Contributions des données de l'assurance à l'étude des risques naturels en France

Application de méthodes d'apprentissage statistique pour l'évaluation de la nature et du coût des dommages assurés liés aux événements naturels en France

Antoine Heranval

sous la direction d'Olivier Lopez et de Maud Thomas

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Préambule

- Contexte **d'événements climatiques extrêmes** de plus en plus commun.
- Thèse CIFRE à la **Mission Risques Naturels**, un groupement technique de **France Assureurs**.

Utilisation des méthodes d'apprentissage statistique dans un **contexte actuariel et d'analyse de données textuelles.**

- Chaque application répond à un besoin et à des contraintes industrielles.

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

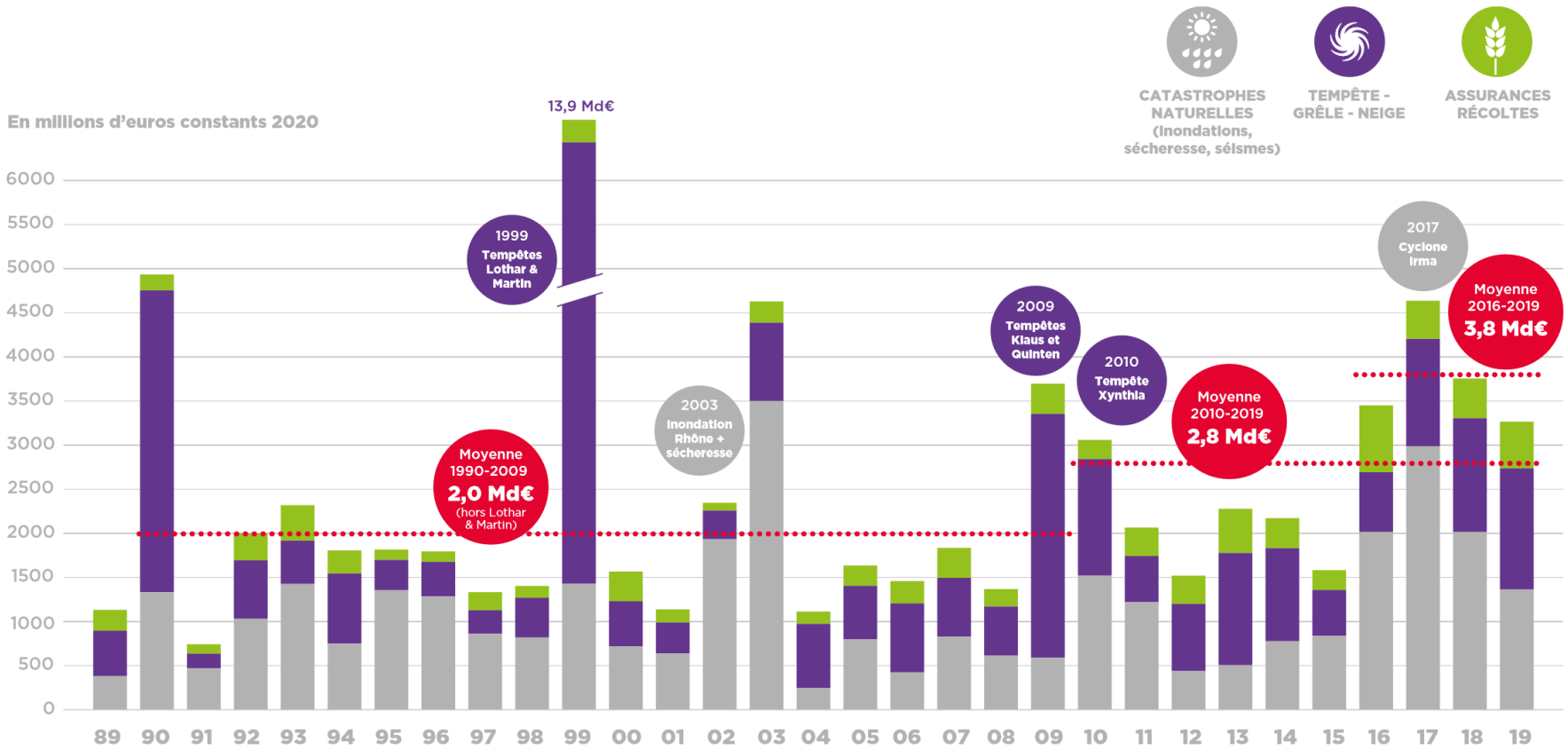
Sommaire

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Introduction

Impact assurantiel des risques naturels en France



MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Mission des sociétés d'assurances pour la connaissance et la prévention des Risques Naturels

FRANCE ASSUREURS
FAIRE AVANCER LA SOCIÉTÉ EN CONFIANCE

MISSION RISQUES NATURELS

Une centaine des sociétés adhérentes à FFA opérant en branche dommages participent à son budget

12 sociétés actuellement représentées au CA MRN

Allianz
AXA
MACIF
PACIFICA
MAAF GMF
Groupama
MAIF
GENERALI GROUP
SMA
GROUPE BPCE
Assurances Crédit Mutuel
SMACL ASSURANCES GROUPE VYV

MISSION RISQUES NATURELS
SORBONNE UNIVERSITÉ CRÉATEURS DE FUTURS DEPUIS 1257

Base de données événements inondations

- Un événement est défini par **une date de début, une date de fin et un ensemble de communes impactées.**

Nous avons regroupé près de 140 000 arrêtés CatNat inondation en plus de 4 300 événements distincts entre 1982 et 2021.

- **99% des communes sont touchés** par au moins un événement inondation, soit **225 000 lignes.**
- Les **10 événements de plus grande ampleur représentent 35%** de la base.

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

La base de données SILECC

- La MRN récolte les sinistres de **tous les périls CatNat et climatique auprès de 12 grandes compagnies d'assurance françaises, soit 70 % du marché.**
- Le coût des sinistres est **actualisé, selon l'indice de la Fédération Française du Bâtiment (FFB).**
- La base renseigne pour chaque sinistre, la date de survenance, le péril, le segment de risque, la localisation et le coût.

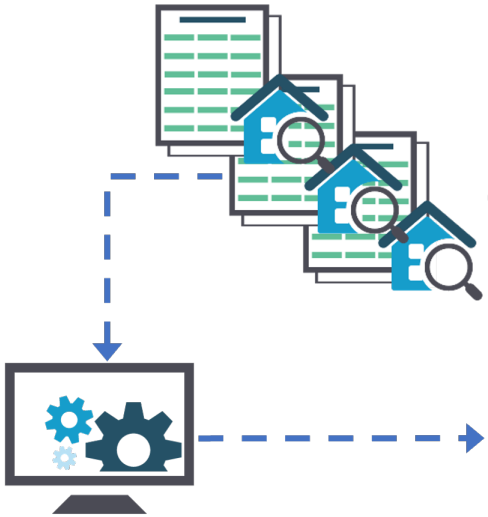
MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Analyse de la sinistralité à l'échelle fine du bâti

Analyse de la sinistralité à l'échelle fine du bâti

La base de données SILEX

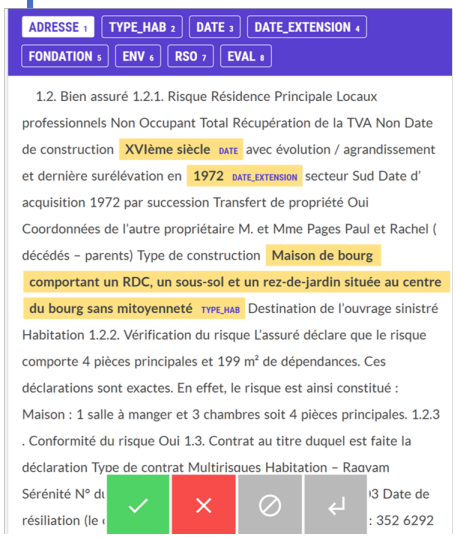


Bases existantes des données d'expertise

→ Application aux données issues des SI des réseaux d'experts de méthodes de classification textuelle

BD « SILEX » : SInistres Liés aux événements EXpertisés catnat et climatiques

→ Une base de données constituées avec les données des réseaux d'expertise contributeurs



Base de données constituée à partir de rapports PDF

→ Application de l'état de l'art en traitement de traitement automatique du langage naturel pour extraire les informations pertinentes dans les rapports d'expertise PDF

Données utilisées

Commentaire final
Abri piscine Piscine
les dommages concernent Embellissements : et Immobilier :
Tempête Couverture
Tabliers de porte-fenêtres doubles Remplacement de tabliers de volets roulants et de brises-soleil Volet roulants alu et brises-soleil Façade Ouest et Sud-Ouest
Serre de jardin 240 x 240 CONTENU Dommages suites aux chocs des grelonssur: -Toiture en polycarbonate de la verandaservant de serre aux végétaux.Verandad'environ 12m^2 au sol installée en 1996. -Anémometrede fermeture d'urgence du storre. -Serre exterieurede jardin 240x240.
TOITURE DEPENDANCES TOITURE
VEGETAUX : SELON FACTURE L'ARBRE ET LA CIME
Tempetesur brise vue (installation extérieure mobilière) mur de cloture
Revêtements intérieurs touchés par les infiltrations Cuisine

