

Exploitation d'images de cartes dans la modélisation de la sinistralité en assurance habitation

Yapi KOUASSI

- Introduction

INTRODUCTION

- L'évaluation précise des risques en assurance s'appuie sur des modèles statistiques nécessitant une quantité importante de données.
- **Objectif :** Proposer un outil algorithmique pour rendre exploitables des données géographiques disponibles sous la forme d'images de cartes

- Introduction

SOMMAIRE

1. Outil algorithmique : conception

- Introduction

SOMMAIRE

1. Outil algorithmique : conception

2. Cas pratique : zonier simplifié pour le risque sécheresse

- Introduction

SOMMAIRE

1. Outil algorithmique : conception

2. Cas pratique : zonier simplifié pour le risque sécheresse

3. Conclusion

- Outil algorithmique : conception

1. Outil algorithmique : conception

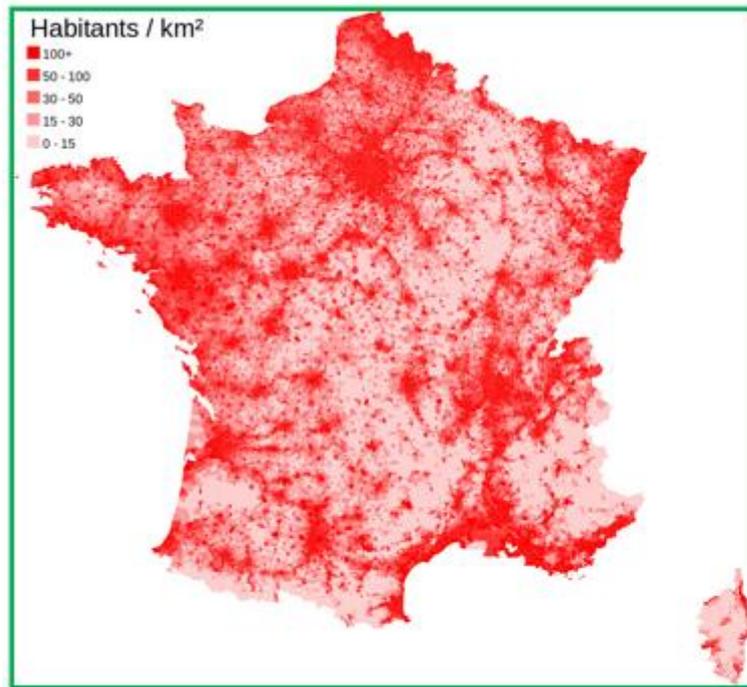
2. Cas pratique : zonier simplifié pour le risque sécheresse

3. Conclusion

- Outil algorithmique : conception

Rappel de l'objectif

Développer un outil d'extraction de données géographiques à partir d'images de cartes.



Carte (au format image)



<u>Codes postaux</u>	<u>Habitants / km²</u>
01000	100+
01090	100+
01100	100+
01110	50 - 100
01120	100+

Base de données

Figure 1 : Passage d'une image de carte à une base de données

- Outil algorithmique : conception

Rappel de l'objectif

Développer un outil d'extraction de données géographiques à partir d'images de cartes.

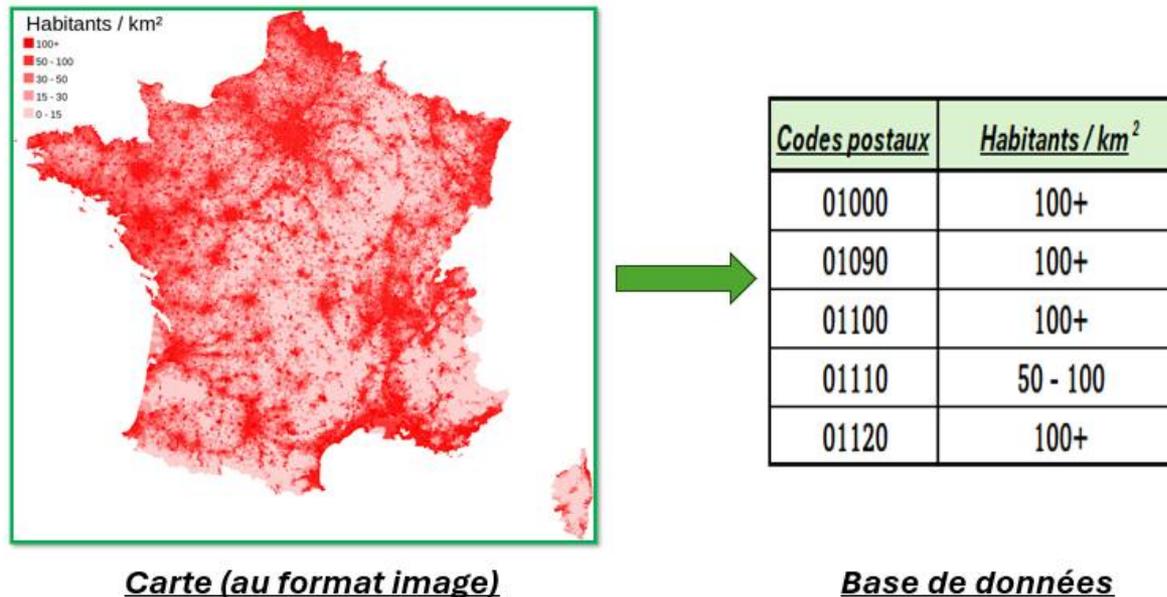


Figure 1 : Passage d'une image de carte à une base de données

Etape préalable : validation légale et définition du cadre d'utilisation de l'outil en consultant le service juridique.

- Outil algorithmique : conception

Premiers tests et constats

- Identifier les pixels correspondant aux localités associées à chaque code postal.
- Extraire la couleur de ces pixels (RGB).
- Comparer ces couleurs avec celles de la légende.

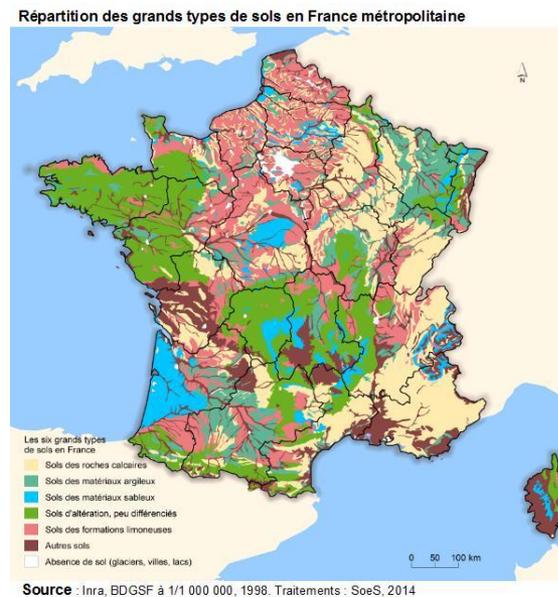
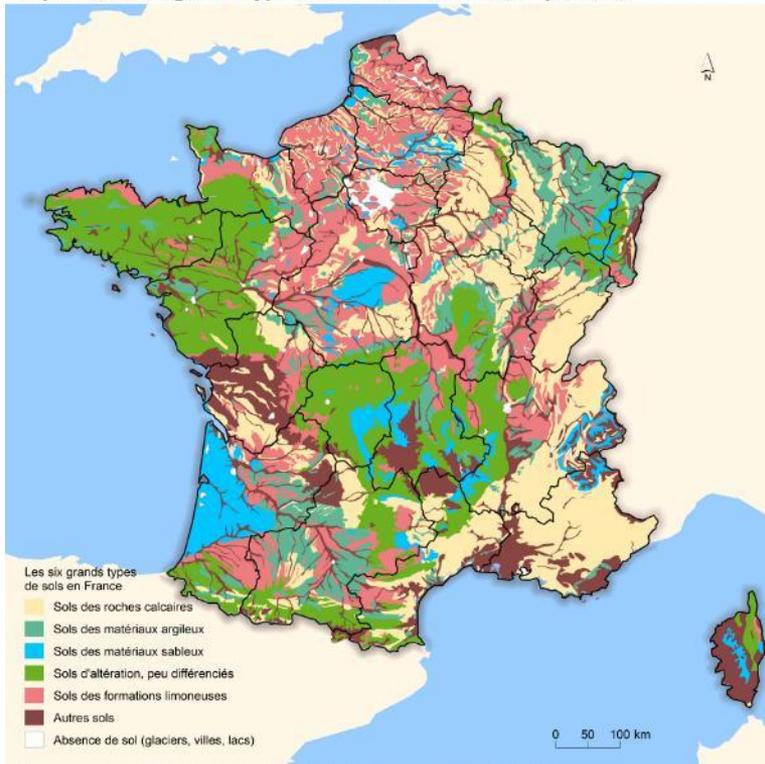


