

Congrès des actuaires 2023

# Valorisation des données spatiales : une réponse aux défis actuels ?

22 juin 2023 – 14h00-14h45



**Eve TITON**  
Responsable risques techniques &  
Senior Consultante  
**Milliman**



**Antoine RAINAUD**  
Analyste risques climatiques &  
Consultant  
**Milliman**



**Bruno BECHA**  
Responsable Actuariat Data  
Science  
**Akur8**



# Données spatiales et Assurance

# Données spatiales et Assurance

## Définition et enjeux



L'Open Data : des données que tout le monde peut consulter, utiliser et partager



De nouvelles données et opportunités



Identification des cas d'usage avec impacts business



Vers le développement de nouvelles compétences et techniques actuarielles



La France, leader européen de l'Open Data



Diversité de types et de formats



Multiplicité des sources



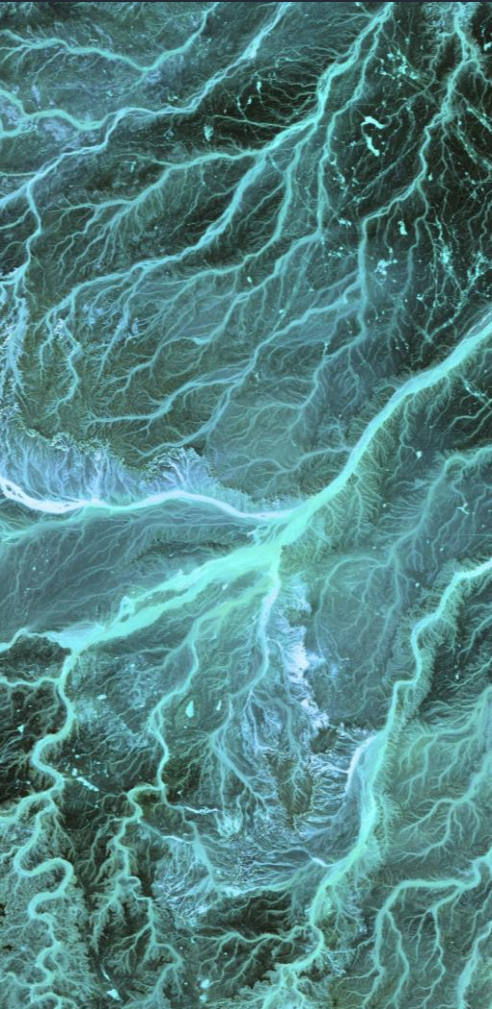
Différents niveaux de qualité des données



Fréquence de mise à jour

# Données spatiales et Assurance

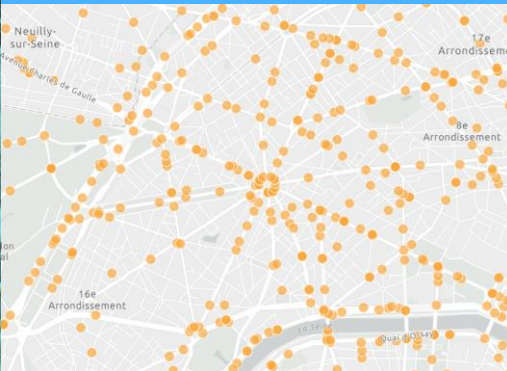
## Présentation des données spatiales



### Points



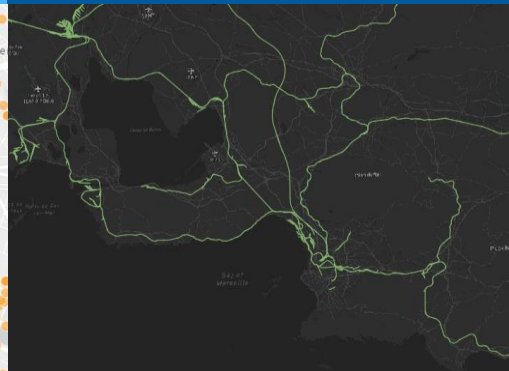
- Localisation des assurés
- Point GPS
- Lieu d'un accident
- Centroïde d'une commune



### Lignes



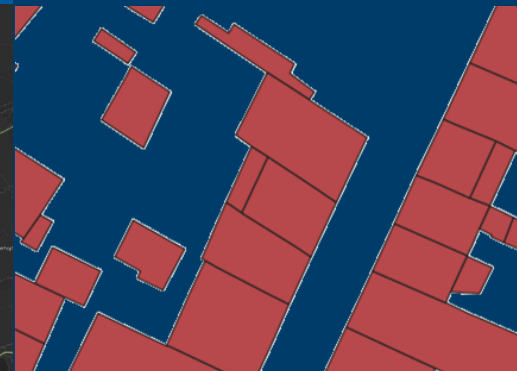
- Fleuves et rivières
- Réseau routier
- Frontières



### Polygones



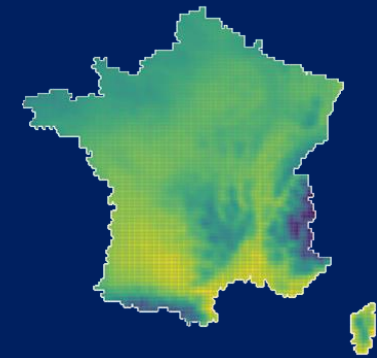
- Parcelle
- Surface d'un bâtiment
- Territoire
- Zone de risque