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Catégories recherchées

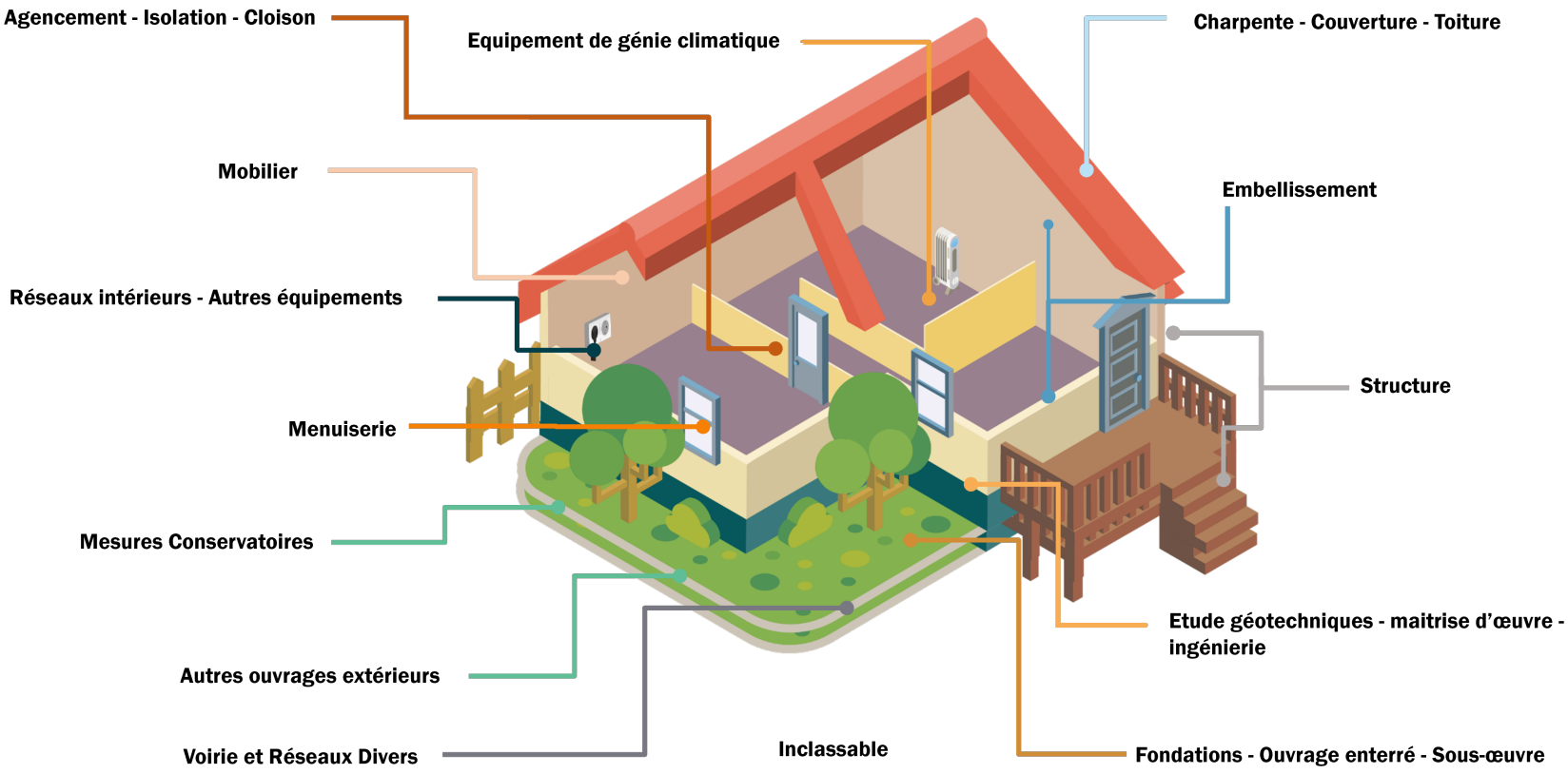


Figure: 15 composantes principales et 75 composantes secondaires.

Méthode d'analyse

- Après avoir expérimenté plusieurs méthodes nous nous sommes tournés vers **les réseaux de neurones avec une couche de contextualisation.**
- On utilise deux architectures classiques en analyse textuelle, **les réseaux de convolution et les réseaux récurrents**, en particulier les Long Short-Term Memory (LSTM)

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Méthode d'analyse

Nous avons utilisé un réseau de neurones convolutionnel (CNN).

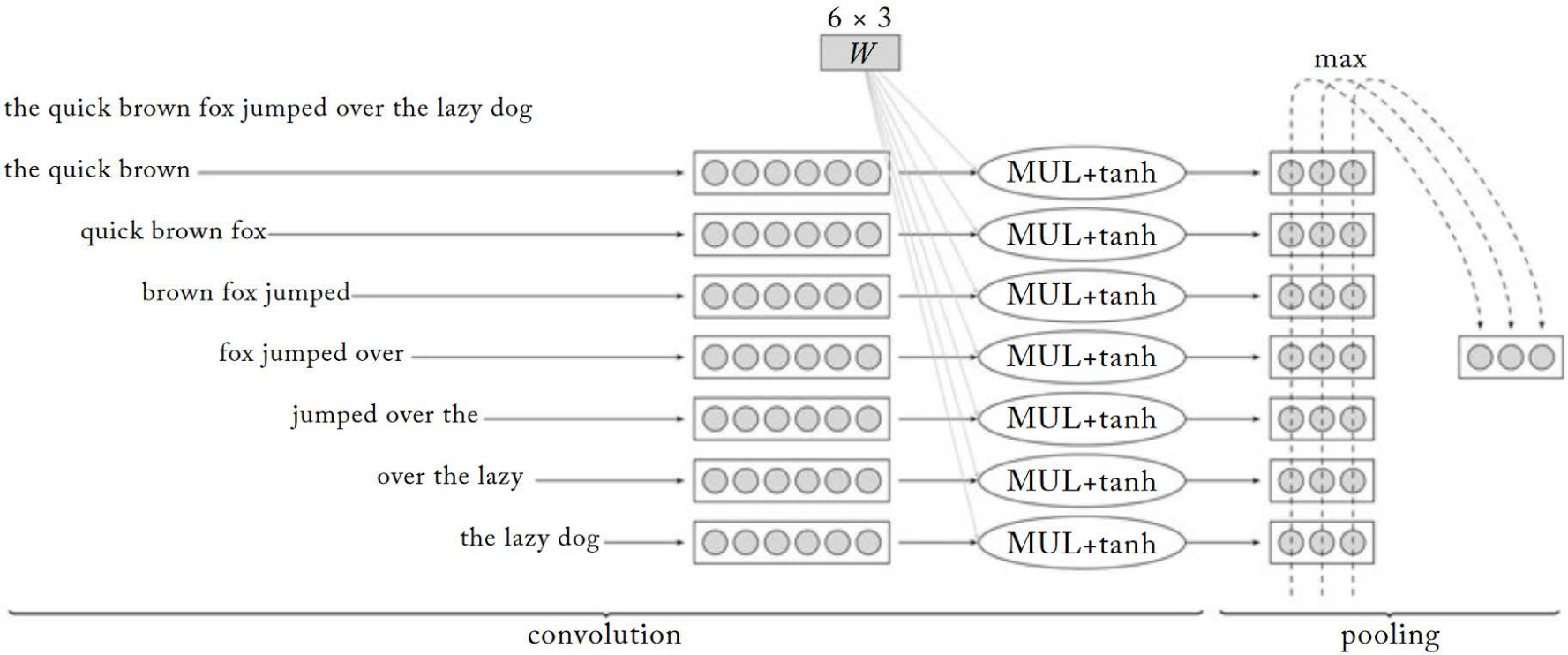


Figure: Illustration d'un CNN, (Source : (Goldberg 2017))

Méthode d'analyse

Nous avons aussi utilisé un « Long short-term memory » (LSTM).

