Spatiotemporal modeling of extreme-wildfire risk

Thomas Opitz

Biostatistics and Spatial Processes, INRAE Avignon (France)

WECC workshop – Weather extremes and Climate Change

19 January 2022

Overview of this talk

1. Wildfire risk and wildfire data
2. Bayesian spatiotemporal regression modeling
3. Statistical inference
4. Results and their discussion

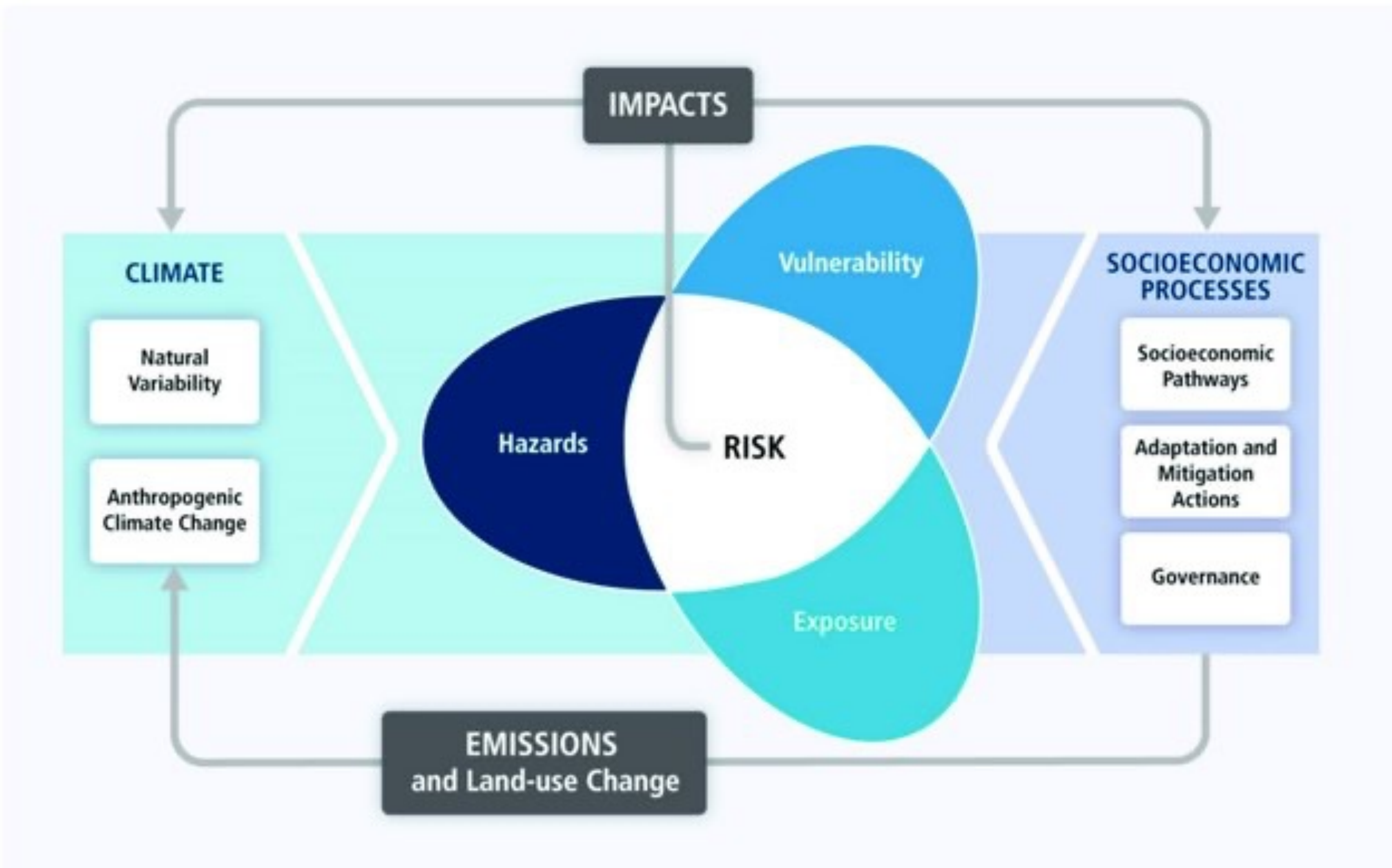
Different notions of risk

<u>**Finance, insurance:** Risk = an uncertain adverse outcome (i.e., a random variable!)</u>

<u>Climate, environment:</u> (IPCC, UN Sendai Disaster risk reduction framework):

<u>Risk = Concomitance of three components over a given space-time window</u>

(where different risk components and drivers could correspond to random variables)

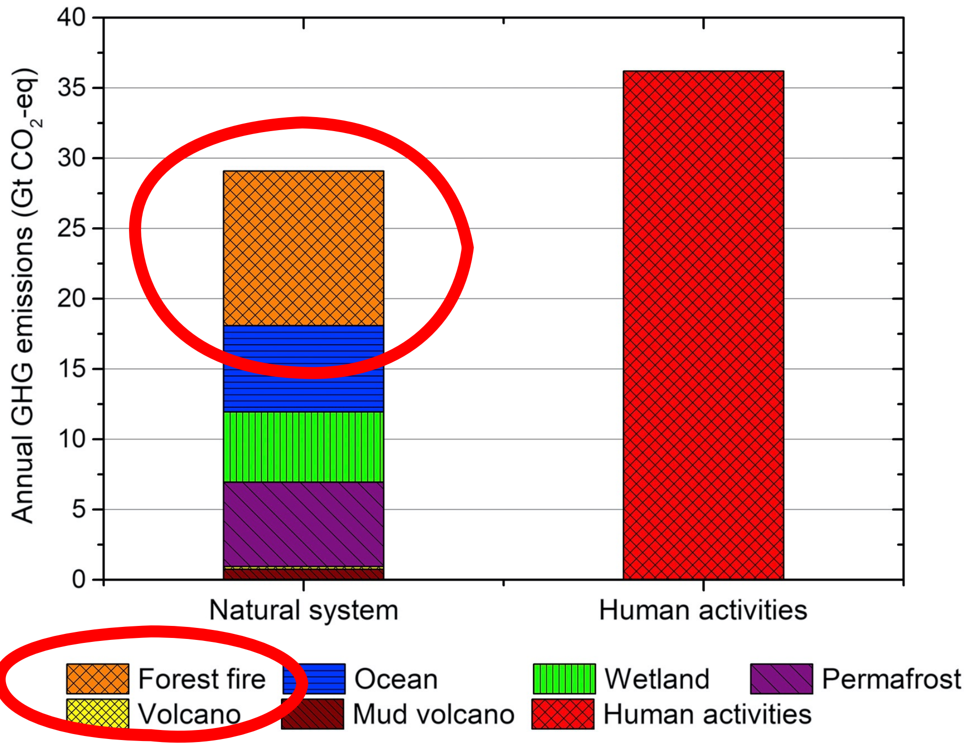


Wildfires : one of the major global environmental risks

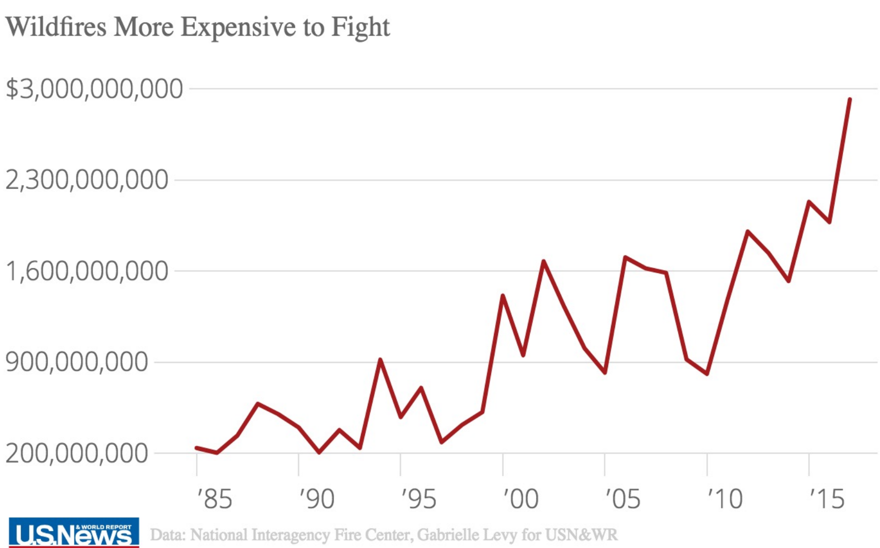
Wildfire = uncontrolled fire of natural vegetation

- First known wildfire around 400 million years ago, soon after evolution of terrestrial plants on Earth
- Major sanitary, economic and ecological damages through wildfires
- Substantial contribution to global greenhouse gas (GHG) emissions

GHG emissions



Wildfire fighting costs in the US



Conceptualization of risk components of wildfire

Different ways to conceptualize wildfire risk are possible.

<u>Exposure:</u> What is at stake and where?

- Forest and the ecosystem services it provides
  (Biodiversity, carbon stock, clean air, timber industry, leisure activities...)
- Measure for exposition: **Surface area covered by forest**

<u>Climate-related vulnerability:</u>

- Forest is vulnerable when exposed to high climatic stress
- Main ***climatic drivers:*** humidity, precipitation, temperature and wind

Their interaction in wildfire risk is complex

- We use the (Canadian) ***Fire Weather Index (FWI)*** (van Wagner, 1971), applied worldwide and also in France

<u>Hazard:</u> occurrence of wildfires, especially of very large wildfires

Wildfires are triggered by human activity (accidents, negligence, arson) or natural causes (lightning)

→ Which datasets are available for France?