Figure 2 : Première carte étudiée

- Outil algorithmique : conception

Premiers tests et constats

Répartition des grands types de sols en France métropolitaine



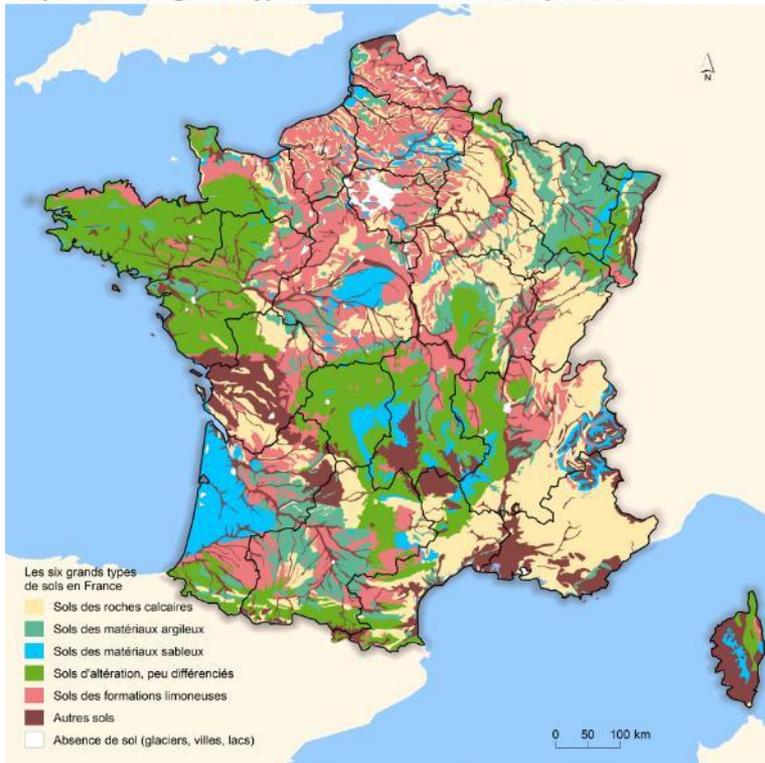
3 principaux constats

Figure 2 : Première carte étudiée

- Outil algorithmique : conception

Premiers tests et constats

Répartition des grands types de sols en France métropolitaine



Source : Inra, BDGSF à 1/1 000 000, 1998. Traitements : SoeS, 2014

3 principaux constats

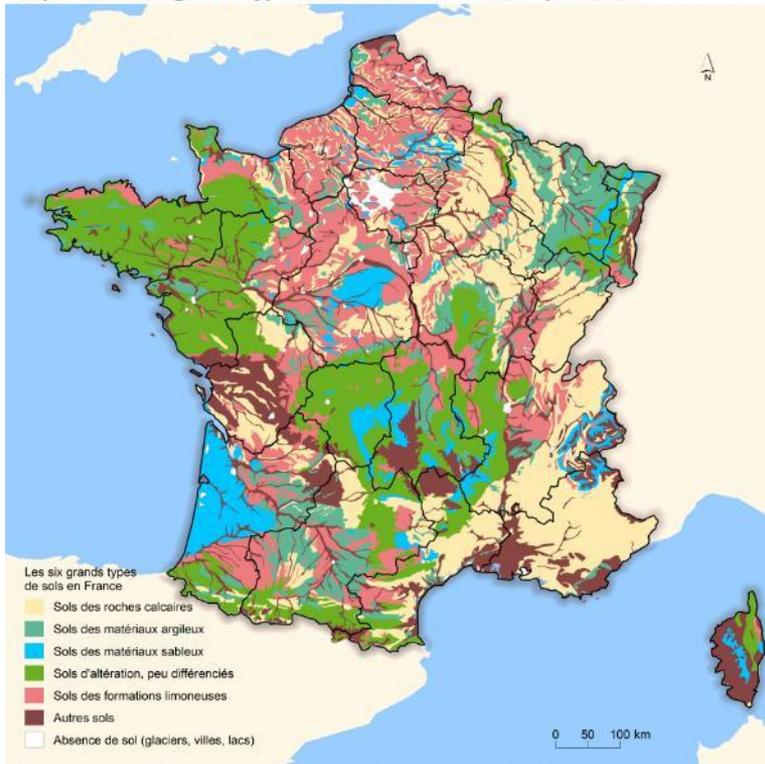
- ❖ Différence entre les couleurs de la légende et celles sur la carte.

Figure 2 : Première carte étudiée

- Outil algorithmique : conception

Premiers tests et constats

Répartition des grands types de sols en France métropolitaine



3 principaux constats

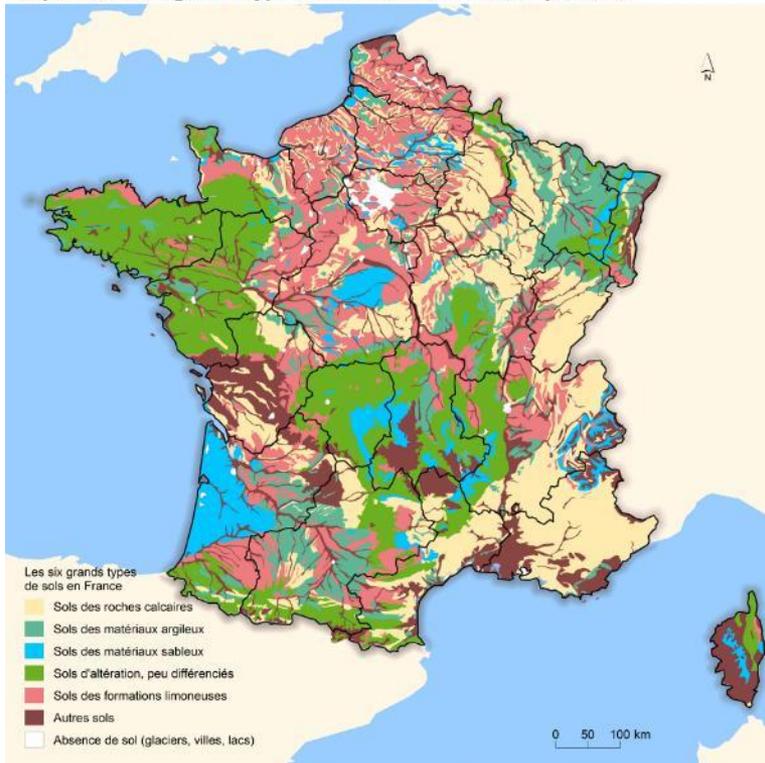
- ❖ Différence entre les couleurs de la légende et celles sur la carte.
- ❖ Présence d'obstacles à la récupération des couleurs des pixels.

Figure 2 : Première carte étudiée

- Outil algorithmique : conception

Premiers tests et constats

Répartition des grands types de sols en France métropolitaine



Source : Inra, BDGSF à 1/1 000 000, 1998. Traitements : SoeS, 2014

3 principaux constats

- ❖ Différence entre les couleurs de la légende et celles sur la carte.
- ❖ Présence d'obstacles à la récupération des couleurs des pixels.
- ❖ Besoin d'uniformisation des positions des codes postaux sur les cartes.

Figure 2 : Première carte étudiée

- Outil algorithmique : conception

Fonctionnement de l'outil

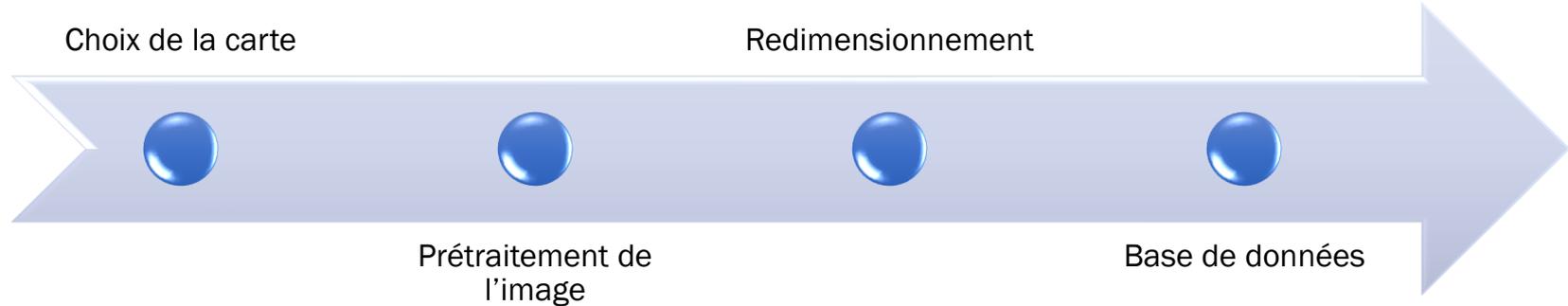


Figure 3 : Fonctionnement de l'outil algorithmique

- Outil algorithmique : conception

Fonctionnement de l'outil

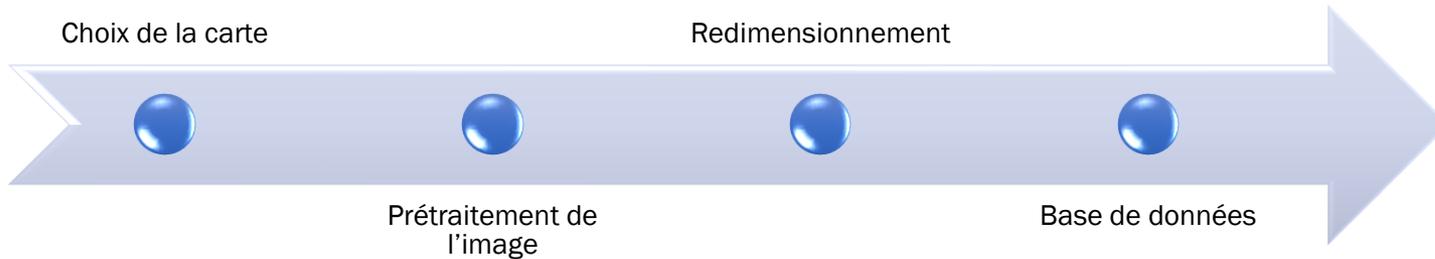


Figure 3 : Fonctionnement de l'outil algorithmique

- **Prétraitement de l'image** : réduire le bruit et séparer les couleurs

- Outil algorithmique : conception

Fonctionnement de l'outil

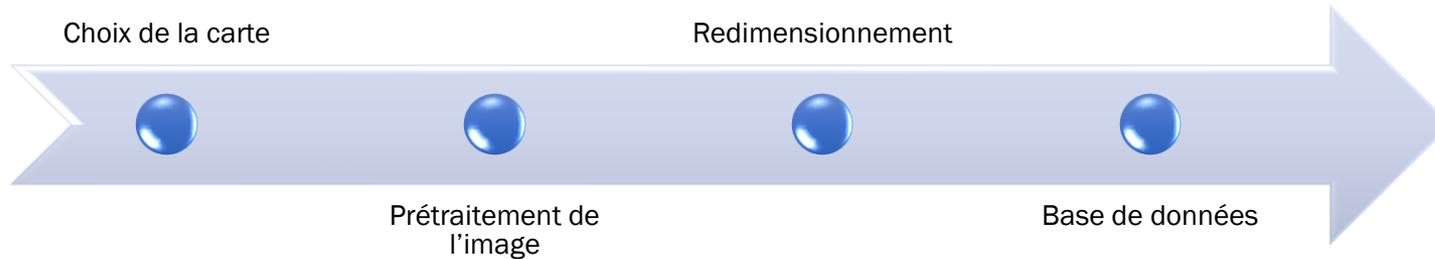


Figure 3 : Fonctionnement de l'outil algorithmique

- **Prétraitement de l'image** : réduire le bruit et séparer les couleurs
- **Redimensionnement** : uniformiser les positions des codes postaux sur toutes les cartes en utilisant une carte de référence.