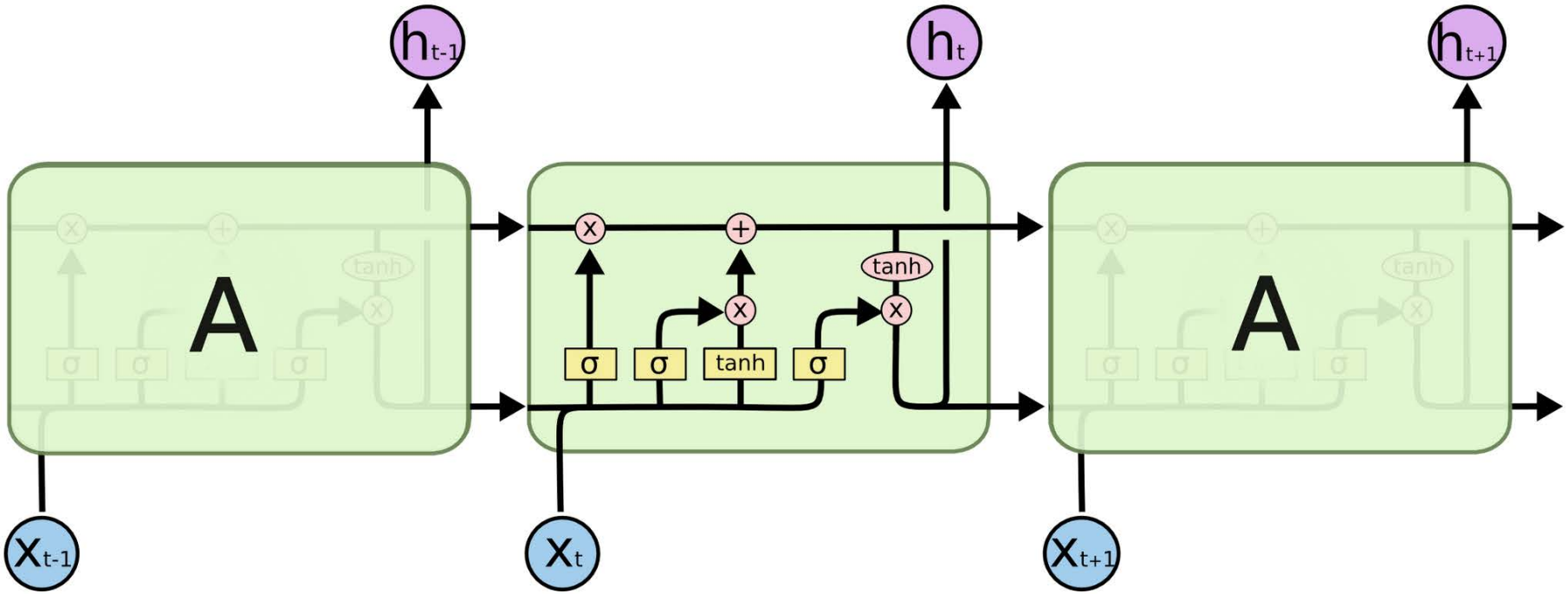
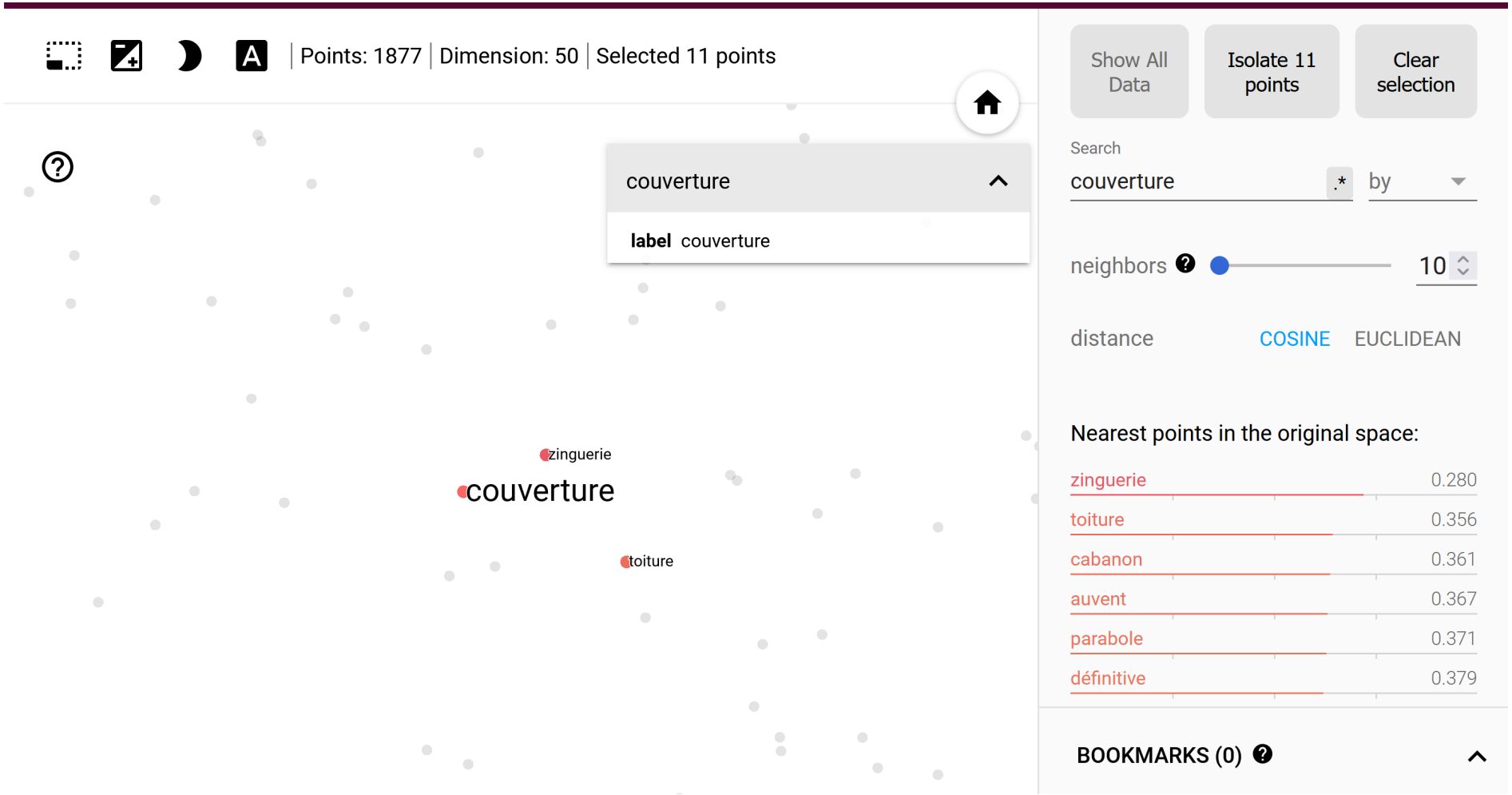


Figure: Illustration d'un LTSM, (Source : https://colah.github.io/posts/2015-08-Understanding-LSTMs/)

Illustration de l'embedding via TensorBoard

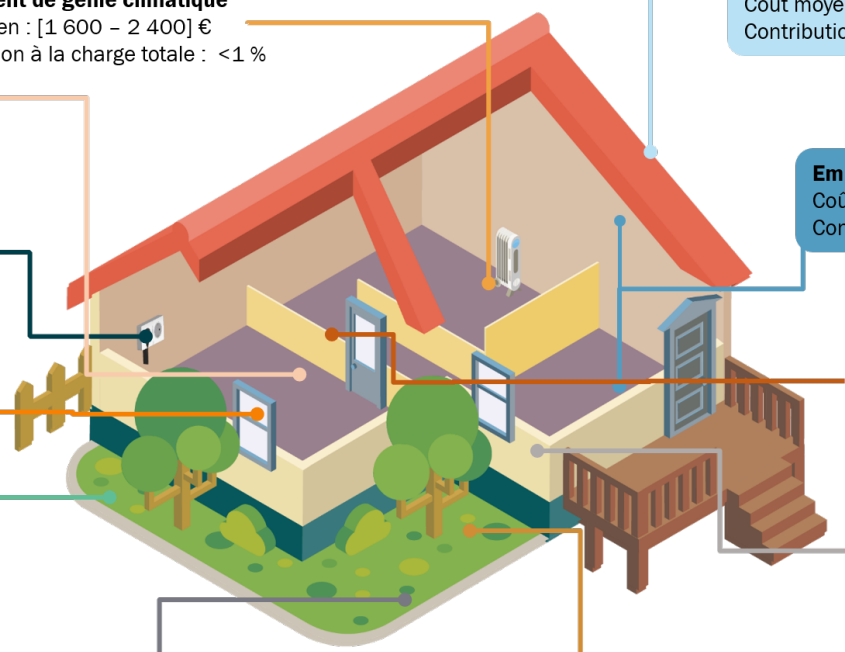


Résultats

- On a analysé les données de deux réseaux, **de 2014 à 2019 pour la grêle et pour la tempête.**
- Base d'apprentissage réutilisable.
- Les modèles pré-codés dans Keras sont **adaptés à nos données.**

On obtient des bons scores avec une **précision de 0.75**

Résultats : Tempête



Equipement de génie climatique
Coût moyen : [1 600 – 2 400] €
Contribution à la charge totale : <1 %

Charpente - Couverture – Toiture
Coût moyen : [4 400 – 4 800] €
Contribution à la charge totale : 37 %

Mobilier
Coût moyen [1 600 – 3 400] €
Contribution à la charge totale : <1 %

Embellissement
Coût moyen : [2 200 – 2 600] €
Contribution à la charge totale : 12 %

Réseaux intérieurs - Autres équipements
Coût moyen : [1 000 – 1 600] €
Contribution à la charge totale : < 1 %

Agencement - Isolation – Cloison
Coût moyen : [1 300 – 2 300] €
Contribution à la charge totale : 1 %

Menuiserie
Coût moyen : [2 600 – 2 900] €
Contribution à la charge totale : 6 %

Mesures Conservatoires
Coût moyen : [2 200 – 3 100] €
Contribution à la charge totale : 3 %

Structure
Coût moyen : [2 600– 3 100] €
Contribution à la charge totale : 4%

Voirie et Réseaux Divers
Coût moyen : [3 500 – 6 000] €
Contribution à la charge totale : 2 %