- <u>Weather:</u> *SAFRAN* reanalysis data from Météo France at 8km resolution
- <u>Forest and vegetation:</u> *Corine Land Cover,* databases of IGN, ONF
- <u>Wildfire occurrences:</u> *Prométhée* database

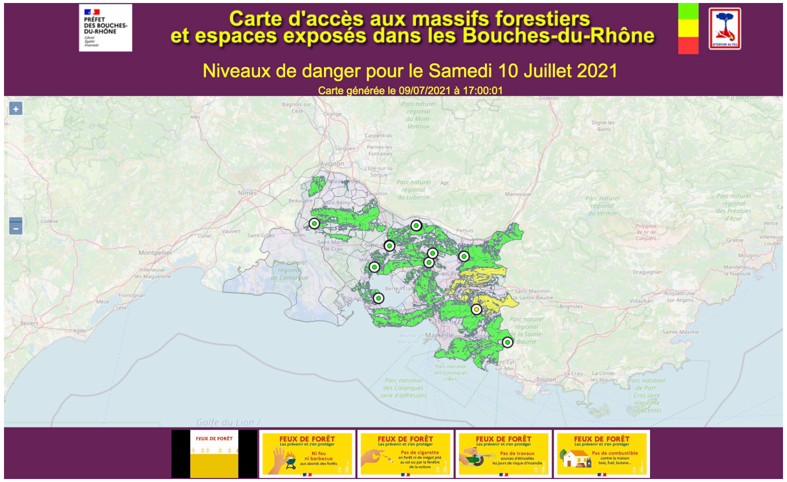
Major goals of wildfire research

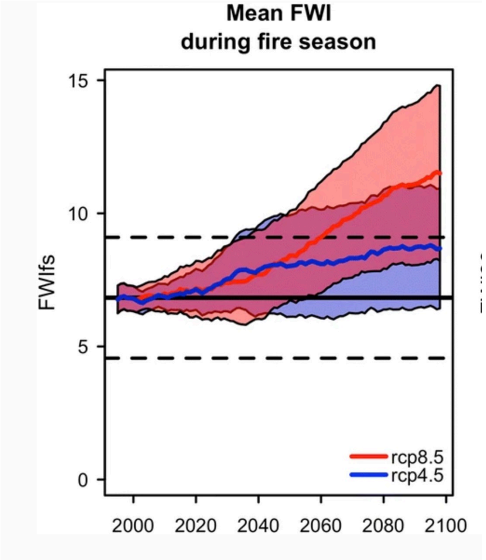
Id	Type	Category	Covariate	Estimate	CI
1	Climate		Precipitation (square root)	−3.15	[−3.66,−2.65]
2			Temperature anomaly	0.09	[0.08,0.1]
3	Land	Topography	Altitude (av.)	−1.48	[−1.64,−1.33]
4			Altitude (sd)	−1.56	[−2.66,−0.46]
5			Slope (av.)	1.2	[0.5,1.9]
6		Urban	Building cover (av.)	−5.21	[−6.71,−3.7]
7			Building cover (sd)	2.71	[1.38,4.04]
8			Path length (av.)	−0.89	[−1.64,−0.14]
9			Path length (sd)	1.49	[0.83,2.15]
10			Road length (av.)	2.45	[2,2.91]
11			Road length (sd)	−1.87	[−2.45,−1.29]
12			Secondary road length (av.)	−1.28	[−1.81,−0.76]
13			Secondary road length (sd)	2.69	[2.11,3.27]
14		Vegetation	Coniferous cover (av.)	0.36	[0.17,0.55]
15			Coniferous cover (sd)	0.29	[0.04,0.54]
16			Forest cover (sd)	0.77	[0.49,1.04]
17			Moorland (sd)	0.21	[0,0.43]
18			Protected zone cover (av.)	0.14	[0.05,0.22]
19			Shrubland (sd)	0.33	[0.05,0.6]
20			Water (av. coverage)	−1	[−1.21,−0.8]
21		Interfaces	Forest cover + building cover	4.53	[2.27,6.79]
22			Forest cover + paths	−2.54	[−4.06,−1.02]
23	Time		Time	−0.48	[−0.91,−0.05]

<u>Attribution to risk drivers</u>

- Identifying/quantifying contributions of risk drivers
- Different behavior for very large wildfires?
- Efficiency of wildfire management

<u>Risk mapping</u> and **<u>forecasting</u>**





<u>Long-term projections</u>: potential impacts of

- **Climate change**
- Land-Use Land-Cover change
- Other dynamics, e.g. related to wildfire management

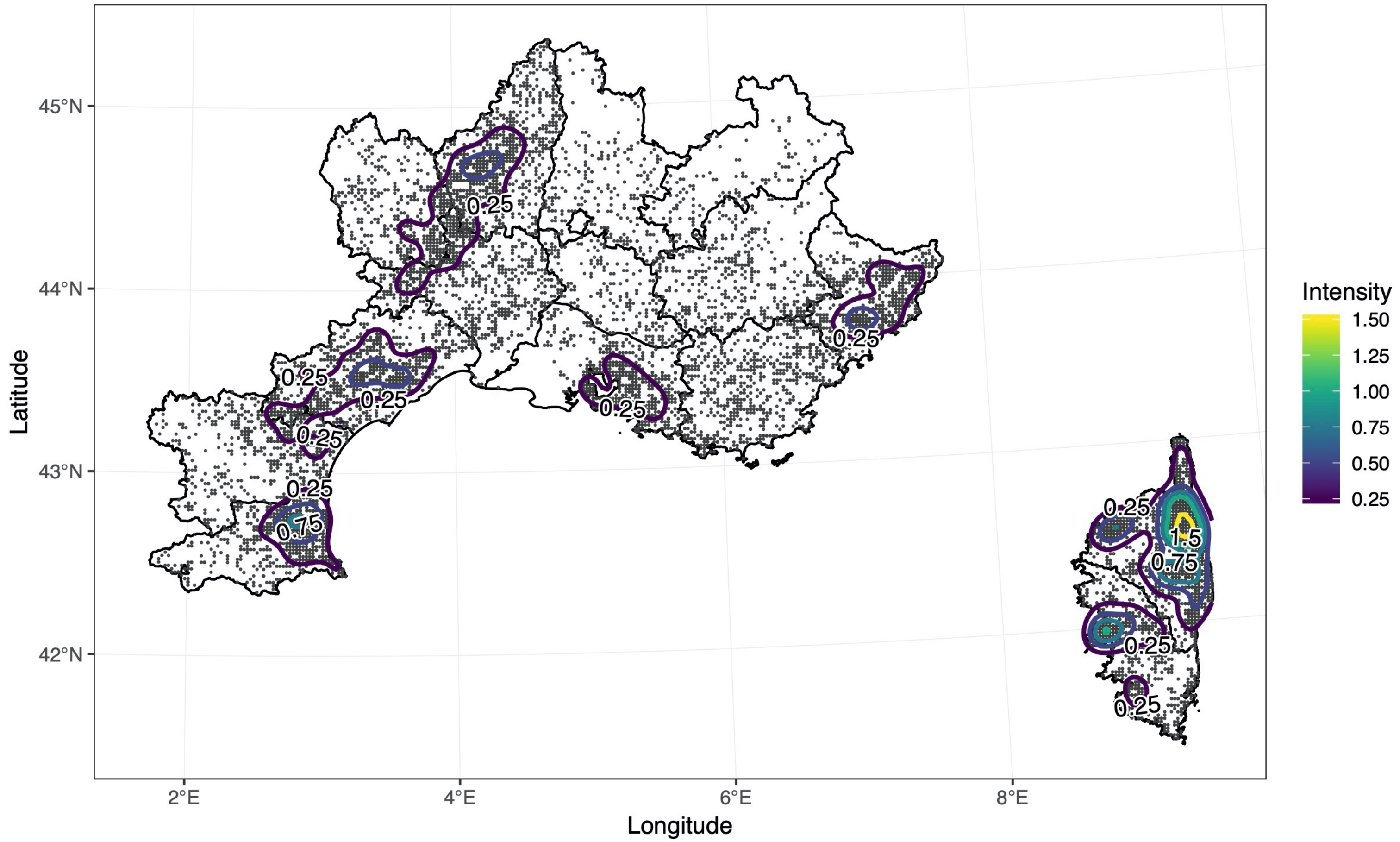
Wildfire occurrence data in Southern Frace

***Prométhée* database (since 1970s):** position of ignition (at 2km resolution), burnt area, etc.

~12000 wildfires larger than 1ha since 1995

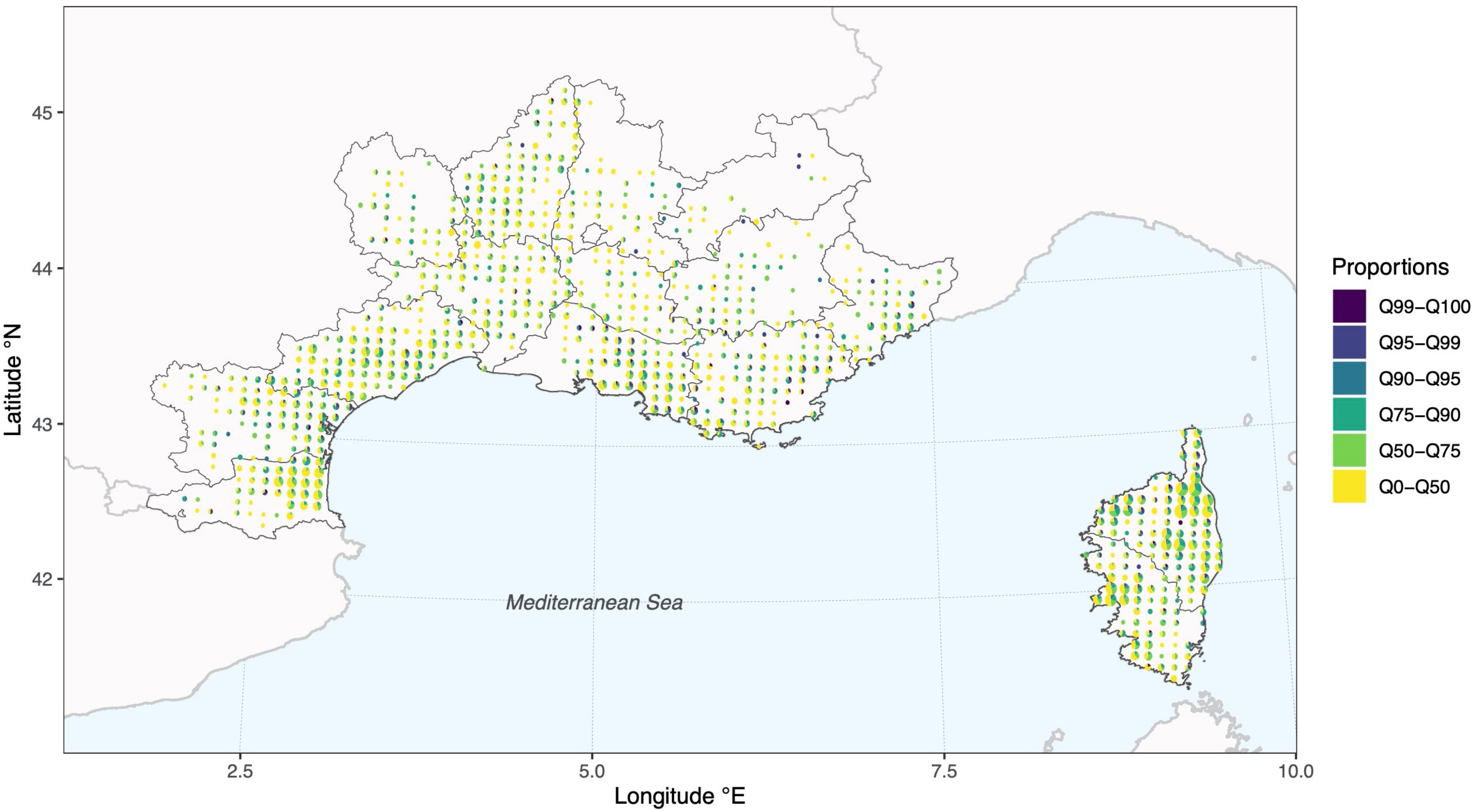
→ Data can be viewed as a ***pattern of points marked with burnt areas***

