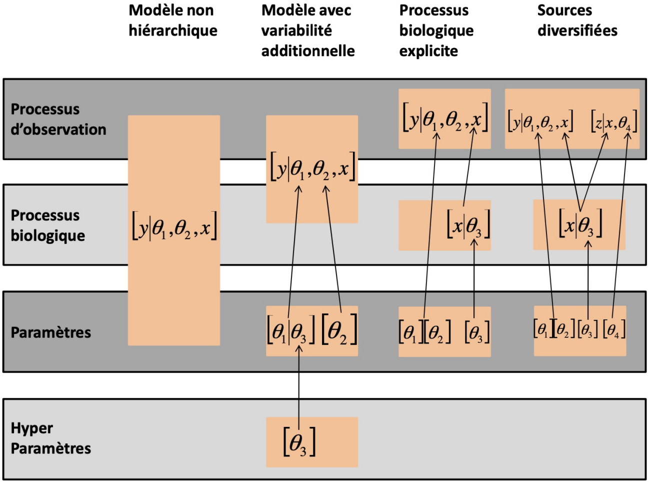
$$\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta}) \qquad \textbf{Hyperparameters}$$

⚠ $\pi(\cdot)$ is used generically for conditional/unconditional probability densities.

Many variants of hierarchical models

Different processes / hierarchies / data sources.

Illustration (with slightly different notations):
variants involving a biological process, for example a mechanistic model for population dynamics or disease spread



D'après Clark, 2005

The class of Latent Gaussian models (LGMs)

A popular class of models widely used for spatial and spatiotemporal data.

Latent Gaussian Models

- *A priori*, $\boldsymbol{x}$ is an m-dimensional **multivariate Gaussian vector**

$$\boldsymbol{x} \mid \boldsymbol{\theta} \sim \mathcal{N}_m(\boldsymbol{\mu}, \Sigma(\boldsymbol{\theta}))$$

- **Observation matrix** (or **projection matrix**) $A \in \mathbb{R}^{n \times m}$ defines a **linear predictor**

$$\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n) = \boldsymbol{\eta}(\boldsymbol{x}) = A\boldsymbol{x}$$

  (Observation matrix A is known and fixed in the model)

- *A priori*, the linear predictor $\boldsymbol{\eta}$ is also multivariate Gaussian:

$$\boldsymbol{\eta} \sim \mathcal{N}_n(A\boldsymbol{\mu}, A\Sigma(\boldsymbol{\theta})A^T)$$

$\Rightarrow$ **Flexible models and fast estimation with relatively large n and m**

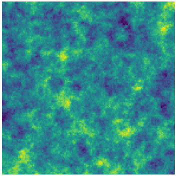
Recall: Gaussian vectors and fields

Gaussian distributions are characterized by their mean vector and covariance matrix:

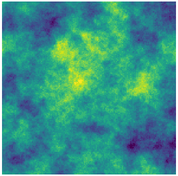
$$\boldsymbol{x} = (x_1, \dots, x_m) \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma), \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{11} & \dots & \sigma_{mm} \\ \sigma_{21} & \dots & \sigma_{2m} \\ \vdots & \vdots & \vdots \\ \sigma_{m1} & \dots & \sigma_{mm} \end{pmatrix}$$

Gaussian fields $x(s)$ for locations s: characterized by a mean function $\mu(s)$ and a covariance function $C(s_1, s_2) = \mathrm{Cov}(x(s_1), x(s_2))$

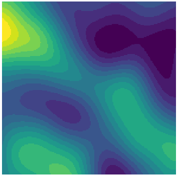
Small range



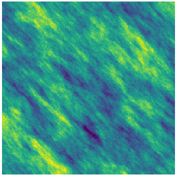
Large range



Smooth surface



Anisotropy



Toy example: Bayesian linear model (two covariates)

How to cast the classical linear model in the latent Gaussian framework?

$$y_i = \beta_0 + \beta_1 v_{1,i} + \beta_2 v_{2,i} + \varepsilon_i, \quad \varepsilon_i \overset{\text{ind.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$$

- We have $y_i \mid (\boldsymbol{x}, \boldsymbol{\theta}) \sim \mathcal{N}(\eta_i(\boldsymbol{x}), \sigma_\varepsilon^2)$

  with linear predictor $\eta_i(\boldsymbol{x}) = \beta_0 + \beta_1 v_{1i} + \beta_2 v_{2i}$ and $A = \begin{pmatrix} 1 & v_{11} & v_{21} \\ \vdots & \vdots & \vdots \\ 1 & v_{1n} & v_{2n} \end{pmatrix}$

- Latent Gaussian $\boldsymbol{x} = (\beta_0, \beta_1, \beta_2)$ with prior

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \Sigma), \quad \Sigma = \begin{pmatrix} \sigma_0^2 & 0 & 0 \\ 0 & \sigma_\beta^2 & 0 \\ 0 & 0 & \sigma_\beta^2 \end{pmatrix}$$