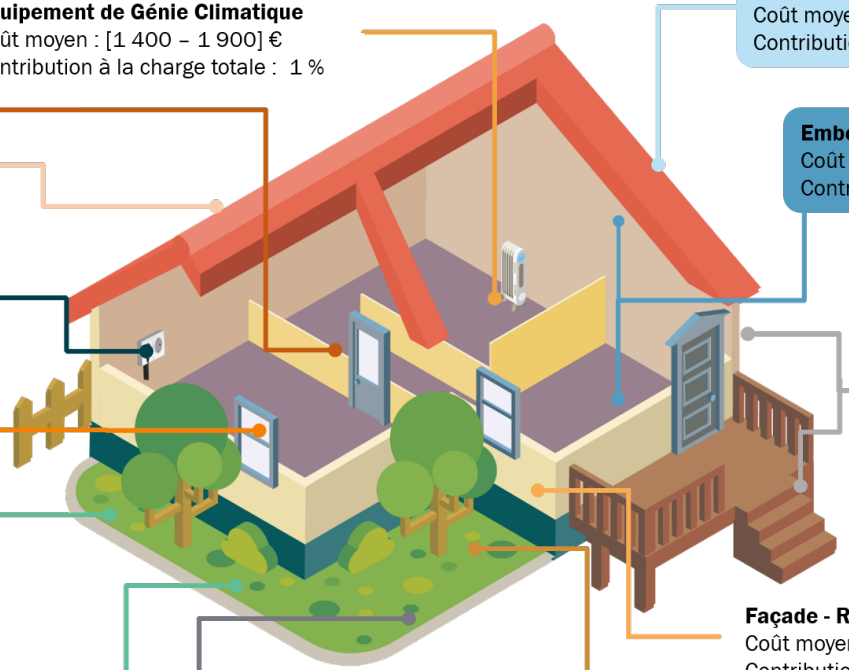
Autres ouvrages extérieurs
Coût moyen : [3 800 – 4 200] €
Contribution à la charge totale : 27 %

Inclassable
Coût moyen : [2 800 – 3 300] €
Contribution à la charge totale : 6 %

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Résultats : Grêle



Cloisons isolation agencement
Coût moyen : [2 000 – 2 700] €
Contribution à la charge totale : 1 %

Equipement de Génie Climatique
Coût moyen : [1 400 – 1 900] €
Contribution à la charge totale : 1 %

Charpente – Couverture
Coût moyen : [4 800 – 5 300] €
Contribution à la charge totale : 37 %

Désamiantage
Coût moyen [8 500 – 13 000] €
Contribution à la charge totale : 2 %

Embellissement – Finition
Coût moyen : [2 000 – 2 300] €
Contribution à la charge totale : 8 %

Réseaux intérieurs - Autres Equipements
Coût moyen : [1 000 – 1 500] €
Contribution à la charge totale : < 1 %

Menuiserie
Coût moyen : [2 700 – 3 000] €
Contribution à la charge totale : 12 %

Structure
Coût moyen : [900– 3 800] €
Contribution à la charge totale < 1%

Mesure conservatoire
Coût moyen : [1 800 – 2 300] €
Contribution à la charge totale : 3 %

Façade - Revêtement Extérieur
Coût moyen : [3 300 – 3 800] €
Contribution à la charge totale : 5 %

Viabilité - Réseau Extérieur – Jardin – Piscine
Coût moyen : [2 800 – 3 200] €
Contribution à la charge totale : 6 %

Véranda
Coût moyen : [3 000 – 3 400] €
Contribution à la charge totale : 3 %

Inclassable
Contribution à la charge totale : 16 %

Abris divers
Coût moyen : [2 800 – 3 400] €
Contribution à la charge totale : 4 %

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Estimation du coût d'un épisode de sécheresse

Régime CatNat

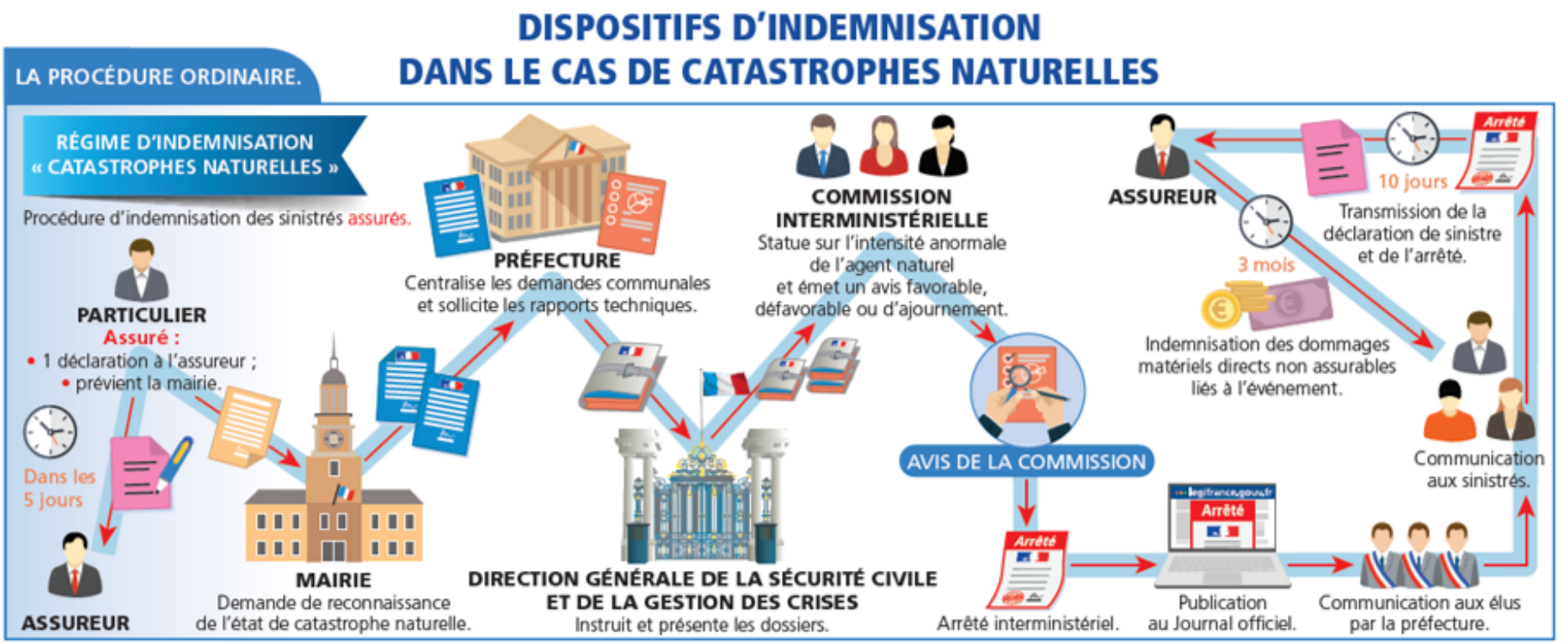
- Il repose sur une **extension de garantie obligatoire** impliquant la solidarité entre les Français.
- **Partenariat Public Privé**, les assureurs se chargent de la gestion des sinistres et l'État régule les caractéristiques clés du contrat.
- L'état apporte sa **garantie illimitée** au régime CatNat par l'intermédiaire de la **Caisse Centrale de Réassurance (CCR)**.
- Il couvre les **inondations, la sécheresse, les mouvements de terrain, les secousses sismiques, les cyclones, les avalanches...**
- C'est un régime qui permet un bon niveau d'assurance en France malgré un caractère déresponsabilisant et un processus de gestion complexe.

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Régime CatNat

Pour recevoir une indemnisation, un **arrêté reconnaissant la commune en état de catastrophe naturelle doit être publié au Journal Officiel**.



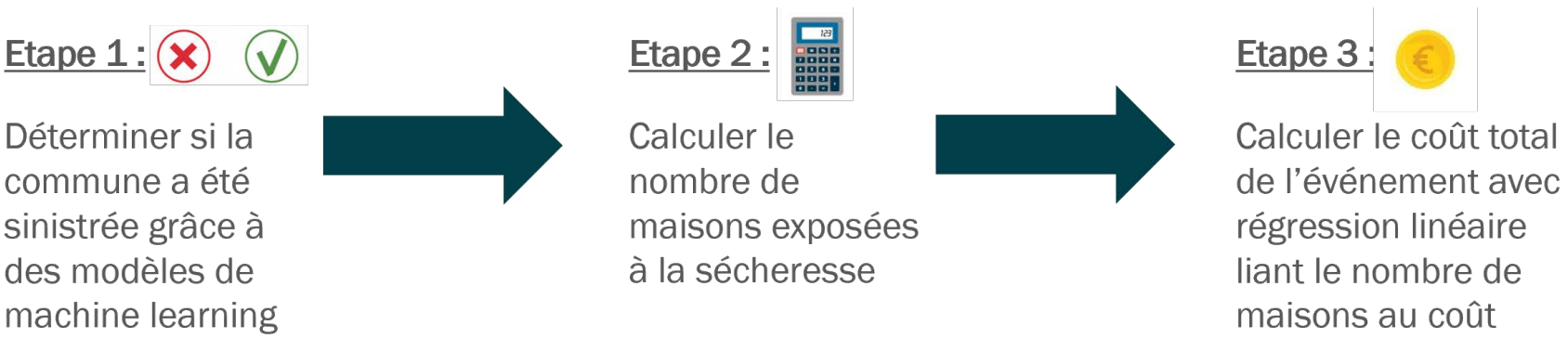
Sécheresse

- Le risque sécheresse, lié au retrait et au gonflement de l'argile (RGA), est responsable d'environ **30% du montant total du régime CatNat.**
- Les dommages liés à ce risque représentent près de **14 milliards d'euros sur la période 1989-2019.**
- Ces montants sont dus à une exposition importante du territoire : **10,5 millions de maisons en zone susceptibilité moyenne ou forte.**
- Les sinistres ont un coût moyen très élevés avec **16 300€.**
- Sur les neuf dernières années **50% des demandes de reconnaissance CatNat n'ont pas été acceptées**

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Méthode générale pour l'estimation du coût pour tout le marché



Deux travaux académiques récents traitent aussi de l'évaluation de la sécheresse en France : Ecoto, Bibaut, and Chambaz 2021; Charpentier, James, and Ali 2021.

La base de données SILECC RGA

- On agrège **les sinistres de la BD SILECC** à la commune et à l'année.
- La période va de **2003 à 2018.**

La base de données est **très déséquilibrée**, 6% des communes sont concernées par un sinistre.