- Outil algorithmique : conception

Fonctionnement de l'outil

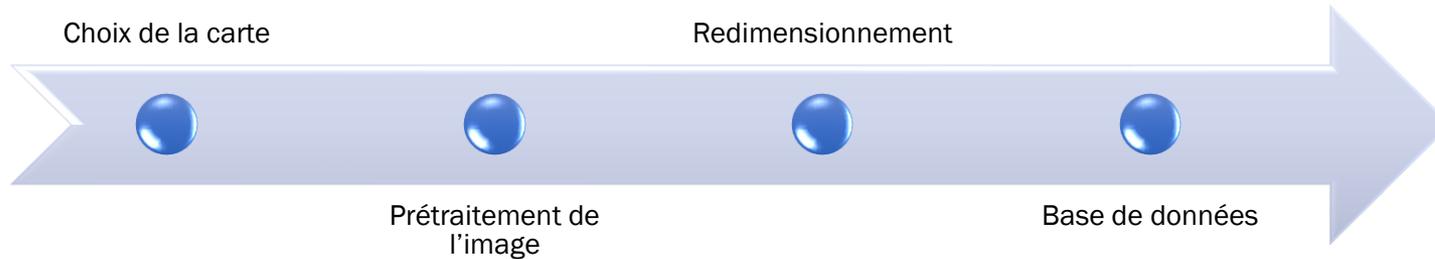


Figure 3 : Fonctionnement de l'outil algorithmique

- **Prétraitement de l'image** : réduire le bruit et séparer les couleurs
- **Redimensionnement** : uniformiser les positions des codes postaux sur toutes les cartes en utilisant une carte de référence.
- **Bases de données** : constituer la base de données finale qui attribue aux codes postaux la légende correspondante.

- Outil algorithmique : conception

Prétraitement des images : clustering colorimétrique (1/2)

2 outils statistiques sont utilisés pour le clustering colorimétrique : K-Means et ACP

- **K-Means** : Regroupement direct des pixels en K groupes (clusters) de couleurs.

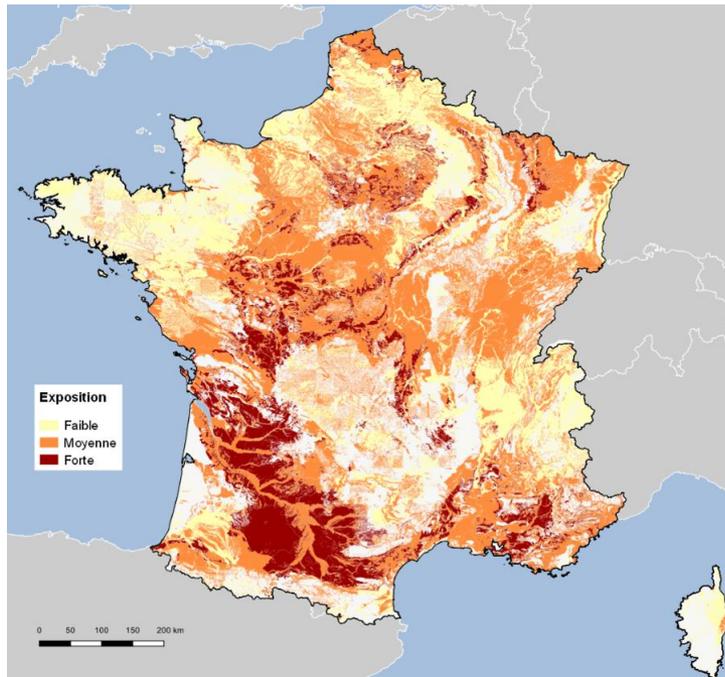


Figure 4 : Exemple de carte n° 1

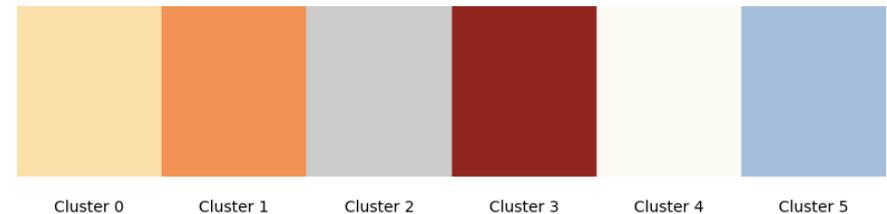


Figure 5 : Clusters avec K-Means

- Outil algorithmique : conception

Prétraitement des images : clustering colorimétrique (1/2)

2 outils statistiques sont utilisés pour le clustering colorimétrique : K-Means et ACP

- **K-Means** : Regroupement direct des pixels en K groupes (clusters) de couleurs.

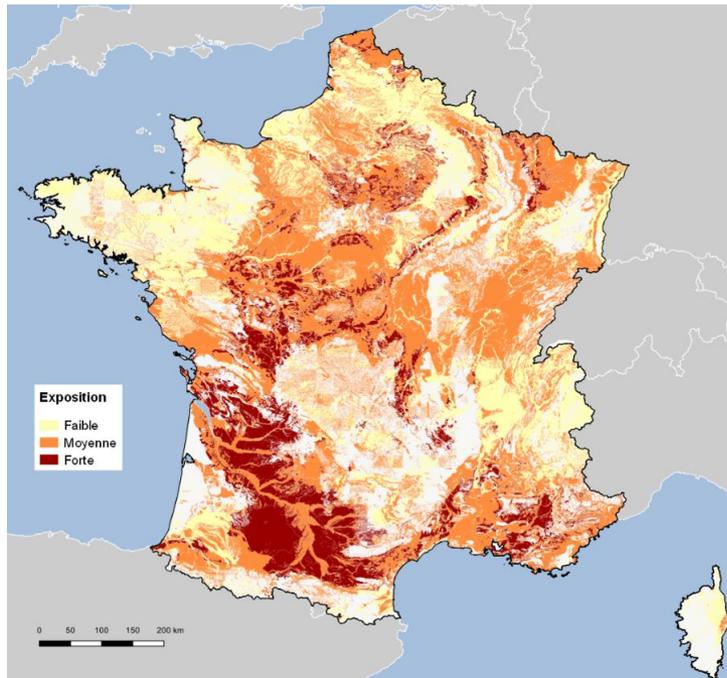


Figure 4 : Exemple de carte n° 1

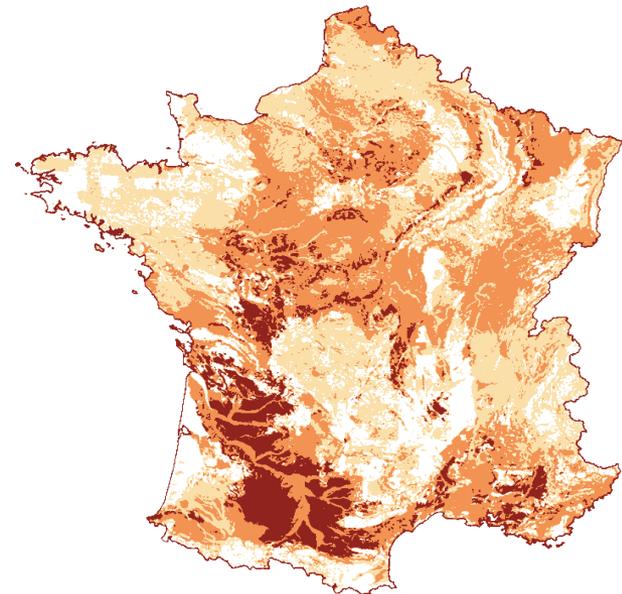


Figure 6 : Carte n° 1 prétraitée avec K-Means

- Outil algorithmique : conception

Prétraitement des images : clustering colorimétrique (2/2)

- **ACP** : Représentation des pixels dans un plan et constitution manuelle des clusters.