- Gaussian likelihood in the data layer:

$$\pi(y_i \mid \eta_i(\boldsymbol{x}), \sigma_\varepsilon^2) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left(\frac{1}{2} \frac{(y_i - \eta_i(\boldsymbol{x}))^2}{\sigma_\varepsilon^2}\right)$$

- Hyperparameters $\boldsymbol{\theta} = (\sigma_\varepsilon^2, \sigma_0^2, \sigma_\beta^2)$

Example of an LGM: Poisson–lognormal model

- Models for **count data** $\Rightarrow$ **Poisson likelihood**:

$$y_i \mid \lambda_i \sim \text{Poisson}(\lambda_i), \quad i = 1, \dots, n$$

  (**note:** there are no hyperparameters in the Poisson distribution)

- Poisson intensity parameters are assumed to follow the log-Gaussian distribution:

$$\boldsymbol{\eta} = \log \boldsymbol{\lambda} = (\log \lambda_1, \dots, \log \lambda_n)^T = A\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, A\boldsymbol{\Sigma}A^T)$$

  with $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ and $\boldsymbol{\Sigma}$ the $m \times m$ variance-covariance matrix.

- Here, we use a **log-link function** between the linear predictors $\boldsymbol{\eta}$ and the likelihood parameters $\boldsymbol{\lambda}$.

Example: Firelihood model for wildfire counts

- **Observations:** wildfire counts y_i in voxel i (pixel-days with $8km \times 8km$ pixels)
- **Data likelihood:**

$$y_i \mid \lambda_i \sim \mathrm{Poisson}(\lambda_i)$$

- **Linear predictor:**

$$\begin{aligned}\eta_i = \log \lambda_i = & \beta_0 + f_{\mathrm{FWI}}(\mathrm{FWI}_i) + f_{\mathrm{FA}}(\mathrm{FA}_i) + f_{\mathrm{DOY}}(\mathrm{DOY}_i) \\ & + f_{\mathrm{YEAR}}(\mathrm{YEAR_i}) + f_{\mathrm{PIXEL}}(\mathrm{PIXEL}_i),\end{aligned}$$

with nonlinear **additive components** (using **basis functions** a_k^{COMP})

$$f_{\mathrm{COMP}}(\mathrm{COMP}_i) = \sum_{k=1}^{K_{\mathrm{COMP}}} \beta_k^{\mathrm{COMP}} a_k^{\mathrm{COMP}}(\mathrm{COMP}_i),$$

where the observation matrix A has entries $a_k^{\mathrm{COMP}}(\mathrm{COMP}_i)$ in row i.

- **Latent Gaussian vector:**

$$\boldsymbol{x} = (\beta_0, \boldsymbol{\beta}^{\mathrm{FWI}}, \boldsymbol{\beta}^{\mathrm{FA}}, \boldsymbol{\beta}^{\mathrm{DOY}}, \boldsymbol{\beta}^{\mathrm{YEAR}}, \boldsymbol{\beta}^{\mathrm{PIXEL}}).$$

Example: representation of $f_{\mathrm{PIXEL}}(\mathrm{PIXEL}_i)$

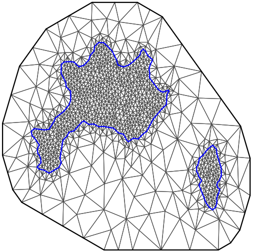
A priori, f_{PIXEL} is a Gaussian 2D random field.

SPDE approach (Lindgren et al. (2011); Krainski et al. (2018))

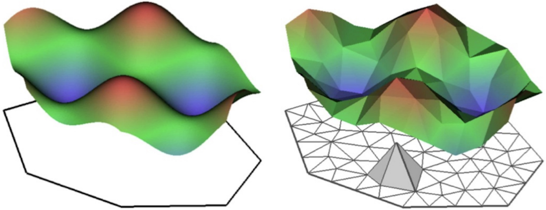
1. We discretize geographic space using a triangulation with m_{PIXEL} nodes.
2. On each node s_k, we put a Gaussian variable x_k.
3. Latent Gaussian vector $\boldsymbol{x} = (x_1, \ldots, x_{m_{\mathrm{PIXEL}}})$ with Matérn covariance Σ.
4. Basis functions a_k^{PIXEL} are finite elements ("pyramids").

