

**Mémoire présenté pour la validation de la Formation  
« Certificat d'Expertise Actuarielle »  
de l'Institut du Risk Management  
et l'admission à l'Institut des actuaires  
le 24 mars 2023**

Par : *Pauline ROY*

Titre : *Assurance habitation et garantie sécheresse : contributions à la prévention et au provisionnement par l'utilisation de données géographiques*

Confidentialité :  NON  OUI (Durée :  1an  2 ans)  
*Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus*

*Membres présents du jury de l'Institut des actuaires :*

---

---

---

Entreprise : COVEA – MMA

Nom : Ewen WILCZYK

Signature et Cachet :



Directeur de mémoire en entreprise :

Nom : Ewen WILCZYK

Signature :



*Membres présents du jury de l'Institut du Risk Management :*

---

---

---

---

---

---

---

---

Invité :

Nom : \_\_\_\_\_

Signature :

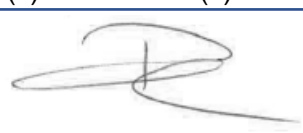
**Autorisation de publication et de mise en ligne sur un site de diffusion de documents actuariels**

*(après expiration de l'éventuel délai de confidentialité)*

Signature du responsable entreprise



Signature(s) du candidat(s)



Secrétariat :

Bibliothèque :



## Résumé

Les assureurs sont confrontés à des risques climatiques de plus en plus fréquents et difficiles à prévoir dans le domaine de l'assurance habitation. Ces événements contribuent de manière significative aux pertes subies par les assureurs. Parmi les risques liés aux catastrophes naturelles, la sécheresse commence à peser de plus en plus sur les sinistres observés.

La sinistralité due à la sécheresse présente des spécificités qui ajoutent d'autres aléas à la survenance même du sinistre. Comme tout sinistre lié aux catastrophes naturelles, l'intervention de l'assureur dépend de la reconnaissance de l'événement comme catastrophe naturelle, soumise à des règles administratives changeantes. En parallèle, les dommages causés par la sécheresse ne sont identifiables qu'après un délai de plusieurs années. Les facteurs temps et réglementation complexifient donc la maîtrise de la sinistralité sécheresse.

Afin de permettre aux assureurs de se prémunir contre ce risque croissant, ce mémoire cherche à améliorer la connaissance de la sécheresse en s'intéressant tout particulièrement à l'identification des phénomènes climatiques à l'origine des sinistres et à la compréhension des spécificités réglementaires régissant la prise en charge. Deux applications basées sur des analyses géographiques viennent compléter cet exposé afin d'améliorer la connaissance territoriale du risque par le biais d'un zonier et d'affiner la connaissance du portefeuille sinistré sur l'exercice le plus récent.

**Mots-clés :** Sécheresse, RGA (Retrait Gonflement des Argiles), SWI (Soil Wetness Index), changements climatique, arrêtés de catastrophes naturelles, analyse géographique, zonier, prévention, régression logistique, modèle d'apprentissage supervisé, provisionnement

## Abstract

Insurers are facing increasingly frequent and difficult-to-predict climate events in the field of home insurance. These events significantly contribute to the losses suffered by insurers. Among the risks associated with natural disasters, drought is starting to weigh more and more on the observed claims.

The claims resulting from drought have specificities that add further uncertainty to the occurrence of the claim itself. Like any claim related to natural disasters, the insurer's intervention depends on the recognition of the event as a natural disaster, subject to changing administrative rules. In parallel, damages caused by drought are only identifiable after a delay of several years. Therefore, time and regulatory factors complicate the management of drought-related claims.

In order to enable insurers to protect themselves against this growing risk, this presentation seeks to improve the understanding of drought, with a particular focus on the identification of climate phenomena that cause claims and the comprehension of the regulatory specificities governing coverage. Two applications based on geographical analysis complete this study to enhance the territorial understanding of the risk through geographic zoning and to refine the knowledge of the claims portfolio in the most recent period.

**Key-words** : Drought, shrinkage and swelling of clay, SWI (Soil Wetness Index), climate change, decrees of natural disasters, geographical analysis, geographic zoning, prevention, logistic regression, supervised learning model, provisioning

## Remerciements

Ce mémoire n'aurait pu voir le jour sans la contribution et le soutien d'un certain nombre de personnes.

En tout premier lieu merci à l'entreprise COVEA de m'avoir donnée la possibilité d'intégrer la formation du CEA.

Je tiens également à remercier l'ensemble de mes collègues de la direction IARD pour leur soutien au quotidien. Un merci tout particulier à mon manager Ewen WILCZYK pour m'avoir suivi tout au long de ce mémoire, pour les échanges instructifs et les conseils apportés.

Enfin un grand merci au Pôle Risques Majeurs et tout particulièrement à Romain MARTEAU pour m'avoir fait découvrir le risque sécheresse et pour sa grande aide dans la fourniture et l'utilisation des données spécifiques à ce phénomène.



## Introduction

Dans le cadre de l'assurance habitation, les assureurs sont tenus d'intervenir auprès de leurs assurés lors de la survenance de risques standards mais également de risques climatiques. Ces différents événements difficilement prévisibles prennent une part de plus en plus conséquente dans la sinistralité des assureurs et sont malheureusement de plus en plus fréquent. Pour continuer à maintenir une rentabilité, l'ACPR préconise une hausse de prime de 130% à 200% d'ici 30 ans sur l'ensemble des contrats habitation.

Parmi l'ensemble des risques dits de catastrophes naturelles, la sécheresse commence à peser de plus en plus dans la sinistralité observée. L'ensemble des acteurs ont révisé leurs estimations à horizon 2050 affichant des constats bien plus qu'alarmant concernant la charge inhérente à ce risque. France Assureurs<sup>1</sup> a par exemple révisé largement à la hausse sa projection en 10 ans passant d'une charge estimée à 21 Mrd€ sur la période 2014-2040 à 43Mrd€ sur la période 2020-2050. En se ramenant à un cout moyen annuel, la dernière estimation est désormais 1,8 fois plus importante que celle faite quelques années auparavant.

La sinistralité sécheresse présente des spécificités qui la font différer des autres sinistres classiques venant ajouter de l'aléa complémentaire à la simple survenance d'un sinistre. En effet, un sinistre sécheresse n'est identifiable qu'après un différé de quelques années ce qui correspond au temps nécessaire pour que les stigmates de la sécheresse soient visibles sur le bâtiment. Un facteur temps vient donc s'ajouter à l'aléa de la sinistralité. Puis comme tous sinistres de type catastrophes naturelles, l'intervention de l'assureur va dépendre de la reconnaissance de l'événement comme catastrophes naturelles par un arrêté reposant sur des règles administratives assez mouvantes. Un facteur réglementaire est à ajouter là encore à l'aléa de survenance du sinistre.

Pour faire face à l'accroissement certain de ce risque, le gouvernement réagit également en proposant de nouveaux textes de lois visant à protéger au mieux l'assuré. Le PPL<sup>2</sup> Baudu en est le dernier exemple en date, pour autant de nombreux points restent à éclaircir notamment sur la taxe obligatoire visant à alimenter le fonds de catastrophes naturelles. Cette taxe aujourd'hui à 12% sur les contrats habitation n'est pas suffisante pour faire face aux engagements, le fonds étant déficitaire. Un accroissement de ce taux n'est pas exclu mais engendrerait un fort impact du côté des assureurs.

Malgré une exposition stabilisée du portefeuille habitation, les différents changements à venir et notamment l'impact du réchauffement climatique ont un impact non négligeable sur le risque assuré. Il est donc essentiel pour les assureurs de pouvoir se prémunir face à ce risque en améliorant leur connaissance de la sécheresse.

L'élément primordial permettant de maîtriser ce risque est déjà d'identifier le phénomène climatique à l'origine de la sinistralité. Ce mémoire traitant d'un risque de catastrophes naturelles, la compréhension des spécificités réglementaires qui régissent la prise en charge est également un point clé. Une fois cette étape effectuée, deux applications basées sur ces informations permettront

---

<sup>1</sup> Impact du changement climatique sur l'assurance à l'horizon 2040 et 2050

<sup>2</sup> PPL : Proposition de loi

d'affiner la connaissance du risque et mieux maîtriser l'exposition du portefeuille face à la sinistralité sécheresse.

Une première étude permettra d'affiner la maîtrise géographique du risque par la définition d'une cartographie des communes sous la forme d'un zonier. Ce dernier doit permettre d'identifier de manière exhaustive l'ensemble des communes qui peuvent potentiellement subir des sinistres liés à la sécheresse. Cet outil permettra à la fois une meilleure connaissance de l'exposition du territoire et donc du risque couvert par l'assurance, mais également de réfléchir dans les zones les plus à risques aux moyens de prévention à proposer aux assurés afin de limiter les impacts et l'occurrence d'un sinistre sécheresse. Sa construction s'appuiera sur un large panel de données exogènes et permettra de comparer diverses techniques de modélisation.

Une seconde analyse s'intéressera à affiner la connaissance du portefeuille sinistré sur le dernier exercice. Là encore la résolution de ce besoin se fera au travers d'une approche géographique afin de prédire les risques impactés par la sécheresse avant même la déclaration de ces sinistres.

Ces deux applications seront illustrées sur le portefeuille de la branche habitation de la marque MMA et permettront d'accentuer la prise de conscience sur l'impact de l'exposition face au risque sécheresse.



## Table des matières

Chapitre 1 – Risque sécheresse en assurance habitation.....	11
1.1. Sinistralité sécheresse, origine et impact chez les assureurs .....	11
1.1.1. Sinistres catastrophes naturelles .....	11
1.1.2. Origine du risque sécheresse.....	12
1.1.3. Impact de la sécheresse dans la sinistralité des assureurs.....	13
1.1.4. Sécheresse, risque croissant du portefeuille de l'assureur.....	14
1.2. Evolution de l'exposition du portefeuille MMA face à la sécheresse.....	15
1.2.1. Evolution interne de la structure du portefeuille MMA .....	16
1.2.2. Impact de la migration de la population .....	18
1.2.3. Impact de l'évolution face au climat.....	24
Chapitre 2 – Particularités de la sinistralité sécheresse.....	29
2.1. Spécificités du risque sécheresse.....	29
2.1.1. Saisonnalité et localisation de la sinistralité sécheresse .....	29
2.1.2. Prise en charge de la sinistralité sécheresse .....	31
2.1.3. Réglementation mouvante des arrêtés de catastrophe naturelle pour la sécheresse .....	32
2.2. Description des données utilisées pour définir la sinistralité sécheresse.....	33
2.2.1. Conformité de la maille géographique.....	33
2.2.2. Données expliquant le facteur de prédisposition.....	34
2.2.3. Données expliquant le facteur déclenchant .....	38
2.2.4. Données à l'origine des facteurs aggravants .....	41
2.2.5. Données internes.....	42
2.2.6. Particularité des données externes.....	44
Chapitre 3 – Application au zonier .....	47
3.1. Préambule à la modélisation.....	47
3.1.1. Choix de la variable à modéliser.....	47
3.1.2. Sélection des variables explicatives.....	48
3.1.3. Analyse et retraitement des variables explicatives.....	48
3.1.4. Base de modélisation – base de test .....	50
3.1.5. Choix des indicateurs de performance .....	52
3.2. Modèle classique - Régression logistique.....	53
3.2.1. Théorie du modèle.....	53
3.2.2. Choix des variables explicatives.....	55
3.2.3. Sélection des variables explicatives.....	57
3.2.4. Calage des coefficients et effets des variables .....	57
3.2.5. Utilisation de la validation croisée .....	59
3.2.6. Performance du modèle .....	60

3.3. Modèle d'apprentissage supervisé – Forêt aléatoire .....	60
3.3.1. Arbre seul.....	61
3.3.2. Forêt aléatoire par bagging.....	62
3.3.3. Forêt aléatoire par gradient boosting .....	63
3.3.4. Performance des modèles .....	64
3.4. Comparaison des modèles.....	65
3.4.1. Performance sur la base de test .....	65
3.4.2. Performance sur la base complète .....	66
3.5. Comparaison du zonage.....	67
3.5.1. Classification par la méthode de Jenks.....	67
3.5.2. Comparaison avec le zonier actuel.....	68
3.5.3. Comparaison avec la carte argile .....	70
3.5.4. Illustration du nouveau zonier face à l'exposition.....	71
Chapitre 4 – Application au provisionnement.....	73
4.1. Process actuel.....	73
4.1.1. Utilisation des triangles de liquidation .....	74
4.1.2. Spécificité de la sécheresse .....	75
4.1.3. Choix d'une méthode d'estimation géographique.....	77
4.2. Identification des communes à risque.....	79
4.2.1. Application des règles des arrêtés de catastrophe naturelle.....	79
4.2.2. Modélisation des communes à risque .....	82
4.3. Identification du portefeuille à risque.....	85
4.3.1. Périmètre produits sélectionnés .....	85
4.3.2. Définition du critère de répartition .....	86
4.4. Comparaison avec le process actuel .....	87
4.4.1. Charge économique à partir des triangles de liquidation .....	88
4.4.2. Charge économique à partir de la modélisation .....	89
4.4.3. Pertinence de la modélisation .....	89

## Chapitre 1 – Risque sécheresse en assurance habitation

L'assurance habitation propose aux assurés une couverture plus ou moins complète pour assurer leurs biens qu'ils soient immobiliers ou mobiliers. Pour autant, quel que soit le niveau de couverture souhaitée, la sinistralité des catastrophes naturelles est systématiquement incluse et ce afin de fournir aux assurés une assurance en cas de sinistres majeurs pouvant aller jusqu'à la disparition complète du risque.

Peu fréquent, les risques de catastrophes naturelles sont pourtant une source forte de sinistralité car ils sont la plupart du temps associés à des phénomènes d'ampleur. Parmi ces événements on retrouve la sécheresse. Ce risque a souvent été minoré par rapport à d'autres catastrophes notamment l'inondation et les tempêtes. Mais désormais le poids de la sécheresse prend de plus en plus de place et certaines prévisions les estiment même tous les trois à part égale dans les années à venir.

Afin d'analyser au mieux le risque sécheresse dans les études présentées par la suite, il est primordial de s'intéresser avant tout à l'origine de ce phénomène.

### 1.1. Sinistralité sécheresse, origine et impact chez les assureurs

Au même titre que les inondations, les séismes, les glissements de terrains, les coulées de boue, les éruptions volcaniques, les raz-de-marée et les avalanches ; la sécheresse fait partie des événements dits de **catastrophes naturelles**. Non couvert de manière obligatoire dans la plupart des périmètres d'assurance, seule l'assurance habitation fait exception à cette règle en se devant d'inclure la sinistralité des catastrophes naturelles dans la couverture de base proposée aux assurés.

#### 1.1.1. Sinistres catastrophes naturelles

Une **catastrophe naturelle** est par définition liée à un phénomène imprévisible engendrant des événements de faibles fréquences d'occurrence pour un coût sinistre parfois conséquent. Cette particularité induit une difficulté d'estimation de ces risques que ce soit en terme de survenance mais également en terme de calibrage de la charge.

Les sinistres catastrophes naturelles sont d'ailleurs explicitement définis dans l'article L125-1 du Code des Assurances « Sont considérés comme les effets des catastrophes naturelles, au sens du présent chapitre, les dommages matériels directs non assurables ayant eu pour cause déterminante l'intensité anormale d'un agent naturel, lorsque les mesures habituelles à prendre pour prévenir ces dommages n'ont pu empêcher leur survenance ou n'ont pu être prises. ».

Ce type de sinistre se distingue également lors de son indemnisation. En effet, contrairement aux sinistres classiques enregistrés sur le périmètre dommages aux biens, la déclaration seule d'un assuré couvert par le risque ne suffit pas à déclencher le remboursement des dégâts par l'assureur. L'indemnisation par la compagnie d'assurance est régie par une commission interministérielle qui contrôle la prise en charge des sinistres via l'obtention d'**arrêté de catastrophe naturelle**. Cette commission regroupe le ministère de l'Intérieur, le ministère de l'Economie et des

Finances, le ministère de l'Ecologie et du Développement Durable ainsi que la Caisse Centrale de Réassurance (CCR).

La figure 1.1 ci-dessous décrit la procédure actant ou non de la prise en charge du sinistre sécheresse par la compagnie d'assurance.

Figure 1.1 : Extrait de la réglementation concernant les arrêtés de catastrophe naturelle

#### Arrêté de catastrophe naturelle publié

Même si vous êtes assuré contre les catastrophes naturelles, cela ne suffit pas pour obtenir l'indemnisation de votre sinistre par l'assurance.

Il faut en plus qu'un arrêté interministériel de catastrophe naturelle ait été adopté et publié par le gouvernement au *Journal officiel*. L'arrêté précise les zones touchées par la catastrophe, les périodes auxquelles les faits se sont produits et la nature des dommages causés par la catastrophe naturelle.

Les maires des communes touchées par la catastrophe naturelle demandent à la préfecture que leur communes soient visées par l'arrêté de catastrophe naturelle.

Vous devez donc faire une déclaration de sinistre à votre assureur le plus rapidement possible et envoyer une copie à la mairie. Cela permettra au maire de votre commune de demander à la préfecture le classement en zone de catastrophe naturelle.