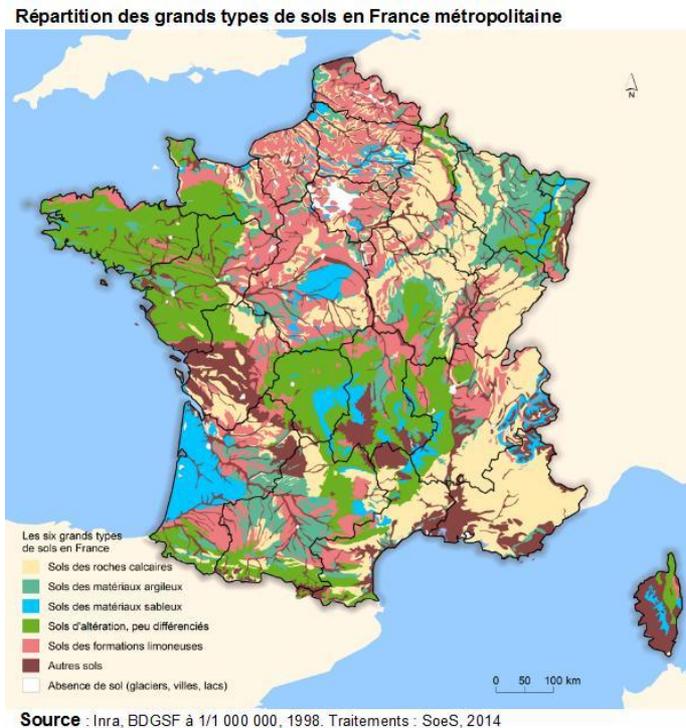


Figure 7 : Exemple de carte n°2

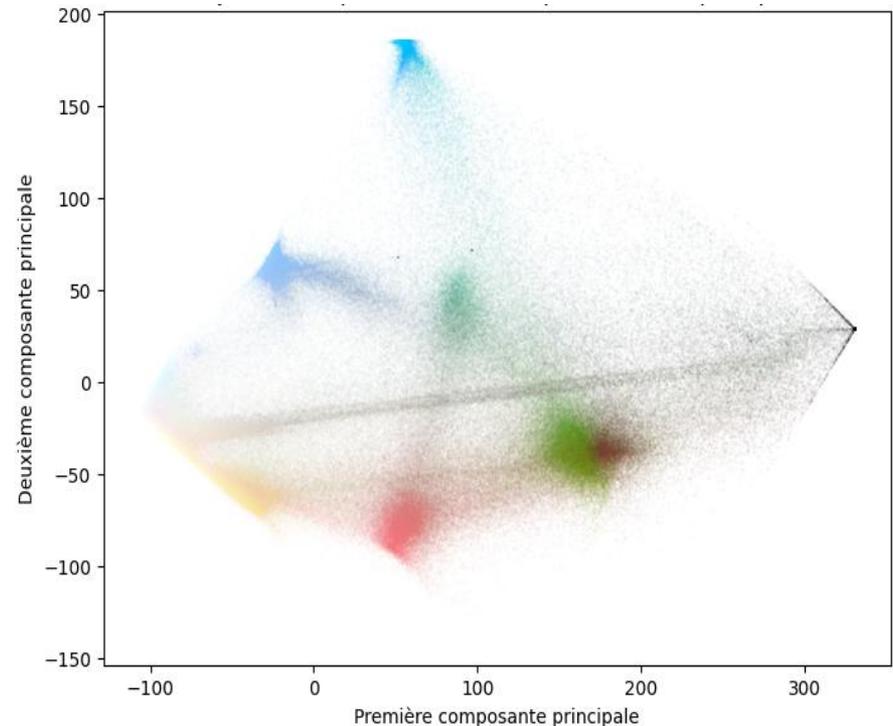


Figure 8 : Projection des pixels (ACP)

- Outil algorithmique : conception

Prétraitement des images : clustering colorimétrique (2/2)

- **ACP** : Représentation des pixels dans un plan et constitution manuelle des clusters.

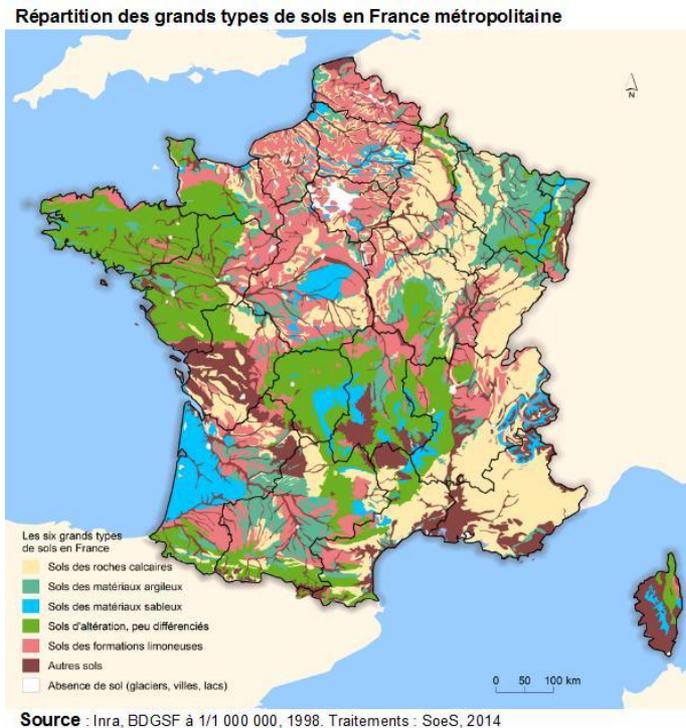


Figure 7 : Exemple de carte n°2

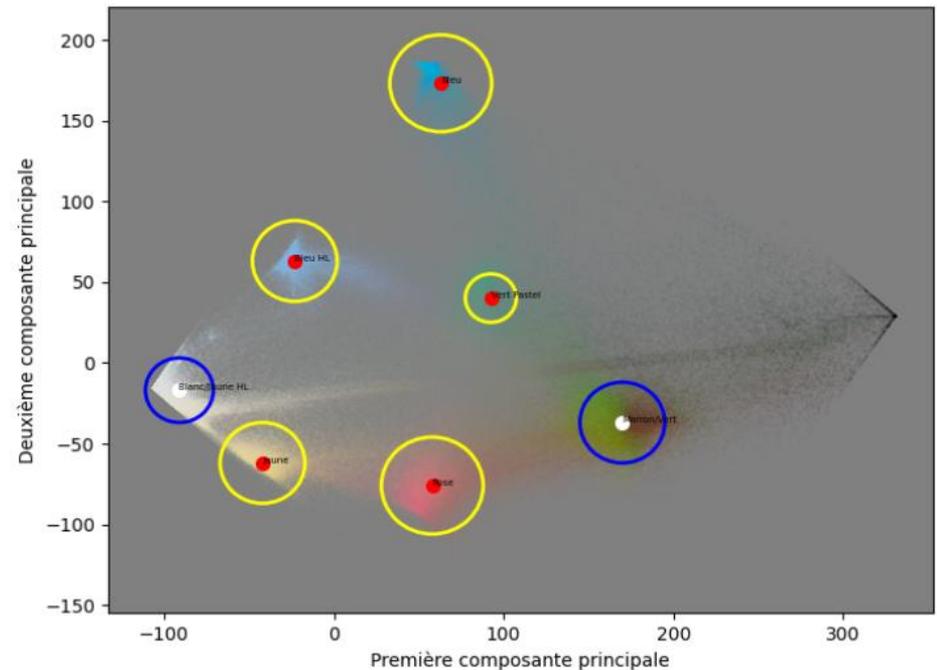


Figure 9 : Projection des pixels (ACP)

- Outil algorithmique : conception

Prétraitement des images : clustering colorimétrique (2/2)

- **ACP** : Représentation des pixels dans un plan et constitution manuelle des clusters.

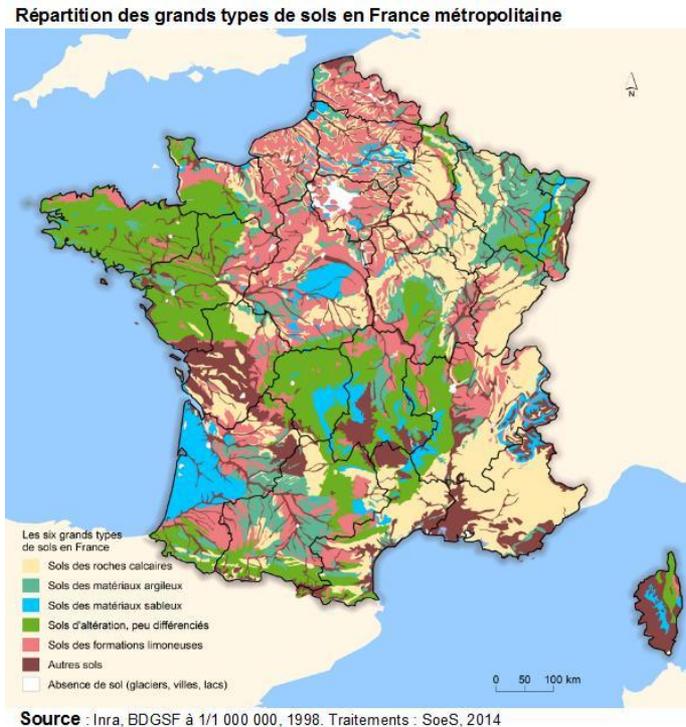


Figure 7 : Exemple de carte n°2

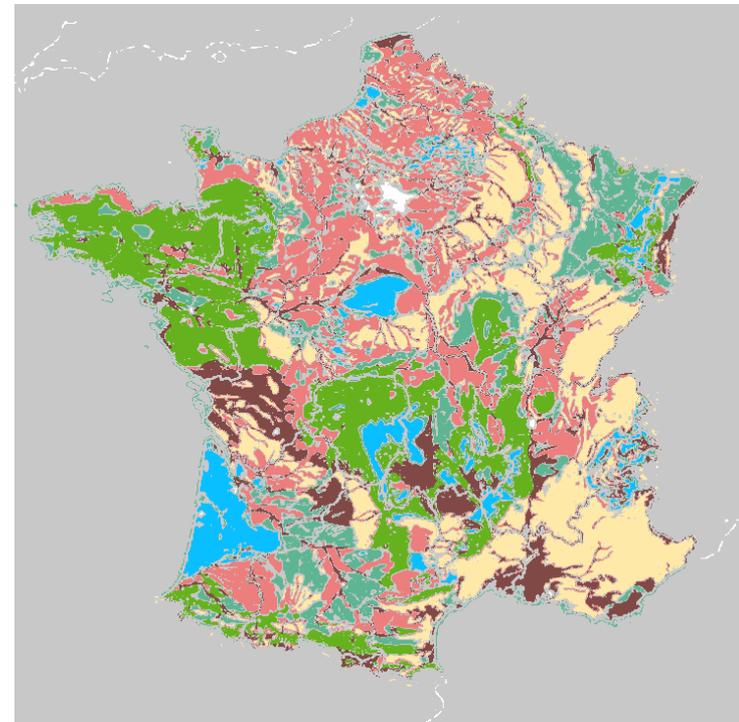
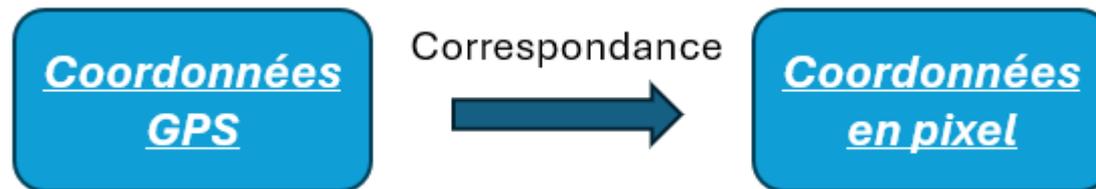