→ Mathematical representation as ***marked spatiotemporal point processes***



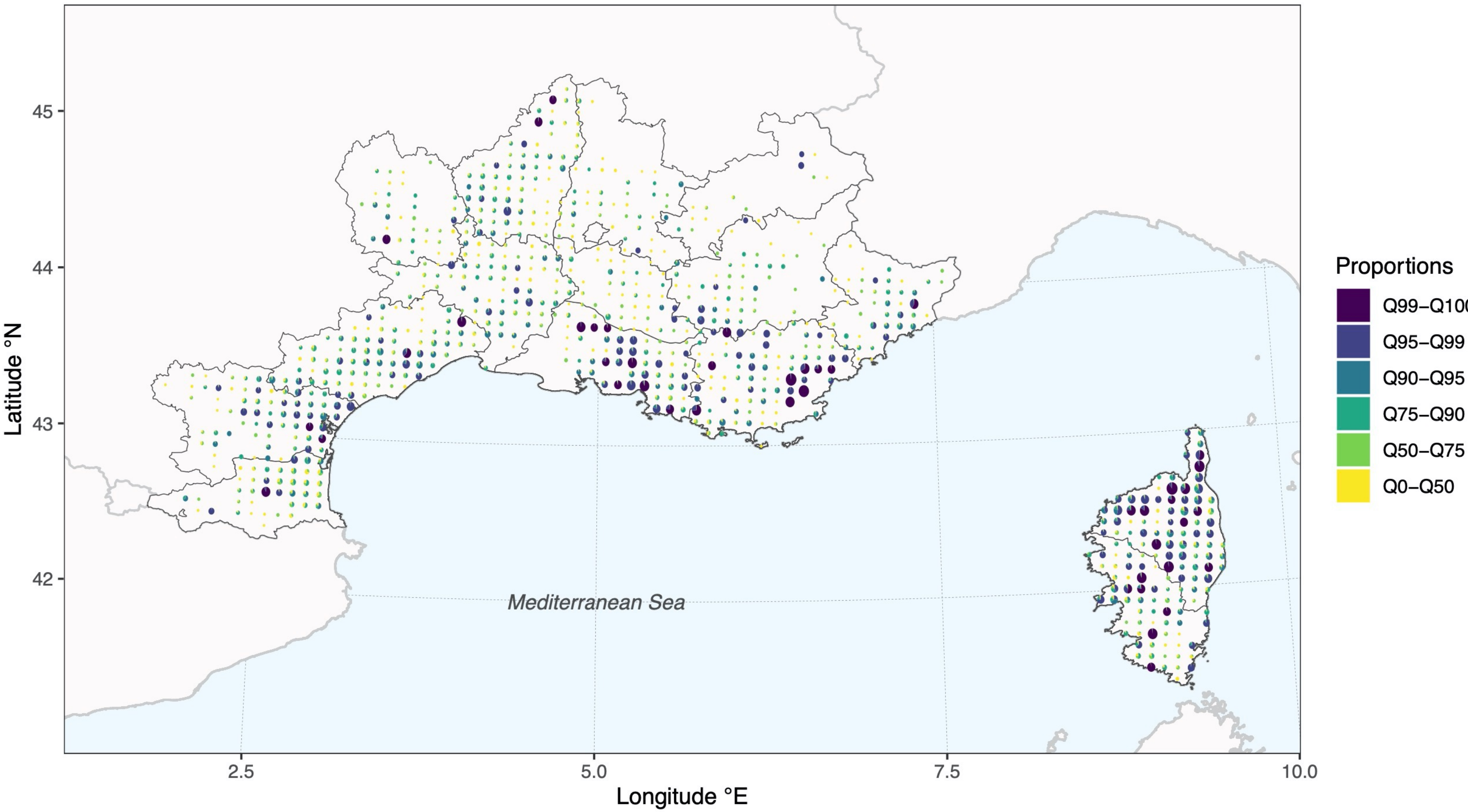
Map of wildfire counts

- Wildfire counts on **8km *SAFRAN* grid** of Météo France (for weather reanalysis data)
- Colors indicate 6 classes of wildfire sizes



Map of wildfire burnt areas

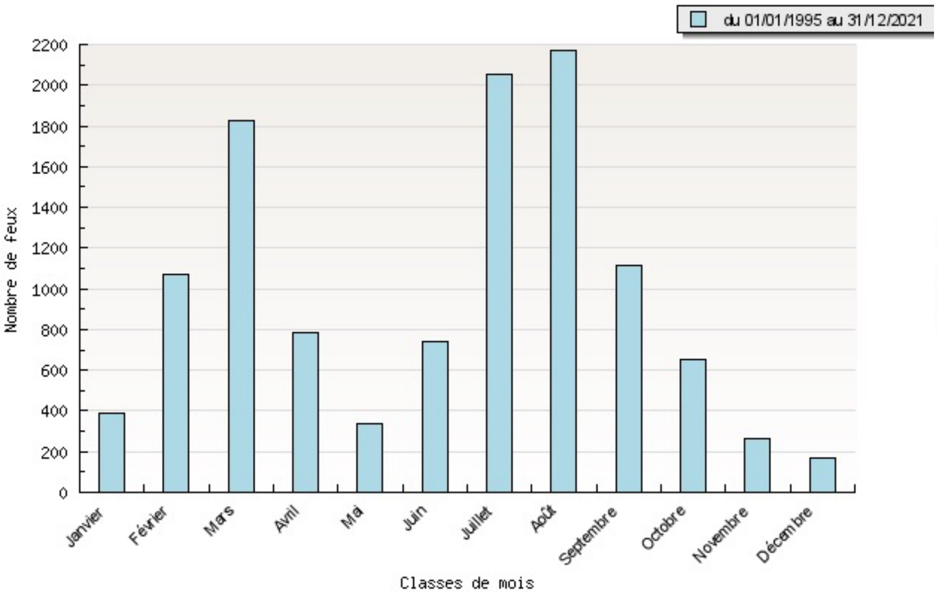
- Wildfire burnt areas on **8km *SAFRAN* grid** of Météo France
- Colors indicate 6 classes of wildfire sizes



Strong seasonalities in wildfires

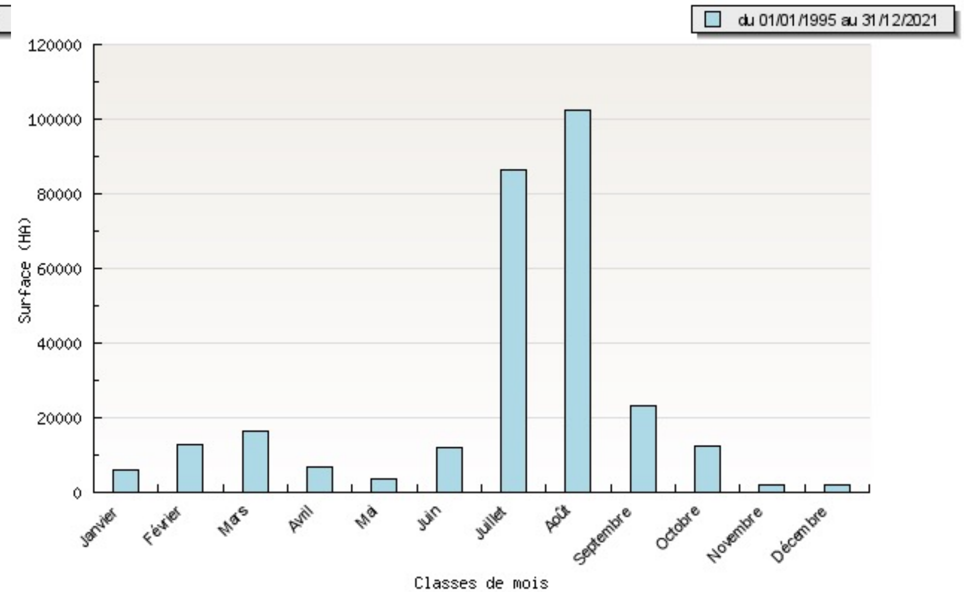
Strong correlation with weather drivers, especially for large wildfires

Counts



Indicateur : Répartition mensuelle des nombres de feux, Type de feu : Forêt, Apartir du : 01/01/1995, jusqu'au : 31/12/2021, Toute la zone Prométhée (15 départements), de : 1.01
Source des données : www.promethee.com
Données calculées le