Source : Site [service-public.fr](http://www.service-public.fr)<sup>3</sup>

### 1.1.2. Origine du risque sécheresse

Contrairement à la majorité des sinistres couverts dans l'assurance habitation, la sécheresse est un événement qui dépend majoritairement de **phénomènes climatiques et géologiques** venant altérer le volume du sol par un phénomène de retrait et gonflement à l'origine des dégâts matériels.

Cette instabilité du sol n'est pas homogène sur l'ensemble du territoire et dépend de certains facteurs de risques.

- Premièrement le **facteur géologique** est une condition primordiale à l'origine de la sécheresse. Les sols argileux étant plus exposés aux changements volumiques, la présence d'argile en quantité dans le sol est le critère de prédisposition à la sinistralité sécheresse. Sa structure en couches de minéraux superposés confère à l'argile une forte plasticité, cette typologie est donc fortement sensible à la dilatation et à la rétractation selon l'humidité environnante.
- Deuxièmement le **facteur climatique** va lui jouer le rôle de critère déclencheur. Pour modifier la densité du sol argileux il est nécessaire d'observer des variations conséquentes d'humidité dans le sol qui causeront son retrait et son gonflement. Les variations de pluviométries seront donc à l'origine de la survenance du sinistre.

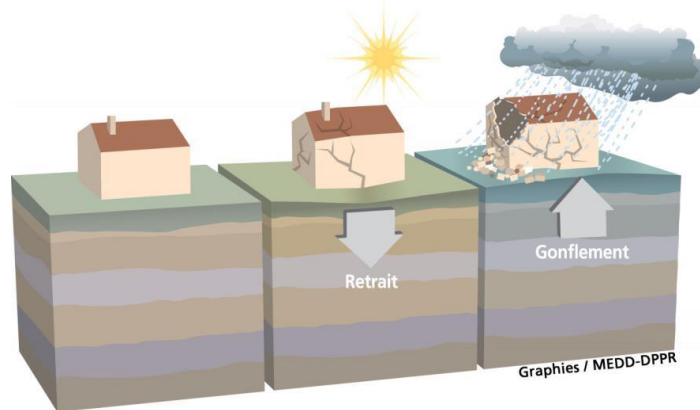
---

<sup>3</sup> <https://www.service-public.fr/particuliers/vosdroits/F3076>

- Dernièrement on peut citer quelques **facteurs aggravants** qui peuvent intervenir en complément des deux premiers cités. Par exemple la mauvaise qualité des fondations peut aggraver le phénomène de même que l'occupation des sols alentours par d'importants végétaux ou par des dispositifs fragilisant le sol.

Ce phénomène est illustré en figure 1.2 où l'on observe bien la rétractation du sol lors de période de sécheresse suite à l'évaporation de l'eau contenue dans la couche argileuse. A contrario, lors de forte période pluvieuse, l'argile absorbe en grande proportion l'eau et sa dilatation engendre un gonflement du sol.

Figure 1.2 : Phénomène de retrait et gonflement du sol



La succession de ces modifications sur le volume du sol est à l'origine de la fragilisation des fondations des bâtiments qui entraîneront l'apparition de dommages majoritairement visibles sous forme de fissures.

Ces dommages impactant la structure même de l'habitation, la sinistralité résultante est essentiellement concentrée sur la partie immobilière du risque assuré sauf cas extrême d'effondrement du bâtiment.

### 1.1.3. Impact de la sécheresse dans la sinistralité des assureurs

Outre le fait d'appartenir à la catégorie des sinistres événements naturels impliquant une gestion spécifique de l'indemnisation du sinistre, la sécheresse se distingue également au moment de la déclaration du sinistre.

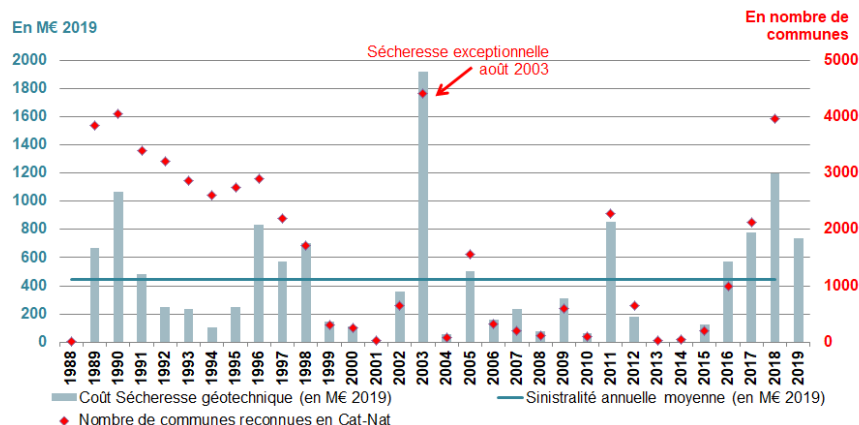
Contrairement à la majorité des sinistres en assurance habitation, la déclaration du sinistre ne peut se faire au moment de la survenance réelle du sinistre car la constatation des dommages nécessite un **effet différé** à moyen voire long terme qui se chiffre en année. C'est en effet malheureusement le facteur temps qui va rendre visible les dommages causés par la sécheresse sur les bâtiments.

Cet élément est handicapant pour les assurés mais également pour la compagnie d'assurance. La déclaration sinistre par l'assuré est complexifiée car pour obtenir la prise en charge il doit mettre en lien son sinistre avec un arrêté de catastrophe naturelle en moyenne 2 ans après l'épisode de sécheresse enregistré.

### 1.1.4. Sécheresse, risque croissant du portefeuille de l'assureur

Là où depuis les années 2000 la sécheresse était synonyme d'événement d'amplitude rare avec quelques grandes sécheresses historiques, on observe désormais une **amplification** de l'aléa depuis 2016 comme l'illustre la figure 1.3.

Figure 1.3 : Évolution de la sinistralité liée à la sécheresse et à la réhydratation des sols

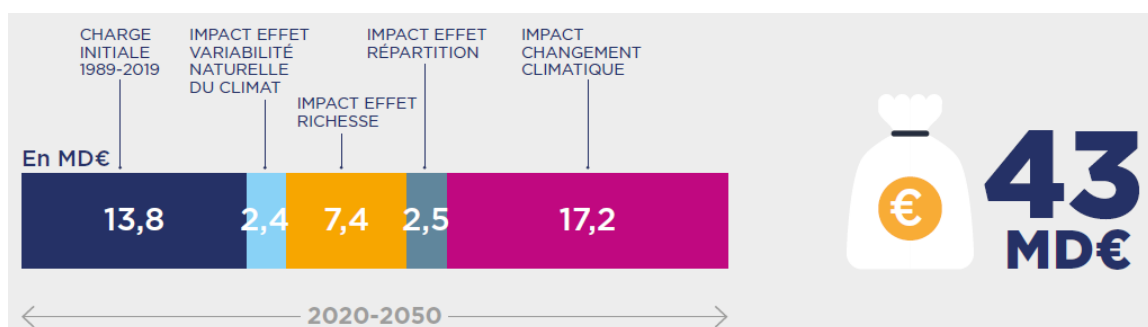


Source : CCR 2020, Site notre-environnement.gouv.fr<sup>4</sup>

Pour ne citer que les périodes les plus touchées, la sécheresse a fortement impacté les années 1989-1990, 2003 ainsi que 2011. Plus récemment, c'est l'année 2018 qui est apparue comme sécheresse exceptionnelle malgré le contexte haussier de ce phénomène. C'est d'ailleurs suite à ces différents épisodes de sécheresses historiques que des évolutions ont pu être menées par l'état. Par exemple suite à l'épisode de sécheresse de 1989-1990 le risque sécheresse a été intégré au système d'arrêté de catastrophe naturelle déjà existant pour les autres événements.

Par ailleurs, l'impact du **changement climatique** sur les risques de catastrophes naturelles est au cœur de nombreuses études actuelles. Dans la dernière étude publiée par France Assureurs<sup>5</sup> en octobre 2021, ils estiment à 43 milliards d'euros la charge sécheresse cumulée sur la période 2020-2050 soit 3 fois plus que la charge observée entre 1989-2019. Comme illustré sur la figure 1.4, 40% de cette hausse du coût est liée au changement climatique.

Figure 1.4 : Distinction des effets dans l'estimation de la charge sécheresse à horizon 2050



Source : France Assureurs 2021

<sup>4</sup> <https://notre-environnement.gouv.fr/donnees-et-ressources/ressources/graphiques/article/graphique-27>

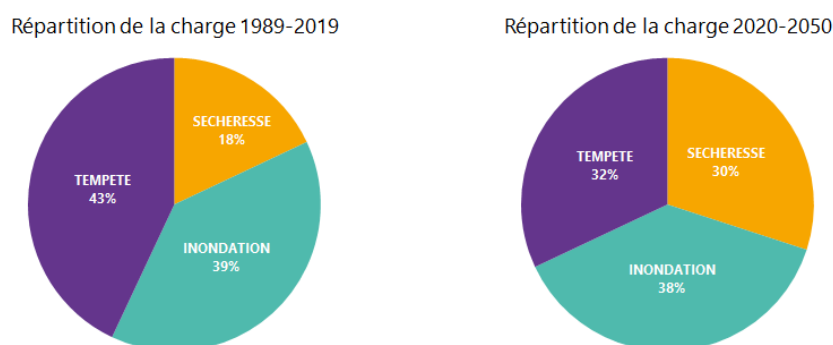
<sup>5</sup> Impact du changement climatique sur l'assurance à l'horizon 2050, France Assureurs, Octobre 2021

Au-delà du simple effet risque, les **facteurs sociaux-économiques** ont aussi leurs impacts sur la charge sinistre à venir. La hausse du niveau de vie moyen traduite par l'impact effet richesse serait à l'origine de 17% de l'augmentation du coût. Les migrations de populations avec la modification des parts de logement sont également estimées avec une tendance haussière de 6% à travers l'impact effet répartition.

L'étude étant faite sur 3 périls : l'inondation, la tempête et la sécheresse ; c'est le risque sécheresse qui se démarque le plus en affichant une évolution de la charge +215% sur la période 2020-2050 par rapport à la période 1989-2019.

En plus de la forte hausse du coût sécheresse sur les 30 années à venir, ce risque prendrait de plus en plus d'ampleur parmi les risques climatiques. La figure 1.5 illustre la distorsion de cette répartition entre les sinistres observées jusqu'à 2019 et à la charge estimée à horizon 2050.

Figure 1.5 : Evolution de la répartition de la charge climatique entre 1989-2019 et 2020-2050



Source : France Assureurs 2021

Ces différents éléments militent donc pour s'intéresser davantage au risque sécheresse afin d'accroître la vigilance que l'on a sur ce périmètre afin d'améliorer la connaissance et la maîtrise de ce risque.

## 1.2. Evolution de l'exposition du portefeuille MMA face à la sécheresse

La marque MMA propose des solutions d'assurances depuis près de 200 ans pour couvrir les biens particuliers mais aussi professionnels de ces clients. Le périmètre habitation est donc un marché largement développé chez MMA depuis de nombreuses années avec un portefeuille dont la structure est stabilisée.

Au cours du temps l'exposition d'une compagnie évolue de par des changements structurels liés à l'implantation de ses points de ventes, l'installation de concurrents ou suite à des effets externes modifiant le parc assurantiel comme la migration de population ou même des évolutions réglementaires. Avant toute analyse on a donc souhaité identifier l'exposition de la marque MMA face à cette sinistralité en comparaison avec le marché immobilier.

Au-delà même du risque couvert, le portefeuille d'un assureur peut être source d'une exposition plus ou moins accrue. Quand on parle de risque sécheresse on s'intéresse principalement à la cible des propriétaires de maisons qu'ils soient occupants ou non occupants. Les locataires n'étant pas

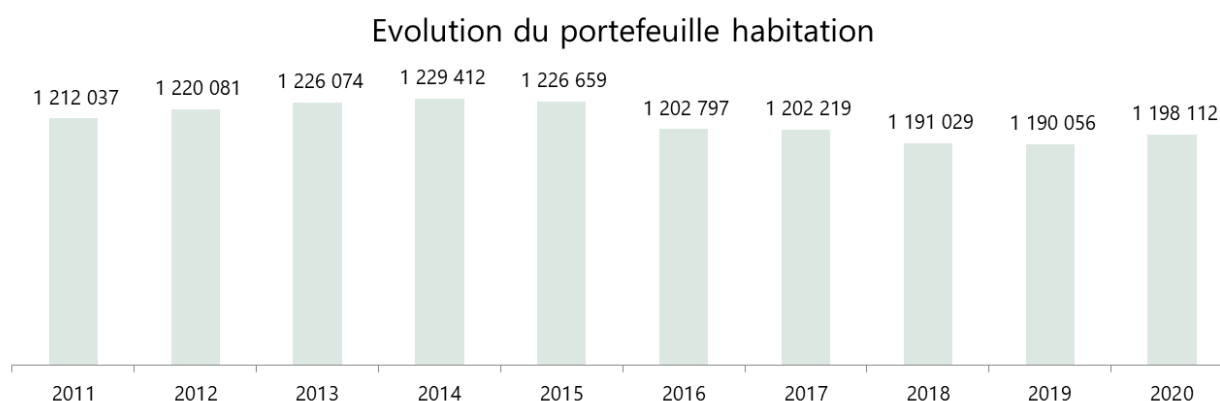
responsable des biens aux bâtiments ils sont exclus du périmètre analysé. De même, la sinistralité appartements est très rarement prise en charge directement par le propriétaire puisque les dommages aux bâtiments sont normalement à la charge du syndic ou de la copropriété.

Les propriétaires de maisons étant une cible assez stable : les contrats sont conservés 8 ans en moyenne chez MMA sans modification du risque ; le principal risque porté par la compagnie sur le risque sécheresse est lié à l'exposition de son portefeuille bien plus qu'à sa production.

### 1.2.1. Evolution interne de la structure du portefeuille MMA

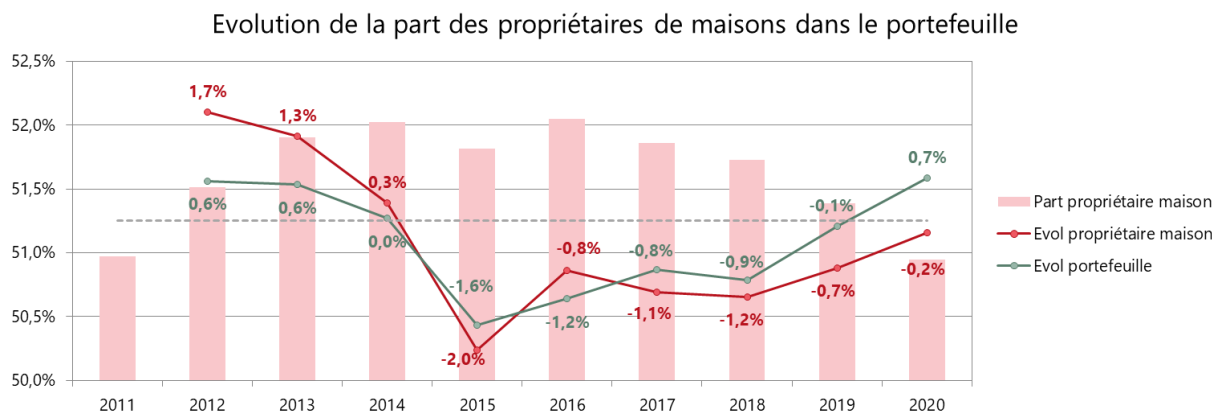
Sur une période de 10 ans le portefeuille habitation a subi une forte décroissance liée à des tensions sur le marché ayant affaibli la production ainsi qu'à une stratégie marque orientée sur le développement des produits professionnels. Une nouvelle dynamique sur le particulier a permis dernièrement de retrouver une légère croissance du portefeuille. La figure 1.6 représente son évolution sur la période 2011 – 2020 dont le nombre de contrats a diminué de 2.7% sur la période.

Figure 1.6 : Evolution du portefeuille habitation sur 10 ans



Pour autant la décroissance du portefeuille ne s'est pas faite de manière complètement homogène selon le profil de risque. En se focalisant sur les propriétaires de maison, la baisse du portefeuille n'a pas engendré une évolution identique. La figure 1.7 présente l'évolution de la répartition de cette cible parmi l'ensemble des contrats.