Figure 10 : Carte n°2 prétraitée avec ACP

- Outil algorithmique : conception

Redimensionnement : carte et base de données de référence

- **But** : Connaître la position de tous les codes postaux sur la carte de référence.



- Pour faire la correspondance, une régression linéaire est utilisée suivant deux approches :

$$1. \begin{cases} Pixel_X = \beta_0 + \beta_1 \cdot Latitude + \beta_2 \cdot Longitude + W \\ Pixel_Y = \alpha_0 + \alpha_1 \cdot Latitude + \alpha_2 \cdot Longitude + Z \end{cases}$$

$$2. \begin{cases} Pixel_X = \beta'_0 + \beta'_1 \cdot Xlambert + \beta'_2 \cdot Ylambert + W' \\ Pixel_Y = \alpha'_0 + \alpha'_1 \cdot Xlambert + \alpha'_2 \cdot Ylambert + Z' \end{cases}$$

- Outil algorithmique : conception

Redimensionnement : modification / déformation des cartes étudiées (1/2)

- Cette étape sert à modifier la carte étudiée pour qu'elle se superpose à l'image de référence.

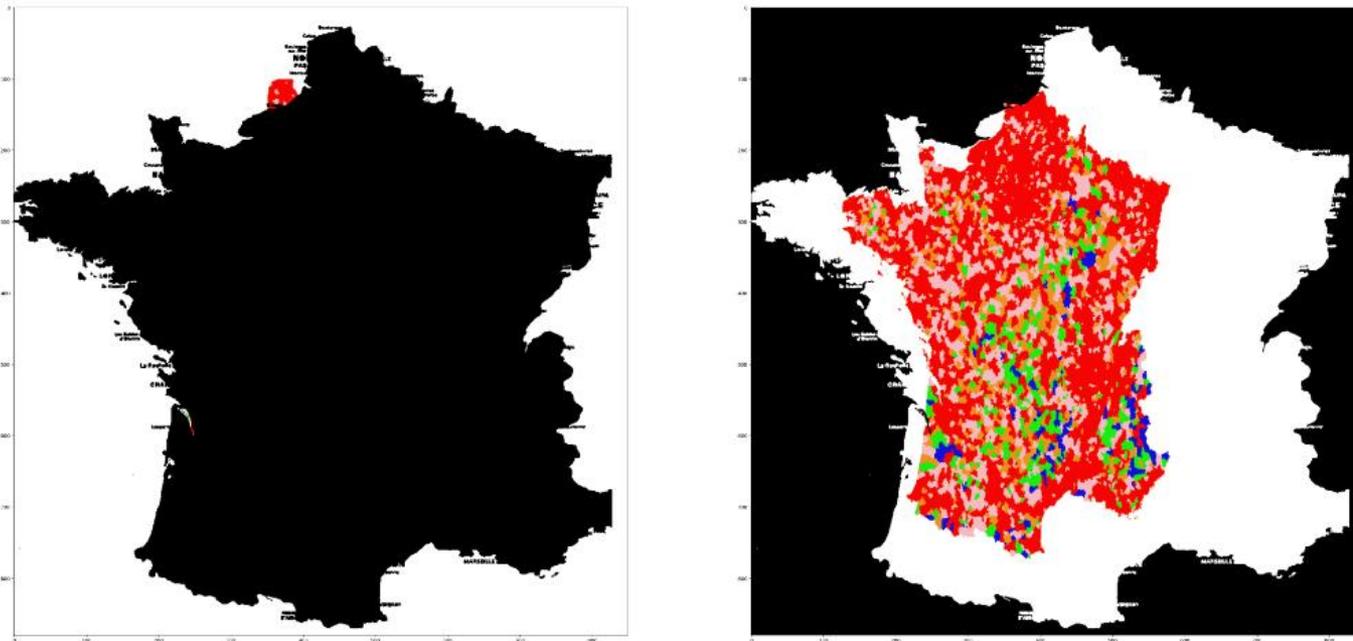


Figure 11 : Superposition initiale d'une carte et du masque de référence

- Outil algorithmique : conception

Redimensionnement : modification / déformation des cartes étudiées (2/2)

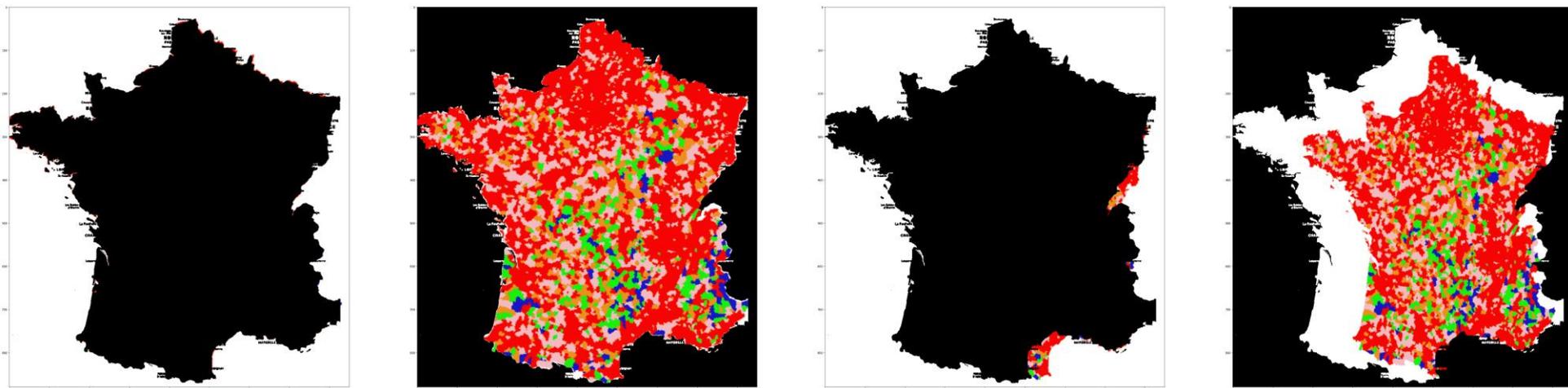


Figure 12 : Redimensionnement : optimal vs non optimal

- Outil algorithmique : conception

Redimensionnement : remplissage des trous par k-NN

- **k-NN** : On attribue à un pixel « vide », la couleur des pixels autour de lui.

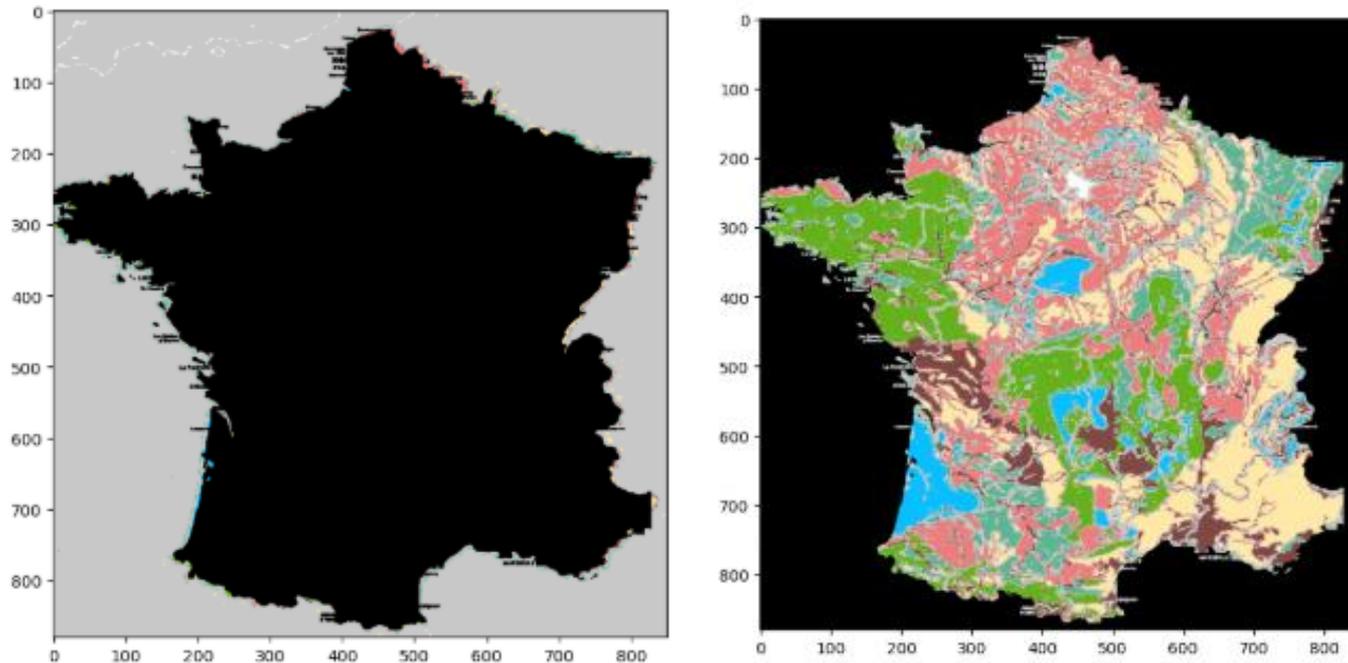


Figure 13 : Carte n°2 : redimensionnement + remplissage

- Outil algorithmique : conception

Redimensionnement : remplissage des trous par k-NN

- **k-NN** : On attribue à un pixel « vide », la couleur des pixels autour de lui.