Burnt areas



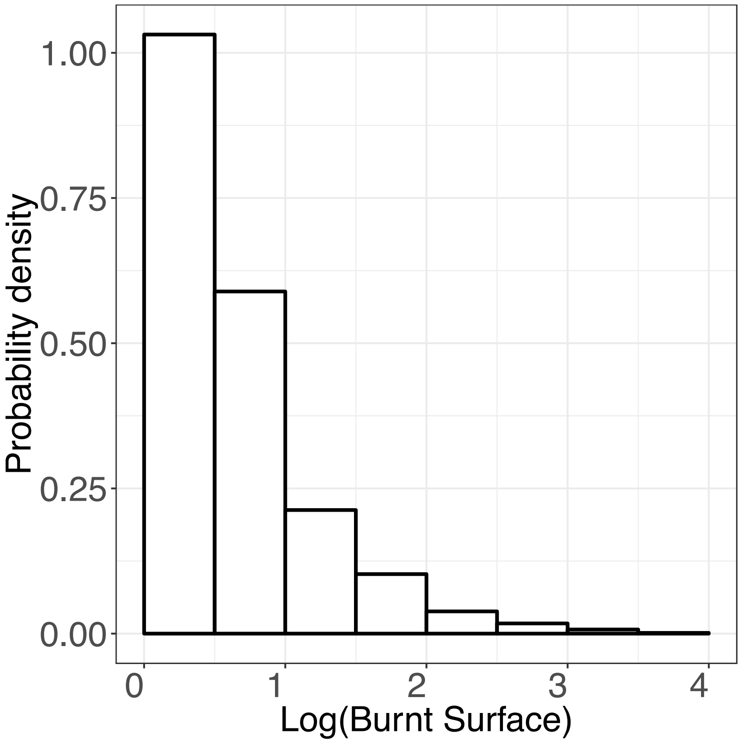
Indicateur : Répartition mensuelle des superficies brûlées, Type de feu : Forêt, Apartir du : 01/01/1995, jusqu'au : 31/12/2021, Toute la zone Prométhée (15 départements), de : 1.01
Source des données : www.promethee.com
Données calculées le

Areas burnt by wildfires have heavy-tailed distribution

"1% of fires do 99% of the damage"

- Burnt area (as a risk metric) is a proxy for wildfire damages (air pollution, biodiversity, timber loss, greenhouse gases…)
  → **Accurate modeling of extreme wildfires (with very large burnt area) is crucial**
- Data of extreme wildfires are scarce by definition (several hundreds for Prométhée zone)
- Accurate joint modeling of moderate and extreme wildfires requires including model components tailored to the extremes!

Histogram of logarithm of burnt area



The Generalized Pareto distribution for threshold exceedances

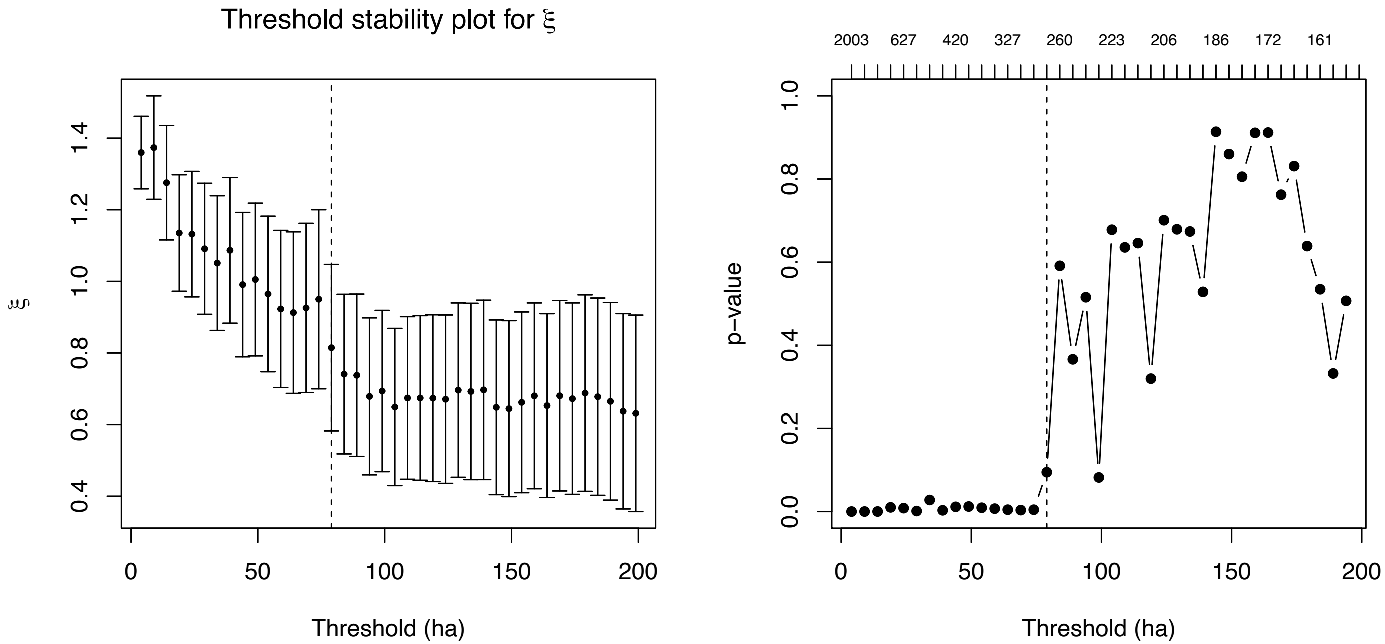
The ***Generalized Pareto distribution (GPD)*** arises asymptotically for the positive excesses of a random variable $X \sim F$. For a high threshold u, we assume that

$$\Pr(X > x + u \mid X > u) \approx 1 - \mathrm{GPD}_{\sigma,\xi}(x) = (1 + \xi x/\sigma)_+^{-1/\xi}$$

with <u>shape parameter</u> ξ and <u>scale parameter</u> $\sigma > 0$

<u>**Threshold choice for burnt areas**</u>: smallest u for which the GPD is not rejected by a statistical test

→ u = 80 ha



Large wildfire occurrence as a thinning operator

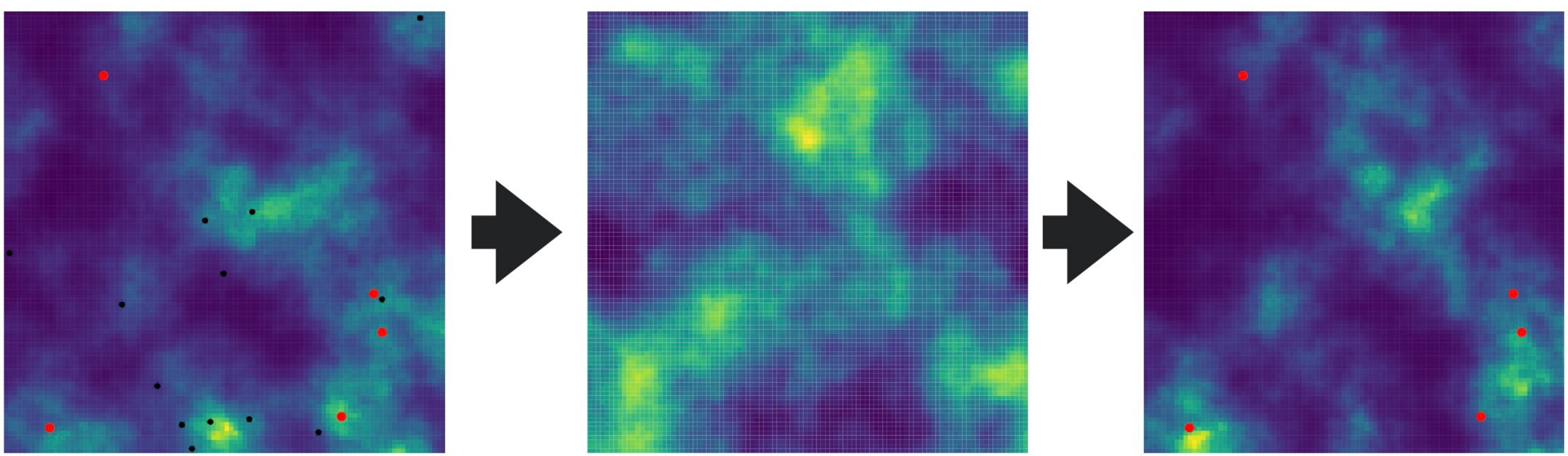
Some fires cannot be quickly extinguished and exceed the severity threshold u for extreme wildfires.