Variables

- Les données doivent être disponibles rapidement après un événement.
- Un indicateur résultant de la cartographie publiée par le **BRGM**.
- Un indice météorologique spatio-temporel, produit par Météo-France, **l'indice standardisé d'humidité des sols (SSWI).**

Dans l'ensemble, notre base de données contient **154 variables et 522 600 observations.**

Catégorie de données	Nombre
Variables relatives au SSWI	96
Description des événements de sécheresse	37
Critère utilisé par la commission	4
Susceptibilité au retrait et au gonflement de l'argile	11
Population dans la commune	1
Déclarations de catastrophe naturelle passées	4

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ CRÉATEURS DE FUTURS DEPUIS 1257

Un problème de classification binaire

- $Y \in \{0,1\}$ la variable réponse,

$$Y_{ij} = \begin{cases} 1, & \text{si un sinistre dans la commune i, l'année j} \\ 0, & \text{sinon} \end{cases}$$

- $X \in \mathbb{R}^p$ les variables

Notre objectif est d'estimer $p(X) = \mathbb{P}[Y_{ij} = 1 \mid X]$.

On utilise :

- Le **modèle linéaire généralisé (GLM)** ;
- Les **Forets aléatoires (RF)** ;
- **Extreme Gradient Boosting (XGBOOST)** ;
- Nous considérons également l'**agrégation** selon la moyenne des probabilités obtenues de ces trois modèles

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Modèle linéaire généralisé pénalisé

On suppose ici que $Y|X \sim \mathscr{B}(p(x))$

- $g(p(X)) = X\beta$
- $\beta \in \mathbb{R}^p$ est le vecteur des paramètres inconnus
- g une certaine fonction, $g(y) = \text{logit}(p(X)) = \log(p(X)/(1-p(X)))$.

Soit $f_\beta(y,x)$ **la vraisemblance du modèle. L'estimateur GLMNET est défini comme** :

$$\widehat{\beta} = \arg\max_{\beta} \sum_{i=1}^{n} \log(f_\beta(Y_i, X_i)) - \lambda\{\alpha\|\beta\|_1 + (1-\alpha)\|\beta\|_2\},$$

avec λ qui est trouvé par cross validation.

Forets aléatoires

Les forêts aléatoires (Random Forets, RF) reposent **sur l'agrégation d'arbres de régression (Breiman 2001).** On estime $p(x)$ par

$$\widehat{p}(x) = \sum_{j=1}^{K} p_j R_j(x),$$

où, pour tout x, $R_j(x) = 0$ pour tout j sauf un.
L'estimation des valeurs p_j est faite pour chaque région de l'espace $R_j(x)$.

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Extreme Gradient Boosting

- L'Extreme Gradient Boosting (XGBOOST) est une méthode alternative aux RF qui repose sur **le boosting et non plus le bagging.**
- $\widehat{p}^{(t)}(x) \Leftarrow \widehat{p}^{(t-1)}(x) + \pi_t(x)$
- $\pi_t(x)$ est un **arbre de régression sélectionné de manière à faire diminuer la fonction de perte autant que possible.**
- À l'étape t, l'algorithme tente de trouver π_t qui minimise

$$\sum_{i=1}^{n} \partial_2 \ell(y_i, \widehat{p}^{(t-1)}(x_i)) \times \pi_t(x_i) + \frac{1}{2} \partial_2^2 \ell(y_i, \widehat{p}^{(t-1)}(x_i)) \times \pi_t^2(x_i) + pen(\pi_t).$$

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Évaluations des résultats pour des données déséquilibrées

Avec p_c, le seuil de discrimination,

$$\text{Precision}(p_c) = \frac{\text{vrais positifs}}{\text{vrais positifs} + \text{faux positifs}}$$

et

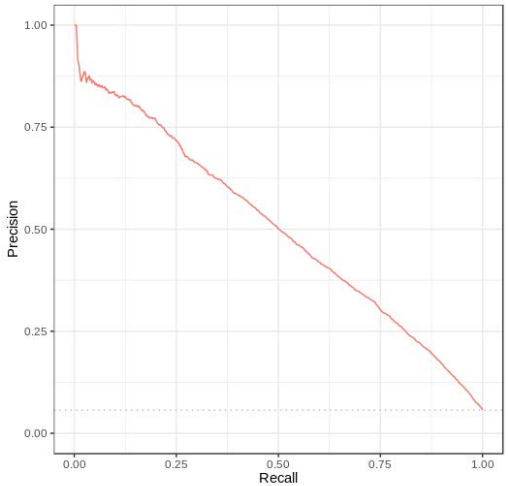
$$\text{Recall}(p_c) = \frac{\text{vrais positifs}}{\text{vrais positifs} + \text{faux positifs}},$$

Le F_1–score combine ces deux mesures:

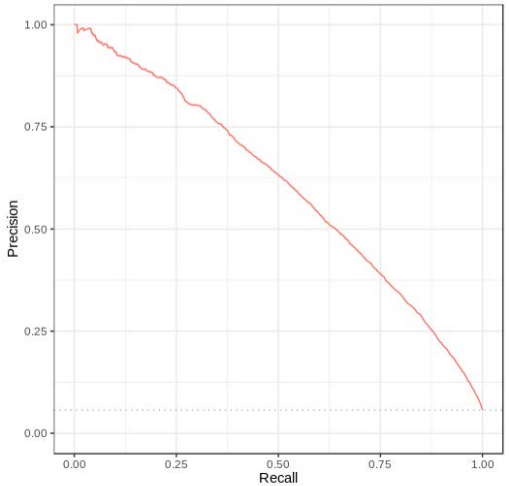
$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

La courbe PRC affiche les valeurs de Précision et de Recall lorsque p_c varie de 0 à 1.

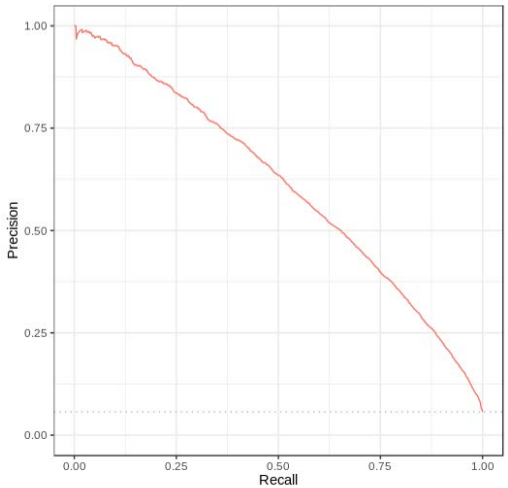
Résultats: Courbes PRC



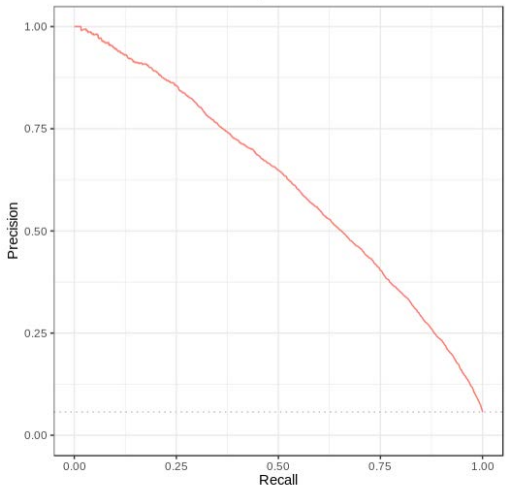
GLMNET



RF



XGBOOST



AGGREGATE

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Résultats

Modèle	AUC ROC	AUC PRC
GLMNET	0.907	0.503
RF	0.933	0.604
XGBOOST	**0.936**	**0.609**
AGGREGATE	**0.936**	**0.615**

On fait varier p_c pour maximiser le F1-Score.

Modèle	F1-score	Seuil
GLMNET	0.503	0.221
RF	0.570	0.306
XGBOOST	**0.573**	**0.291**
AGGREGATE	**0.576**	**0.264**

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Mesure des contributions des variables

Regarder quelles sont les variables qui contribuent le plus à la prédiction permet de mieux interpréter les modèles « boîte noire » et leurs performances.

- Pour GLMNET, on utilise **la valeur du coefficient associé à la variable**, après normalisation des données.
- Pour RF, l'importance de la variable est mesurée grâce à **l'indice de Gini.**
- Pour XGBOOST, on regarde **la contribution fractionnelle de chaque caractéristique au modèle**, sur la base du gain total des splits de la caractéristique.