$\Rightarrow$ **Sparse precision matrices** $Q = \Sigma^{-1}$ allow working with large m

Spatial mesh



Finite-element representation



Estimating Bayesian hierarchical models

⚠ Usually we cannot calculate the **posterior estimations** in closed form:

- **Posterior densities** $\pi(\theta_j \mid \boldsymbol{y})$, $\pi(x_k \mid \boldsymbol{y})$, $\pi(\eta_i \mid \boldsymbol{y})$
- **Posterior mean estimates** $\mathbb{E}(\theta_j \mid \boldsymbol{y})$, $\mathbb{E}(x_k \mid \boldsymbol{y})$, $\mathbb{E}(h(\eta_i) \mid \boldsymbol{y})$

⚠ Latent Gaussian vector $\boldsymbol{x}$ is often very high-dimensional (up to $m \approx 10^5$)

$\Rightarrow$ Use **numerical approximation of complicated integrals**, such as

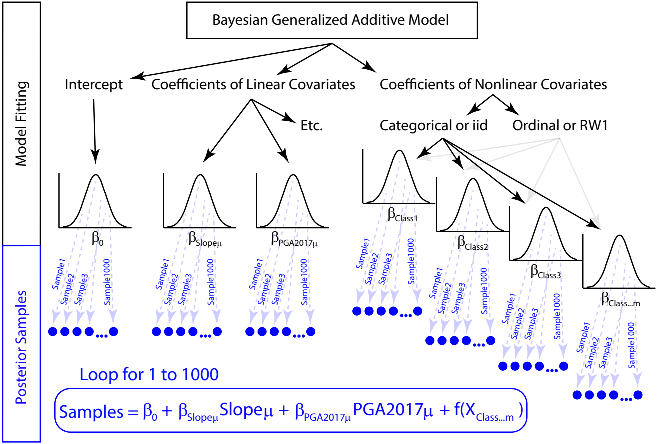
$$\pi(\theta_j \mid \boldsymbol{y}) = \int \int \pi(\boldsymbol{\theta}, \boldsymbol{x} \mid \boldsymbol{y}) \, \mathrm{d}\boldsymbol{x} \mathrm{d}\boldsymbol{\theta}_{-j}$$

- **Markov-Chain Monte-Carlo**: simulate from posterior density $\pi(\boldsymbol{\theta}, \boldsymbol{x} \mid \boldsymbol{y})$
- **Integrated Nested Laplace Approximation:** astute numerical integration for Latent Gaussian Models (Rue et al. 2009)
  - $\mathrm{d}\boldsymbol{x}$: variants of Laplace approximation
  - $\mathrm{d}\boldsymbol{\theta}$: variants of classical numerical integration

Simulation-based posterior inferences

Monte–Carlo estimation of posterior quantities based on a large sample of the estimated posterior model ⇒ Accurate propagation of posterior incertainties.

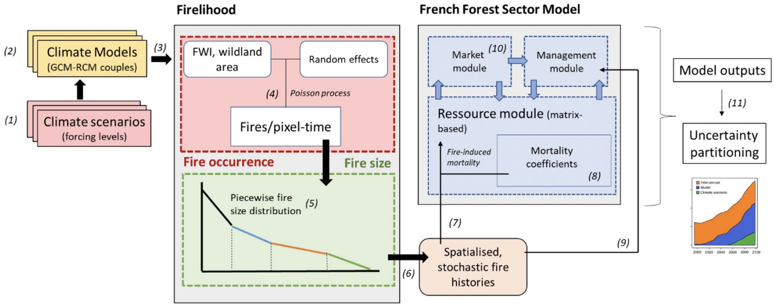
Illustration : (here for a model fitted to landslides)



Case study: Bioeconomic projections of wildfire risk

Soft-coupling of three models:

- Climate model ⇒ Deterministic simulations EURO-CORDEX
- **Statistical wildfire model ⇒ Bayesian hierarchical model Firelihood**
- Economic forest sector model ⇒ Deterministic model FFSM



Firelihood model for wildfire occurrences and sizes

- A system of latent Gaussian models for several response variables
- Our model combines several **regression equations** :
  - **Wildfire counts** per pixel-day (**Poisson**)
  - **Size exceedance** above thresholds $10, 100, 1000$ha (**Bernoulli**)
  - **Sizes** in the intervals $[1, 10], [10, 100], [100, 1000] ha$ (**truncated Pareto**)
  - **Extreme sizes** exceeding $1000 ha$ (**Generalized Pareto Distribution**)
- Each regression equation has a linear predictor with form

$$\eta_i = \beta_0 + f_{\mathrm{FWI}}(\mathrm{FWI}_i) + f_{\mathrm{FA}}(\mathrm{FA}_i) + \\ + f_{\mathrm{YEAR}}(\mathrm{YEAR_i}) + f_{\mathrm{DOY}}(\mathrm{DOY}_i) + f_{\mathrm{PIXEL}}(\mathrm{PIXEL}_i)$$