→ The pattern of extreme fires is a ***thinning*** of the overall pattern, with ***thinning probability p(s,t)***:

$$\boxed{\lambda^{\text{extreme}}(s,t) = \mathrm{p}(s,t) \times \lambda^{\text{full}}(s,t)}$$

(**Recall:** intensity $\lambda(s,t)$ = average number of points per space-time unit around (s,t))

Full point pattern **Thinning probability** **Extreme-event pattern**



A spatiotemporal stochastic modeling framework

***Firelihood*: A modeling framework developed since 2018**

→ Core developments at INRAE Avignon (URFM, BioSP)

→ National and international collaborators

In this talk: model version of *Koh, Pimont, Dupuy, Opitz (2022+)*

We jointly model the following three aspects:

- **Wildfire counts** for wildfires larger than 1 ha for each pixel-day (using SAFRAN pixels)
- **Wildfire size** (given its pixel-day of occurrence)
- Wildfires are extreme if they exceed the ***severity threshold*** **u = 80 ha**

A spatiotemporal generalized additive regression system

We use a <u>system of four regression equations</u> with the following response distributions:

- **Poisson distribution** for the pixel-day wildfire counts:

$$N_i \sim \text{Pois}(\lambda_i) \quad \text{where} \quad log\,\lambda_i = \mu_i^{\text{POIS}}$$

- **Bernoulli distribution** for the probability of occurrences becoming extreme:

$$1(Y_i > u) \sim \text{Bernoulli}(p_i^u) \quad \text{where} \quad \log\left(p_i^u/(1 - p_i^u)\right) = \mu_i^{\text{BERN}}$$

- **Beta distribution** for burnt areas of moderate-sized wildfires:

$$Y_i \mid (Y_i < u) \sim u \times \text{Beta}(\zeta_i, \nu) \quad \text{where} \quad \log\left(\zeta_i/(1 - \zeta_i)\right) = \mu_i^{\text{BETA}}$$

- **Generalized Pareto distribution** for burnt area exceedances above the threshold:

$$(Y_i - u) \mid (Y_i \geq u) \sim \text{GPD}(\sigma_i, \xi) \quad \text{where} \quad log\,\sigma_i = \mu_i^{\text{GPD}}$$

<u>General additive structure of the linear predictors:</u>

$$\mu_i^{\text{COMP}} = \beta_0^{\text{COMP}} + \sum_{k=1}^{K} g_k^{\text{COMP}}(z_k(s_i, t_i))$$

where

- COMP = { POIS, BERN, BETA, GPD}
- Functions g allow capturing nonlinear effects of predictor variables $z_k(s_i, t_i)$ (e.g., of FWI, Forest Area, Spatial location, Year, Month)

Examples of linear predictors μ

For flexible modeling, predictor components can be **nonlinear** and **stochastic**

→ **Random effects with multivariate Gaussian representations**

Types of predictor contributions:

- Spatial fields specific to a component, such as $g_1^{\mathrm{POIS}}(s_i)$
- **Spatial fields shared between components to allow for cross-correlation**, such as

$$g^{\mathrm{POIS-BETA}}(s_i),\ g^{\mathrm{POIS-BERN}}(s_i),\ g^{\mathrm{BERN-GPD}}(s_i)$$

- Nonlinear effects for Fire Weather Index (FWI), Forest Area (FA) and year (a(t))

Example: Wildfire counts

$$\mu_i^{\mathrm{POIS}} = \beta_0^{\mathrm{POIS}} + g_1^{\mathrm{POIS}}(s_i) + \beta^{\mathrm{POIS-BETA}} g^{\mathrm{POIS-BETA}}(s_i) + \boldsymbol{\beta}^{\mathbf{POIS-BERN}} \boldsymbol{g}^{\mathbf{POIS-BERN}}(\boldsymbol{s_i}) + \\ + g_2^{\mathrm{POIS}}\left(z_{\mathrm{FA}}(s_i, t_i)\right) + g_3^{\mathrm{POIS}}(z_{\mathrm{FWI}}(s_i, t_i); m(t_i)) + g_4^{\mathrm{POIS}}\left(a(t_i)\right) + g_5^{\mathrm{POIS}}\left(m(t_i)\right)$$

Example: Exceedance probability of severity threshold u

$$\mu_i^{\mathrm{BERN}} = \beta_0^{\mathrm{BERN}} + \boldsymbol{g}^{\mathbf{POIS-BERN}}(\boldsymbol{s_i}) + \beta^{\mathrm{BERN-GPD}} g^{\mathrm{BERN-GPD}}(s_i) \\ + g_1^{\mathrm{BERN}}\left(z_{\mathrm{FWI}}(s_i, t_i)\right) + g_2^{\mathrm{BERN}}\left(z_{\mathrm{FA}}(s_i, t_i)\right) + g_3^{\mathrm{BERN}}(a(t_i))$$

Background: Generalized additive models

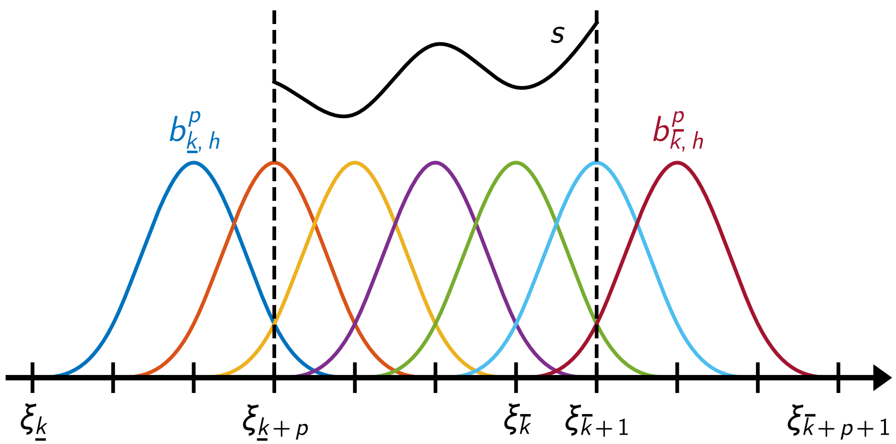
- The response distribution F depends on a parameter μ through a link function
- The ***linear predictor*** $\mu = \mu(\mathbf{z})$ can vary according to predictors $\mathbf{z} = (z_1, \dots, z_K)$,
- **Additive structure** of the linear predictor μ :

$$\mu(\mathbf{z}) = \beta_0 + \sum_{k=1}^{K} g_k(z_k)$$

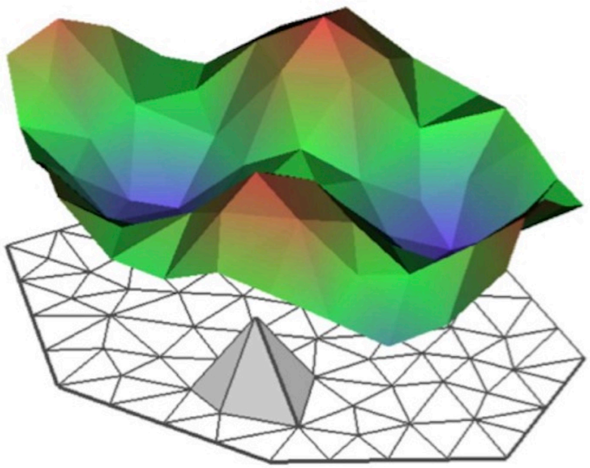
using ***basis representations* with coefficients *β*** to be estimated:

$$g_k(z_k) = \sum_{l=1}^{L_k} \beta_{k,l} b_{k,l}(z_k)$$

Cubic spline basis for scalar *z_k*



Finite-element basis for spatial locations *z_k*



Background: Gaussian random effects and the SPDE approach

1. Our models have several thousands of coefficients β to be estimated
   → We need to **control the global and local variability of estimated functions and fields**
2. For accurate modeling of **uncertainty**, we can assume a stochastic behavior of coefficients

→ A priori, we assume that the vector of coefficients $\boldsymbol{\beta}$ follows a ***multivariate normal distribution:***

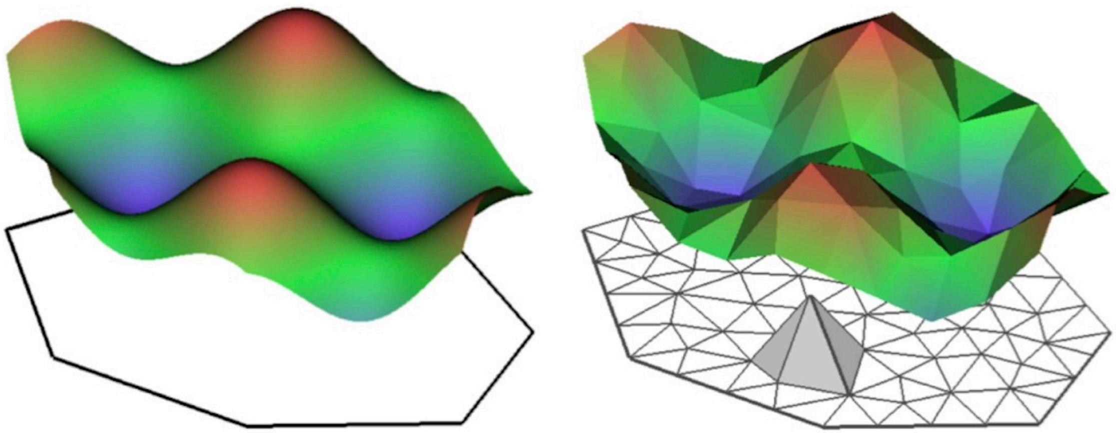
$$\boldsymbol{\beta} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{Q}^{-1}\right)$$

→ We use ***Matérn covariance functions*** for 1D-functions or 2D-fields

→ For efficient calculations with large-dimensional $\boldsymbol{\beta}$, we need **sparse precision matrices $\boldsymbol{Q}$**

Lindgren et al. (2011) obtain sparse $\boldsymbol{Q}$ for the Matérn:

- Gaussian Matérn fields are solution of a Stochastic Partial Differential Equation (SPDE)
- Solving this SPDE approximately using a finite-element basis yields sparse and explicit $\boldsymbol{Q}$



Background: Bayesian inference with INLA

→ Gaussian random effects = Gaussian process priors for $\boldsymbol{\beta} \in \boldsymbol{R}^{m}$

→ We also specify **penalized complexity prior distributions for *hyperparameters*** (variances, correlation ranges, GPD shape parameter)

Bayesian estimation requires computing the **posterior densities** of all parameters of interest.

<u>Example:</u> Posterior density of a coefficient β_i

$$\pi(\beta_i \mid \mathbf{y}) = \int_{\Theta} \int_{\mathrm{R}^{m-1}} \pi(\boldsymbol{\beta}, \boldsymbol{\theta} \mid \mathbf{y}) \mathrm{d}\boldsymbol{\beta}_{-i} \mathrm{d}\boldsymbol{\theta}$$

Such posterior densities have **very complicated form with high-dimensional integrals**
Note: Joint posterior density $\pi(\boldsymbol{\beta}, \boldsymbol{\theta} \mid \mathbf{y}) \propto \boldsymbol{\pi}(\boldsymbol{\beta}|\boldsymbol{\theta}, \boldsymbol{y}) \times \boldsymbol{\pi}(\boldsymbol{\theta})$
→ it is easy to compute but not useful in itself!