### Raster

	4	4	3
4	4	4	4
4	3	3	1
4	3	3	1

- Altitude
- Utilisation du sol
- Images satellites

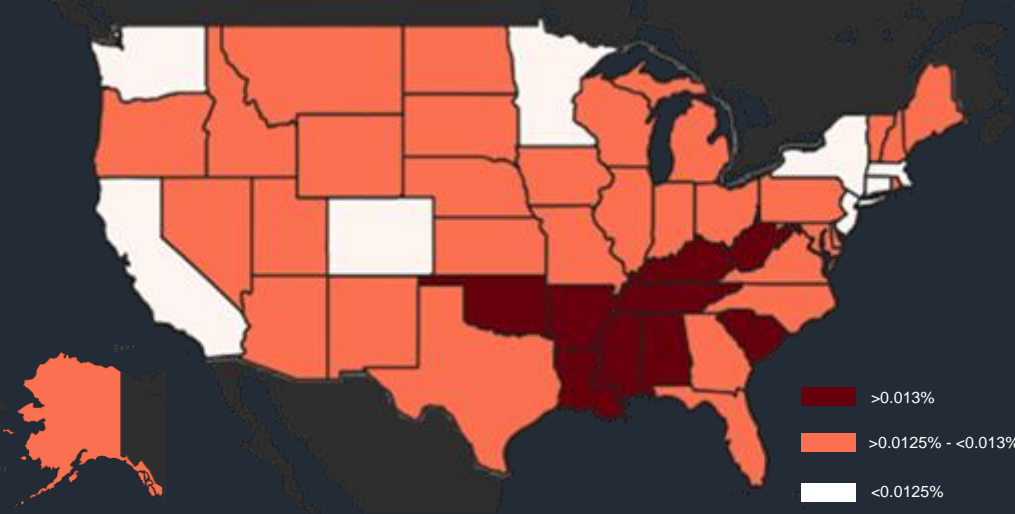




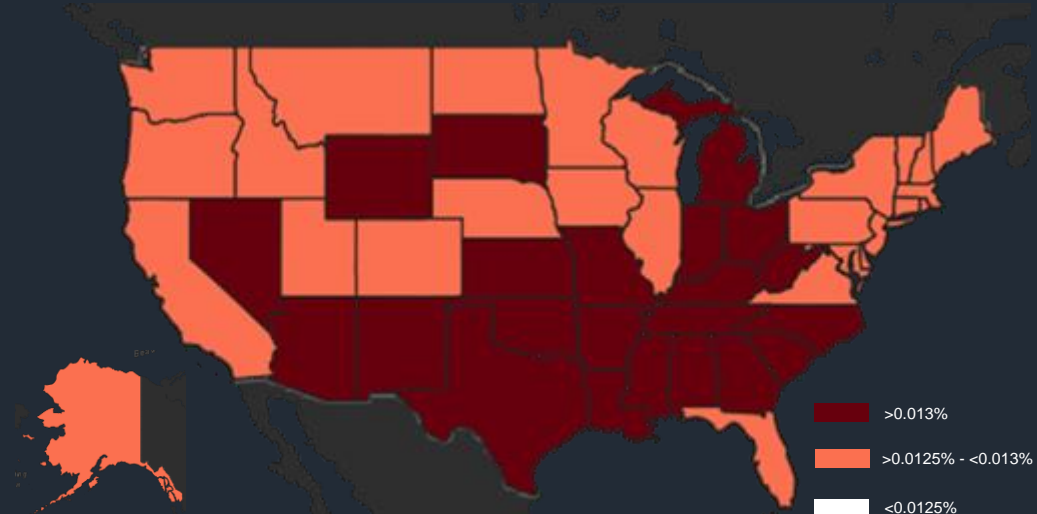
# Données spatiales et Assurance

Première visualisation des données spatiales : la carte choroplèthe

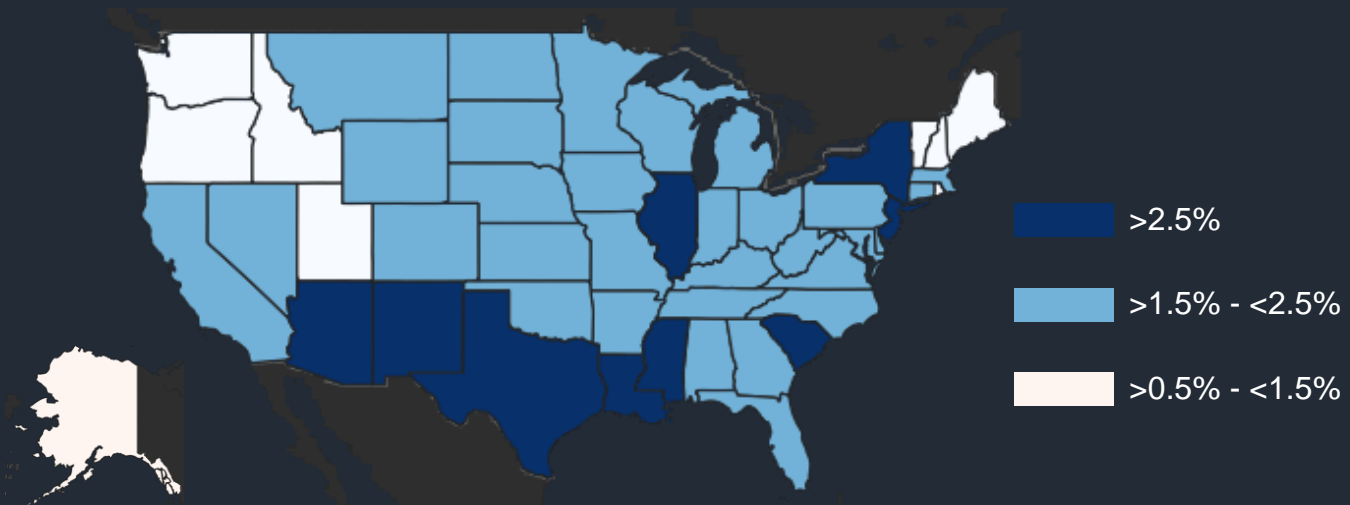
## Taux de mortalité 2019



## Taux de mortalité 2020



## Taux évolution 2019-2020

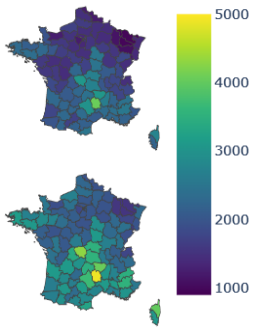


Sources : Census, ESRI et USMDB

# Données spatiales et Assurance

Des exemples de données structurées...

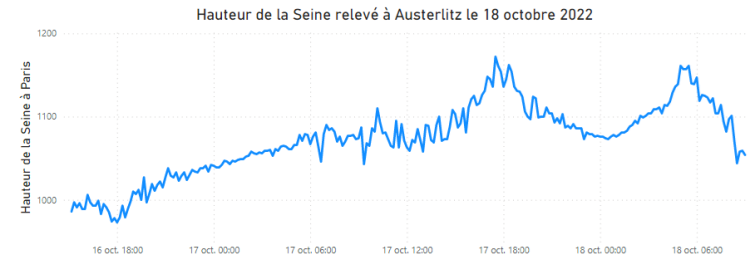
## Maladies mentales en France



La base de données des ALD d'AMELI renseigne sur l'évolution du **nombre de malades**.

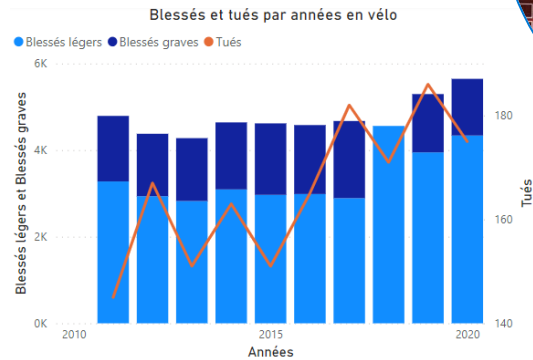
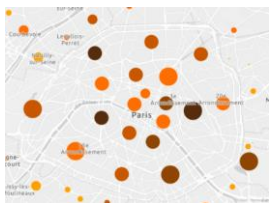
## Débit d'un cours d'eau

L'API d'Eau France donne le **niveau et débit des cours d'eau** en France avec 4200 stations de mesures.



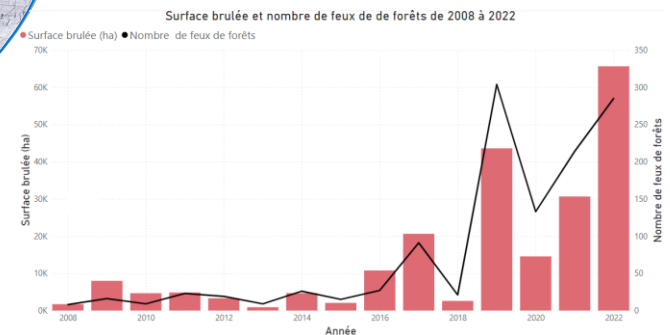
## Accidents corporels

La **BAAC** répertorie les **accidents corporels de la circulation**.



## Surface brûlée par les feux de forêts

La plateforme **EFFIS** renseigne en détail les informations sur les **feux de forêts en Europe**. Elle intègre notamment un **calcul de risque de feux de forêts**.



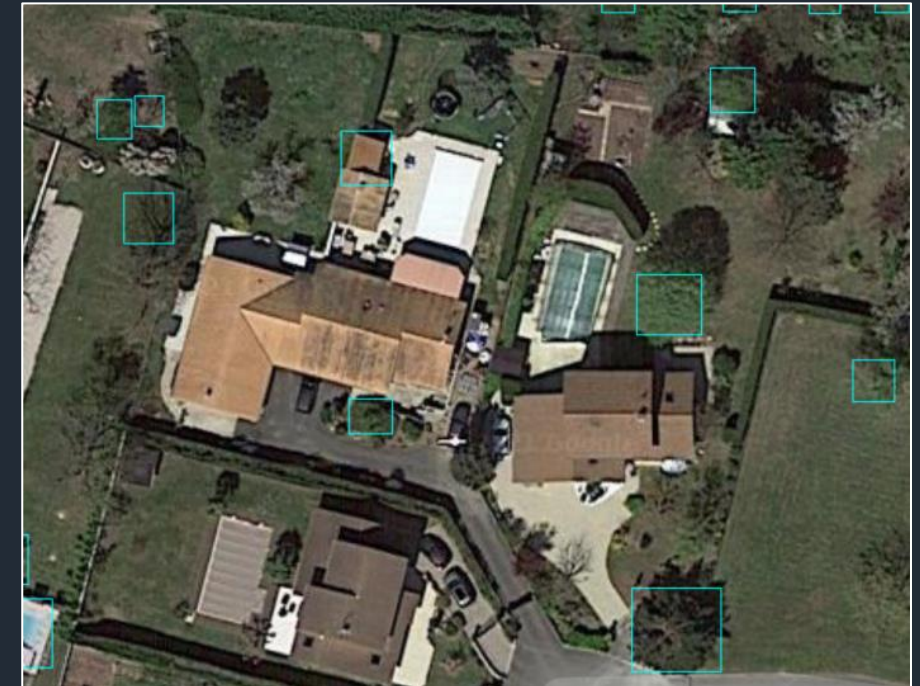
# Données spatiales et Assurance

... mais aussi non structurées

- Avec de nombreux cas d'usage dans l'analyse du risque, l'aide à la souscription, l'enrichissement de modèles de suivi de risque, la détection de fraude, etc.