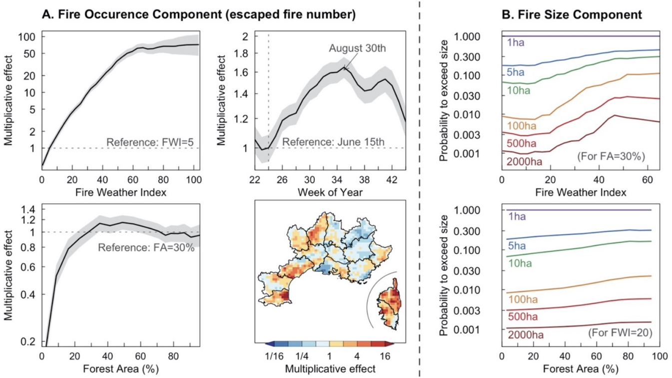
(some components may be left out)

Estimated predictor components

Pimont et al. (2021, Ecological Applications).

Model was fitted to main fire season (June–September)

Posterior effects of Fire-Weather Index, Forest Area, Seasonal, Spatial



Spatial/seasonal effects account for strong residual variability not explained by biophysical predictors.
Tail index of $\hat{\xi} \approx 0.4$ of extreme fire sizes confirms very heavy tails.

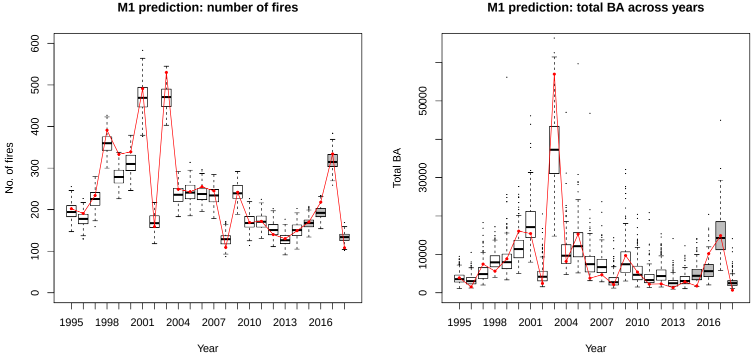
Model-based predictions of fire numbers and sizes

We perform posterior simulation and aggregate results, here to annual scale.

Boxplots of posterior samples (dark grey = validation period 2015–2018)

vs.

Observations (red lines)

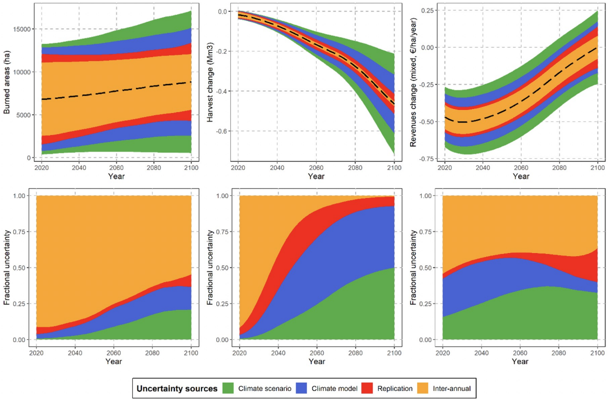


Bioeconomic projections for the forest sector (–2100)

Riviere et al. (2022, Earth's Future).

- For coniferous forests
- Climate scenarios : RCPs 4.5 et 8.5
- Climate projections EURO-CORDEX : 5 pairs GCM-RCM
- 30 simulations of Firelihood posterior per triplet scenario-GCM-RCM

Trend and uncertainty (top row) ; decomposition of variance (bottom row)
Burnt areas (left); Δ Harvested timber (middle); Δ Earnings (right)



Discussion: Good statistical practices

Noël Cressie (2021). A few statistical principles for data science. *Australian & New Zealand Journal of Statistics.*

- Geophysicists conserve energy but what do data scientists conserve?

  In any well-defined statistical model, there is conservation of variability.

- The holy grail: all scales of variation are additive

  Seek a transformation of the scientific process where all components of variation act and interact additively.

- Patches in close proximity are commonly more alike ...

  Everything is related to everything else, but near things are more related than distant things (Tobler 1970).

Discussion: Bayesian hierarchical models

- Bayesian hierarchical models put focus on:
  - **Parameter inference**
  - **Uncertainty assessment**
  - **Interpretability**

  and allow including **prior (expert) knowledge**

- ⚠ Bayesian models often less scalable to very big datasets (but frameworks such as R-INLA can handle millions of observations)

- Many implementation frameworks (BUGS, R-INLA, JAGS, Stan, TMB...) and training opportunities

Thanks for your attention!

Some relevant resources:

- RESSTE network (INRAE and beyond): `https://reseau-resste.mathnum.inrae.fr/`

  Risques, Extrêmes et Statistique Spatio-TEmporelle

  → Workshops, Tutorials, Mailing list

- R-INLA for Latent Gaussian Models: `https://www.r-inla.org/`