<u>Laplace approximation:</u>

- Approximate computation of integrals having general form $\int_{\mathrm{R}^m} \exp(\mathrm{g}(\boldsymbol{\beta}))\, \mathrm{d}\boldsymbol{\beta}$
- Idea: replace $\mathrm{g}(\boldsymbol{\beta})$ by its second-order Taylor development
  → $\exp(\mathrm{g}(\boldsymbol{\beta}))$ is approximated by a (scaled) multivariate Gaussian density
  → The approximate integral easy to calculate!

  → Astute use of Laplace approximations in **<u>INLA (Integrated Nested Laplace Approximation</u>**)

Model selection and validation

- We have explored different model structures:
  - Our burnt-area model (BETA-BIN-GPD) vs various alternatives (log-Gaussian, gamma…)
  - Models with and without shared random effects

How to choose and check the best model?

Prediction scores, especially for extremes

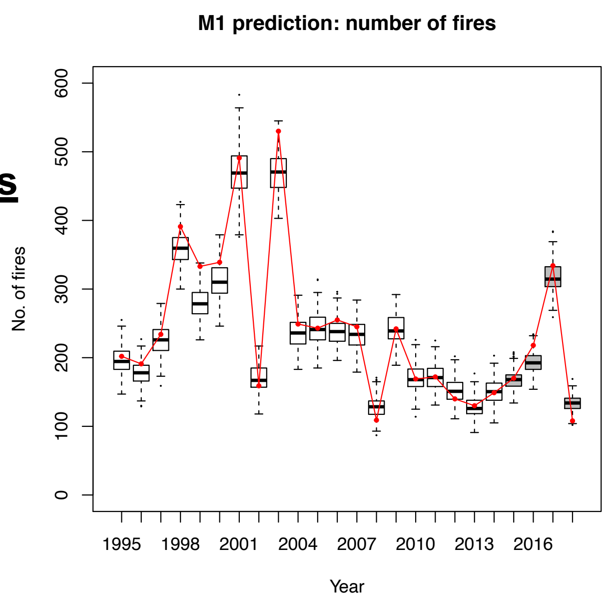
→ preference for a relatively complex model M1

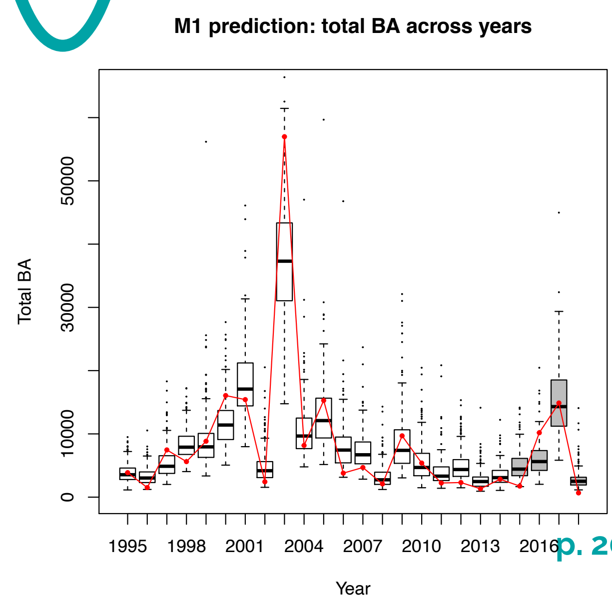
	Score	Model				
		M1	M2	M3	M4	M5
Individual fires, $n = 823$	sCRPS	2.74	2.87	2.94	2.84	3.19
	p-value	-	$< 5\%$	$< 1\%$	$< 5\%$	$< 1\%$
	Brier_{q90}	0.0855	0.0868	0.0866	0.0944	0.0967
	p-value	-	$< 5\%$	6%	$< 1\%$	$< 1\%$
	$1 - \text{AUC}_{q90}$	0.3052	0.3502	0.3516	0.3184	0.3122
	p-value	-	$< 5\%$	$< 5\%$	40%	41%
Dép-month, $n = 75$	sCRPS	3.55	3.62	3.64	3.62	3.58
	p-value	-	7%	7%	9%	39%

Predictions vs observations

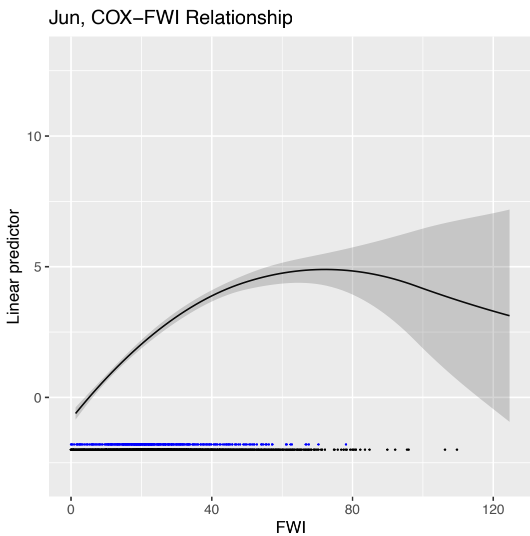
(within-sample 1995–2014, out-of-sample 2015–2019)

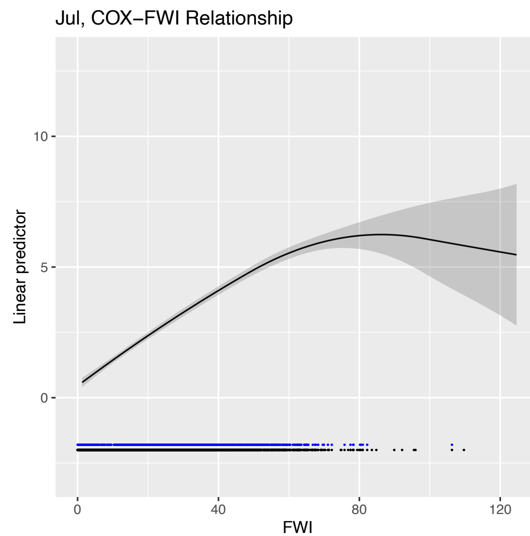
→ Good performance!

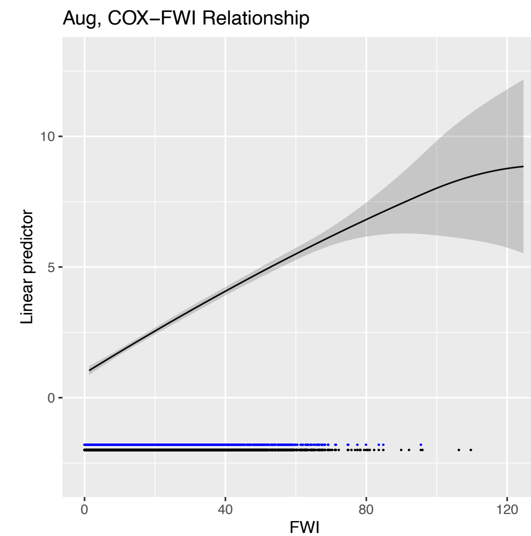


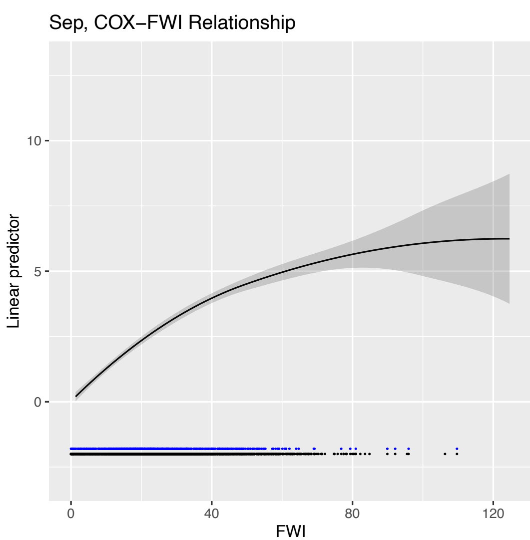


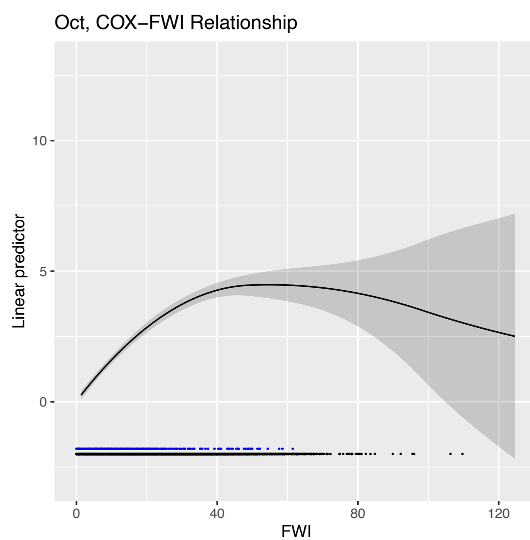
Results: Monthly FWI effect on numbers of fires