Google Street View  
Utilisation d'OpenCV (Python)



Google Map  
Utilisation de DeepForest (Python)



# Données spatiales et Assurance

## Cas d'usage en assurance

- Tous les domaines sont concernés : Assurance Vie, Non-Vie, Santé et Prévoyance.



### Tarification

Enrichissement des variables tarifaires et des modèles

Etablissement de zoniers



### Souscription

Allègement du parcours de souscription et des questionnaires

Aide à la souscription avec des outils transformant une adresse en un KPI de risque



### Prévention et suivi des risques

Mise en œuvre d'alertes en cas de météo à risque

Mise à disposition de fiches descriptives des risques et d'actions de prévention



### Provisionnement

Estimation de charges ultimes anticipée sur des risques complexes (comme la sécheresse)

Aide au provisionnement individuel



### Gestion des sinistres

Détection de fraude via des analyses complémentaires

Versements automatiques et assurance paramétrique



### Risques climatiques

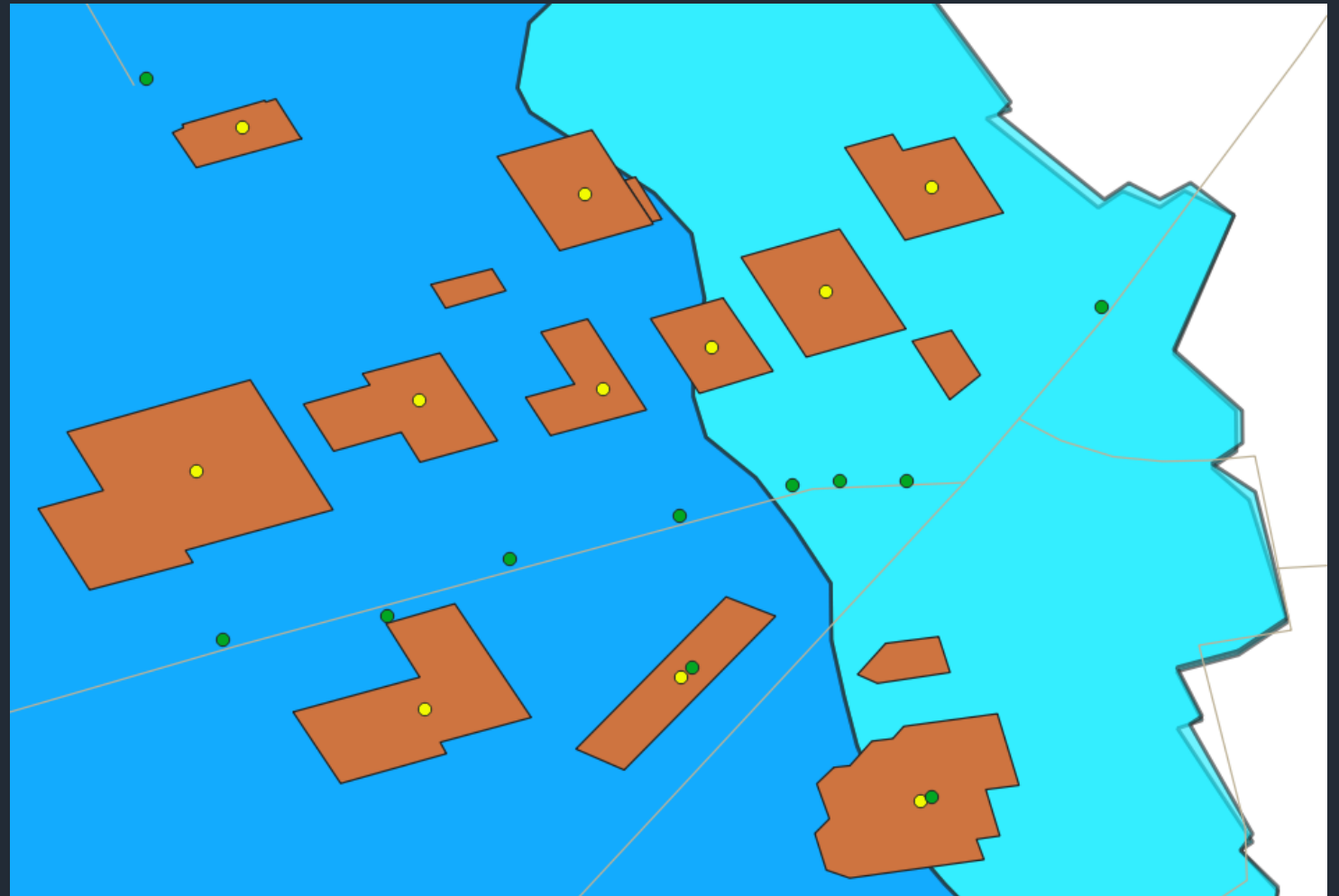
- Modélisation des impacts du changement climatique pour les risques **santé et mortalité** (exemples: vagues de chaleur, pollution de l'air, maladies transmises par les moustiques, etc.)
- Modélisation des **impacts du climat en assurance Non-Vie** (récolte, MRH, etc.)
- Modélisation des **pandémies**.(exemples: prise en compte des déplacements de population)



# Données spatiales et Assurance

Le complexe géocodage et les problèmes de précision géographique

- Les **cartographies d'exposition** deviennent plus **précises** permettant de mieux comprendre les risques naturels.
- L'**inondation** est très **localisée** alors que les **tornades** sont très **étendues**.
- Le **géocodage** est une tâche complexe qui peut mener à de nombreuses erreurs dans l'estimation du risque à souscrire.
- Comment évaluer la qualité du **Géocoding** ?





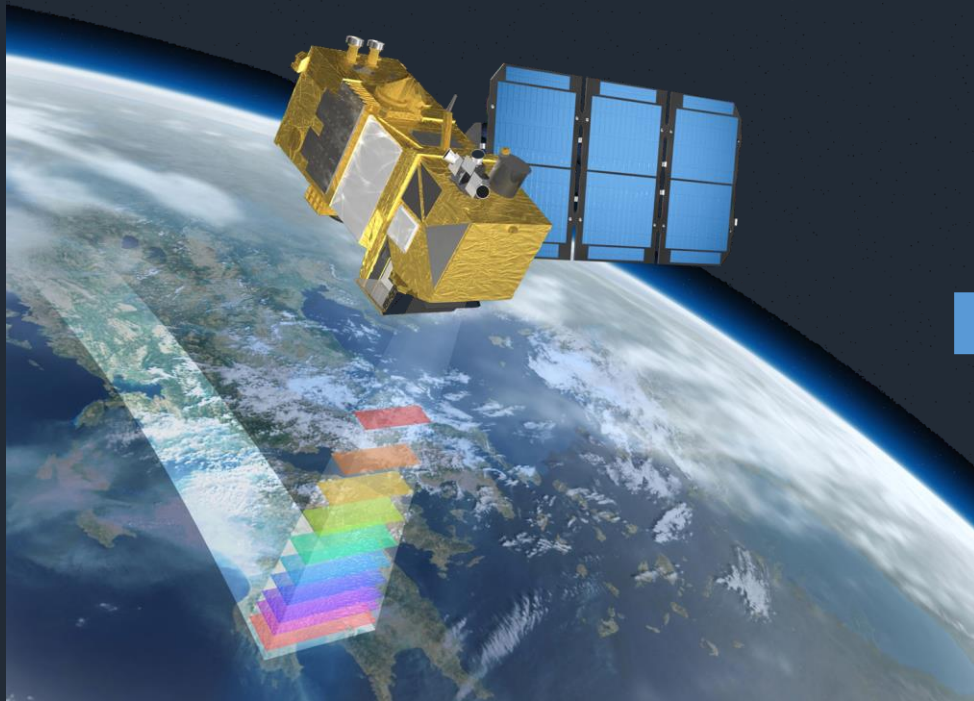
# La géomatique et les risques climatiques