Figure 1.7 : Impact de l'évolution du portefeuille sur les propriétaires de maisons

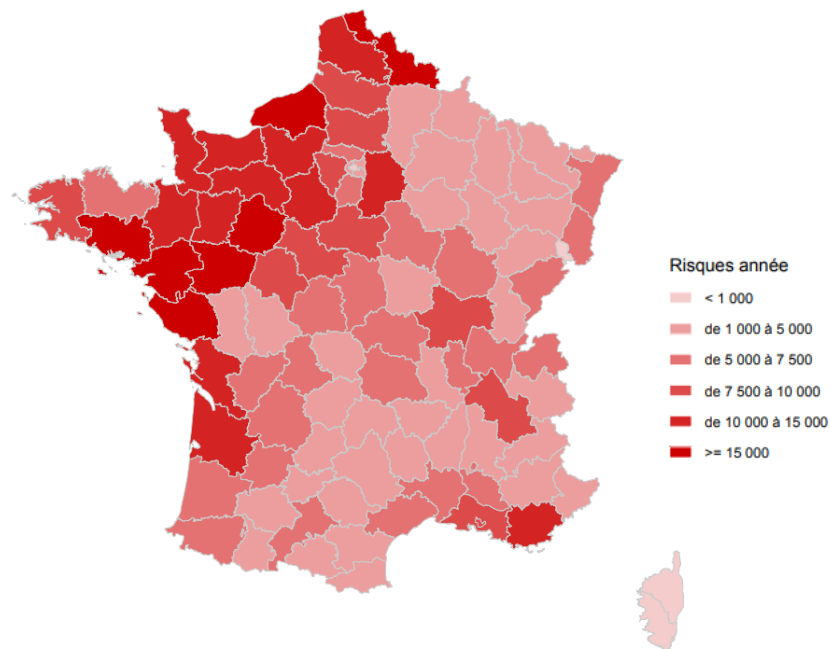




Les évolutions observées entre le portefeuille global et le portefeuille à risque illustrent bien la distorsion du portefeuille face à la production. En effet, on observe fréquemment en assurance habitation un turn-over plus important sur les appartements que sur les maisons. En cas de tensions marchés c'est généralement la part de marché sur les appartements qui est la plus volatile. Ce phénomène est d'ailleurs visible en comparant quelques chiffres : sur l'année 2020 58% de contrats appartement ont été enregistrés en production pourtant en portefeuille leur part se limite à 49%. Le changement de stratégie opéré en 2018 pour relancer la production a effectivement engendré une croissance visible plus rapidement sur les appartements que sur les maisons. Ces différents éléments ont permis de revenir à une répartition quasi équilibrée entre appartement et maison au même niveau que 2011.

En poussant encore plus finement l'analyse, la disparité de l'exposition est bien sûr visible à la maille géographique comme le montre la figure 1.8. L'exposition initiale du portefeuille des propriétaires de maisons présente une couverture du territoire éparse.

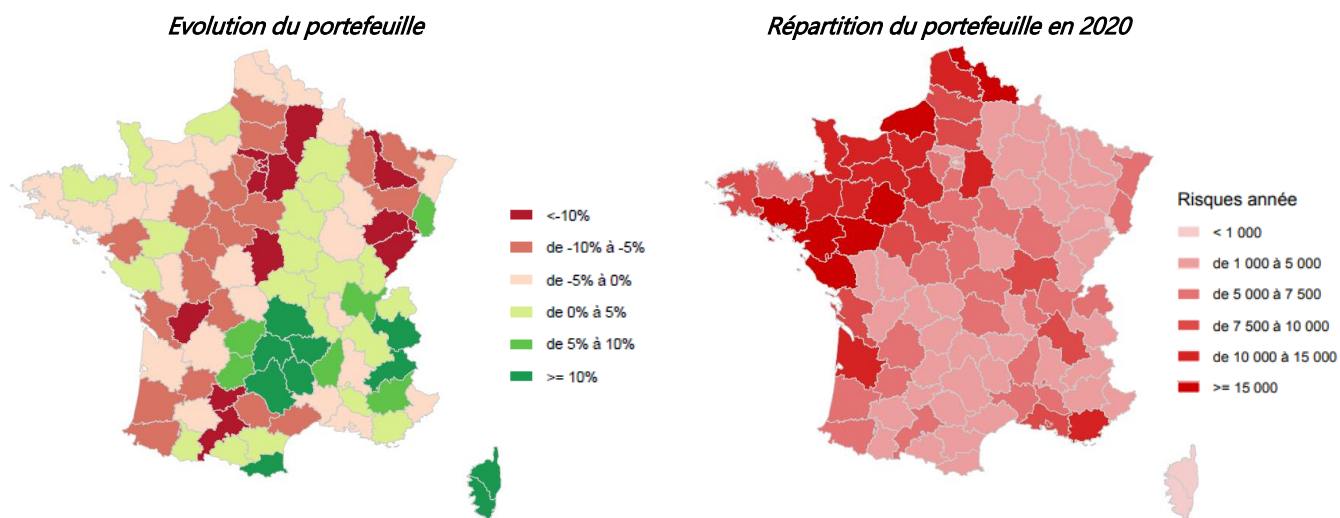
Figure 1.8 : Répartition du portefeuille propriétaire de maisons en 2011



Du fait de l'implantation historique de la marque et d'un parc immobilier hétérogène France entière, le Nord-Ouest de la France concentre une grande partie du portefeuille MMA. Cette spécificité fait légèrement différer la répartition de la marque par rapport à celle de la population française.

La modification du portefeuille sur les 10 années suivantes n'a par ailleurs pas été identique sur le territoire. Comme illustré sur la figure 1.9, l'Est et le Nord de la France ont subis de plus forte pertes de contrats tandis que le centre et le Sud-Est se sont développés. Malgré ces changements, la structure initiale étant tellement hétérogène département par département, la répartition finale du portefeuille n'a que très peu été impactée et les zones d'exposition fortes ont été conservées.

Figure 1.9 : Evolution par département du portefeuille propriétaire de maisons entre 2011 et 2020



Les produits commercialisés étant ancrés depuis de nombreuses années dans le milieu de l'assurance il n'y a que peu de distorsion du portefeuille historique au niveau France. L'exposition n'est donc pas fortement volatile année après année.

### 1.2.2. Impact de la migration de la population

Au-delà des effets internes de commercialisation d'une marque, nouvelle implantation d'agence, nouvelle stratégie de la marque ... les **effets externes** sont également à prendre en compte dans la distorsion d'une exposition. En effet l'implantation de la population est mouvante ce qui engendre forcément un impact sur le parc immobilier et donc sur la matière assurable.

Parmi les études récurrentes de l'INSEE se trouve l'analyse de la population ainsi que son évolution. Dans le cas de notre analyse nous nous référerons à la publication de novembre 2016 présentant les projections de population 2013-2070 pour la France<sup>6</sup> ainsi qu'à son adaptation département par département à horizon 2050<sup>7</sup>.

Différents scénarii ont été étudiés en s'appuyant sur l'évolution de 3 critères majeurs à l'origine de l'évolution de la population : la fécondité, la mortalité ainsi que les flux migratoires.

- Les **naissances** sont estimées en s'appuyant sur le taux de fécondité projeté sur la population des femmes fécondes (15 à 50 ans) par tranche d'âge. Le nombre global de naissance tient également compte des naissances affectées aux femmes entrant dans le territoire. Les naissances sont ensuite réparties par genre en respectant le poids de 51% en faveur du sexe masculin.
- L'estimation des **décès** s'appuie sur la projection des quotients de mortalité par âge et par sexe. Là encore les décès tiennent bien compte du solde migratoire. Y est également intégré une proportion de décès sur les nouveaux nés.

<sup>6</sup> Projections de population à l'horizon 2070 – Projections de population 2013-2070 pour la France | Insee

<sup>7</sup> Projections de population 2013-2050 pour les départements et les régions | Insee

- Le **solde migratoire** correspond à la différence entre les entrées et les sorties du territoire projetées par sexe et par âge.

Les projections de populations sont réalisées en s'appuyant sur la méthode des composantes c'est-à-dire que l'effet de chaque critère présenté ci-dessus est calculé année après année, par genre, et pour tous les âges. La population pour l'année suivante s'appuie sur la population de l'année précédente modulée des naissances, des décès et de la migration sur cette même année. L'équation suivante illustre mathématiquement l'estimation de la population pour l'année  $N + 1$  :

$$Population_{N+1} = Population_N + Naissance_N - Décès_N + Solde migratoire_N.$$

Sur chacun des critères intégrés dans l'estimation de la population différentes hypothèses sont définies en respectant une vision centrale et par modulation à la hausse et à la baisse des effets. Le tableau exposé en figure 1.10 présente les valeurs qui ont été retenues pour chaque élément selon les 3 variantes étudiées.

Figure 1.10 : Indicateurs associés aux différentes hypothèses

	Situation au 1 <sup>er</sup> janvier 2013 (dernier recensement)	Hypothèse centrale	Variante basse	Variante haute	
Indice conjoncturel de fécondité	1,99 enfant par femme	1,95 enfant par femme	1,8 à partir de 2020	2,1 à partir de 2020	Naissances
Âge moyen à la maternité	30,2 ans	32,0 ans à partir de 2040	32,0 ans à partir de 2040	32,0 ans à partir de 2040	
Espérance de vie à la naissance des femmes	85,0 ans	93,0 ans en 2070	90,0 ans en 2070	96,0 ans en 2070	Décès
Espérance de vie à la naissance des hommes	78,7 ans	90,1 ans en 2070	87,1 ans en 2070	93,1 ans en 2070	
Valeur du solde migratoire	+ 72 000 personnes entre 2012 et 2013	+ 70 000 personnes par an	+ 20 000 personnes par an à partir de 2020	+ 120 000 personnes par an à partir de 2020	Solde migratoire

Source : Insee, projections de population 2013-2070

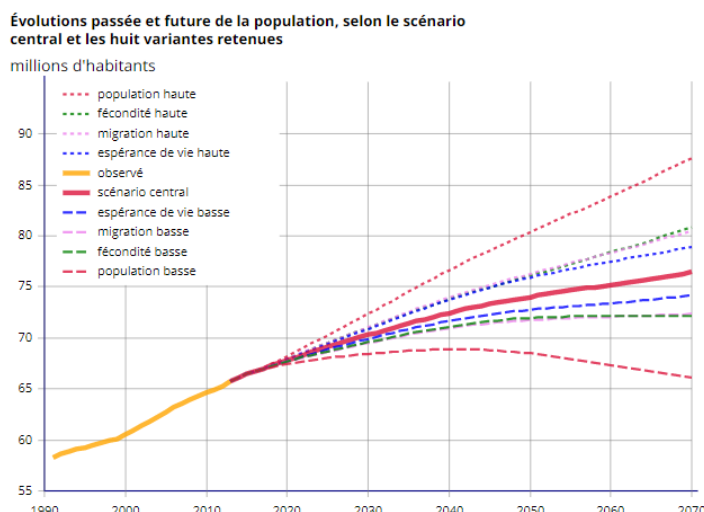
La combinaison des différentes hypothèses pour chacun des critères étudiés permet de définir de nombreux scénarii, pour autant dans la réalité certains sont plus privilégiés que d'autres.

Le premier scénario fréquemment étudié correspond au scénario central pour lequel tous les éléments s'appuient sur l'hypothèse moyenne. Selon la problématique il est également d'usage de s'intéresser aux scénarii pour lesquels 1 seul élément dévie de l'hypothèse centrale à la fois. Cette variation permet de distinguer l'effet seul de ce critère dans l'estimation de la population. Il est par ailleurs courant de distinguer les effets au sein d'une même composante de l'équation en ne faisant varier qu'un unique élément comme par exemple la mortalité des hommes ou des femmes au sein de l'estimation des décès.

D'autres scénarii peuvent également être proposés à des fins d'étude comme par exemple s'intéresser aux cas de la mortalité constante, d'un solde migratoire nul ou encore intégrer la fécondité d'après les observations européennes.

La figure 1.11 présente graphiquement la projection de l'évolution de la population selon différents scénarii étudiés. Le scénario central qui nous intéresse tout particulièrement prévoit d'atteindre 76 millions de personnes en 2070 soit une évolution estimée de +16% par rapport à 2013.

Figure 1.11 : Projections de l'évolution de la population française à horizon 2070

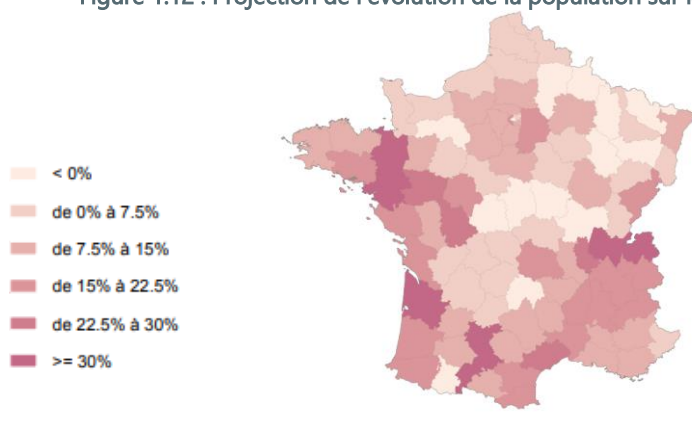


Source : Insee, projections de population 2013-2070

A partir des éléments de cette publication, une discrétisation de ces résultats au niveau départemental a été publiée en juin 2017 dans un complément à cette étude. L'aboutissement de ces travaux a permis d'identifier les départements les plus sujets aux variations migratoires en se limitant à l'horizon 2050.

Les différents indicateurs présentés ci-dessus ont été calculés département par département à partir des données observées sur l'année 2013 (taux de mortalité, quotient migratoire et quotient de fécondité). Ils ont par la suite été revalorisés annuellement des projections établies nationalement en suivant les mêmes scénarii que ceux exposé ci-dessus. L'évolution de la population estimée à +13% d'après le scénario central est donc discrétisée au département comme le montre la figure 4.11.

Figure 1.12 : Projection de l'évolution de la population sur la période 2013 – 2050



L'évolution de la population n'apparaît pas homogène sur le territoire mais bien concentrée sur certaines régions, principalement l'Ouest et le Sud de la France. Avec de telles disparités dans la migration de la population, le parc immobilier devrait lui aussi subir des changements conséquents pour adapter l'offre à la demande.

Les projections nationales ont été révisées en novembre 2021 pour intégrer une mise à jour des années 2020 à 2022 liée à l'effet covid mais surtout à une révision des hypothèses de fécondité. Le scénario central tient désormais compte d'un taux de fécondité à 1.8, hypothèse basse lors de la précédente étude. Comme aucune application de ces résultats n'a été faite au département, l'étude initiale des projections à fin 2050 reste donc appliquée dans la suite de cet exposé. A noter cependant que les estimations de la population présentées dans ce mémoire surestiment le niveau désormais observé.

Les scénarii établis par l'INSEE ont un objectif commun : prévoir l'évolution de la population à horizon 2050. Pour autant ce qui nous intéresse c'est davantage l'évolution du nombre de logements que celui du nombre d'habitants par commune. C'est en effet à partir de l'évolution de cette variable que l'on pourra estimer l'impact de l'effet population sur la part des contrats à risque en sécheresse.

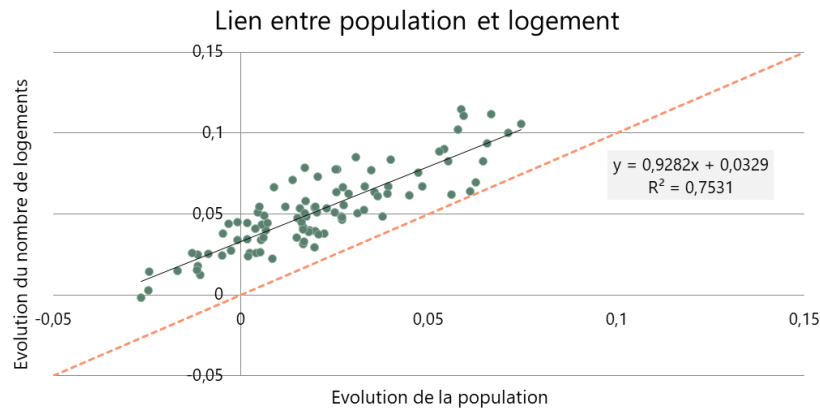
Pour identifier le lien entre l'évolution des logements et l'évolution de la population une analyse a été faite en s'appuyant sur les séries historiques de l'INSEE. Cette base permet de suivre l'évolution de plusieurs indicateurs classiquement publiés par l'INSEE selon les différentes dates de recensement. Sont présentés pour chaque commune la population, le nombre de ménage, les naissances et les décès ainsi que le nombre de logements dissociés entre les résidences principales, secondaires et les logements vacants.

Même si l'intérêt ici se situe principalement dans l'estimation de l'évolution du nombre de maison, l'obtention de cet indicateur suivi dans le temps n'est pas directement fournis par l'INSEE. L'agrégation manuelle des différentes bases logements est fortement complexifiée par la mouvance du découpage administratif surtout sur un historique long. Par simplicité la base fournie par l'INSEE sera conservée, la source de données est ainsi fiabilisé et les retraitements nécessaires à la bonne comparaison des éléments ont déjà été appliqués.

Afin de se rapprocher des estimations départementales données par l'INSEE, la base a été agrégée au niveau département.