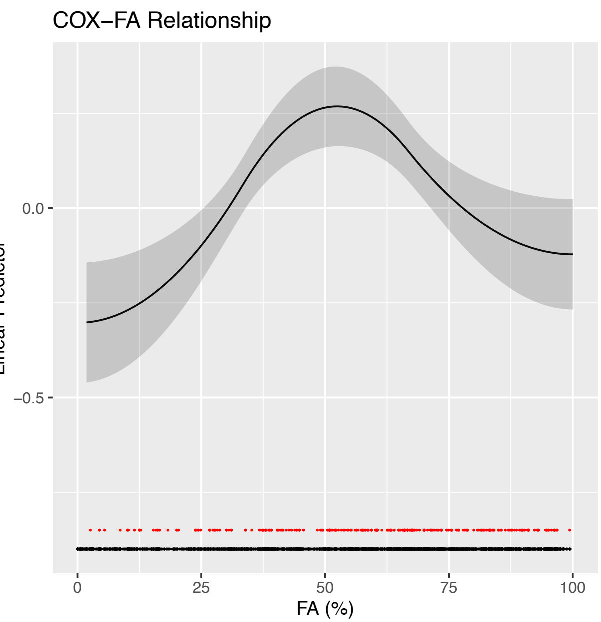




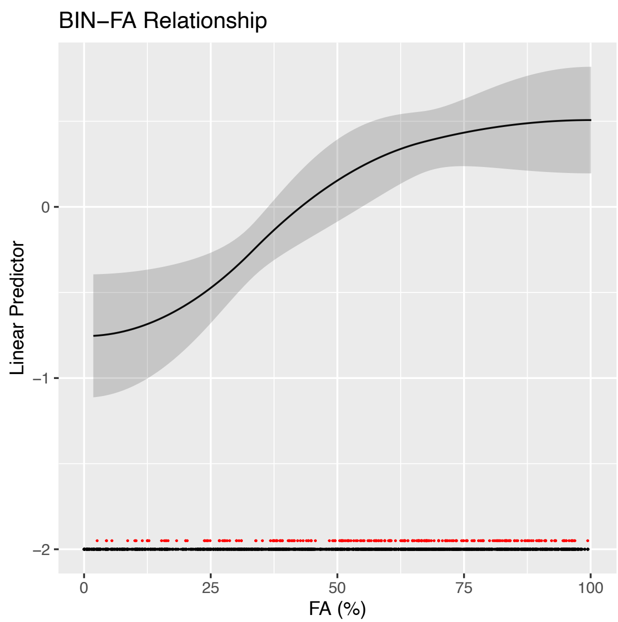
- Strongly nonlinear effect
- Month-specific response
- Strongest effect in August

Results: Effects of Forested Area

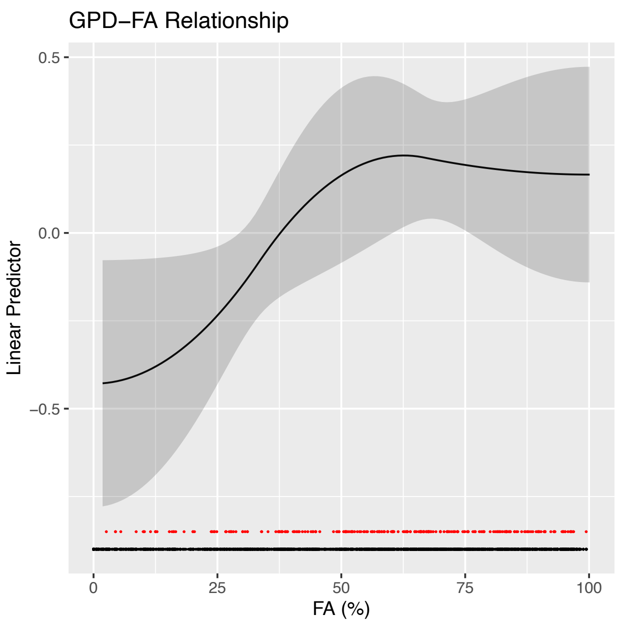
Effect on POIS



Effect on BERN



Effect on GPD

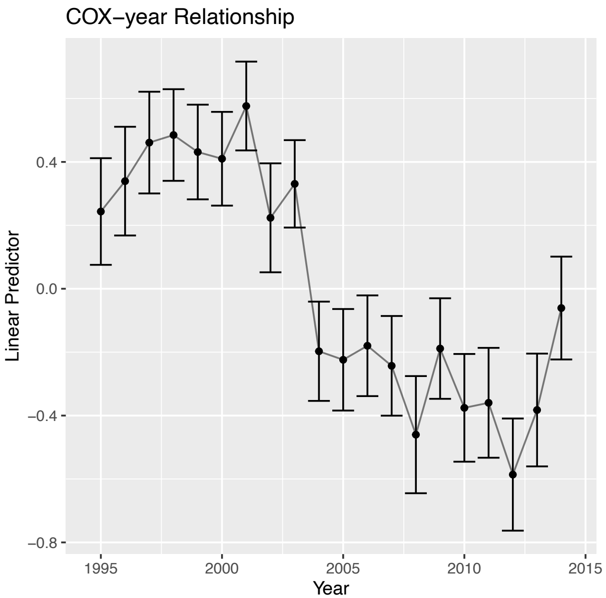


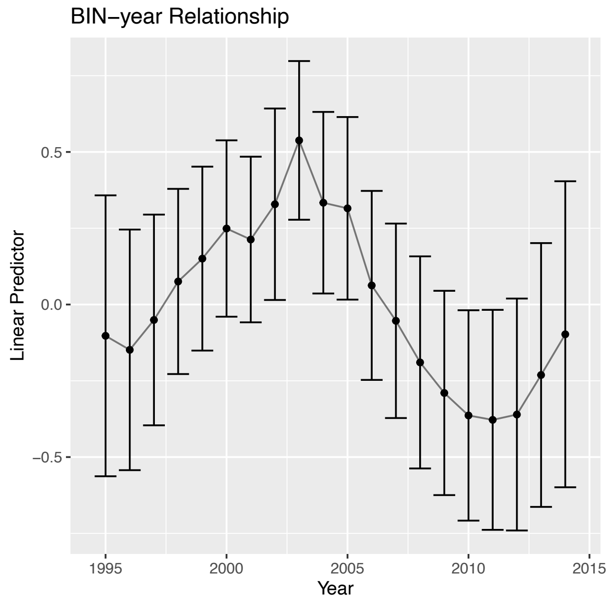
Relatively more wildfires in pixels with intermediate forest cover

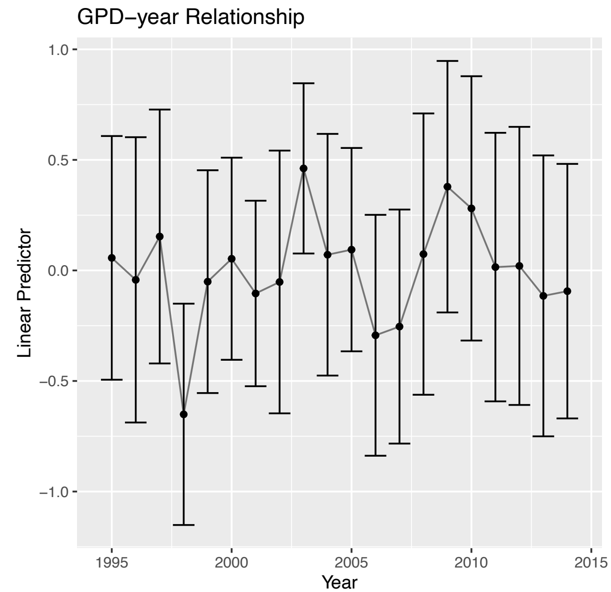
Relatively more „escaping“ wildfires that become very large in dense forest cover

Results: Yearly effect over the study period

- Increasing trend in wildfire activity until the „catastrophic“ year 2003
- Strong drop after 2003, attributed to improved wildfire prevention and fighting
- Post-2012 increase in wildfire numbers (especially extreme fires) should alarm wildfire managers!