Géomatique, images aériennes et feux de forêt

# La géomatique et les risques climatiques

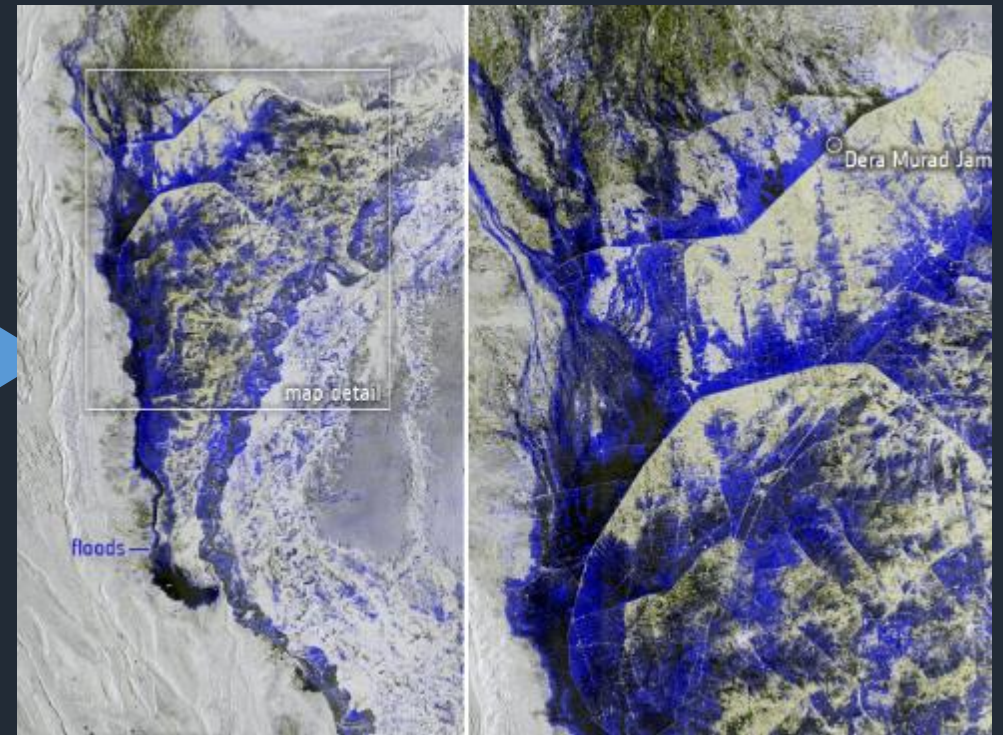
Les images aériennes, la télédétection et leurs cas d'usages

Les **images aériennes** pour cartographier, analyser et appréhender les **risques climatiques**.



Source : ESA

Géomatique



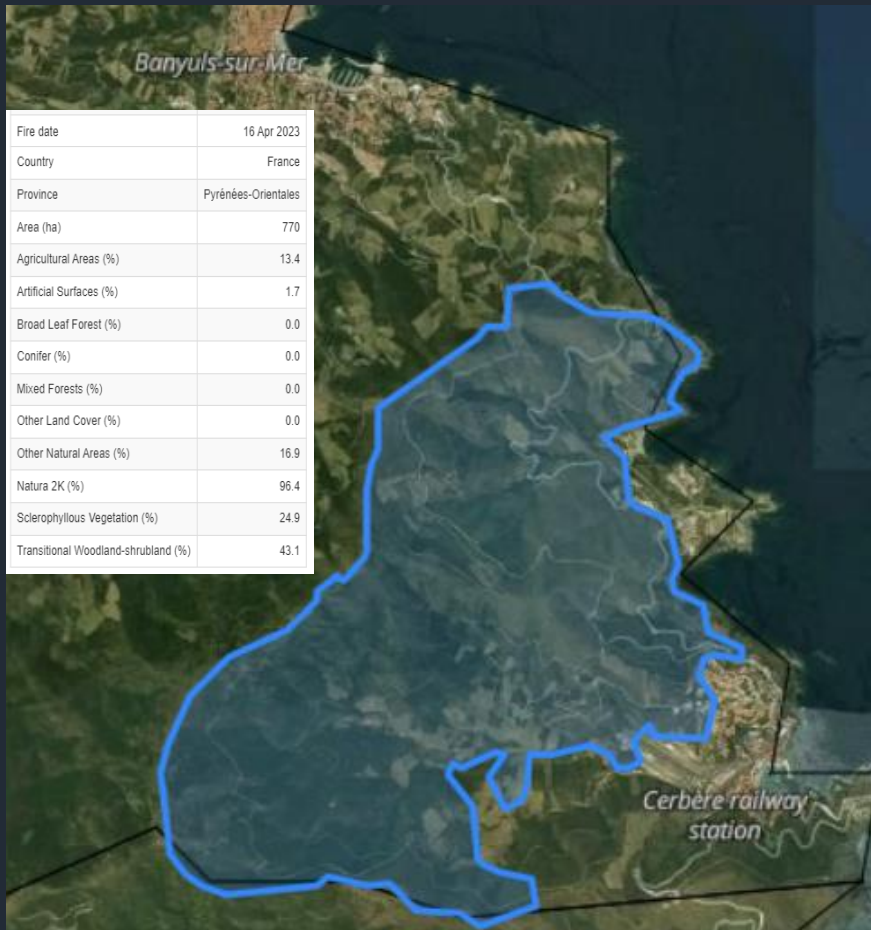
Source : ESA



# La géomatique et les risques climatiques

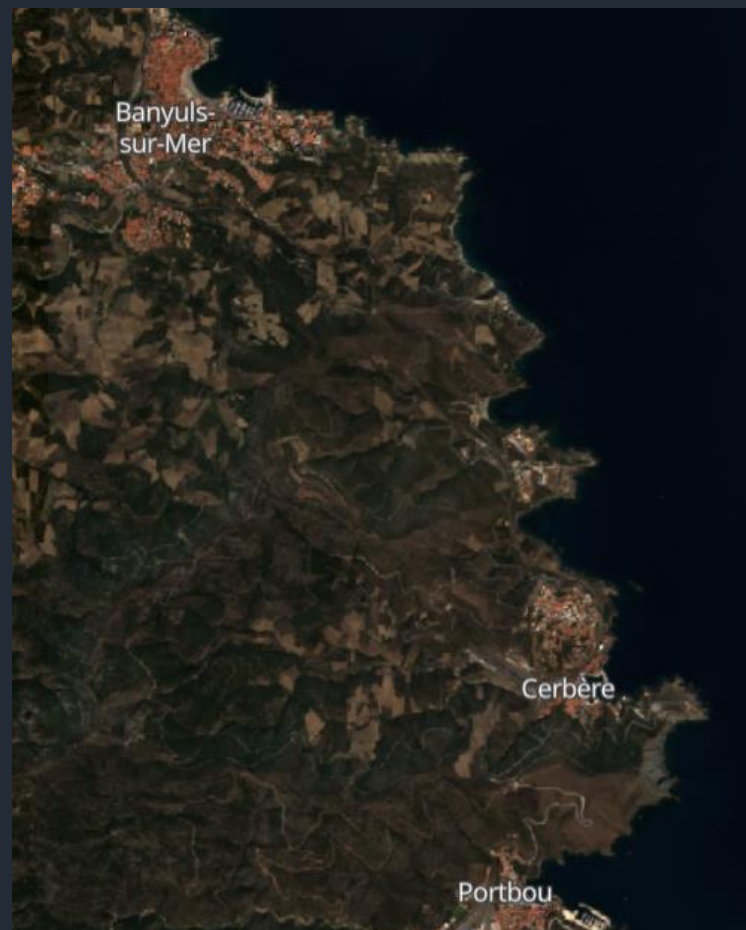
L'exemple du feu dans les Pyrénées orientales (1/5)