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Variables clés selon les modèles

	GLMNET	RF	XGBOOST
1	Valeur maximale de l'indice SSWI 12 pour février	Nombre de déclarations passées de catastrophe naturelle	Nombre de déclarations passées de catastrophe naturelle
2	Valeur maximale de l'indice SSWI 12 pour le mois d'août	Surface sans susceptibilité au retrait gonflement des argiles	Nombre d'événements pour l'année précédente
3	Valeur maximale de l'indice SSWI 6 pour novembre	Nombre de maisons	Nombre de maisons
4	Valeur maximale de l'indice SSWI 12 pour le mois de juin	Proportion de la surface ayant une faible susceptibilité au retrait gonflement des argiles	Surface ayant une faible susceptibilité au retrait gonflement des argiles
5	Valeur maximale de l'indice SSWI 3 pour le mois d'août	Surface en zone urbaine	Valeur minimale de l'indice SSWI 1 pour le mois d'août
6	Classement de la gravité des événements	Surface avec susceptibilité moyenne au retrait gonflement des argiles	Valeur minimale du SSWI 3 pour le mois d'octobre
7	Valeur maximale du SSWI 12 pour le mois de juin	Nombre de maisons ayant une susceptibilité moyenne au retrait gonflement des argiles	Nombre de refus passés avec le calcul effectué avec notre SSWI
8	Valeur minimale de l'indice SSWI 12 pour le mois de juin	Durée totale des épisodes de sécheresse	Surface avec une propension moyenne au retrait et au gonflement de l'argile
9	Valeur maximale de l'indice SSWI 12 pour janvier	Nombre d'événements pour l'année précédente	Nombre de maisons ayant une susceptibilité moyenne au retrait gonflement des argiles
10	Valeur maximale de l'indice SSWI 6 pour le mois de janvier	Valeur minimale de l'indice SSWI 6 pour le mois de novembre	Durée totale des épisodes de sécheresse

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Résultats des prédictions pour 2018

Selon France Assureurs, le **coût de la sécheresse en France pour 2018 est de 900 millions d'euros**, et elle pourrait atteindre 1 200 millions. **En 2022 son estimation est de 1 300 millions**.

Modèle	**Estimation**	Borne inf	Borne sup
GLMNET	**579 350 811**	461 125 885	697 575 737
RF	**1 618 225 685**	1 396 432 680	1 840 018 69
XGBOOST	**977 086 655**	839 262 189	1 114 911 122
AGGREGATE	**965 750 651**	796 820 728	1 134 680 547

Conclusion et discussion

- Nous avons développé une méthodologie permettant **d'estimer le coût de la sécheresse**, avec des résultats encourageants bien que plusieurs incertitudes subsistent.
- Une difficulté vient du processus des **arrêtés de catastrophes naturelles.**
- Avec les variables météorologiques et géologiques dont nous disposions, nous n'avons abordé **qu'une partie des facteurs à l'origine du risque.**
- Les techniques que nous avons utilisées pourraient être améliorées avec des connaissances supplémentaires sur les **phénomènes de dépendance spatiale entre les villes.**
- Plus d'informations peuvent être trouvées dans [Heranval, Lopez, and Thomas, *European Actuarial Journal*, 2022]

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Estimation du coût des inondations

Contexte

- On cherche à estimer **le coût d'un événement inondation rapidement après son occurrence, en particulier, les événements extrêmes.**
- Les arbres de régression, introduits par (Breiman et al. 1984), font partie des **outils simples et interprétables largement utilisés dans le secteur de l'assurance.**
- Le but est de constituer des **classes d'observations** qui ont un comportement similaire relativement à une variable réponse Y.
- On peut faire varier l'objectif et dans notre cas **se concentrer sur les valeurs extrêmes.**

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Classification And Regression Trees (CART)

Regression tree (Breiman et al., 1984)

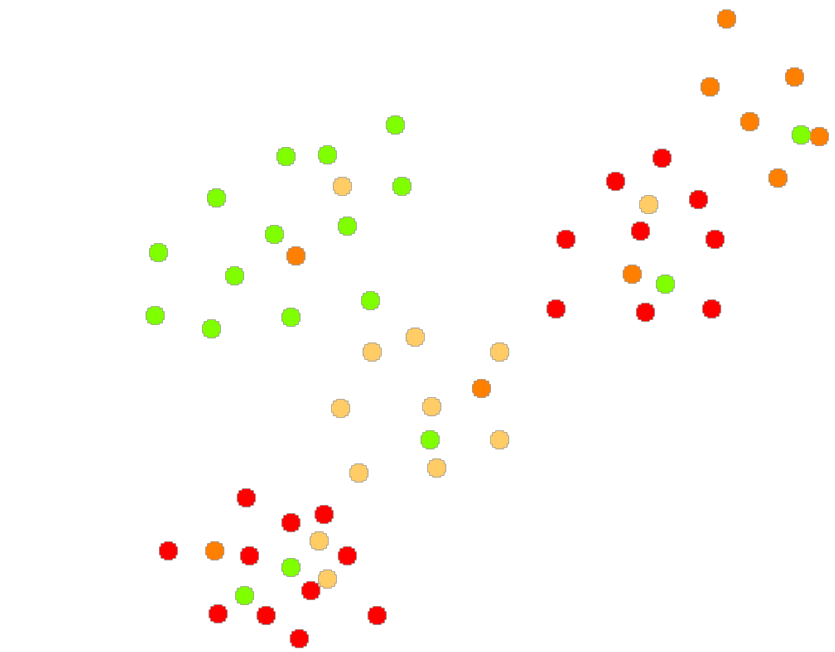
$$\theta^*(X) = \arg\min_{\theta \in \mathscr{F}} \mathbb{E}[\varphi(Z, \theta(X)],$$

- Z est la variable à prédire, le coût d'un événement dans notre cas
- $X \in \mathscr{X} \subset \mathbb{R}^d$ est un ensemble de variables explicatives
- $\mathscr{F}$ est une classe de fonctions cibles sur $\mathbb{R}^d$
- φ est une fonction de perte qui dépend de la quantité que l'on souhaite estimer

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Construction de l'arbre



CART : Step 0

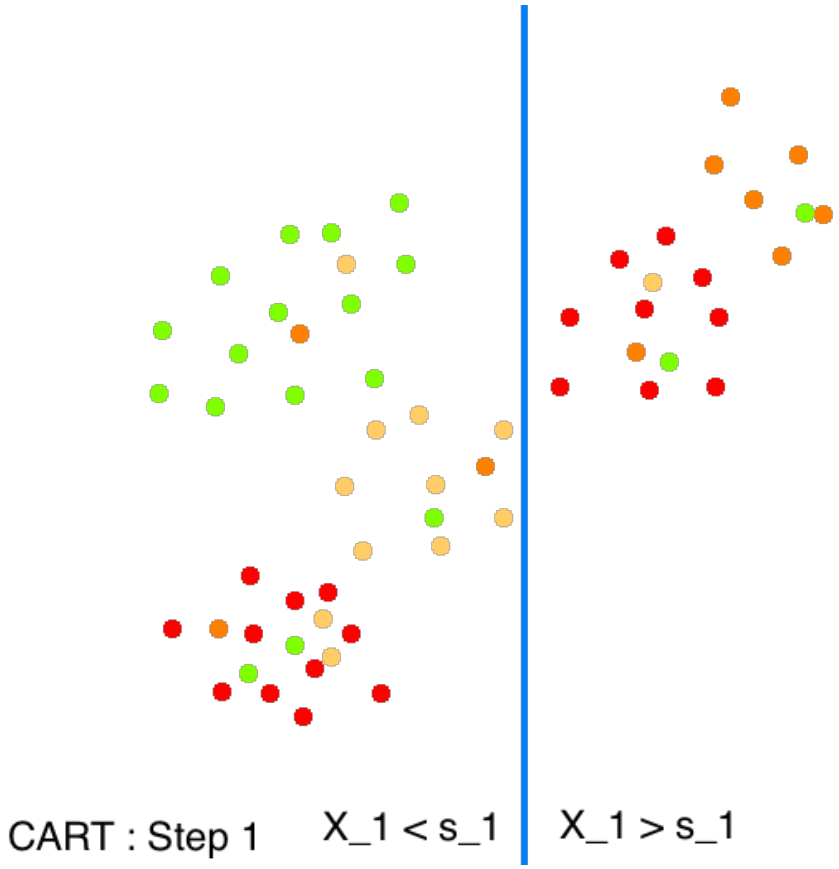
Construction de l'arbre

Règles de partitionnement

$$\mathrm{x} = (x^{(1)}, \ldots, x^{(d)}) \longrightarrow R_j(\mathrm{x})$$

avec

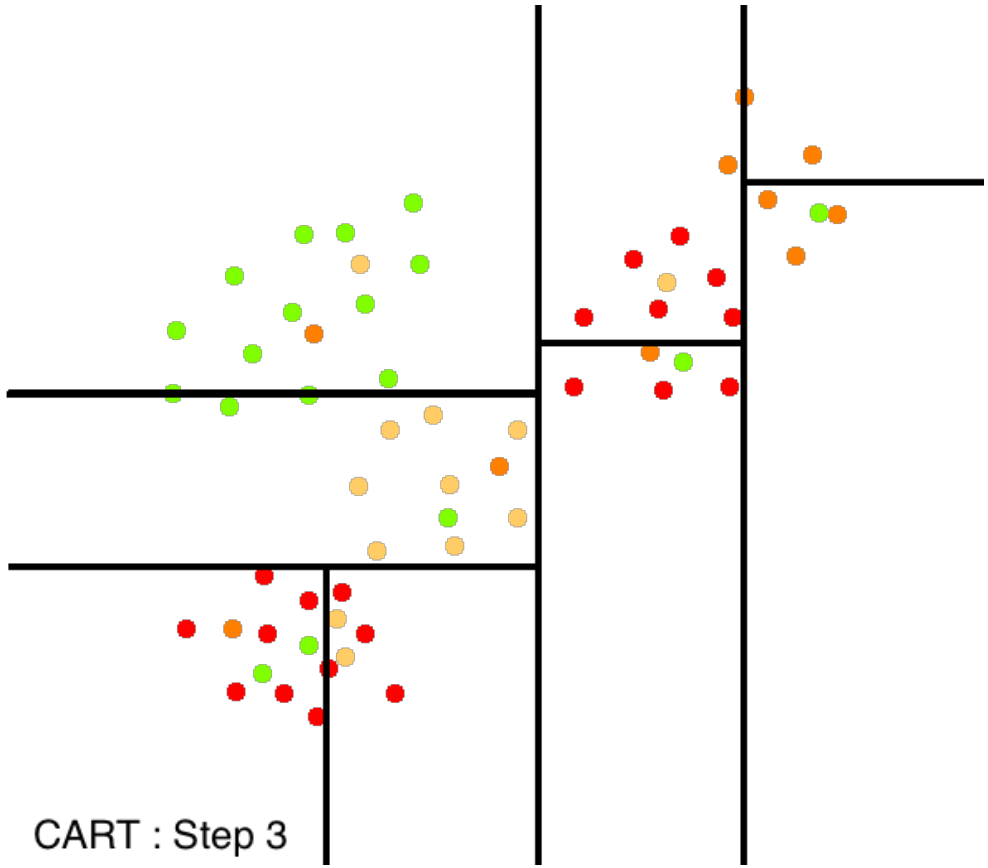
$$\begin{cases} R_j(\mathrm{x}) & = 0 \text{ ou } 1 \\ R_j(\mathrm{x}) R_{j'}(\mathrm{x}) & = 0 \text{ pour } j \neq j' \\ \sum_j R_j(\mathrm{x}) & = 1 \end{cases}$$