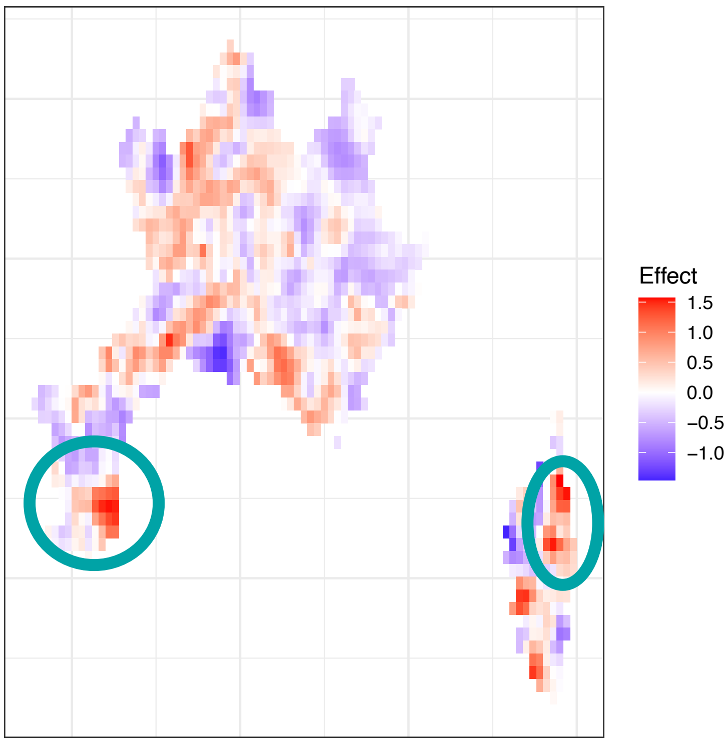




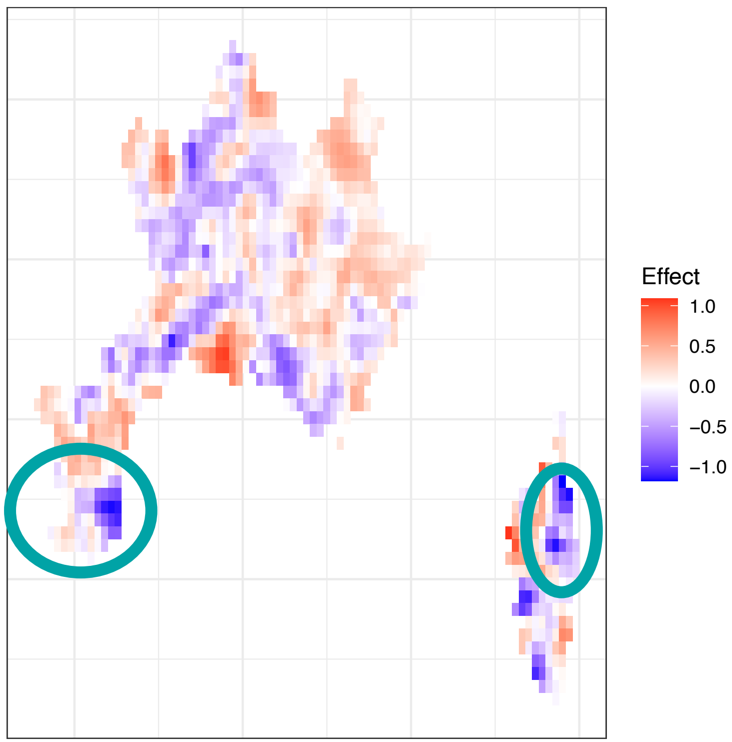
Results: Shared spatial effects

Example: Spatial effect shared between fire numbers and exceedance probabilities

Contribution to fire numbers



Contribution to exceedance probabilities



→ Highlighted zones incur relatively frequent but non-extreme wildfires

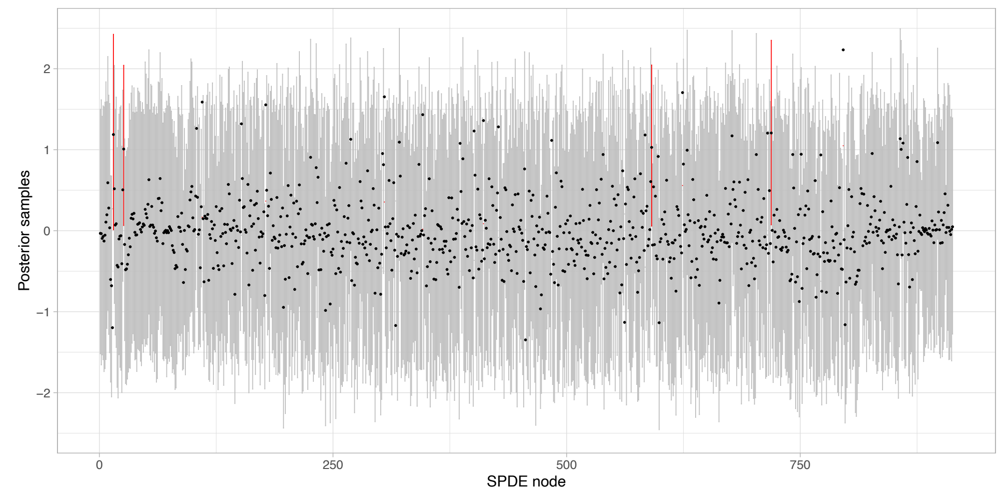
Results: Decreased uncertainty with shared random effects

Sharing can reduce estimation uncertainties for components with „weak data signal“ (e.g. extremes)

→ We let the model decide if this is possible!

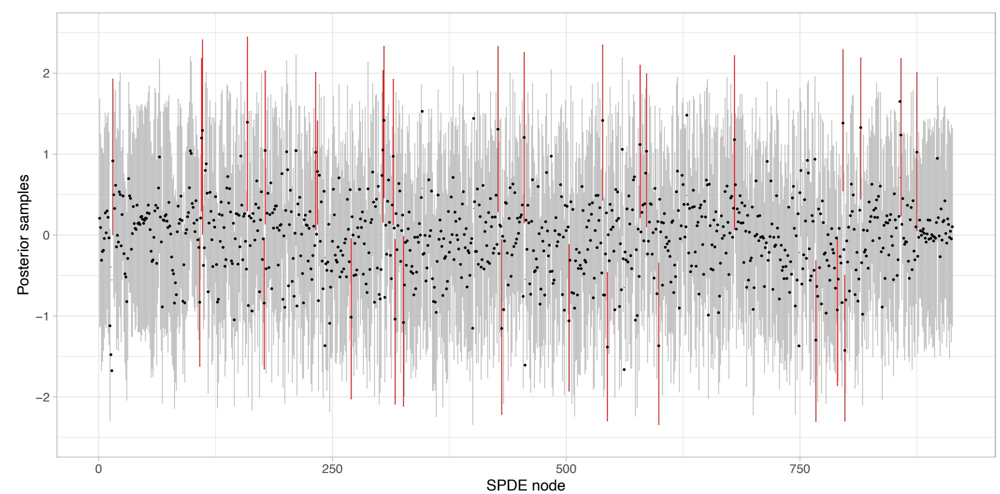
<u>Example:</u> Credible intervals for combined spatial effects in exceedance probabilities

No sharing



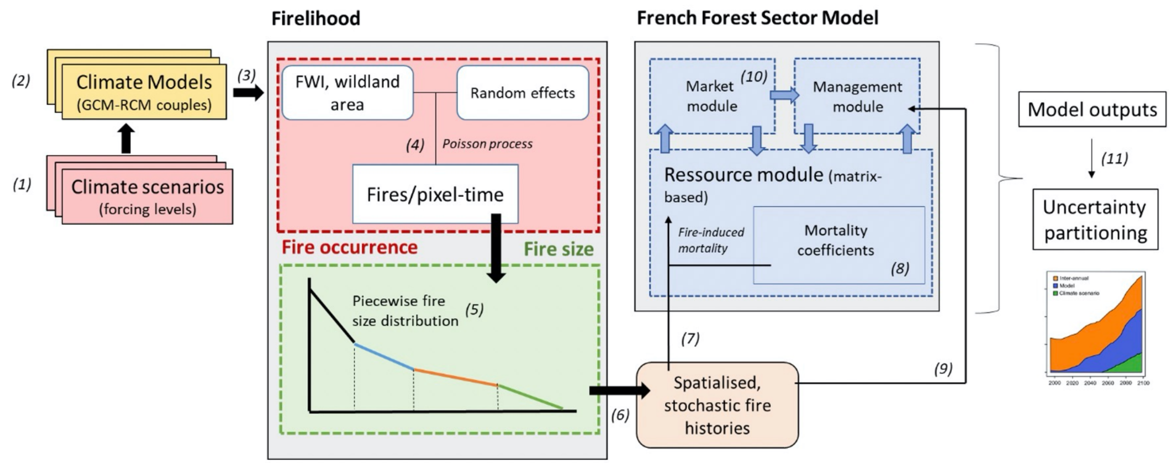
With sharing:

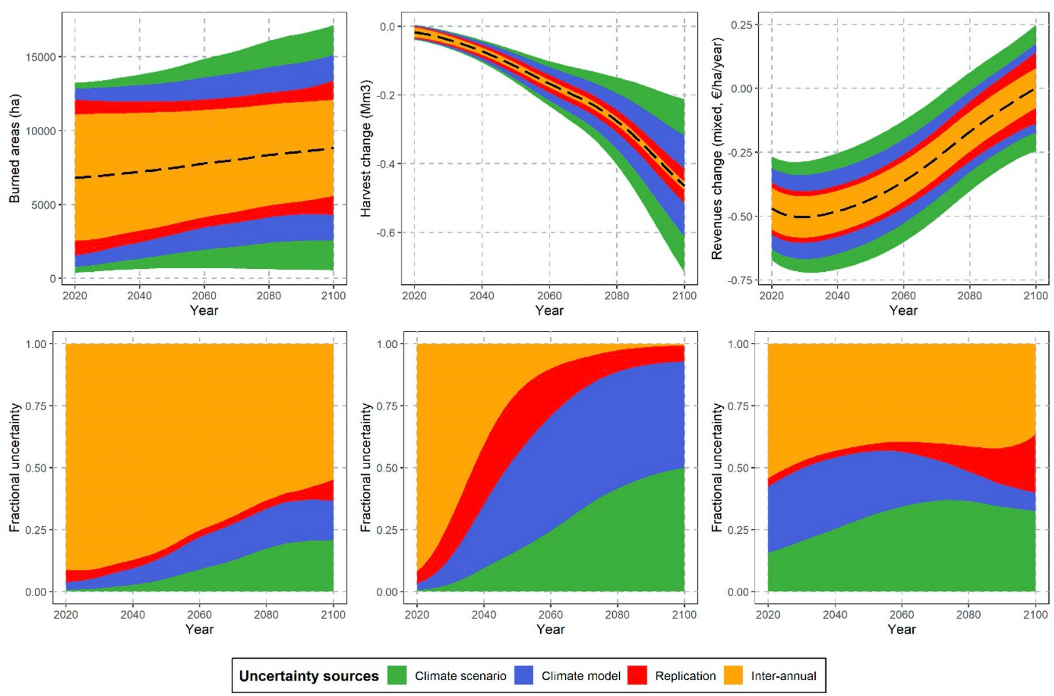
- (Slightly) narrower CIs
- More “significant“ pixels (in red)



Application: bio-economic forest projection (climate change)

Rivière et al. – A Bioeconomic Projection of Climate-induced Wildfire Risk in the Forest Sector. *Preprint.*





Projected change (–2100) for:

- Burnt areas (left)
- Harvests (middle)
- Expected revenues (right)

Uncertainty decomposition

Concluding remarks

- Complex **Bayesian spatiotemporal models** jointly enable **attribution, prediction and projection**
- Models detect **strong residual spatial-temporal trends** not well explained by „physical“ predictors
- **Shared random effects** improve modeling of extreme wildfires by borrowing information from moderate wildfires (where possible)
- Climate change will strongly increase vulnerability but adaptation measures can substantially mitigate wildfire risk

<u>Ongoing projects:</u>

(PhD of Jorge Castel-Clavera and H2020 project FIRE-RES)

→ Better inclusion of Land-Use Land-Cover and wildfire management variables

→ Construction of new wildfire-danger indices, more accurate than Canadian FWI

→ Better characterization and forecasting of extreme wildfires

→ Extension to other environments of other European countries

<u>References:</u>

- Pimont et al. (2021). Prediction of regional wildfire activity in the probabilistic Bayesian framework of Firelihood. *Ecological applications*, e02316
- Fargeon et al. (2018). Assessing the increase in wildfire occurrence with climate change and the uncertainties associated with this projection. In *8. International conference on forest fire research.*
- Opitz et al. (2020). Point-process based Bayesian modeling of space–time structures of forest fire occurrences in Mediterranean France. *Spatial Statistics*, *40*, 100429.
- Gabriel et al. (2017). Detecting and modeling multi-scale space-time structures: the case of wildfire occurrences. *Journal de la Société Française de Statistique*, *158*(3), 86-105.
- Koh et al. Spatiotemporal wildfire modeling through point processes with moderate and extreme marks. *arXiv preprint arXiv:2105.08004*.
- Rivière et al. A Bioeconomic Projection of Climate-induced Wildfire Risk in the Forest Sector. *Preprint.*