Surface brûlée par un feu de forêt le  
**16 avril 2023**



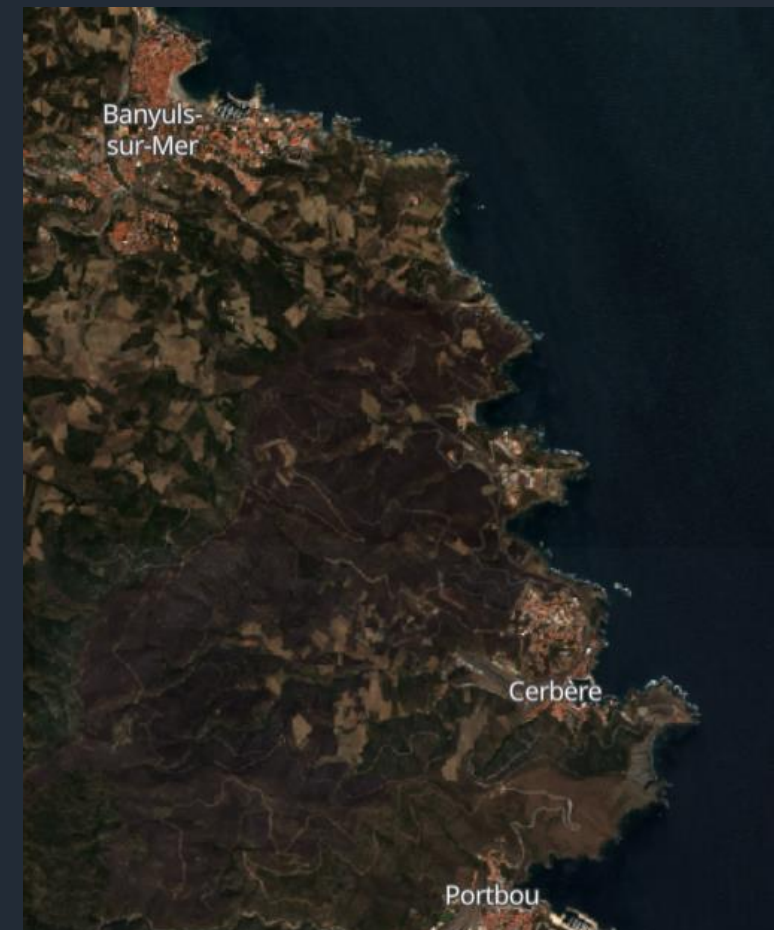
Source : EFFIS, Burned Area Locator

Image aérienne Sentinel 2 du  
**12 avril 2023**



Source : Sentinel Hub

Image aérienne Sentinel 2 du  
**17 avril 2023**

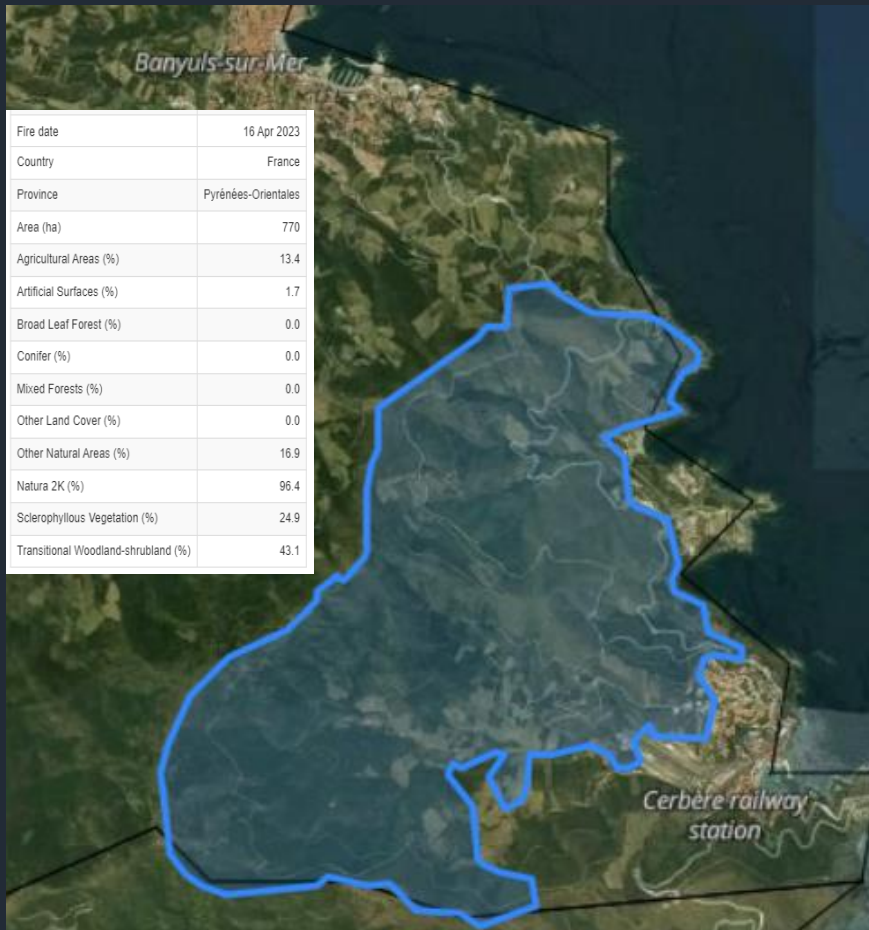


Source : Sentinel Hub

# La géomatique et les risques climatiques

L'exemple du feu dans les Pyrénées orientales (2/5)

Surface brûlée par un feu de forêt le  
**16 avril 2023**



Source : EFFIS, Burned Area Locator

Image aérienne Sentinel 2 du  
**12 avril 2023 - SWIR**



Source : Sentinel Hub

Image aérienne Sentinel 2 du  
**17 avril 2023 - SWIR**



Source : Sentinel Hub



# La géomatique et les risques climatiques

L'exemple du feu dans les Pyrénées orientales (3/5)

Image aérienne Sentinel 2 du  
17 avril 2023 - **SWIR**



Source : Sentinel Hub

Image aérienne Sentinel 2 du  
17 avril 2023 **segmentée**

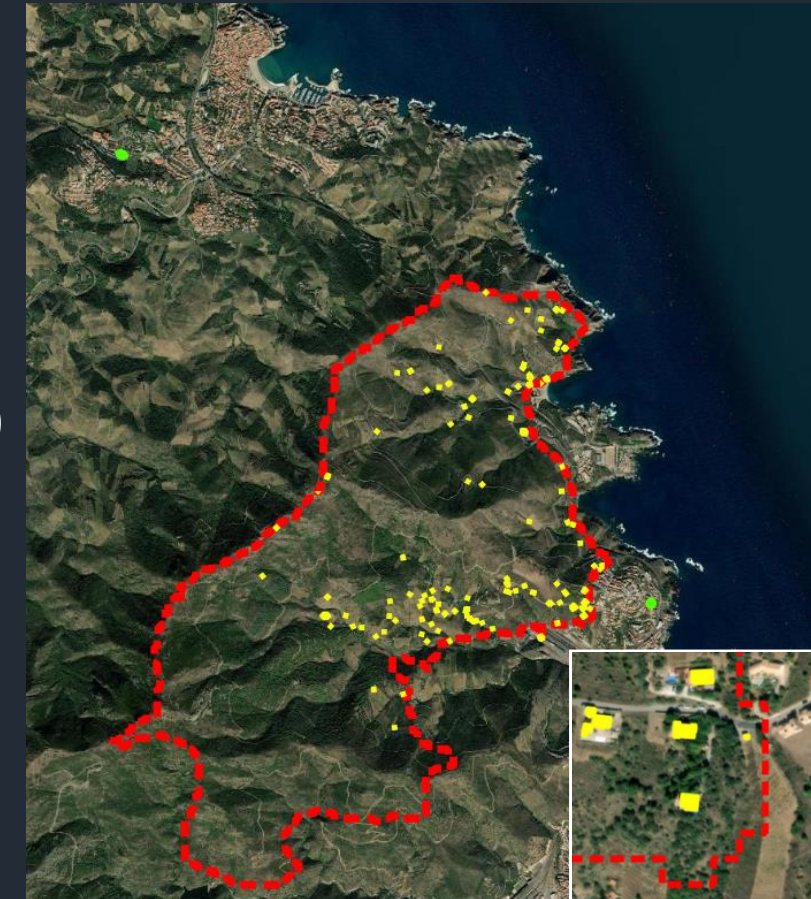


Segment  
Anything  
Model (SAM)