Construction de l'arbre

Estimateur de la fonction de régression $\widehat{\theta}(X)$ donné par

$$\widehat{\theta}(x) = \sum_{\ell=1}^{K} \widehat{\theta}(R_j) R_j(x) = \sum_{\ell=1}^{K} \widehat{\theta}_\ell 1_{x \in \mathcal{T}_\ell}$$



Élagage : sélection de modèle

- Soit $T_{\max}$ l'arbre maximal obtenu dans la première phase et $K_{\max}$ le nombre de ses feuilles.
- Extraire de $T_{\max}$ un sous-arbre qui réalise un **compromis entre simplicité et bonne adéquation.**
- **Critère pénalisé** : soit T_K un arbre à K feuilles $(\mathcal{T}_\ell)_{\ell=1,\dots,K}$

$$\sum_{\ell=1}^{K} \sum_{i=1}^{n} \varphi(Y_i, \theta(X_i)) \mathbb{1}_{X_i \in \mathcal{T}_\ell} - \lambda K$$

- $\lambda > 0$ est choisi par **cross-validation.**

Théorie des valeurs extrêmes

Méthode "Peaks-over-Threshold" (PoT)

- Événement extrême, Y_i qui dépasse u, seuil fixé au préalable.
- Sachant que $Y_i > u$, un excès est défini par $Z_i = Y_i - u$.
- La loi des excès :

$$\overline{F}_u(z) = P[Y_1 - u > z \mid Y_1 > u] = \frac{\overline{F}(u+z)}{\overline{F}(u)}, \ z > 0.$$

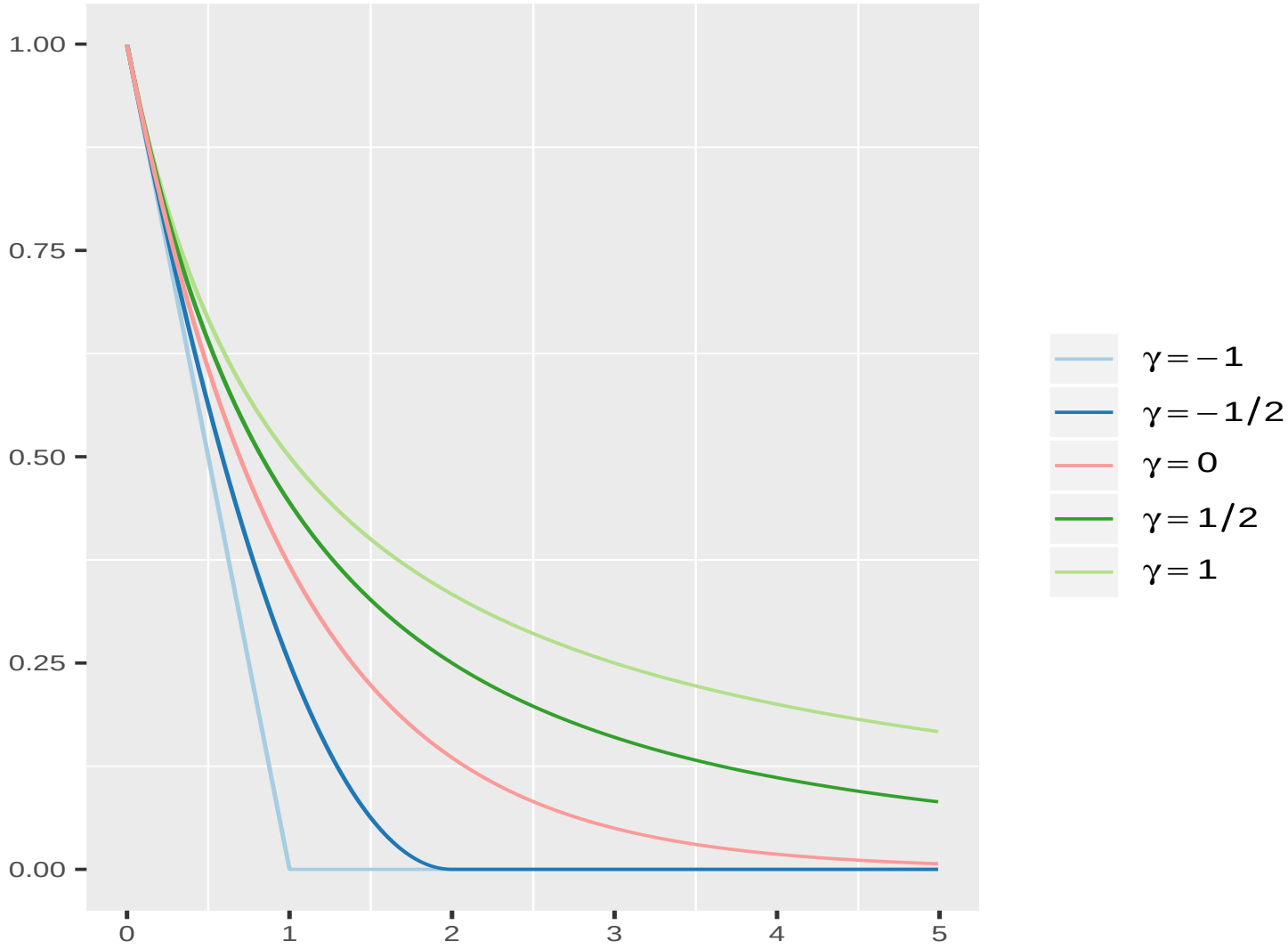
Balkema et de Haan (1974), Pickands (1975)

Sous certaines conditions, la loi des excès F_u converge vers une loi de Pareto généralisée (GPD) dont la fonction de répartition est

$$H_{\sigma,\gamma}(z) = \begin{cases} 1 - \left(1 + \frac{\gamma}{\sigma} z\right)^{-1/\gamma} & \text{si } \gamma \neq 0 \\ 1 - \exp\left(-\frac{z}{\sigma}\right) & \text{si } \gamma = 0 \end{cases}$$

Théorie des valeurs extrêmes

Fonctions de survie des GPD



Arbres de régression avec une loi de Pareto généralisée

- Ici on suppose que $u(x) = u \in [u_{\min}, u_{\max}]$.
- k_n : nombre moyen de $Y_i \geq u$.
- $\widehat{\theta}_\ell^K$ = valeur estimée du paramètre dans la feuille $\mathcal{T}_\ell$, c'est la valeur qui maximise la log-vraisemblance de la feuille ℓ

$$L_n^\ell(\theta) = \frac{1}{k_n} \sum_{i=1}^{n} \varphi(Y_i - u, \theta) 1_{Y_i > u} 1_{X_i \in \mathcal{T}_\ell},$$

- T_K : arbre avec K feuilles notées $\mathcal{T}_\ell, \ell = 1, \ldots, K$ avec les paramètres $\widehat{\theta}_\ell^K$.
- Cet estimateur devrait être proche de

$$\theta_\ell^{*K} = \arg\max_\theta L^\ell(\theta) \quad \text{avec} \quad L^\ell(\theta) = k_n n^{-1} \mathbb{E}[L_n^\ell(\theta)].$$

- T^* = arbre avec les mêmes feuilles que T_K mais avec les paramètres θ_ℓ^{*K}.

Arbres de régression avec une loi de Pareto généralisée

- ${\theta^*}^K(x)$ n'est pas exactement notre cible, on cherche : $\theta_{0,\ell} = (\sigma_0(\mathcal{T}_\ell), \gamma_0(\mathcal{T}_\ell))$, tel que

$$\lim_{t\to\infty} \sup_{z>0} |\overline{F}_t(z \mid \mathcal{T}_\ell) - \overline{H}_{\sigma_0(\mathcal{T}_\ell,t),\gamma_0(\mathcal{T}_\ell)}(z)| = 0,$$

  avec $\overline{F}_t(z \mid \mathcal{T}_\ell) = \mathbb{P}(Y - t \geq z | X \in \mathcal{T}_\ell, Y \geq t)$.

- T_0 : arbre avec les mêmes feuilles que T_K mais avec les paramètres $\theta_{0,\ell}$.