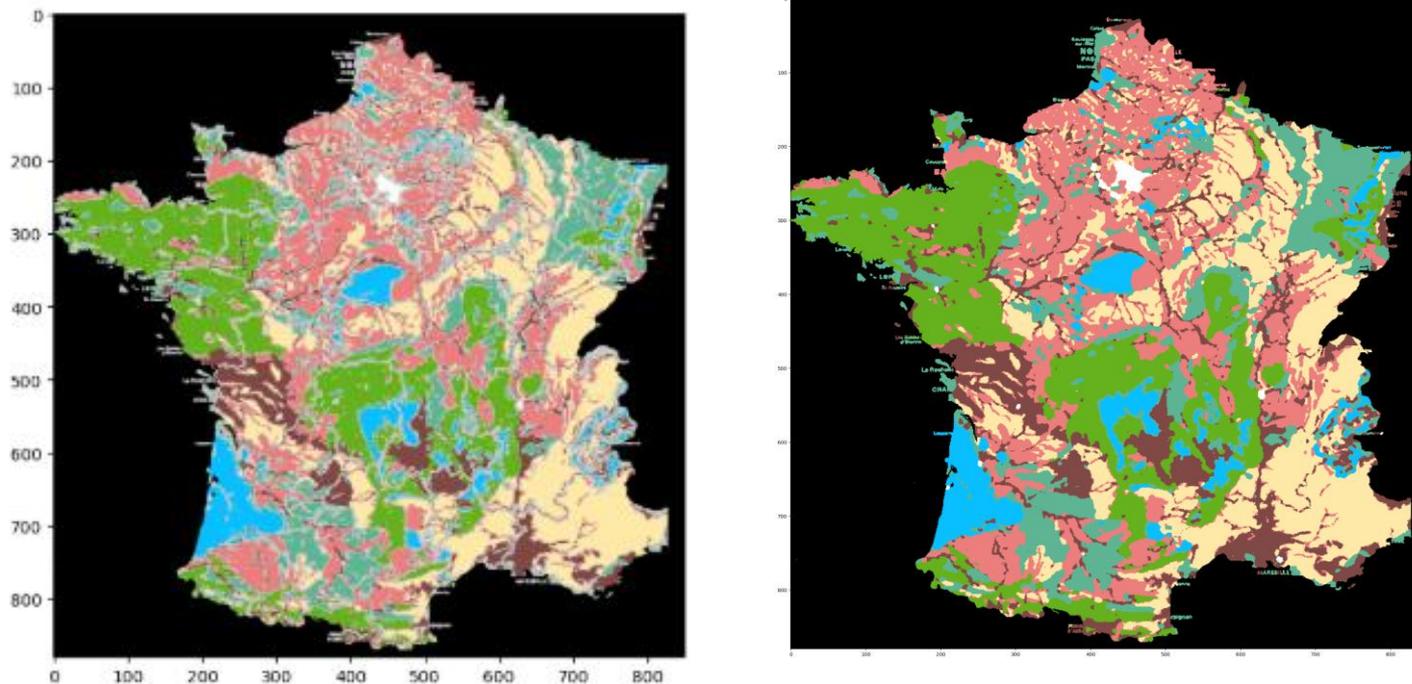
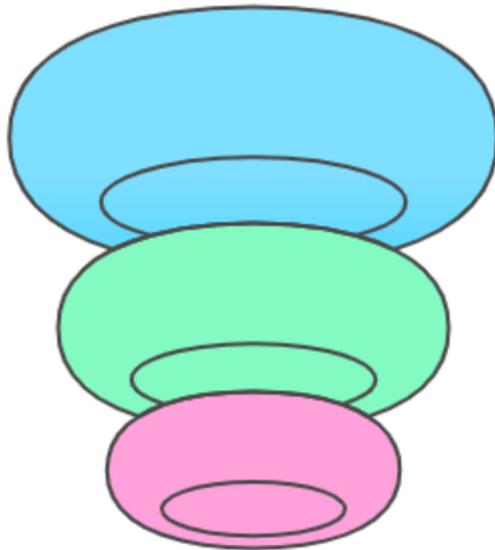


Figure 13 : Carte n°2 : redimensionnement + remplissage

- Outil algorithmique : conception

Base de données finale



Extraction des couleurs

Extraction des couleurs des pixels de carte



Correspondance des couleurs

Correspondance des couleurs extraites avec celles de la légende



Attribution des légendes

Attribution des légendes correspondantes aux codes postaux

EXPLOITATION D'IMAGES DE CARTES DANS LA MODÉLISATION DE LA SINISTRALITÉ EN ASSURANCE HABITATION

- Outil algorithmique : conception

Résultats et analyses

- Outil algorithmique : conception

Résultats et analyses

- **Performance de l'outil** : taux de classification correcte allant jusqu'à 89,63 %

- Outil algorithmique : conception

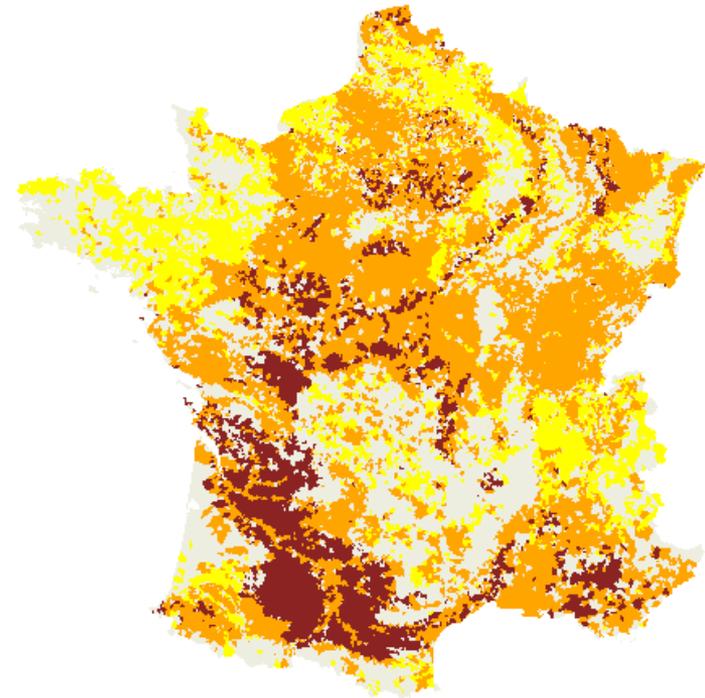
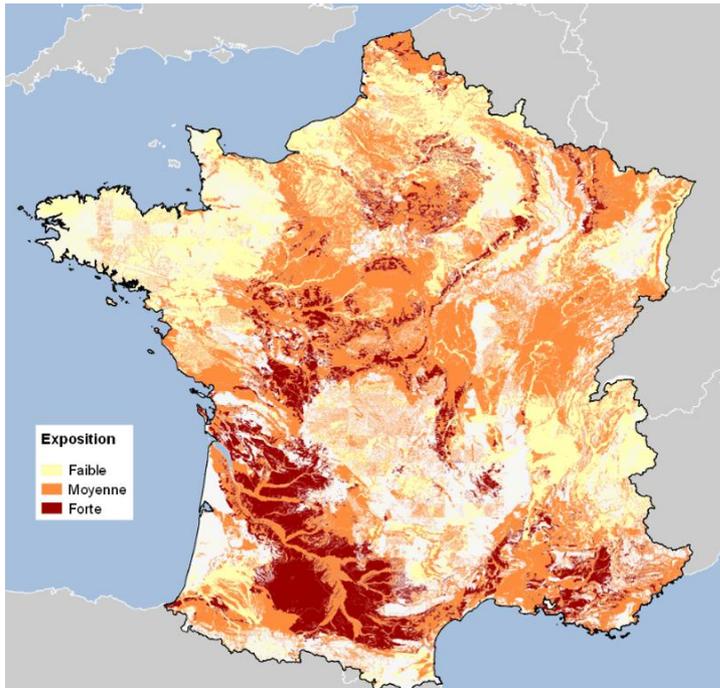
Résultats et analyses

- **Performance de l'outil** : taux de classification correcte allant jusqu'à 89,63 %
- **Limites** : absence d'évaluation numérique et sensibilité à la complexité des cartes utilisées.

- Outil algorithmique : conception

Résultats et analyses

- **Limites** : absence d'évaluation numérique et sensibilité à la complexité des cartes utilisées.



- Outil algorithmique : conception

Résultats et analyses

- **Performance de l'outil** : taux de classification correcte allant jusqu'à 89,63 %
- **Limites** : absence d'évaluation numérique et sensibilité à la complexité des cartes utilisées.
- **Perspectives d'amélioration** : techniques avancées de détection de la France sur les images ; filtres pour réduire le bruit.