BD TOPO  
IGN

Cartographie de la **zone brûlée** et  
du **bâti** en zone brûlée



Zone brûlée Bâti en zone brûlée Caserne de pompiers

Sources : IGN, ESRI, Sentinel Hub

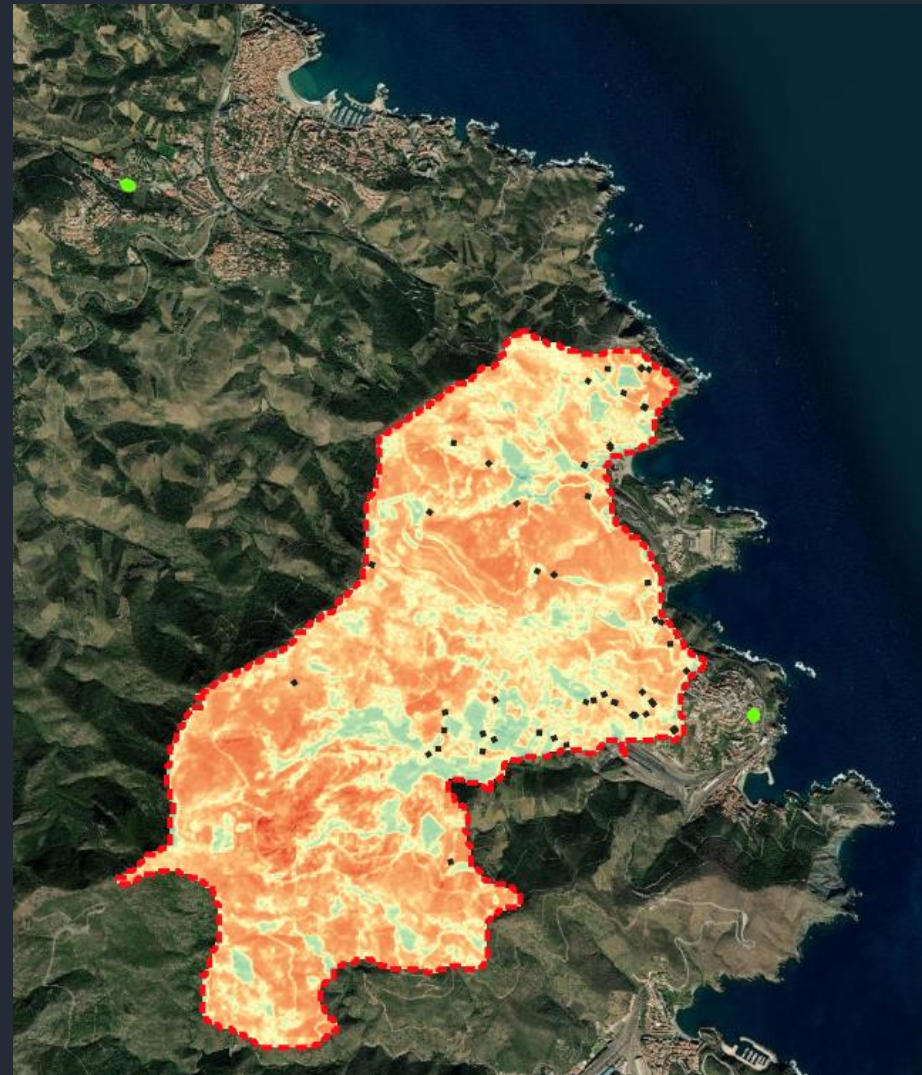


# La géomatique et les risques climatiques

L'exemple du feu de forêt (4/5)

Cartographie de la **sévérité du feu de forêt**

Zone	Nombre de « bâti »	Part
Sévérité critique	0	0%
Sévérité élevée	11	9%
Sévérité modérée	35	29%
Faible sévérité	43	36%
Non brûlé	32	26%



- Zone brûlée
- Sévérité critique
- Sévérité élevée
- Sévérité modérée
- Faible sévérité
- Non brûlé
- Caserne de pompiers
- Bâti en zone de sévérité moyenne - haute

Sources : IGN, ESRI, Sentinel Hub

# La géomatique et les risques climatiques

L'exemple du feu de forêt (5/5)



## Limites

- La **précision géographique** de ces données satellites l'exactitude des estimations.
- Ces images satellites sont **sensibles aux nuages**.
- La BD TOPO, produit de l'IGN, a aussi une **précision**, évaluée, limitée réduisant la qualité des calculs.

## Perspectives

- Des images satellites payantes **plus précises** et **plus fréquentes** existent.
- D'autres cas d'usages peuvent être développés avec ces ressources (**sécheresse**, **inondation** ...) afin de faire des **premières estimations rapides** et d'indemniser rapidement les assurés.
- **Créer des nouvelles variables géographiques** pour enrichir les modèles actuariels.

# La géomatique et les risques climatiques

L'ajout de nouvelles variables spatiales



## Enrichir les modèles avec les données spatiales

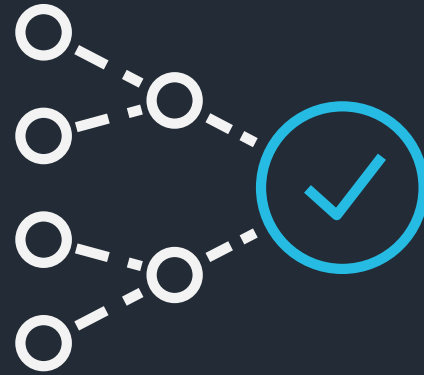
Avec la multiplicité des sources de données externes spatiales et des techniques pour les traiter, des nouvelles variables émergent pour les modèles actuariels.



## Zoniers climatiques

- Créer un zonier historique par risque climatique à partir de l'historique de sinistralité observé
- Créer un zonier prospectif par risque à l'aide des données calculées par le GIEC





# Données spatiales et Tarification

L'apport des données spatiales dans le processus tarifaire

# Discussion autour de l'intégration de variables géographiques externes

- **Périmètre :**
  - Modèles Fréquence / Coût / Charge attritionnels (catastrophes naturelles exclues)
  - Données de type géographique externe :
    - Densité de population, densité de magasins, de bars, revenu moyen...
    - Températures moyennes ou extrêmes sur le territoire
    - Distance à la rivière la plus proche, altitude...
- **Comment exploiter au mieux les données géographiques externes ?**
  - Compréhension des difficultés inhérentes à ce type de données pour les modèles statistiques.
  - Techniques pour adresser ces difficultés.

# Corrélations : fallacieuses ou pas ?

- **Corrélation :**

- Une relation ou une connexion entre deux ou plusieurs variables.
- La causalité est un concept toujours problématique mais c'est ce vers quoi l'on veut tendre.

- **Corrélation fallacieuse :**

- Une relation *apparente* entre variables qui ne découlerait pas d'une relation de cause à effet et qui ne serait pas susceptible de généralisation.