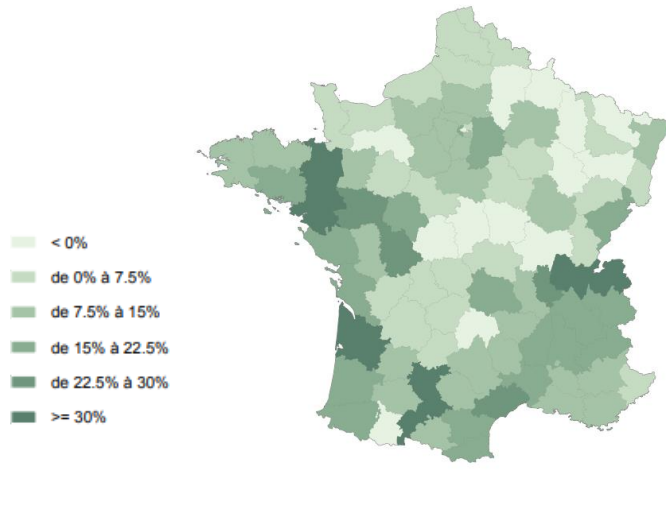
La figure 1.13 représente l'évolution du nombre de logements par rapport à la population sur la période 2008-2013 pour chaque département. Le nuage de points ne semble pas suivre l'équation  $y = x$ , il est donc censé de chercher à définir une équation entre les 2 facteurs plutôt que d'appliquer directement l'évolution de la population sur les logements. A partir de ce nuage une équation linéaire a donc été définie afin de permettre d'estimer l'évolution du nombre de logements par rapport à la connaissance de la population. En affichant un  $R^2$  au-delà de 75% l'équation est donc bien pertinente face aux données présentes.

Figure 1.13 : Lien entre l'évolution de la population et celle du nombre de logements entre 2008 et 2013



En appliquant cette équation sur la projection de la population selon le scénario central on obtient la répartition au département visible en figure 1.14.

Figure 1.14 : Projection de l'évolution du nombre de logements sur la période 2013 – 2050



Pour une évolution de la population à +13% le nombre de logements devraient augmenter de +15% à horizon 2050. L'évolution du parc immobilier suit donc l'évolution de la population avec des hausses concentrées sur l'Ouest et le Sud de la France. La volumétrie est simplement légèrement haussière. Cet effet s'explique par un constat observé depuis quelques années sur la répartition des ménages : le nombre de ménage d'une unique personne est en forte hausse et vise à poursuivre cette évolution d'où un besoin de logement plus important.

Pour identifier l'effet population dans l'impact de l'exposition de la marque face à la sécheresse diverses hypothèses de stabilité ont été prises afin de pouvoir appliquer le scénario d'évolution :

- L'évolution de logement estimée a été appliquée uniformément pour toutes les communes du département. L'évolution du portefeuille est donc considérée comme homogène et ne tient pas compte des spécificités communales qui pourraient engendrer une modification du nombre de logements différente entre les zones urbaines et rurales par exemple.
- La structure interne du portefeuille est supposée inchangée c'est-à-dire que la part des propriétaires de maisons reste constante quelle que soit l'évolution de logements observée.
- L'évolution population-logement est supposée homogène annuellement afin de pouvoir appliquer par département un effet annuel. Comme l'impact a été appliqué sur le



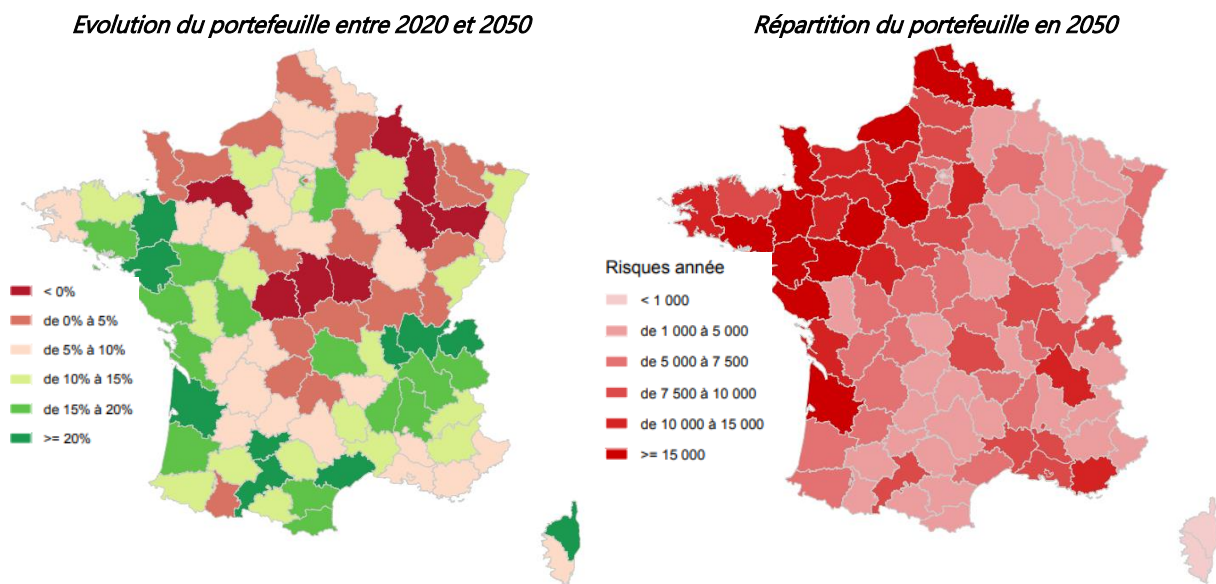
portefeuille 2020 pour conserver la structure la plus récente de l'exposition, la projection 2013-2050 est ainsi proratisée pour devenir une projection sur la période 2020-2050.

Dans les faits il est très peu probable que ces différents éléments restent stables dans le temps, pour autant une telle démarche permet de révéler si l'évolution de la population va drastiquement modifier l'exposition du portefeuille de la marque.

Les projections d'après le scénario central proposé par l'INSEE se concrétiseraient par une hausse du portefeuille de + 11% en 30 ans entre 2020 et 2050 (après avoir linéarisée l'évolution de +13% sur la période 2013 à 2050). De par l'hypothèse de répartition constante faite précédemment, cette évolution porte aussi bien sur le portefeuille dans sa globalité que sur la cible des contrats potentiellement concernés par le risque sécheresse.

En couplant l'exposition initiale non homogène des propriétaires de maisons au niveau du territoire ainsi que la projection de la population, la répartition du portefeuille de la compagnie France entière se voit être modulée de ces effets. A fin 2020 l'exposition initiale du portefeuille est principalement orientée sur le Nord-Ouest de la France ce qui correspond à l'implantation historique de la marque (cf figure 1.9.). La figure 1.15 présente les évolutions liées à l'effet population ainsi que le résultat global à fin 2050.

Figure 1.15 : Projection de la répartition des propriétaires de maisons



Contrairement aux évolutions observées sur le portefeuille sur la période 2011 à 2020 ce sont désormais la côte Atlantique et la région Auvergne / Rhône-Alpes qui évoluent le plus. Pour autant l'exposition initiale étant si hétérogène, le portefeuille reste quasiment constant avec une forte concentration de son exposition sur le Nord-Ouest et la côte Atlantique.

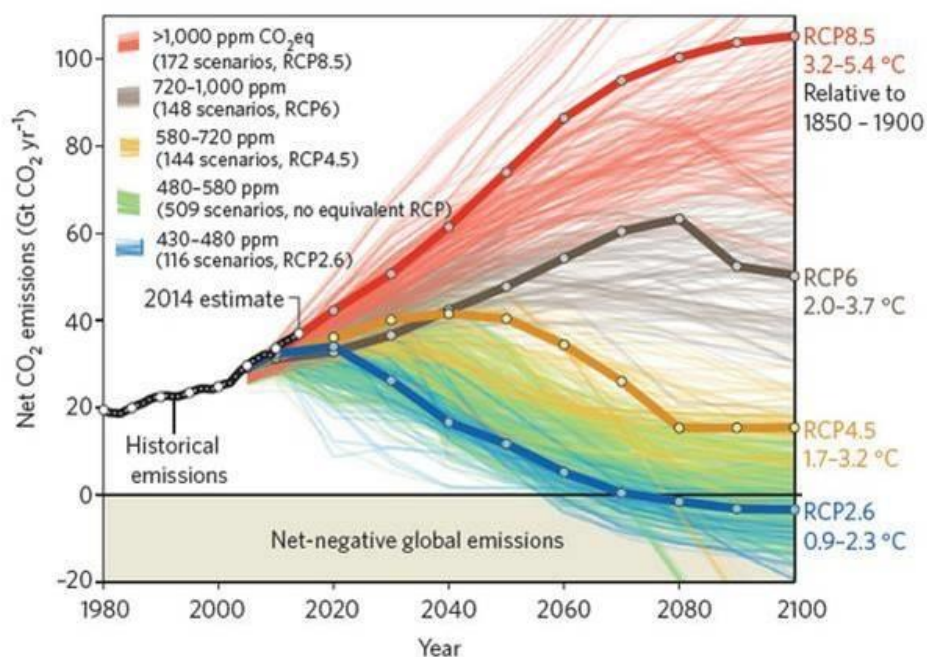
Le peu de distorsion causée par la migration de la population sur l'exposition du portefeuille MMA ne semble pas mettre en péril l'équilibre existant de la marque. Ce n'est cependant pas cet effet qui est défini comme responsable d'une forte hausse du risque sécheresse mais bien la partie climatique.

### 1.2.3. Impact de l'évolution face au climat

Les dernières projections de la sécheresse publiée par la CCR et France Assureurs sont plus qu'alarmistes et dans les effets identifiés comme accroissant du risque se trouve l'évolution du climat. En effet quel que soit le scénario retenu, les experts s'accordent à dire que le réchauffement climatique aura un impact non négligeable sur les différents risques climatiques dans les prochaines années.

Dans la littérature le réchauffement climatique est analysé en s'appuyant sur les émissions de gaz à effet de serre. La figure 1.16 présente les 4 scénarii principaux qui sont le plus souvent étudiés.

Figure 1.16 : Scénarii d'évolution du gaz à effet de serre



Ces différents scénarii diffèrent de par l'estimation des différents effets humains qui pourraient venir moduler les tendances.

- Le scénario **RCP 8.5** correspond à la projection la plus pessimiste dans le cas où l'évolution du gaz à effet de serre augmenterait selon le rythme actuel et donc sans aucune adaptation de l'homme pour en maintenir le niveau.
- Le scénario **RCP 6** correspond à une projection tenant compte d'une stabilisation des émissions de gaz à effet de serre à un niveau moyen correspondant à des efforts mis en place avant la fin du 21<sup>ème</sup> siècle.
- Le scénario **RCP 4.5** correspond là encore à une stabilisation des émissions mais à un niveau faible lié à des efforts plus conséquents mis en place sur l'activité humaine.
- Le scénario **RCP 2.6** correspond à la projection la plus optimiste pour laquelle de gros effort doivent être rapidement mis en place pour que l'année 2020 devienne le point culminant des émissions.

Dans les faits, 2 scénarii sont le plus souvent étudiés. Le scénario RCP 4.5 est le plus souvent retenu comme étant la projection future la plus plausible car ses hypothèses coïncident avec les objectifs actés lors de la COP 21 en décembre 2015. Le scénario RCP 8.5 est également souvent mis à



contribution afin de définir une vision maximale des effets du réchauffement climatique. Pour autant dès lors que l'étude se fait à horizon moyen terme, l'effet des différents scénarii ne diffèrent que de peu.

En septembre 2018 la CCR a publié en partenariat avec Météo France une étude illustrant l'impact du changement climatique sur les risques de catastrophes naturelles en France à horizon 2050<sup>8</sup>. Cette étude a pour objectif d'identifier sur les risques climatiques les plus importants la hausse des coûts face aux risques climatiques majeurs à savoir les inondations, la submersion et la sécheresse.

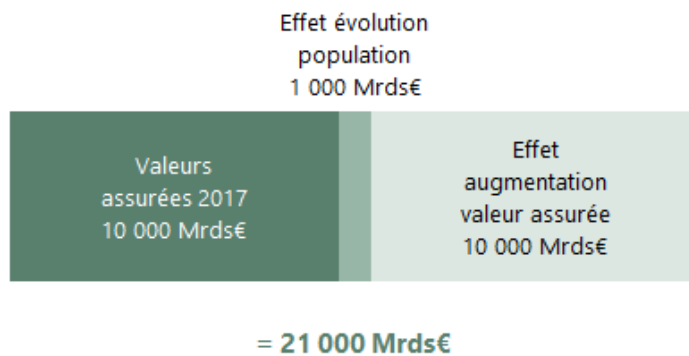
Au-delà de l'utilisation de ces conclusions par les différents organismes d'assurance dans leur prévision de charge, ces études ont également pour vocation d'acculturer sur un risque probable et de favoriser l'engagement individuel et collectif pour essayer au maximum de limiter les effets exposés. Pour cela, l'intégralité de l'étude a été menée en s'appuyant sur le scénario climatique le plus pessimiste. Gardons cependant en mémoire qu'à horizon 2050 il n'y a encore que peu d'écart sur les différents scénarii et que selon les dernières estimations les tendances actuelles suivraient actuellement plus la courbe du scénario RCP 8.5 que du scénario RCP 4.5.

Afin de maximiser la fiabilité des résultats, les projections tiennent compte de différentes hypothèses sur des éléments exogènes au climat :

- L'évolution du nombre de risques à assurer est prise en compte en s'appuyant sur la projection centrale de migration de la population évaluée par l'INSEE (cf étude présentée précédemment).
- L'évolution de la valeur assurée des biens a également été anticipée en s'appuyant sur l'évolution des prix à la consommation et des prix à la construction

Ainsi, avant même de regarder l'impact du changement climatique, les hypothèses exposées ci-dessus sont déjà à l'origine d'un doublement de la valeur assurée entre 2017 et 2050 comme l'illustre la figure 1.17.

Figure 1.17 : Evolution du coût de la valeur assurée à horizon 2050



En cohérence avec les constats obtenus sur la partie population, l'effet lié à la partie migration de la population est fortement limité avec une hausse de 1% par rapport à la charge actuelle. Par contre la partie coût du risque explose en multipliant par 2 la charge de 2017. Ce constat est d'ailleurs

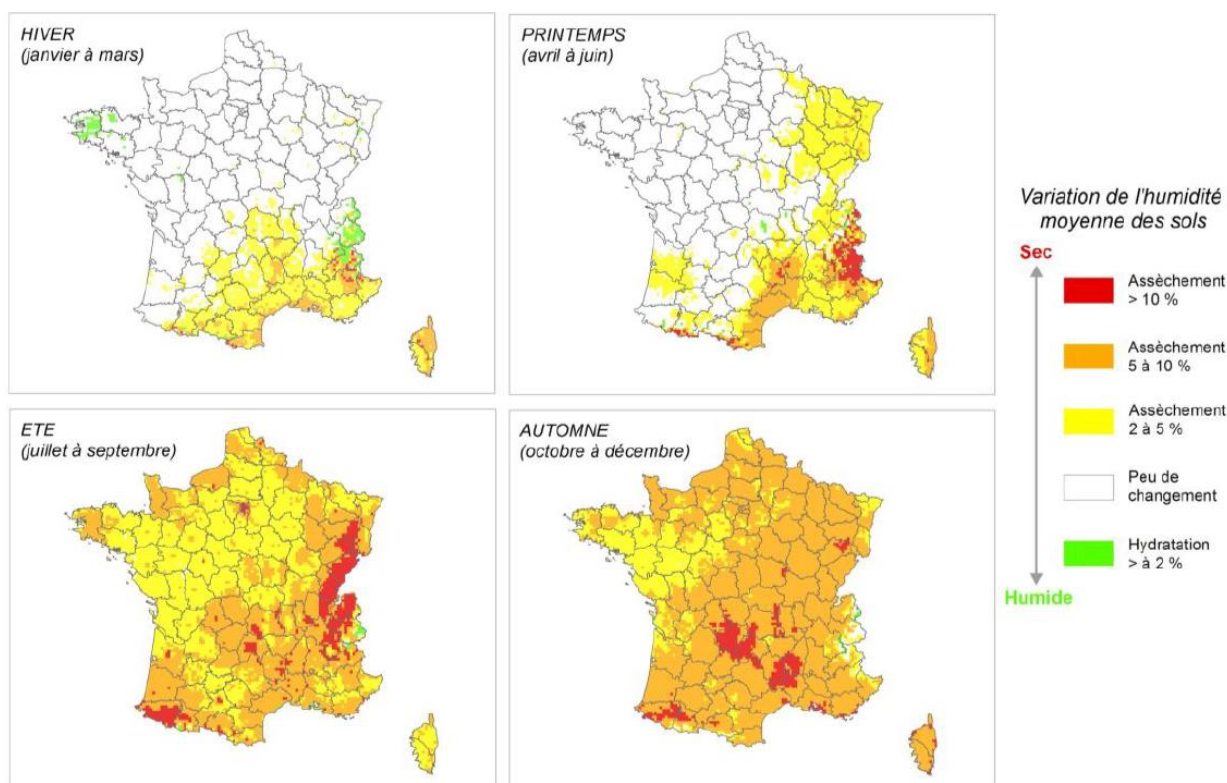
<sup>8</sup> Conséquences du changement climatique sur le coût des catastrophes naturelles en France à horizon 2050, CCR et Météo France, Septembre 2018

partagé par France Assureurs qui dans son étude identifie bien un effet non négligeable lié à l'évolution des richesses dans ses prévisions de la sinistralité sécheresse sur la période 2020 – 2050.