- Cas pratique : zonier simplifié pour le risque sécheresse

1. Outil algorithmique : conception

2. Cas pratique : zonier simplifié pour le risque sécheresse

3. Conclusion

- Cas pratique : zonier simplifié pour le risque sécheresse

Problématique

Cardif IARD souhaite construire un modèle de risque, qui à chaque contrat fournira une estimation de la prime pure technique et de la rentabilité technique.

- Cas pratique : zonier simplifié pour le risque sécheresse

Problématique

Cardif IARD souhaite construire un modèle de risque, qui à chaque contrat fournira une estimation de la prime pure technique et de la rentabilité technique.

Objectif

Concevoir un premier modèle de risque pour la sécheresse sur la base de variables géographiques externes.

- Cas pratique : zonier simplifié pour le risque sécheresse

Comprendre le phénomène de sécheresse en assurance habitation

Le phénomène de sécheresse en assurance habitation est principalement lié aux mouvements de terrain causés par l'alternance de périodes de dessèchement et de réhydratation des sols argileux (retrait gonflement d'argile).

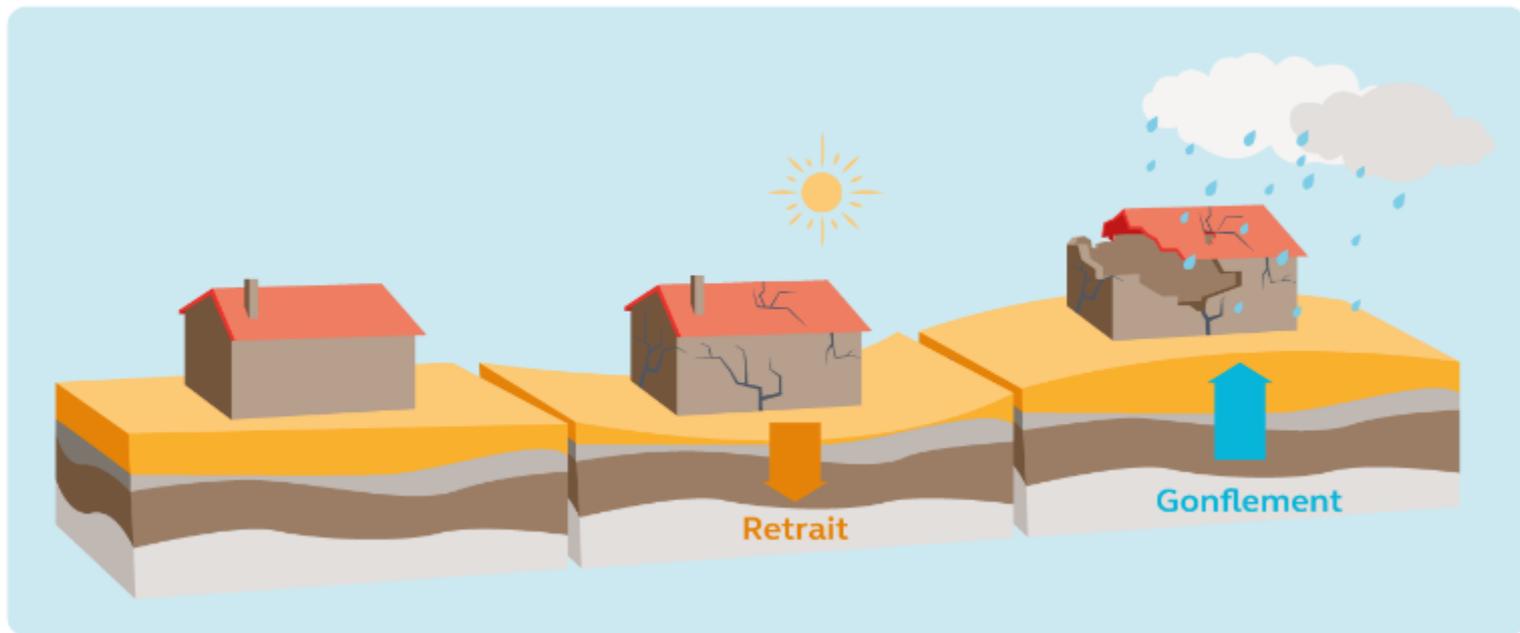


Figure 14 : Retrait-gonflement d'argile

- Cas pratique : zonier simplifié pour le risque sécheresse

Extraction de variables géographiques externes

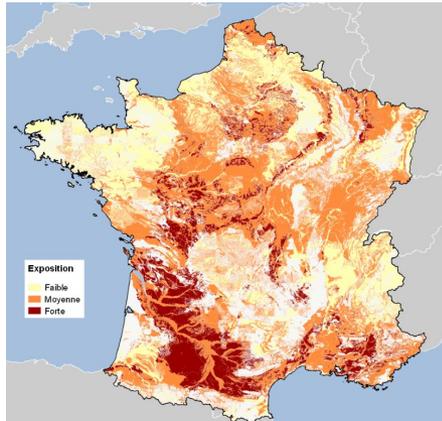


Figure 15 : Carte_rga

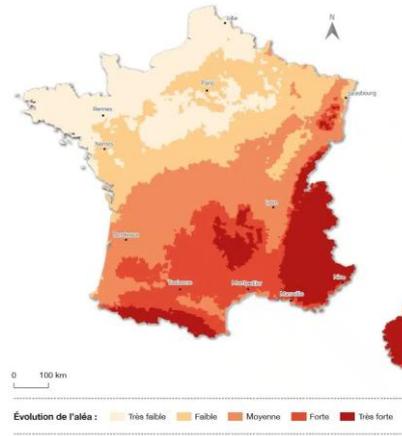


Figure 16 : Carte_chaleur

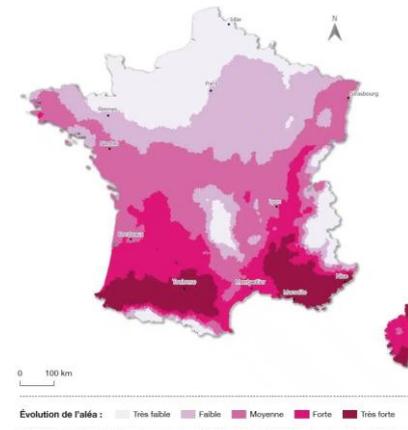


Figure 17 : Carte_irrad

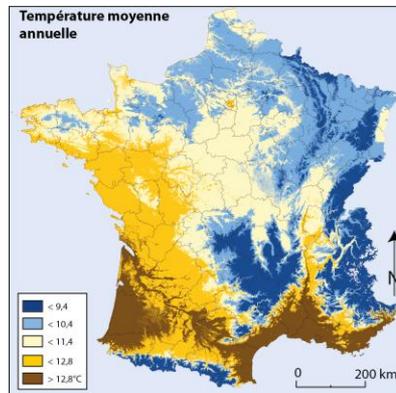


Figure 18 : Carte_temp

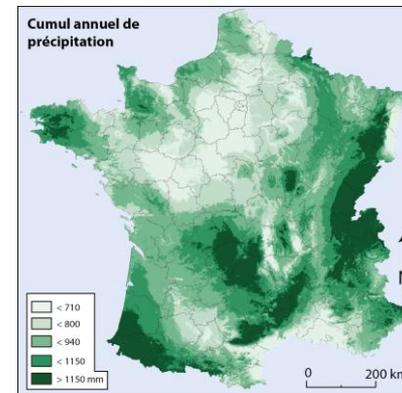


Figure 19 : Carte_pluie

- Cas pratique : zonier simplifié pour le risque sécheresse

Modélisation de la fréquence des sinistres : modèle 1C (1 carte)

Simplifications sur la base de risque :

- Focus sur les propriétaires de maisons
- Période analysée : 2020-2022

Carte_rga	Expos (%)	Fréquences (%)
Faible (A)	43,46	0,13
Sans info (B)	19,71	0,22
Moyenne (C)	27,82	0,34
Fort (D)	9,01	0,74

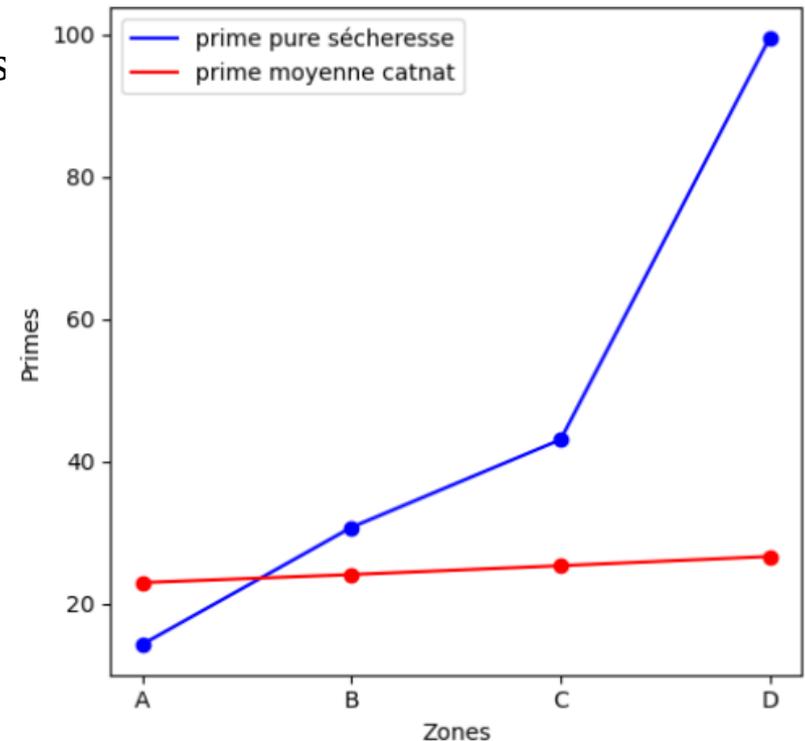


Figure 20 : Retrait-gonflement d'argile

- Cas pratique : zonier simplifié pour le risque sécheresse

Modélisation de la fréquence des sinistres : modèle 1C (1 carte)

Simplifications sur la base de risque :

- Focus sur les propriétaires de maisons
- Période analysée : 2020-2022

Carte_rga	Expos (%)	Fréquences (%)
Faible (A)	43,46	0,13
Sans info (B)	19,71	0,22
Moyen (C)	27,82	0,34
Fort (D)	9,01	0,74

La fréquence des sinistres croît avec le risque de retrait-gonflement des argiles.