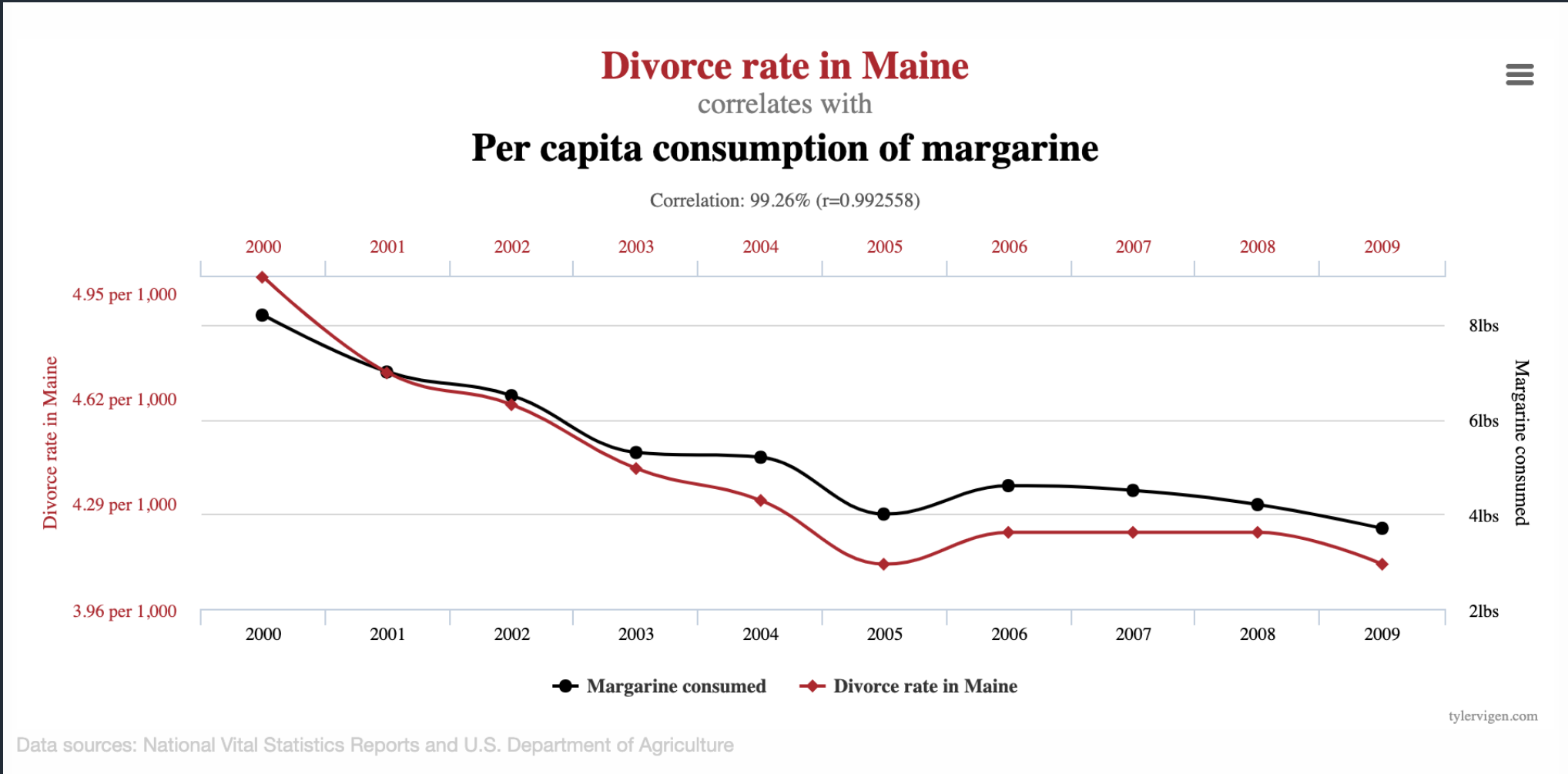
- **Pourquoi le problème est plus marqué sur les variables géographiques :**

- Autocorrélation spatiale : les variables géographiques externes sont spatialement autocorrélées.
- Ce phénomène :
  - rend la sélection de variables plus complexe.
  - peut dégrader la qualité prédictive du modèle.



# Corrélations : fallacieuses ou pas ?

Un lien de cause à effet ? Probablement pas...



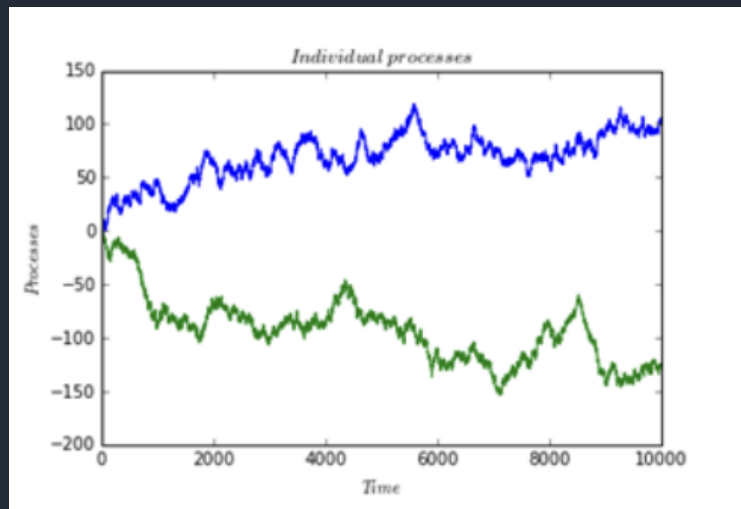
# Corrélations : fallacieuses ou pas ?



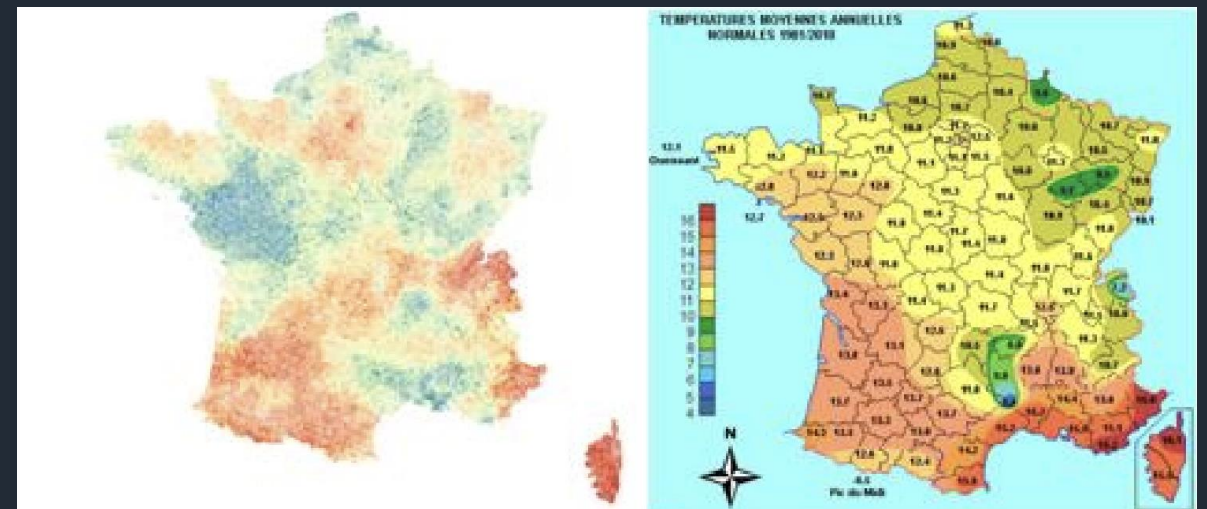
- On veut prédire les ventes de glaces et l'on a deux prédicteurs potentiels :
  - La température
  - Les ventes de lunettes de soleil
- Quelles variables seront prédictives ?
  - Les deux !
  - Mais la température seulement est intéressante...