L'identification des potentiels sinistres sécheresse s'est faite par croisement de plusieurs variables. Les caractéristiques du sol ont été intégrées en s'appuyant sur la cartographie du retrait et gonflement des argiles en provenance du BRGM et les données d'humidité du sol ont été estimées par Météo France. Pour tenir également compte de l'impact de la réglementation dans l'acceptation de la prise en charge de la sinistralité par les compagnies d'assurance, un critère d'éligibilité a également été simulé en s'appuyant sur l'analyse de l'historique du territoire afin d'identifier le caractère exceptionnel de la sécheresse.

En se focalisant sur l'identification de l'aléa, l'objectif a été de mettre en avant trimestre par trimestre les différences entre le climat actuel et futur concernant l'humidité dans le sol. Ces évolutions permettent d'identifier les zones à risque de générer plus ou moins de sécheresse et donc de dommages sur les habitations. La figure 1.18 présente ces évolutions estimées sur les 4 saisons. Au vu des représentations cartographiques l'intégralité du territoire français serait touchée par un assèchement plus élevé du sol.

Figure 1.18 : Evolution de l'humidité des sols d'après le climat futur (scénario RCP 8.5)



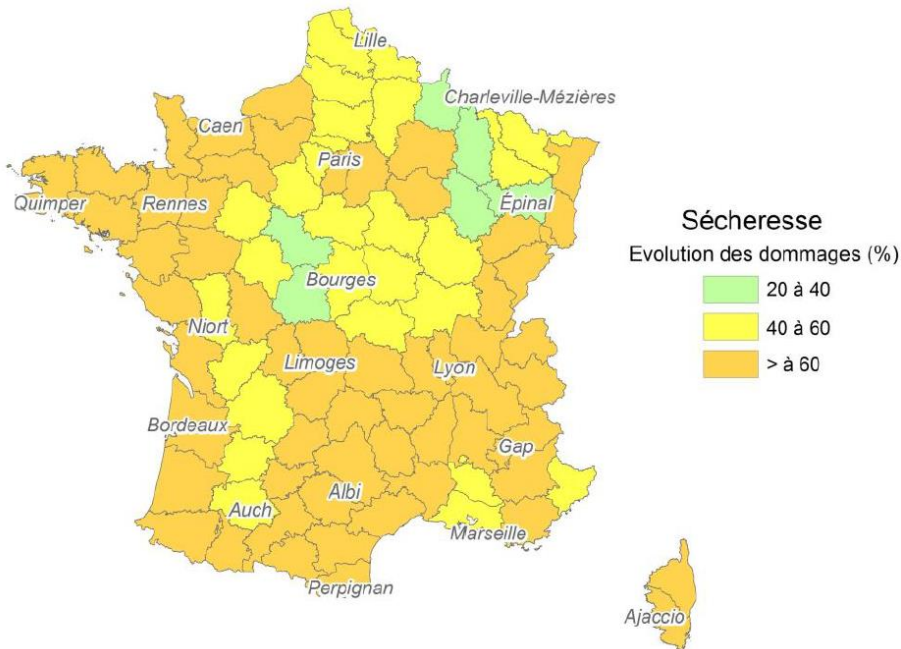
Source : Etude CCR

En s'intéressant dans un premier temps uniquement à l'effet risque pur lié au climat, l'impact estimé par la CCR sur la charge assurantiel correspondrait à une augmentation de +23% annuelle sur le territoire français. En intégrant en plus les effets migratoires cette évolution engendrerait une hausse de près de +75% sur le coût annuel. La dernière variable à intégrer correspond à l'inflation, pour autant son estimation est largement controversée du fait des derniers indices d'inflation publié. Par prudence la CCR estime simplement que cet effet ne sera certainement pas neutre et pourrait vraisemblablement être responsable à minima d'un doublement du coût annuel.

Par ailleurs cette étude datant de 2018, les effets haussiers en termes de fréquence et de coût instaurés par la dernière réforme ne sont pas du tout pris en compte. Les tendances évoquées sont donc probablement inférieures aux évolutions réelles qui seront constatées à horizon 2050.

En se plaçant hors effet inflation la CCR estime que l'intégralité du territoire français verra ses coûts augmenter à minima de 20% par an pour tous les départements. Les plus touchés étant le sud de la France du fait d'un assèchement plus atypique du sol ainsi que la côte atlantique liée à la migration importante de la population sur cette zone. La figure 1.19 présente de manière cartographique le cumul de ces effets par rapport à la sinistralité actuelle.

Figure 1.19 : Evolution de la sinistralité sécheresse annuelle



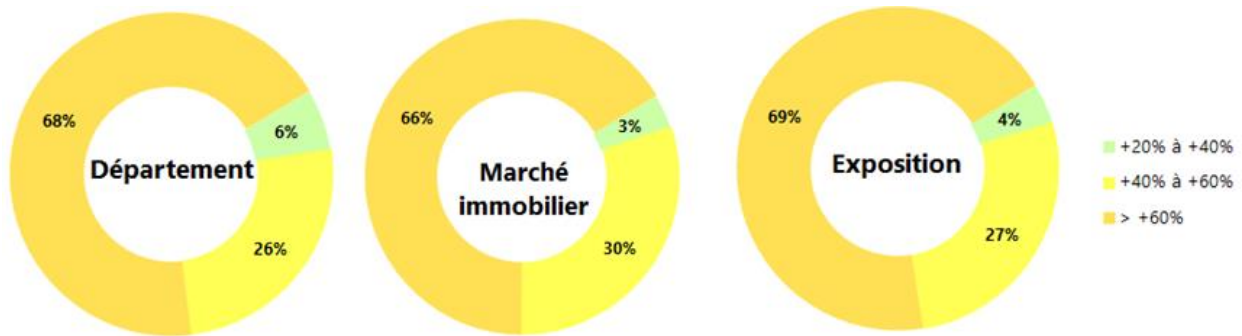
Source : Etude CCR

Cette même étude avait déjà été initiée en 2015 sur la base du scénario médian du réchauffement climatique. A hypothèse équivalente l'effet climatique n'était alors estimé qu'à +3% contre +23% ici. Cependant vu les récentes sécheresses et l'ensemble des événements climatiques arrivés depuis, les positions de l'ensemble des acteurs ont été largement revu à la hausse concernant les estimations. Il apparaissait donc comme plus pertinent de présenter un scénario sans doute pessimiste mais plus en adéquation avec la réalité.

En écludant l'effet inflation qui est actuellement incertain, l'étude présentée ci-dessus a été utilisée afin de mettre en corrélation l'exposition de notre portefeuille avec les évolutions annuelles des dommages estimées.

La figure 1.20 représente la répartition de différents indicateurs selon les évolutions dommages estimées par la CCR. D'après leur estimation 6% des départements n'auraient qu'une évolution modérée pour plus des 2/3 avec un impact au-delà de 60%. Le parc immobilier n'étant pas identique département par département le deuxième camembert donne la répartition des maisons d'après cette même classification. Enfin, le dernier graphique donne la répartition de l'exposition du portefeuille à risque (propriétaire de maison) du groupe.

Figure 1.20 : Répartition selon l'évolution du coût sécheresse



La compagnie ayant une répartition hétérogène de ces risques assurés sur le territoire français, les résultats diffèrent entre la répartition des logements France entière et ceux spécifiques de la marque. Avoir une exposition plus marquée sur l'Ouest et le Sud-Est expose donc le portefeuille de MMA à de fortes hausses du coût sécheresse dans les années à venir et qui pourraient s'avérer plus conséquentes que pour des acteurs plus homogènes d'après leur couverture du territoire.

*Malgré le caractère exceptionnel de la sinistralité sécheresse, il s'avère que ce phénomène s'amplifie d'année en année et qu'il est désormais défini comme un risque à forte croissance sur les années à venir principalement à cause de l'impact du réchauffement climatique.*

*En croisant les différentes études d'impact fournies par des sources externes, la stabilité du portefeuille MMA permettrait de limiter la distorsion de ce dernier face aux évolutions exogènes, mais ne lui permettrait pas de se protéger dans le cas d'un scénario alarmiste concernant le réchauffement climatique. Il est donc pertinent de se sensibiliser aux impacts probables de la sinistralité sur le portefeuille en cherchant à mieux maîtriser ce risque.*

*L'analyse plus fine des particularités de la sinistralité sécheresses à travers les caractéristiques du phénomène et les données l'illustrant sera à l'origine de l'amélioration de notre connaissance de ce risque. A partir de ces éléments, nous serons par la suite en capacité de les appliquer dans différentes problématiques actuarielles afin d'améliorer la maîtrise de ce risque.*

## Chapitre 2 – Particularités de la sinistralité sécheresse

Quel que soit le prisme considéré dans les études en assurance non vie le **risque** est à la base de toute analyse. L'objectif pour la compagnie d'assurance est de s'appuyer sur un raisonnement prudent lui permettant de maîtriser les prestations qu'elle aura à payer aux assurés. L'identification des particularités du risque sécheresse est donc primordiale pour pouvoir par la suite le travailler du point de vue géographique. Cette connaissance permet également de déterminer les données les plus enclines à expliquer ce phénomène.

### 2.1. Spécificités du risque sécheresse

Comme d'autres sinistres de catastrophes naturelles, la sinistralité sécheresse ne peut s'appréhender comme tout autre événement garanti par un contrat d'assurance habitation. L'occurrence de la sinistralité est hétérogène au fil du temps de même que sa localisation ; sa prise en charge est conditionnée par une validation administrative ; et son règlement est lui aussi encadré par un système réglementaire.

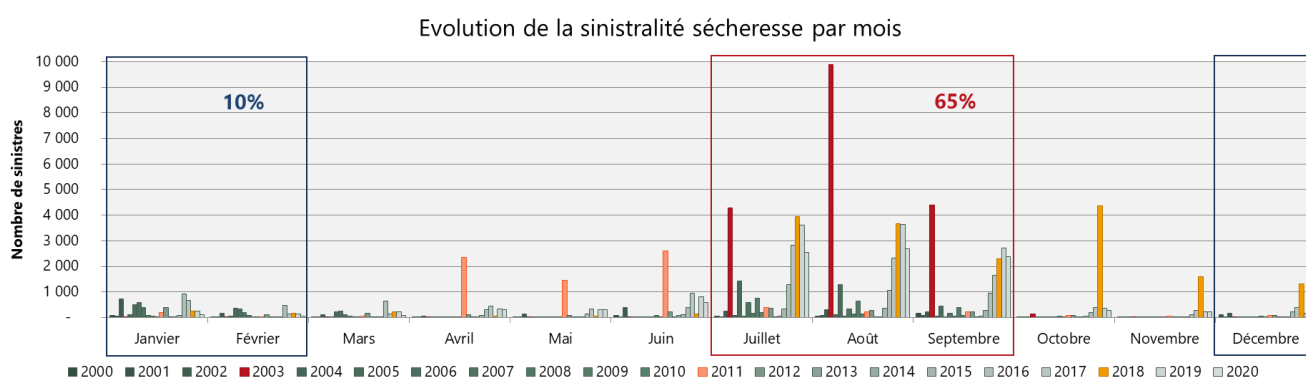
#### 2.1.1. Saisonnalité et localisation de la sinistralité sécheresse

L'analyse du phénomène de la sécheresse dans le chapitre précédent a permis d'identifier les effets mécaniques à l'origine du sinistre. Le retrait et le gonflement de l'argile étant intimement lié à l'humidité, on peut par extension émettre l'hypothèse que la sinistralité est sujette à la saisonnalité.

Le groupe Covéa associe 3 entreprises d'assurances : GMF, MAAF et MMA. Chacune d'elle présente des spécificités de par son exposition, ses orientations stratégiques, ses clients ... pour autant la sinistralité climatique est un point de convergence entre tous les assureurs. Afin de capitaliser sur la force d'un groupe, certaines analyses présentées dans la suite de ce mémoire se feront d'après les données du groupe tout entier, mais leur application finale ne concernera que la marque MMA.

En analysant la répartition mensuelle des sinistres entre 2000 et 2020 sur l'ensemble du groupe Covéa à partir de la figure 2.1, on remarque effectivement une forte proportion de sinistre sur la période estivale.

Figure 2.1 : Illustration de la saisonnalité sécheresse chez Covéa





D'après les experts les sinistres résultent le plus souvent du phénomène de gonflement puis rétractation des sols plutôt que de l'effet inverse. C'est ainsi qu'on obtient 65% de la sinistralité causée par un événement survenu entre **juillet et septembre**. En parallèle, le gonflement suite à la rétractation peut également être la cause de dommage puisque 10% de la sinistralité est néanmoins observable sur les mois de **décembre, janvier et février**.

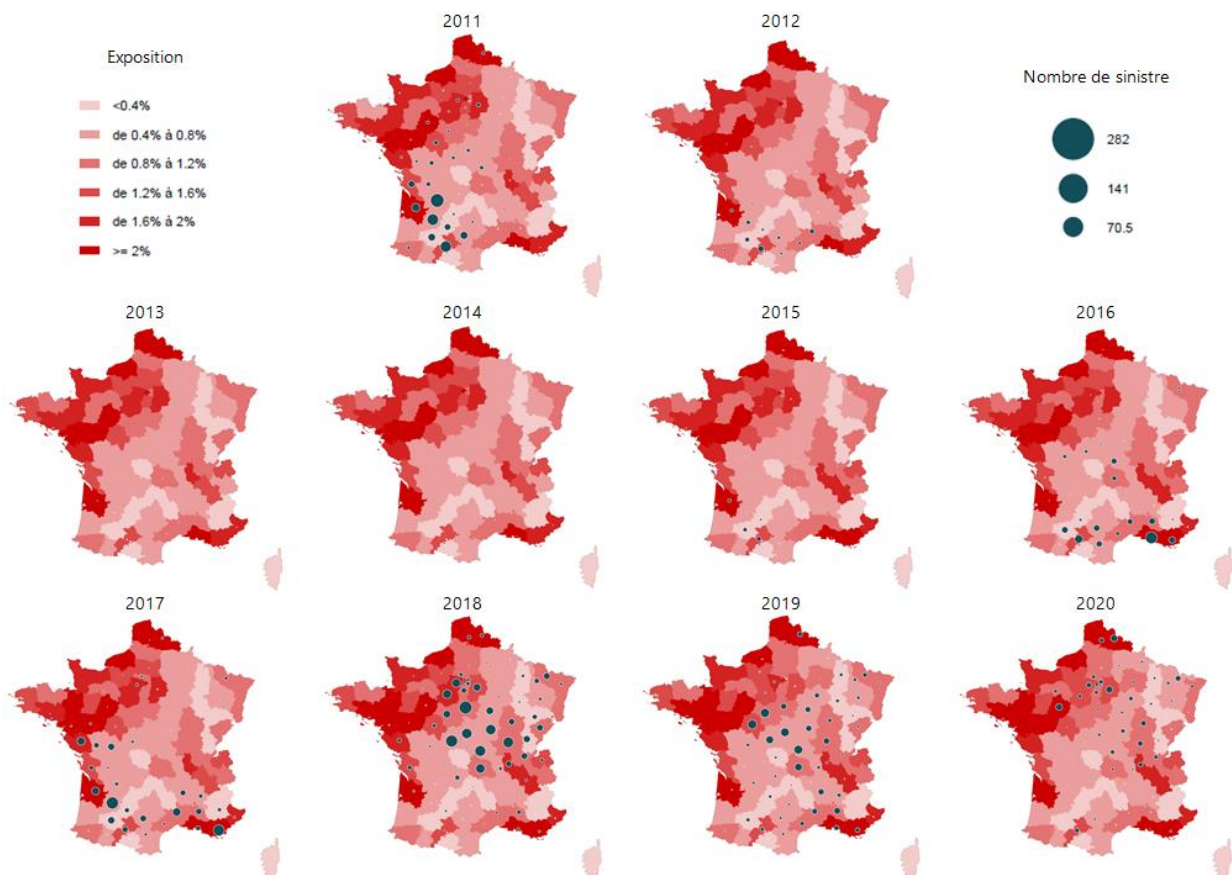
Ce graphique met également en avant l'atypisme de certaines années :

- La sécheresse historique de **2003** ressort de manière flagrante au vu du nombre de sinistres enregistrés ;
- L'année **2011** présente une tendance atypique avec une sinistralité concentrée sur le printemps ;
- L'année **2018** est également particulière avec une sécheresse estivale dont l'ampleur s'est poursuivis durant l'automne.

Au-delà d'un phénomène très volatile, sa localisation est également mouvante. D'une année sur l'autre les zones touchées ne sont pas identiques.

La figure 2.2 présente l'historique de la sinistralité sécheresse par département sur la période 2011 à 2020 selon l'exposition (part du département dans l'exposition France entière). Les sécheresses qui étaient présentes sur l'Ouest de la France et principalement la côte Atlantique sur la période 2011 à 2015 se sont progressivement dirigées vers le Sud en 2016-2017 puis se sont déplacées du centre de la France vers le Nord-Est depuis 2018.