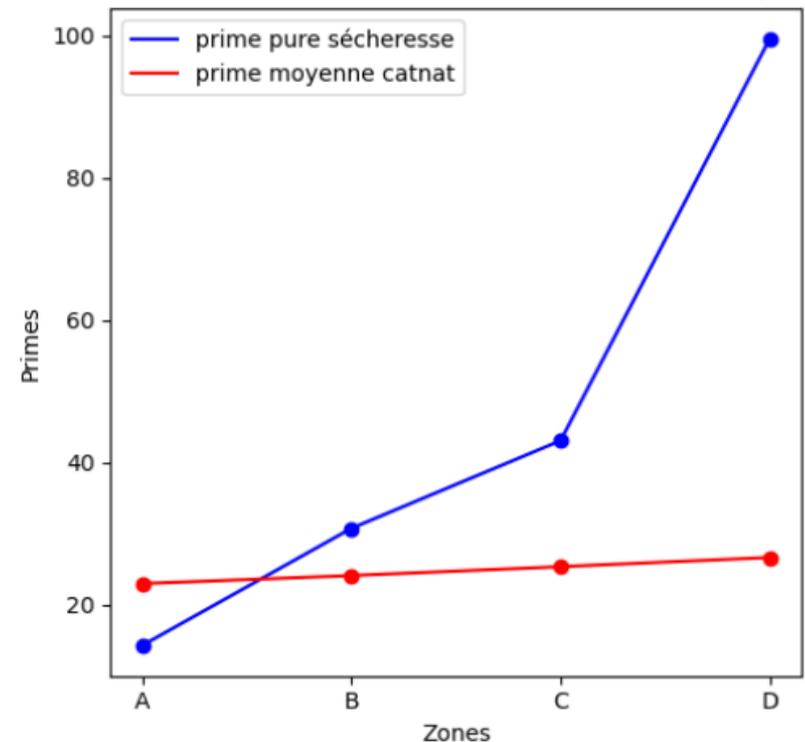


Figure 20 : Retrait-gonflement d'argile

- Cas pratique : zonier simplifié pour le risque sécheresse

Modélisation de la fréquence des sinistres : modèle PC (plusieurs cartes)

Dans ce deuxième modèle, les 5 cartes sont utilisées à travers un GLM Poisson pour modéliser la fréquence des sinistres. Les corrélations, la fréquence des sinistres par modalités de chaque carte et les résultats du GLM sont analysés :

Observations

- ❖ **Carte_rga** : modalités très significatives.
- ❖ **Carte_chaleur** : modalités peu significatives et redondance avec la carte_irrad.
- ❖ **Carte_irrad** : modalités très significatives.
- ❖ **Carte_pluie** : modalités peu significatives.
- ❖ **Carte_temp** : modalités significatives mais instables

Choix

- **Conservation** de la carte_rga et de la carte_irrad
- **Retrait** de la carte_chaleur et de la carte_pluie
- **Scission** du modèle en deux : l'un avec la carte_temp, l'autre sans.

- Cas pratique : zonier simplifié pour le risque sécheresse

Modèles RI et RIT

- **Modèle RI** : Retrait gonflement d'argile et Irradiation solaire. (2 cartes)
- **Modèle RIT** : Retrait gonflement d'argile, Irradiation solaire et Température moyenne. (3 cartes)

Indicateurs	Modèle RI	Modèle RIT	Observation
Déviance	10 143	10 084	
AIC	11 554	11 505	
BIC	- 3 338 271	-3 338 300	
Fréquence (%)	0,323	0,323	0,323
RMSE	0,06508	0,06507	

- Cas pratique : zonier simplifié pour le risque sécheresse

Modèles RI et RIT

- **Modèle RI** : Retrait gonflement d'argile et Irradiation solaire. (2 cartes)
- **Modèle RIT** : Retrait gonflement d'argile, Irradiation solaire et Température moyenne. (3 cartes)

Indicateurs	Modèle RI	Modèle RIT	Observation
Déviance	10 143	10 084	
AIC	11 554	11 505	
BIC	- 3 338 271	-3 338 300	
Fréquence (%)	0,323	0,323	0,323
RMSE	0,06508	0,06507	

Les indicateurs montrent une **légère supériorité** du modèle RIT en termes de précision.

- Cas pratique : zonier simplifié pour le risque sécheresse

Constitution des zones

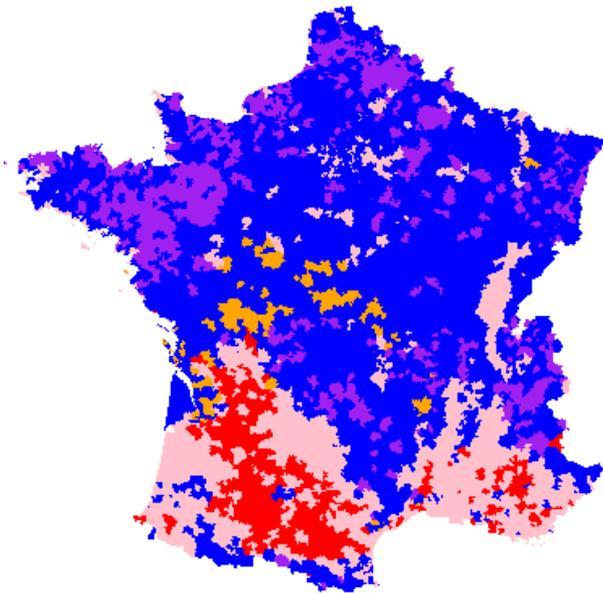


Figure 21 : Zonier RI

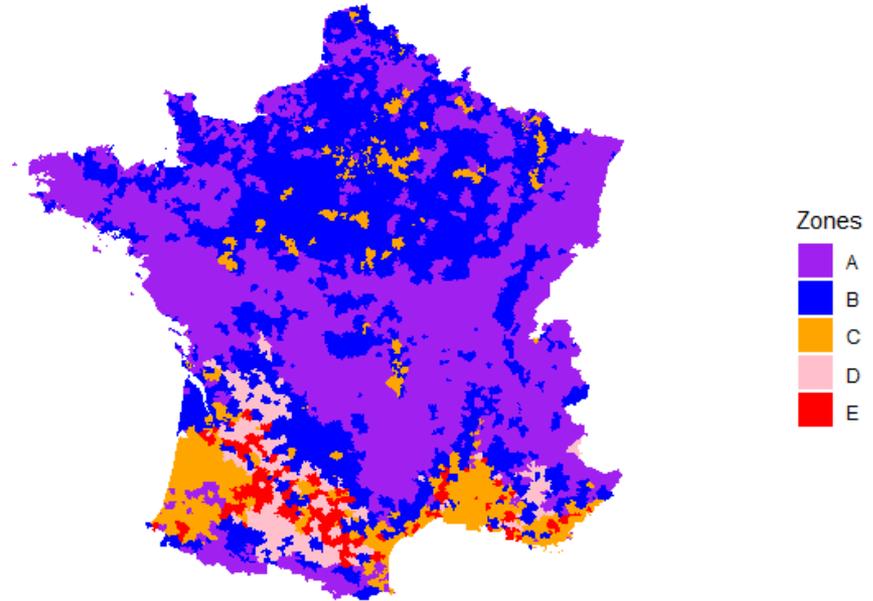


Figure 22 : Zonier RIT

- Le zonier RI est fortement influencé par la carte_rga.
- L'ajout de la carte_temp dans le zonier RIT réduit cette influence.

- Cas pratique : zonier simplifié pour le risque sécheresse

Synthèse des modèles

- Le pouvoir discriminant/segmentant des modèles est mesuré par l'indice de Gini.
- Modèle 1C (**0,275**) < Modèle RI (**0,319**) < Modèle RIT (**0,327**).

Le modèle RIT offre une modélisation **plus précise** du risque de sécheresse mais surtout **la meilleure segmentation**, ce qui en fait le modèle le plus performant de cette étude et celui qui est conservé.

- Conclusion

1. Outil algorithmique : conception

2. Cas pratique : zonier simplifié pour le risque sécheresse

3. Conclusion

EXPLOITATION D'IMAGES DE CARTES DANS LA MODÉLISATION DE LA SINISTRALITÉ EN ASSURANCE HABITATION

- Conclusion

Conclusion

- Conclusion

Conclusion



Objectif du mémoire : Concevoir un outil algorithmique pour transformer des images de cartes en base de données.

- Conclusion

Conclusion



Objectif du mémoire : Concevoir un outil algorithmique pour transformer des images de cartes en base de données.



4 étapes de fonctionnement : le choix de la carte, le prétraitement, le redimensionnement et la constitution de la base finale.

- Conclusion

Conclusion



Objectif du mémoire : Concevoir un outil algorithmique pour transformer des images de cartes en base de données.



4 étapes de fonctionnement : le choix de la carte, le prétraitement, le redimensionnement et la constitution de la base finale.



L'outil présente des **perspectives intéressantes** pour contribuer à la modélisation de divers risques naturels.

MERCI POUR VOTRE ATTENTION !

Exploitation d'images de cartes dans la modélisation de la sinistralité en assurance habitation

Yapi KOUASSI