# Autocorrélation spatiale

- L'équivalent spatial d'un problème courant avec les séries temporelles :
  - La quantité d'information portée par les données est "moindre" que le nombre de points disponibles dans le dataset.
  - Les variables spatiales sont autocorrélées, et corrélées entre elles.
  - Il faut pouvoir stationnariser le signal géographique de manière à pouvoir le traiter de façon efficace et prédictive



Séries temporelles générées aléatoirement mais corrélées



Fréquence vol MRH et températures moyennes en France

# Une exposition accrue au problème

## Big Data

De gros volumes de données externes peuvent désormais être collectés et testés dans le pricing



## Machine Learning

Les technologies actuelles permettent d'incorporer aisément ces volumes de données dans les modèles.



## Exposition accrue aux corrélations fallacieuses

Comment s'assurer que les tendances que l'on détecte dans nos données sont bien capturées et susceptibles de généralisation ?



# Des conséquences métier dommageables

Les corrélations fallacieuses font douter de la qualité du modèle.

Les corrélations fallacieuses peuvent prendre du signal aux variables significatives.

La généralisation aux nouvelles polices et nouvelles zones peut être mise en défaut.

Des corrélations fallacieuses peuvent mener à de mauvaises décisions métier.

# Exemple hypothétique en tarification auto



- Le modèle de base comprend les variables standard (caractéristique clients & véhicules etc.).

- Dans notre portefeuille hypothétique, température et densité de population sont très corrélées (exposition concentrée dans les villes du Sud).

- L'une et l'autre variables sont très prédictives du risque.

- La densité seule est légitimement prédictive, mais utiliser la température permet aussi de bien prédire du fait des corrélations...

- Les zones de haute température mais de faible expo sur le reste du territoire se voit appliquer une majoration.

- Antisélection sur les zones de faibles expos.

# Quelques techniques pour adresser le problème

## Lissage spatial

- Lissage spatial des résidus du modèle de base.
- Exploration des effets résiduels des variables externes.

Le signal purement géographique est capturé avant inclusion des variables externes.

## Modélisation hiérarchique

- Commencer par inclure les variables crédibles.
- Offset du modèle.
- Exploration des résidus du modèle.

Le signal est « forcé » sur des variables a priori crédible avant d'étendre aux autres variables externes.

## Modélisation transparente

- Utiliser des modèles type GAM/GLM permet d'identifier clairement l'impact des variables "fallacieuses".
- Le modélisateur peut choisir d'inclure ou d'exclure directement certains effets.

Certaines structures de modèles facilitent le traitement à dire d'expert.

# Exemple hypothétique en tarification auto avec lissage spatial

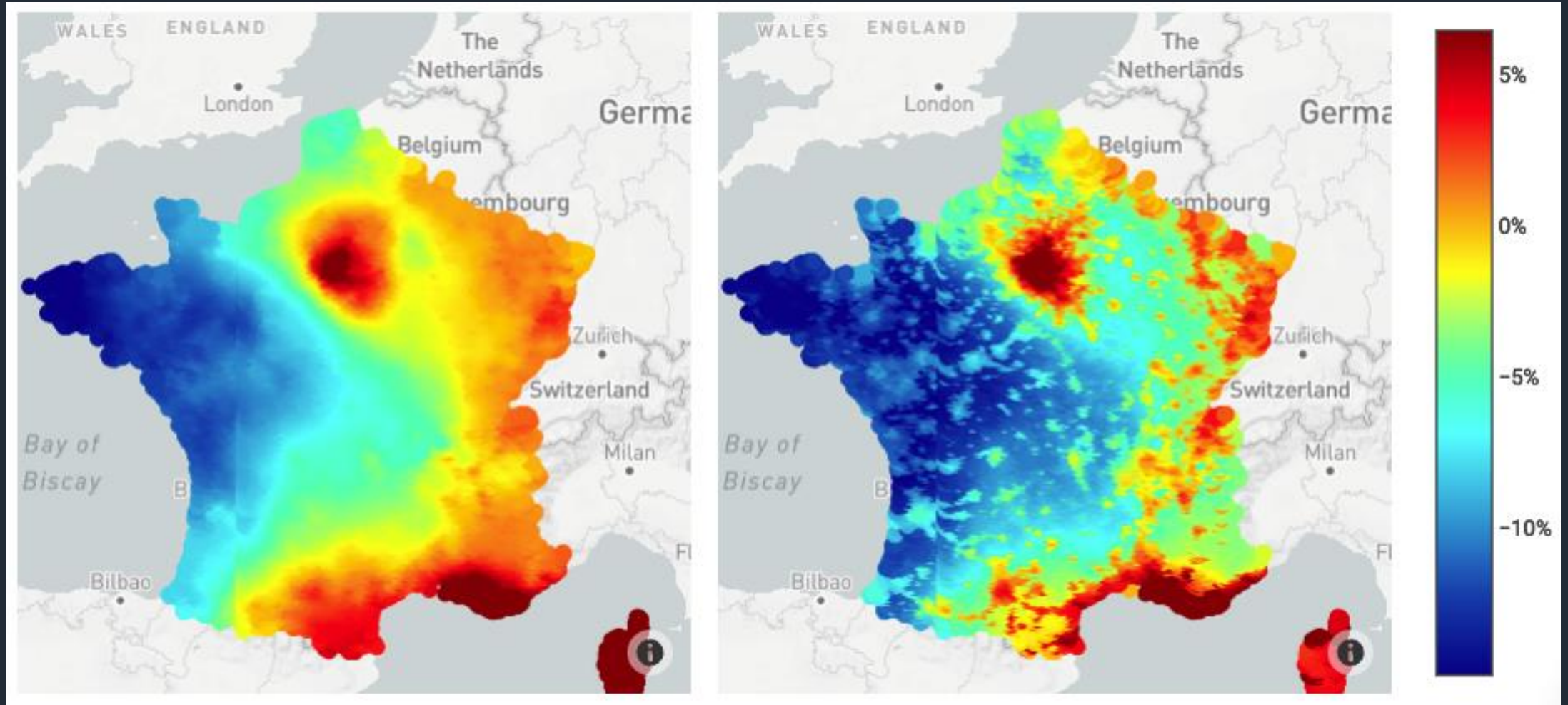


- Le modèle de base comprend les variables standard (caractéristique clients & véhicules etc.).
- Le lissage spatial capture le sur-risque dans le Sud directement sur la composante spatiale “pure”.
- Le signal est capturé et lissé en fonction de la localisation.
- La température n’a plus d’impact sur les résidus du modèle.
- Pas d’antisélection sur les zones chaudes dans le nord du territoire : plus de majorations liées à la température, les coefficients ont été déterminés uniquement par l’expérience effective.



# Une illustration de lissage spatial

Lissage + intégration des variables externes



# Le zonier : un sujet sensible

Le modèle statistique n'est pas tout...

En présence d'un réseau de distribution physique, les problématiques liées au zonier sont notoirement épineuses :

- **Le lissage spatial (structurellement) moyennise différentes zones, par exemple :**
  - Localement il n'est pas toujours évident de comprendre et justifier les variations des zoniers, par exemple des variations importantes d'une rue à l'autre.
  - Sur des échelles plus larges, tout le périmètre d'un agent peut se voir majorer du fait d'une zone attenante plus risquée.

Expliquer ces phénomènes n'est pas évident :

- **La calibration statistique est faite à des échelles supérieures au périmètre d'un seul agent.**
- **L'utilisation de variables externes peut compliquer la tâche lorsqu'il n'apparaît pas de justification causale intuitive.**



**Eve TITON**

Senior Consultante  
**Milliman France, Paris**

[eveelisabeth.titon@milliman.com](mailto:eveelisabeth.titon@milliman.com)



**Antoine RAINAUD**

Analyste risques climatiques & Consultant  
**Milliman France, Paris**

[antoine.rainaud@milliman.com](mailto:antoine.rainaud@milliman.com)



**Bruno BECHA**

Responsable Actuarial Data Science  
**Akur8**

[bruno.becha@akur8.com](mailto:bruno.becha@akur8.com)

The Milliman logo consists of a white square icon with a stylized 'M' shape inside, followed by the word 'Milliman' in a white sans-serif font.The AKUR8 logo features a stylized 'A' icon composed of two overlapping shapes, one light blue and one dark blue, positioned above the text 'AKUR8' in a bold, black, sans-serif font.