Figure 2.2 : Evolution de la sinistralité sécheresse sur 10 ans

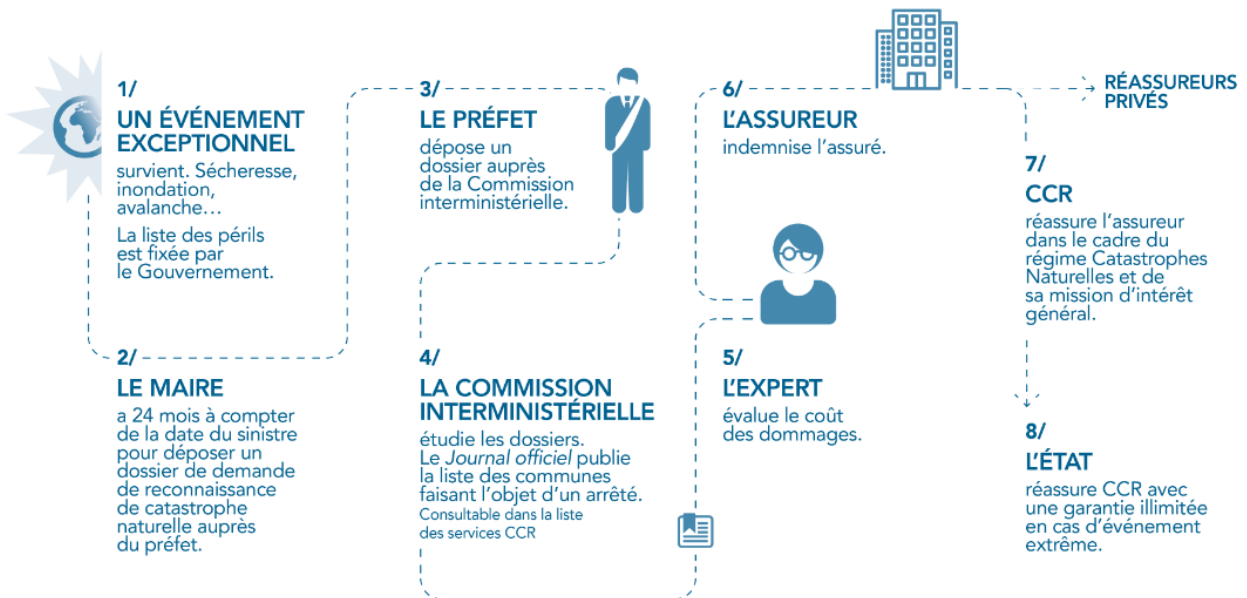


La sécheresse peut s'apparenter aux événements rares avec une fréquence sinistre inférieure à 0,1% sur 10 ans. Pourtant la particularité de la sinistralité sécheresse réside dans l'inconstance du phénomène. L'origine du sinistre étant de nature purement climatique, il est difficilement prévisible. Aucune année n'apparaît comme semblable avec des périodes bien plus exposées que d'autres. On retient par exemple des années de sécheresse intense comme 2003, 2011 ou bien 2018. L'année 2022 semble d'ailleurs propice à intégrer ces fameuses années fortement sinistrées.

### 2.1.2. Prise en charge de la sinistralité sécheresse

Contrairement aux garanties classiquement couvertes par un contrat d'assurance habitation, la déclaration d'un sinistre sécheresse par l'assuré ne déclenche pas forcément l'activation du contrat d'assurance. En effet, comme évoqué dans le chapitre précédent, la prise en charge par l'assureur est strictement encadrée par la législation pour tous risques de catastrophes naturelles. Cette dernière se fait suite à la validation du sinistre par l'obtention d'un arrêté de catastrophe naturelle sur la période de survenance du risque. Pour autant, le remboursement de ces sinistres fait bien partie des impératifs de l'assureur même s'il n'est pas le seul porteur du risque. Une fois le sinistre reconnu par la commission interministérielle, l'assureur prend en charge ce dernier mais profite du système de solidarité nationale en cédant la moitié du coût du sinistre auprès de la CCR. En contrepartie, il doit participer à l'alimentation du fonds de catastrophes naturelles appelé Fonds Barnier en reversant chaque année sous forme de taxe 12% de ses primes dommages aux biens. La figure 2.3 présente l'organisation du remboursement des sinistres de catastrophes naturelles.

Figure 2.3 : Indemnisation des sinistres de catastrophes naturelles



Source : Site de la CCR<sup>9</sup>

Actuellement, la prise en charge de la sinistralité des climatiques n'impacte que peu les compagnies d'assurance puisque que la part restante à charge de la compagnie est la plupart du

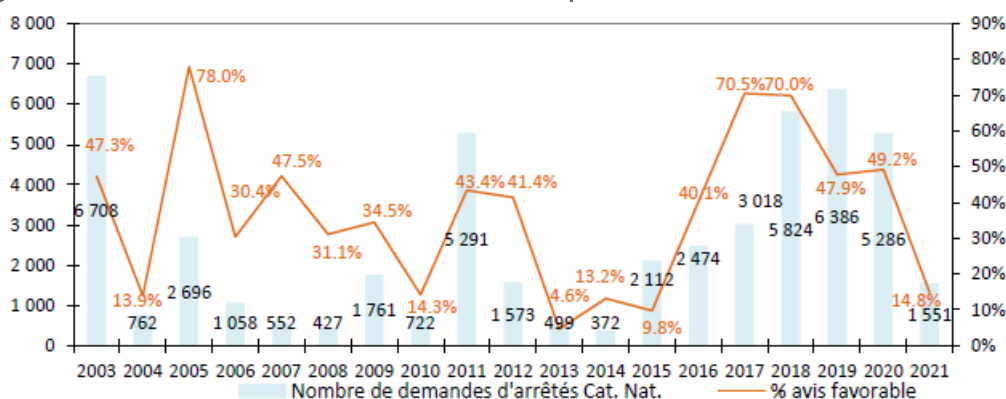
<sup>9</sup> <https://catastrophes-naturelles.ccr.fr/etat-de-catastrophe-naturelle>

temps cédée via les différents traités de réassurance. Cependant, suite à l'atypisme de l'année 2022 concernant les climatiques, de nombreux réassureurs proposent désormais des traités plus strictes afin de limiter leur prise en charge pour conserver leur rentabilité. Il est donc assez probable que la part restant à charge des compagnies d'assurance augmente sur la partie climatique.

### 2.1.3. Réglementation mouvante des arrêtés de catastrophe naturelle pour la sécheresse

Comme évoqué précédemment, l'assureur prend en charge la sinistralité sécheresse uniquement en cas d'obtention par la commune d'un **arrêté de catastrophe naturelle sécheresse** sur la période du sinistre. Cependant comme le montre la figure 2.4, l'obtention de ces arrêtés n'est pas constante dans le temps, ce qui vient complexifier encore plus l'estimation prospective de ce type de sinistre.

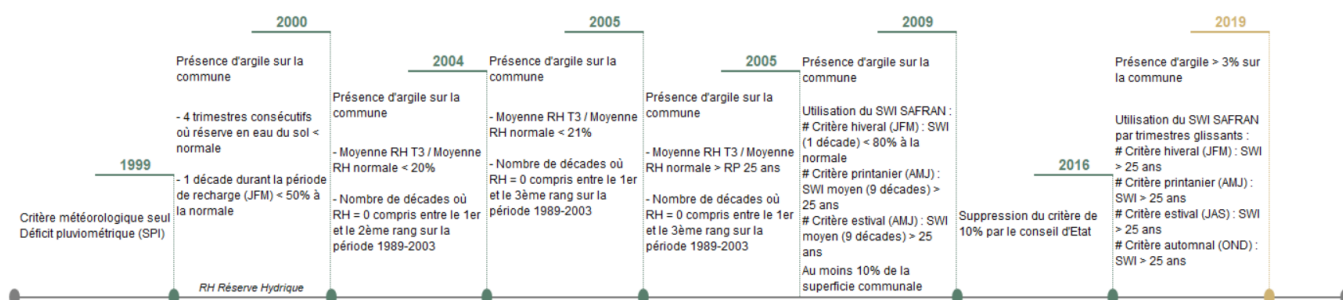
Figure 2.4 : Evolution des demandes d'arrêtés de catastrophes naturelles sécheresse entre 2003 et 2021



Pour rester au plus proche de la sinistralité sécheresse réelle, les critères régissant l'obtention d'un arrêté de catastrophe naturelle sécheresse sont régulièrement actualisés. Par exemple, après les faibles taux d'obtention d'arrêtés sur la période 2013 – 2015 (9,4% en moyenne), les règles ont été revues en 2016.

Depuis 2000, les critères intègrent à la fois le critère de prédisposition aux travers de présence d'argile identifiée sur la commune et le facteur déclenchant en analysant les variations d'humidité dans le sol. Ces éléments sont régulièrement mis à jour ce qui entraîne des modifications dans la prise en charge des sinistres d'une réglementation à une autre. La figure 2.5 représente la chronologie des règles appliquées pour définir un épisode de sécheresse.

Figure 2.5 : Evolution des critères d'obtention d'un arrêté de catastrophe naturelle sécheresse





Plus récemment, l'adoption du texte de la PPL Baudu relatif à la Réforme du Régime de Catastrophe influence également l'avenir de la sinistralité sécheresse. En effet, à partir du 1<sup>er</sup> janvier 2023 le délai concernant la déclaration de la sinistralité est allongée, passant de 2 à 5 ans. De plus, les travaux de réhabilitation des bâtiments seront désormais essentiellement axés sur l'arrêt des désordres existants et non plus sur un simple traitement des fissures apparentes. De telles dispositions vont venir augmenter la fréquence mais également le coût moyen d'un sinistre sécheresse.

Même si l'engagement de l'assureur dépend aujourd'hui de la reconnaissance de l'événement comme catastrophe naturelle, les compagnies doivent rester en alerte face à d'éventuelles modifications de cette réglementation pouvant impacter le délai de reconnaissance, l'élargissement des conditions ... et donc la modification de leur intervention. Au vu de l'historique des mises à jour réglementaires et les prévisions alarmantes sur le risque sécheresse de demain, il est plus que plausible que le nombre de prise en charge par les assureurs tende à augmenter.

## 2.2. Description des données utilisées pour définir la sinistralité sécheresse

L'origine du risque sécheresse a permis d'identifier que la sinistralité était principalement causée par des facteurs exogènes au monde de l'assurance. Une grande partie des éléments qui seront utilisés pour traiter de ce risque vont donc être des données d'ordre climatologique et géologique obtenues avec le concours d'un service interne au groupe Covéa ayant accès à ces éléments.

### 2.2.1. Conformité de la maille géographique

Pour expliquer un aléa tel que la sécheresse, la notion de géographie est un critère fondamental. La plupart des analyses se feront donc selon l'adresse du risque et à partir de critère mettant en avant les différences territoriales.

Pour accorder tous les indicateurs sur une maille unique et exploitable, l'intégralité des données présentées par la suite seront travaillées selon la maille « commune » définie par l'INSEE i.e. le code INSEE. Cette maille correspond au découpage le plus fin qu'il nous a été possible d'obtenir notamment sur les données exogènes.

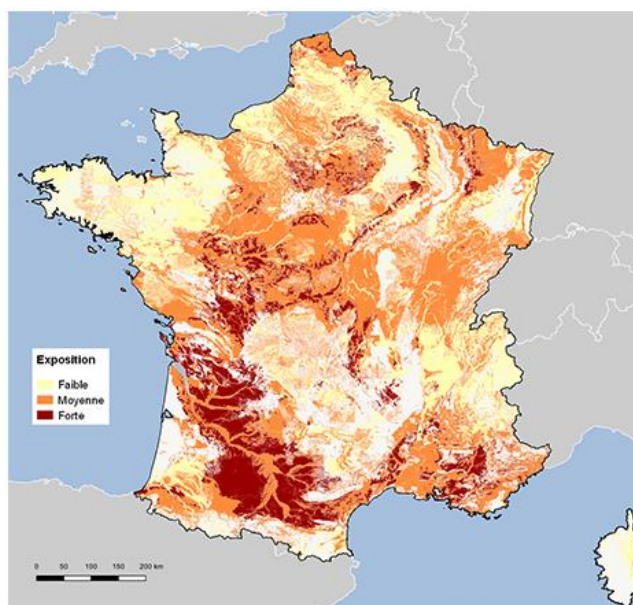
Après définition d'une maille unique, le choix d'une vision harmonisée s'impose également. En effet, la modification assez récurrente du découpage administratif, notamment suite aux nombreuses fusions de communes depuis 2015, complexifie l'utilisation de données internes. Les risques assurés en habitation et notamment les maisons sont des risques assez stables qui peuvent être conservés en portefeuille en moyenne 8 années sans changer d'adresse. Dans nos bases internes les contrats sont enregistrés d'après le référentiel en vigueur au moment du dernier acte de production (à la souscription ou suite à avenant pour changement d'adresse). De nombreuses visions de découpage administratif coexistent donc. Afin de supprimer cet effet, tous les contrats devront être re-géocodés suivant un même référentiel pour s'assurer d'une harmonie au sein du portefeuille mais également avec les autres indicateurs.

## 2.2.2. Données expliquant le facteur de prédisposition

La présence d'argile dans le sol est à la base de toute sinistralité sécheresse car à l'origine du phénomène de retrait et gonflement des sols, fragilisant les bâtiments. Pour identifier cet élément, plusieurs données peuvent nous apporter des éléments de réponses :

- Données **LUCAS** présentant pour chaque commune la proportion d'argile relevée dans les 20 premiers centimètres du sol.
- Données d'**exposition du territoire au phénomène RGA** (Retraitement-Gonflement des Argiles) **issu du BRGM** (Bureau de Recherches Géologiques et Minières) validée par la DGPR (Direction Générale de la Prévention des Risques) qui intègre en plus de la typologie du sol définie en 2007 un retravail lié à la sinistralité marché pour tenir compte des sinistres sécheresse apparus en dehors des zones définies. La carte présentée dans la figure 2.6 en est la représentation officielle.

Figure 2.6 : Représentation du phénomène RGA (Carte officielle)



Carte d'exposition des formations argileuses au phénomène de mouvement de terrain différentiel consécutif à la sécheresse et à la réhydratation des sols (France, 2020). © BRGM

Source : Site [georisques.gouv.fr](https://www.georisques.gouv.fr)<sup>10</sup>

- Données d'exposition du territoire au phénomène RGA après retraitement de la carte du BRGM pour supprimer l'effet du retravail lié à la sinistralité marché.

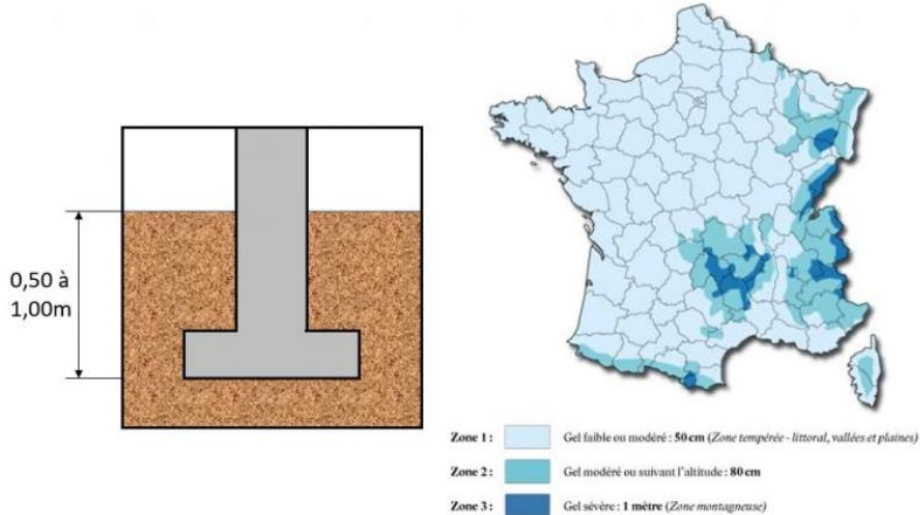
Ce dernier indicateur ne se trouve pas dans la littérature et a été calculé en interne Covéa afin de nous mettre à disposition des données non retraitées. Dans un cadre de modélisation il est toujours plus simple de s'appuyer sur la donnée la plus pure possible afin de bien maîtriser les éléments à l'origine du critère étudié et d'en maîtriser les effets.

La sécheresse étant causée par une modulation de la densité du sol, l'indicateur LUCAS apparaît comme insuffisamment explicite du phénomène. En effet, la sécheresse va venir altérer les

<sup>10</sup> <https://www.georisques.gouv.fr/articles-risques/retrait-gonflement-des-argiles/exposition-du-territoire-au-phenomene>

fondations du bâtiment dont la profondeur répond à des normes spécifiques pour se retrouver hors gel. Les fondations devant être enfouies à minima entre 50 cm et 1 m selon le climat comme le montre la figure 2.7.

Figure 2.7 : Profondeur minimale requise pour la mise hors gel des fondations



Source : Site [toutsurlebeton.fr](http://toutsurlebeton.fr)<sup>11</sup>

L'indicateur d'argile des 20 premiers cm du sol ne permet pas une connaissance de la nature du sol au niveau des fondations. Cette donnée ne sera donc pas retenue dans les travaux menés dans ce mémoire.

Les autres données présentant le phénomène RGA sont obtenues au format cartographique **carreau INSEE** suivant un découpage de la France en carreaux de 200 m<sup>2</sup> dans les zones densément peuplées ou en agglomération comme l'illustre la figure 2.8.