- Si $\theta = (\theta_\ell)_{\ell=1,\dots,K}$ représente les paramètres des K feuilles $(\mathcal{T}_\ell)_{\ell=1,\dots,K}$, on note

$$\theta(x) = \sum_{\ell=1}^{K} \theta_\ell 1_{x\in\mathcal{T}_\ell}.$$

Consistance de l'arbre T_K

- Comparaison de l'arbre T_K et l'arbre T^*
- Distance entre 2 arbres : pour deux arbres T et S,

$$\| T - S \|_2 = \left(\int \| T(x) - S(x) \|_\infty^2 \mathrm{d}\mathbb{P}(x) \right)^{1/2}.$$

avec $\|(a,b)\|_\infty = \max(|a|,|b|)$

Proposition

Sous certaines conditions,

$$\mathbb{E}\left[\| T_K - T^* \|_2^2 \right] \leq \mathscr{C}_1 \frac{K (\log k_n)^2}{k_n}.$$

Biais de modélisation

- La méthode PoT repose sur des résultats asymptotiques.
- Les excès ne sont pas exactement distribués selon une GPD.
- Introduction d'un terme de biais

Proposition

$$\|\theta_0(\mathrm{x}) - \theta^*(\mathrm{x})\|_\infty \quad \leq \quad \mathscr{C}_2(u)\frac{k_n}{n}\left(1 + c\gamma_{\max}\psi(u) + o(\psi(u))\right),$$

avec $\psi(t) \to 0$ lorsque $t \to \infty$ et $\mathscr{C}_2(u)$ une constante dépend de $u, \gamma_{\min}$ et $\gamma_{\max}$

Consistance de l'étape d'élagage de l'arbre

- $$K_0 = \arg \max_{K=1,\dots,K_{\max}} \mathbb{E}\left[\phi(Y-u, \theta^{*K}(X)) 1_{Y>u}\right].$$
- $T^* = T^*_{K_0}$
- Nombre de feuilles sélectionné

$$\widehat{K}(u) = \arg \max_{K=1,\dots,K_{\max}} \left\{ \frac{1}{k_n} \sum_{\ell=1}^{K} \sum_{i=1}^{n} \phi(Y_i - u, \widehat{\theta}^K(X_i)) 1_{Y_i>u} 1_{X_i \in \mathcal{T}_\ell} - \alpha K \right\},$$

- $\widehat{T} = T_{\widehat{K}}$ l'arbre sélectionné correspondant.

Proposition

Sous certaines conditions,

$$\mathbb{E}\left[\| \widehat{T} - T^* \|_2^2\right] \leq \frac{\mathscr{C}_3 K_0 (\log k_n)^2}{k_n}.$$

Arbre de régressions

On utilise les arbres de régressions précédemment décrits et la BD SILECC avec les variables :

- région météorologique,
- nombre d'hydro-écorégions affectées,
- nombre de logements individuels en zone de risque inondation,
- nombre de professionnels en zone de risque inondation.

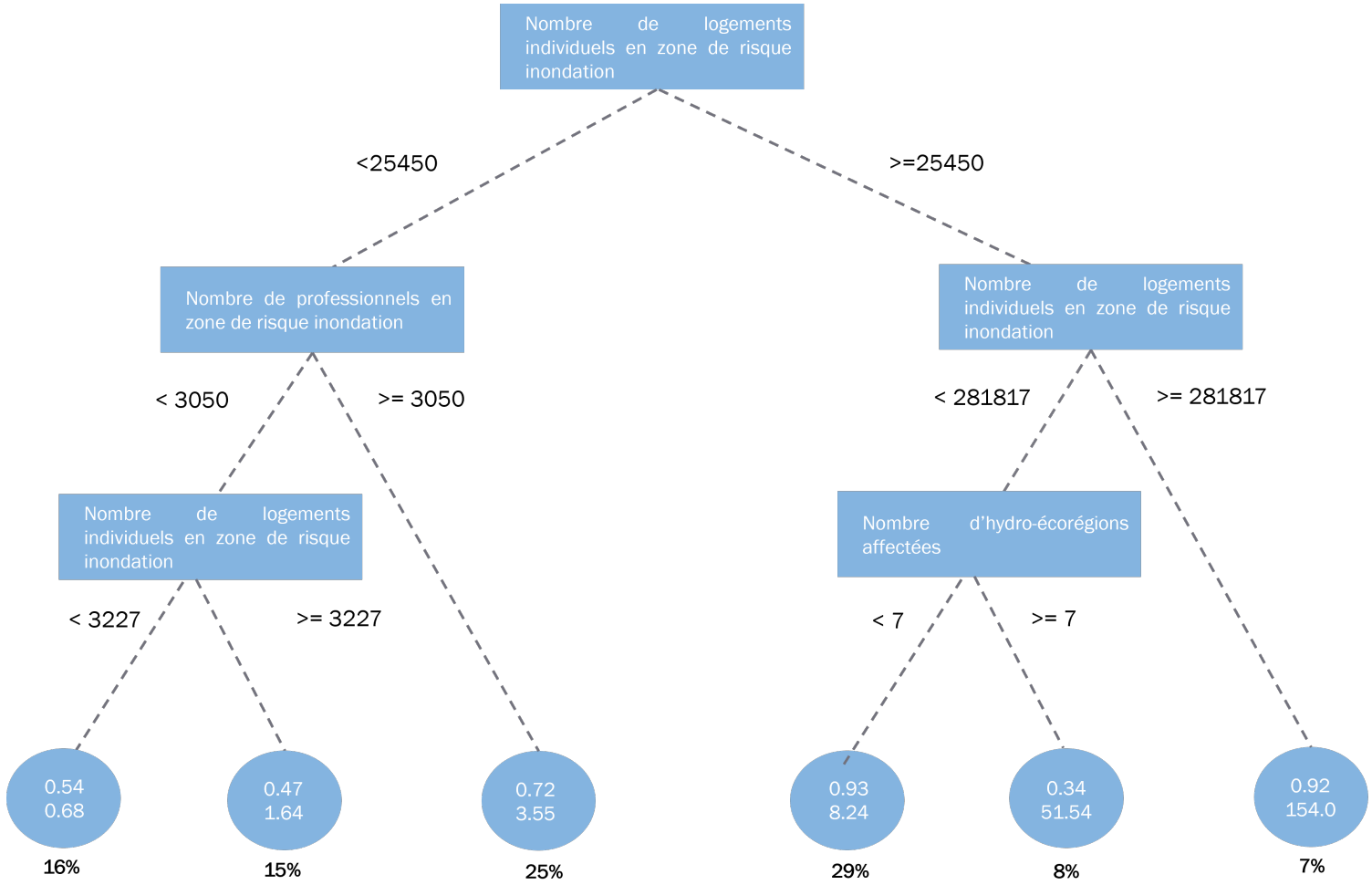
Nous cherchons à **comprendre l'hétérogénéité du coût des événements inondations les plus sévères**, les événements extrêmes.

- **Nous choisissons ici un seuil $u = 100\ 000$ selon des considérations pratiques.**

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Arbre obtenu



Théorie de la crédibilité

- Nous obtenons donc avec cette méthode **une classe avec une distribution pour chaque événement**.
- Mais nous ne pouvons pas donner une estimation fixe pour chaque événement, ce qui est un des attendus pour notre méthode.
- Pour cela nous appliquons la **théorie de la crédibilité bayésienne à l'échelle de la commune.**
- Cela nous permet d'associer les riches informations récoltées à la MRN sur les **coût des événements à l'échelle de la commune** (via BD SILECC) avec les **classes de distributions de Pareto obtenues.**

On cherche $\mathbb{E}[Y_{i,j,n+1} \mid Y_{i,j,1}, ... Y_{i,j,n}]$

Application de la théorie de la crédibilité

On suppose que :

$$Y_{i,j} \sim \mathbb{GPD}(\gamma_j, p_i \sigma_j),$$

avec p_i est la proportion des primes de la commune par rapport au total des primes de l'événement et que :

$$Y_{i,j} \mid \theta_{i,j} \sim \mathbb{EXP}(\theta_{i,j}),$$

avec $\theta_{i,j}$ le profil de risque, $\theta_{i,j} \sim \Gamma(r_j, \lambda_j)$, alors on peut montrer que la loi marginale de $Y_{i,j}$ est une loi de Pareto généralisée.

On peut ensuite trouver que :

$$\mathbb{E}[Y_{i,j,n+1} \mid Y_{i,j,1}, ... Y_{i,j,n}] = \frac{\sum_{k=1}^{n} y_{i,j,k} + \left(\frac{p_i \sigma_j}{\gamma_j}\right)}{n + \frac{1}{\gamma_j} - 1}.$$

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Résultats et discussions

- Nous obtenons des **résultats encourageants avec cette méthode** sur une base de test en comparaison d'une autre approche sévérité fréquence.
- **Incertitudes sur le périmètre touché** le jour J.
- Elle peut sûrement être améliorée en utilisant des informations renseignant **l'intensité de l'aléa.**
- Article en cours de finalisation.

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Discussions et conclusion

Discussions et conclusion

- Une augmentation du coût des événements naturels est à prévoir dans les prochaines années pour maintenir le haut niveau de couverture des dommages par l'assurance, **la réduction des coût est un enjeu essentiel.**
- Une des perspectives pour nos travaux est de participer à terme à **l'amélioration de la résilience des bâtiments.**
- Sur le plan académique, **les travaux sur les arbres de régression ouvrent des perspectives intéressantes pour l'étude des événements extrêmes.**
- L'application de la théorie de la crédibilité à notre problème permet de **considérer le coût à la commune en fonction d'un profil extrême de risque.**

MISSION RISQUES NATURELS

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Merci pour votre attention