Figure 2.8 : Illustration du carroyage sur une commune



Les services de tarification n'ont pas forcément accès aux outils nécessaires à la transformation de ce type de découpage en données facilement exploitables. Le concours d'un service interne au

<sup>11</sup> <https://www.toutsurlebeton.fr/mise-en-oeuvre/quelle-profondeur-pour-vos-fondations/>

groupe Covéa a permis de modifier cette source en un critère facilement exploitable après agrégation de chaque aléa au niveau de la commune.

Pour chaque code INSEE a été défini la proportion de la superficie selon la typographie en 4 classes de l'argile : aléa faible, moyen, fort et risque hors zone si absence d'argile sur le territoire. Chacune de ces proportions a été pondérées selon la **tâche urbaine** (concentration d'habitation sur la commune). La donnée résultante est donc orientée par notre problématique pour correspondre non pas à la superficie de la commune risquée mais à la part des habitations risquées sur la commune.

Pour n'avoir à gérer qu'un unique indicateur par commune, on nous a également proposé une pondération arbitraire de ces proportions afin de définir un classement en 6 catégories des communes allant de l'absence de risque à un risque argile accru via la classification ci-dessous :

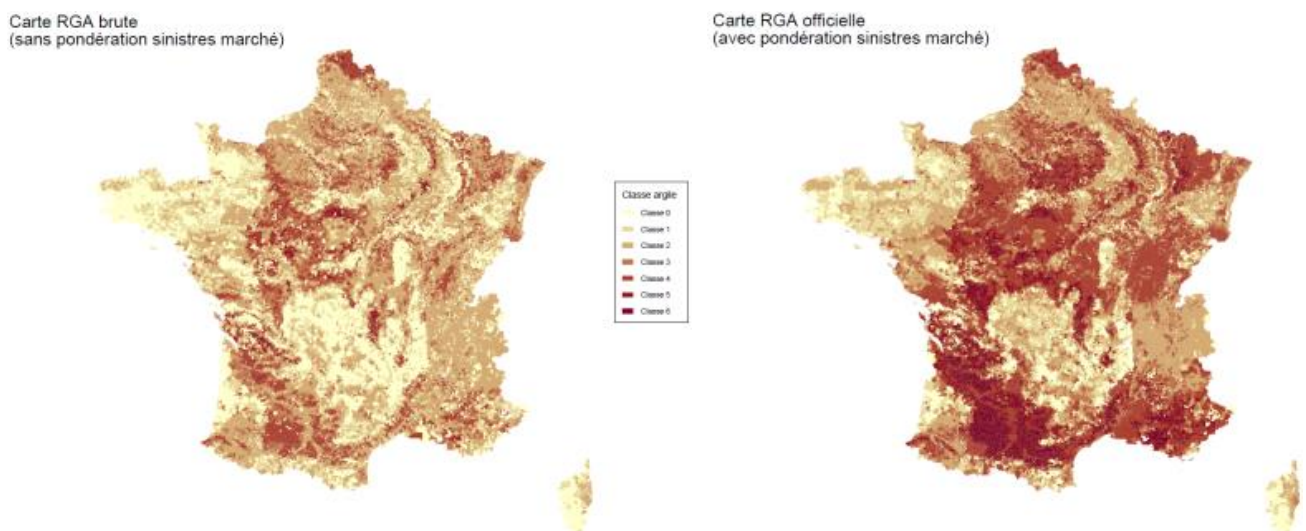
$$Classe_{iSOL6} = \text{arrondi}(2 \times \% \text{Habitat}_{\text{Aléa faible}} + 4 \times \% \text{Habitat}_{\text{Aléa moyen}} + 6 \times \% \text{Habitat}_{\text{Aléa fort}}).$$

Cet indicateur sera challengé dans la suite de cet exposé mais en croisant la fréquence observée sur la sinistralité 2000 à 2018 avec la classe *iSOL6* obtenue, les résultats de cette classification affichent bien une fréquence croissante et segmentante selon le niveau de la zone.

On conserva donc la classification *iSOL6* telle quelle pour effectuer un premier exercice de comparaison entre les deux cartes illustrant le phénomène RGA. Cette analyse a pour objectif d'identifier la source la plus pertinente tout en illustrant l'effet lié à l'intégration de la sinistralité marché dans la carte officielle du BRGM.

A titre illustratif, les cartes présentées en figure 2.9 représentent ces deux indicateurs en appliquant la classification arbitraire évoquée ci-dessus.

Figure 2.9 : Illustration cartographique de l'impact du retraitement par sinistralité marché



Dans une démarche de comparaison des deux indicateurs, une analyse graphique a été effectuée pour illustrer l'intérêt du correctif inclus dans l'indicateur RGA officiel en identifiant les différences.

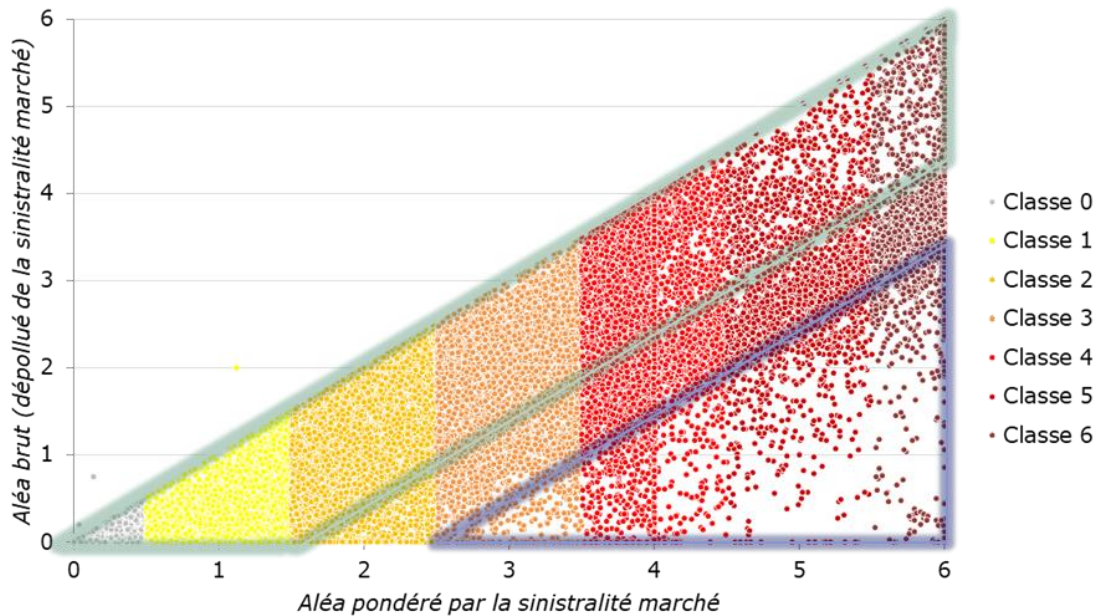
Sur la figure 2.10 sont croisées la valeur de l'aléa catégorisé d'après la carte RGA tenant compte de la sinistralité marché avec la valeur de l'aléa catégorisé dé-biaisé de la sinistralité marché pour



chaque commune. L'aléa catégorisé correspond à la valeur calculée dans la classification *iSOL6* avant passage à l'arrondi soit :

$$Classe_{iSOL6bis} = 2 \times \% Habitat_{Aléa faible} + 4 \times \% Habitat_{Aléa moyen} + 6 \times \% Habitat_{Aléa fort}$$

Figure 2.10 : Croisement de la répartition de l'aléa de chaque commune d'après les deux indicateurs RGA



En présence d'indicateurs fortement corrélés expliquant un phénomène identique avec une légère amélioration du résultat, on devrait obtenir l'ensemble des données condensées dans le **trapèze vert**. La variation entre les deux indicateurs résulterait uniquement d'un ajustement à la marge avec un écart possible d'une classe. 67% des communes affichent des résultats conformes selon les deux bases.

Cependant, près d'un tiers des croisements se trouvent en dehors de cette zone de cohérence et certaines communes affichent même des conclusions drastiquement différentes. Toutes les communes situées dans le **triangle bleu** (2% des communes) sont considérées comme n'ayant aucun risque de sinistre car sans présence d'argile d'après l'indicateur brut pour après ajustement faire partie des communes les plus à risque au vu de la sinistralité passée.

Au-delà d'apporter un léger correctif, le retraitement via la sinistralité marché avait pour objectif d'apporter une correction sur des mauvaises identifications de zones argiles. Ce constat est d'ailleurs extrêmement visible lorsque l'on croise ces indicateurs à la sinistralité observée en interne Covéa. En considérant la sinistralité sécheresse déclarée entre 2000 et 2018 sur l'ensemble du groupe, seuls 59% des sinistres coïncident avec un risque argile moyen ou fort (classes 3 et supérieures) d'après la donnée d'aléa dépollué de la sinistralité marché. Par contre quand on effectue ce même test à partir de l'argile pondéré par la sinistralité marché, on obtient cette fois ci 95% de la sinistralité observée.

Après avoir validé l'utilisation des données d'exposition du territoire au phénomène RGA d'après la carte officielle, une modélisation par régression logistique a été utilisée pour challenger les poids arbitraires appliqués dans la classification précédente *iSOL6*. Deux indicateurs ont été modélisés en compléments en s'appuyant sur un modèle linéaire généralisé. Le premier *CL\_1* correspond au

nombre de sinistres par commune d'après les proportions d'aléa faible, moyen ou fort. Le second *CL\_2* modélise le logarithme du nombre de sinistres par communes toujours aux travers des mêmes variables explicatives. Le passage au logarithme a permis de lisser les effets d'échelle, on a en effet une forte volatilité du nombre de sinistres par communes (maximum 500 sinistres contre 1 unique sur une grande majorité). La modification de la variable initiale est une méthode classique qui permet d'améliorer légèrement les performances prédictives.

Comme illustré sur la figure 2.11, ces trois indicateurs restent pertinents vis-à-vis de la fréquence sinistre sécheresse observée sur la période 2000 à 2018. Ils mettent bien en avant un risque croissant plus la présence d'argile est conséquente.

Figure 2.11 : Analyse de la fréquence sécheresse selon l'indicateur de classification

Classe	Fréquence iSOL6	Fréquence CL_1	Fréquence CL_2
Classe 0	0,009%	0,007%	
Classe 1	0,015%	0,028%	0,012%
Classe 2	0,030%	0,077%	0,043%
Classe 3	0,089%	0,184%	0,142%
Classe 4	0,206%	0,277%	0,232%
Classe 5	0,365%	0,451%	0,475%
Classe 6	0,729%	0,755%	0,770%

Dans la suite de ce mémoire, l'impact du facteur de prédisposition sera analysé grâce aux trois classifications évoquées ci-dessus ainsi qu'avec l'utilisation des parts brutes de typologie d'argile pour chaque commune.

### 2.2.3. Données expliquant le facteur déclenchant

La sécheresse résultant du phénomène couplé argile et variation d'humidité nous avons cherché à inclure dans nos données cette notion de variation d'humidité année après année, mois après mois conformément à notre historique sinistre.

Les données les plus pertinentes correspondent à l'utilisation de l'indice d'humidité des sols SWI (Soil Wetness Index). C'est l'indicateur le plus fréquent dans la littérature liée à la sécheresse, il est d'ailleurs utilisé actuellement après retraitement comme critère d'obtention de l'arrêt de catastrophe naturelle sécheresse.

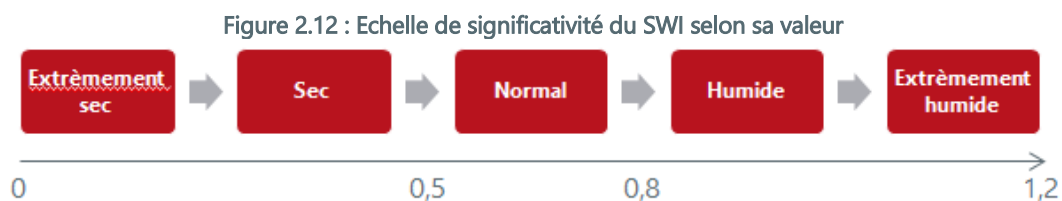
L'indicateur SWI est calculé comme le ratio entre le contenu en eau disponible dans le sol un jour donné et sa valeur maximale :

$$SWI = \frac{W - W_{wilt}}{W_{fc} - W_{wilt}}, \quad \text{avec } \begin{cases} W : \text{contenu intégré en eau du sol} \\ W_{wilt} : \text{contenu en eau du sol au point de flétrissement} \\ W_{fc} : \text{contenu en eau du sol à la capacité au champ} \end{cases}$$

Point de flétrissement : seuil à partir duquel la plante ne peut plus capter l'eau du sol

Capacité au champ : capacité de rétention maximale en eau du sol

Comme l'illustre la figure 2.12, le SWI varie de 0 à 1,2 et chacune de ses valeurs permet de définir si le sol analysé est plus ou moins humide.



Tout comme la donnée argile, la donnée SWI est disponible selon une maille géographique de carré de 64km<sup>2</sup> découpant la France en 8 981 points. Afin de pouvoir intégrer ces éléments dans nos réflexions, les valeurs des SWI ont été retravaillées en amont pour pouvoir nous mettre à disposition une moyenne mensuelle de ces indicateurs pour chaque commune entre 2000 et 2019.

La dernière évolution de la réglementation régissant les critères d'obtention des arrêtés de catastrophe naturelle s'appuie actuellement sur le SWI afin d'identifier trimestre par trimestre si la période analysée présente un épisode de sécheresse atypique vis-à-vis d'un historique de 50 ans (entre 1969 et 2018) Mathématiquement cette analyse se traduit par la comparaison du SWI standardisé 3 mois avec le quantile 0,04.

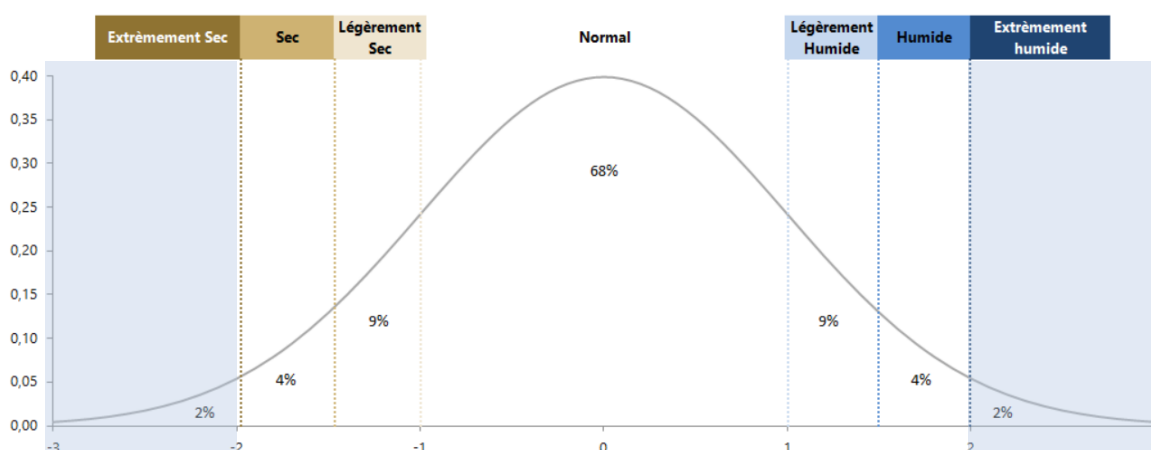
Le **SWI standardisé** permet de s'affranchir de toutes particularités climatiques liées à la localisation de la commune et de définir un indicateur suivant une loi normale centrée réduite  $N(0,1)$ .

$$SSWI = \frac{SWI_m - \overline{SWI}}{\sigma_{SWI}}, \text{ avec } \begin{cases} SWI_m : \text{Humidité du sol vu au mois } m \\ \overline{SWI} : \text{Moyenne du SWI au mois } m \text{ sur la période de référence.} \\ \sigma_{SWI} : \text{Ecart type du SWI à la moyenne} \end{cases}$$

Un SSWI aux alentours de 0 correspond à la moyenne interannuelle. Une valeur négative mettra en avant une anomalie humide alors qu'une valeur positive correspondra à un atypisme sec.

Dès lors que le SSWI 3 mois sur au moins l'un des mois du trimestre affiche un épisode sec ou humide anormal sur les 50 dernières années (soit un SSWI en-deçà de -2,06 ou au-delà de 2,06), alors le critère météorologique est validé pour l'obtention d'un arrêté sur ce trimestre. La figure 2.13 présente la correspondance entre le niveau d'humidité selon la valeur du SSWI. Les rectangles bleus illustrent les seuils de détection conformément aux critères d'obtention d'un arrêté de catastrophe naturelle.

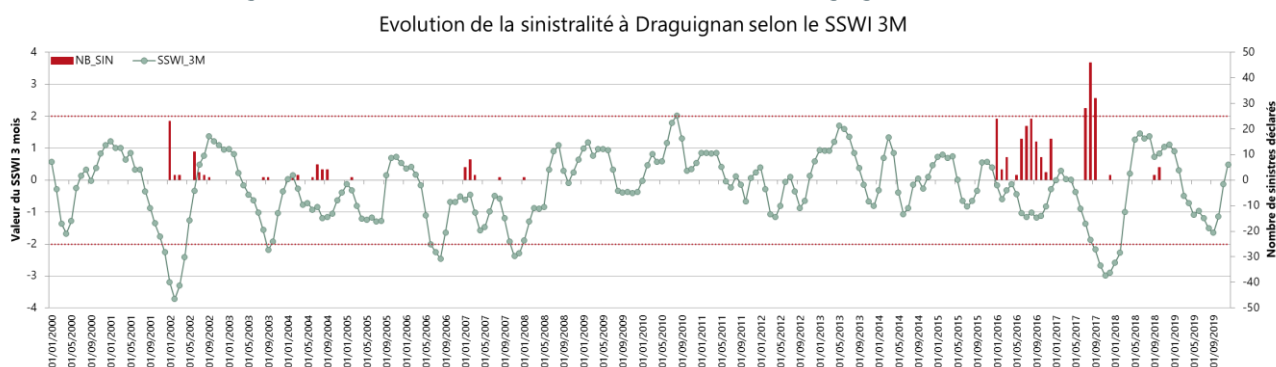
Figure 2.13 : Correspondance SSWI et anomalie humide ou sèche



Pour être en mesure d'approcher le risque sécheresse historique il nous faut utiliser un indicateur explicatif du phénomène sans pour autant être biaisé par la réglementation mouvante des arrêtés de catastrophe naturelle sécheresse.

En testant l'indicateur actuellement utilisé (SSWI 3 mois) sur une des communes les plus sinistrées sur la période 2000 à 2019, on remarque directement l'impact fort de la réglementation sur les arrêtés de catastrophe naturelle sécheresse. En effet, comme le montre la figure 2.14, la sinistralité n'est que très peu liée à la valeur du SSWI. Seuls 43% des sinistres affichent une variation d'humidité anormale sur le trimestre au sens de la réglementation actuelle : SSWI en dehors des bornes [-2,06 ; 2,06].

Figure 2.14 : Corrélation entre la sinistralité déclarée à Draguignan et le SSWI 3



Cet exemple illustre l'impact fort des critères d'arrêt de catastrophe naturelle dans la gestion des sinistres par l'assureur. Un sinistre accepté dans les années 2000 ne le serait peut-être pas au sens de la réglementation actuelle et inversement.

Pour approcher le facteur déclenchant par un critère moins contraint par la réglementation, un indicateur d'évolution du SWI mensuel a été créé sur la période de 2001 à 2019 pour chaque commune.

La création d'un indicateur en écart relatif permet de s'affranchir des spécificités climatiques de chaque commune en ne mettant en avant que les variations d'humidité.

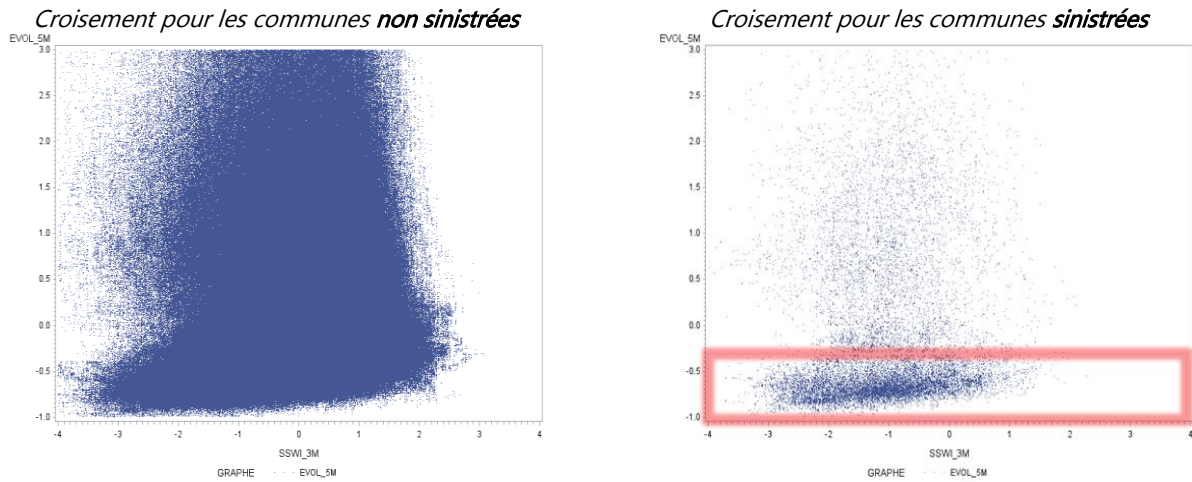
$$Evol_{x\text{ mois}} = \frac{SWI_{brut\text{ mois }auj}}{SWI_{brut\text{ }x\text{ mois }avant}}$$

De manière intuitive il paraît probable que les évolutions au-delà de 3 mois et en-deçà de 9 mois seraient les plus compatibles à une explication du risque. En effet, la forte variation d'humidité est assez liée à la saisonnalité, au sein d'une même saison il est peu probable d'observer une modification climatique extrême. Pour identifier l'indicateur le plus pertinent, une analyse graphique a été utilisée dont l'objectif était d'illustrer la différence entre une commune sinistrée et une commune non sinistrée sur le mois et l'année considérée.

Graphiquement on observe sur la figure 2.15 la représentation des points commune – année – mois en croisant l'évolution du SWI 5 mois et le SSWI 3 mois. Le graphique de gauche correspond à tous les points pour lesquels aucun sinistre n'a été déclaré sur le triplet commune – année – mois. Le graphique de droite correspond au pendant pour des triplets sinistrés.



Figure 2.15 : Croisement entre Evol SWI 5 mois et SSWI 3 mois selon la sinistralité



A partir de ce constat nous avons cherché à définir un seuil expliquant le risque sécheresse d'après le facteur déclenchant qui au vu du graphique se situe aux alentours de -0,4. Ainsi une commune fortement risquée devrait être en dehors de ce seuil donc sur l'intervalle  $]-\infty ; -0,4[$  bien plus souvent qu'une commune à faible risque. En considérant le seuil à -0,4 et l'indicateur d'évolution 5 mois, on observe une concentration de 70% des points commune – année – mois en cas de sinistres contre 30% sans sinistre.

Ces différents constats graphiques s'observent également en considérant l'indicateur d'évolution sur 4 mois. Par ailleurs considérer un seuil à -0,5 est également cohérent après analyse de la concentration des points commune – année – mois.

L'indicateur illustrant le facteur déclenchant pour le risque sécheresse sera donc défini en analysant pour chaque commune la fréquence de dépassement du seuil selon l'évolution du SWI. Cet indicateur met en avant une surexposition des communes sinistrées à être en-deçà du seuil défini.

Suite à cette analyse nous conserverons dans la suite de ce mémoire la possibilité de sélectionner, selon la pertinence, la fréquence par commune d'être en-deçà du seuil choisi selon un l'évolution du SWI sur  $x$  mois. Seront testés les seuils à -0,4 et -0,5 croisés avec les évolutions sur 4 et 5 mois.

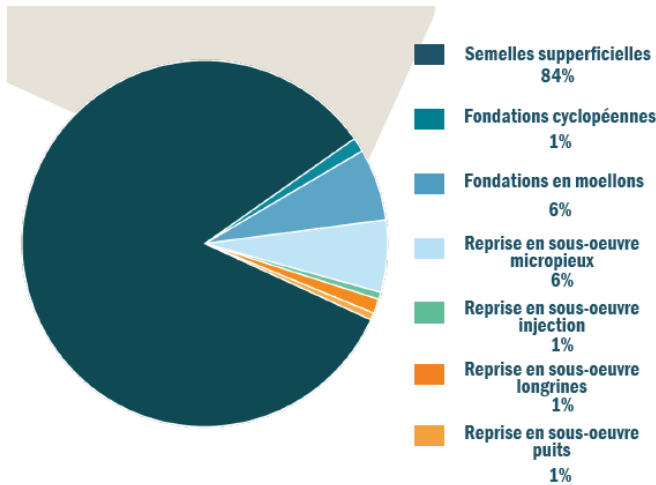
#### 2.2.4. Données à l'origine des facteurs aggravants

De nombreuses études tendent à moduler le risque de la sécheresse selon les **caractéristiques de construction du bâtiment sinistré**.

Une étude a été menée par la MRN<sup>12</sup> (Mission Risques Naturels) en 2018 à partir des rapports d'expertise de différents sinistres. Suite à cette étude ils ont pu mettre en avant la fréquence sinistre selon le type de fondations ainsi que l'ancienneté de construction du bâtiment. Comme l'illustre les figures 2.16 et 2.17, la sinistralité est fortement concentrée sur les habitations n'étant équipées que de semelles superficielles et sur des constructions récentes postérieures à 1975.

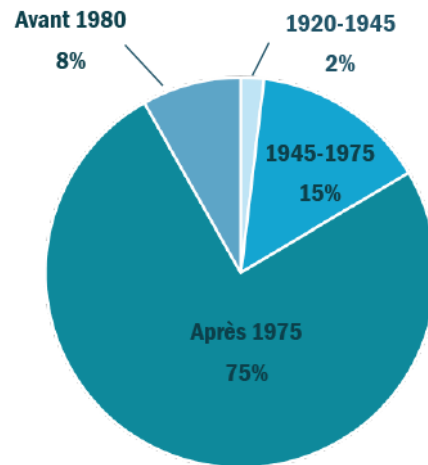
<sup>12</sup> Sécheresse Géotechnique - De la connaissance de l'aléa à l'analyse de l'endommagement du bâti, MRN, Décembre 2018

Figure 2.16 : Types de fondations de l'échantillon étudié



Source : MRN sur 157 dossiers

Figure 2.17 : Date de construction de l'échantillon étudié



Source : MRN sur 70 dossiers

Malgré la faible taille de l'échantillon ayant permis d'arriver à ces conclusions, ce sont pour la plupart des maisons individuelles construites récemment qui sont ainsi identifiées comme à risque accru.

Comme il n'est pas encore possible d'avoir une connaissance détaillée de la profondeur des fondations, des spécifications de construction ainsi que des types de matériaux caractérisant chaque bâtiment assuré, nous utiliserons les données INSEE pour essayer d'approcher certaines de ces caractéristiques commune par commune.

La base des données infra communales – **Logement** mise en ligne le 03/11/2020 permettra d'intégrer dans nos études la répartition des maisons par année de construction en utilisant le nombre de résidence principale de type maison selon les tranches de construction déterminées par l'INSEE.

La base des données infra communales – **Couples – Famille – Ménages** mise en ligne le 19/10/2020 sera utilisée pour définir la répartition des catégories sociaux-professionnelles à partir de l'identification de l'activité de la personne référente dans chaque ménage. Ces données permettront l'approche indirecte de la typologie de matériaux utilisés ainsi que le style de maison habité.

### 2.2.5. Données internes

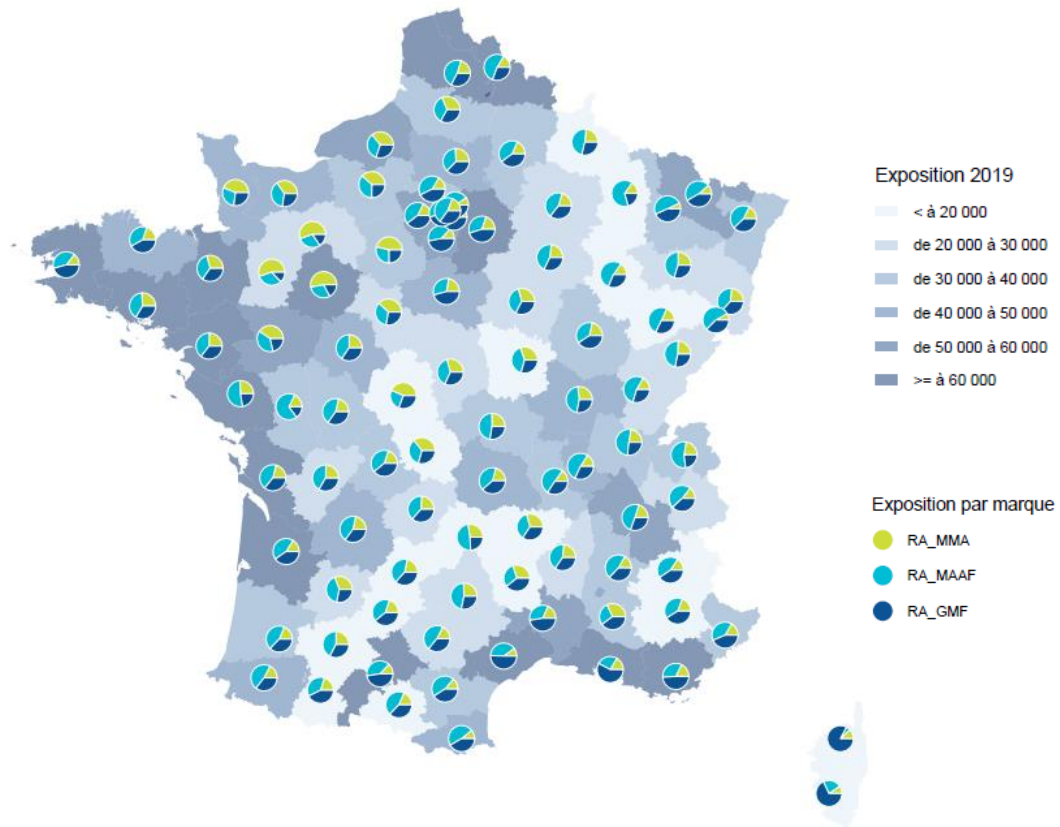
En complément des données purement externes, il est bien évidemment nécessaire de s'appuyer sur des données internes spécifiques à l'assureur pour adapter le risque à notre part de marché.

Même si les applications se feront uniquement sur la marque MMA, il a été préféré d'utiliser les données contrats et sinistres du groupe Covéa tout entier afin de maximiser la connaissance du risque sur un périmètre plus large.

La base sinistre a été constituée à partir des sinistres historiques sécheresse enregistrés entre 2000 et 2018. La sinistralité 2019 n'étant que partiellement connue au moment de la création de la base de données du fait d'un recul insuffisant, cette année a été exclue de la base d'étude. Le périmètre retenu concerne l'ensemble des maisons ayant souscrit une assurance habitation que le souscripteur soit occupant ou non du risque assuré. La base d'étude contient donc près de 76 000 sinistres sur 19 ans ce qui correspond approximativement à une fréquence de 0,1%.

A fin 2019, le portefeuille MMA représentait 22% du portefeuille du groupe Covéa. La marque MMA est donc sous représentée en proportion parmi l'ensemble du portefeuille habitation. Par ailleurs son exposition n'est pas homogène sur le territoire et dépend fortement de l'implantation historique de la marque. Le choix du portefeuille Covéa permet donc de maximiser la représentativité de l'intégralité du territoire français sur tous les départements. Comme le montre la figure 2.18, l'agrégation du portefeuille des 3 marques permet d'augmenter la connaissance sans pour autant gommer les hétérogénéités de souscriptions au global et par marque.

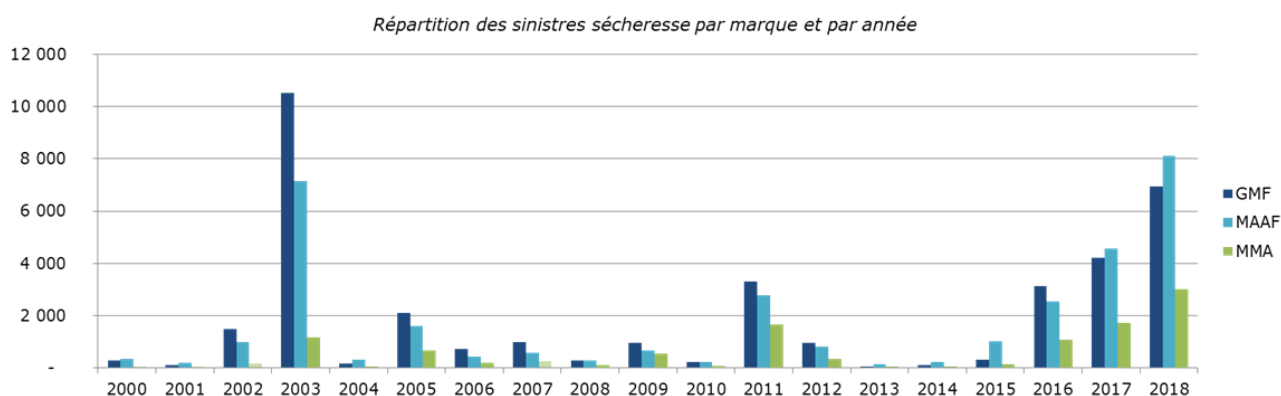
Figure 2.18 : Répartition de l'exposition Covéa par département en 2019



La marque MMA est historiquement moins impactée par le risque sécheresse du fait de profil client / risque différent de MAAF et GMF ainsi que d'une exposition géographique différente. À elle seule, elle ne représente que 14% de la sinistralité sur 19 ans soit une sous-exposition flagrante par rapport au poids de son portefeuille dans le groupe. L'agrégation des données du groupe entier permet donc une meilleure connaissance des sinistres historiques.

La figure 2.19 met en avant la sous-exposition de la sinistralité MMA année après année face aux deux autres assureurs du groupe.

Figure 2.19 : Répartition des sinistres sécheresse déclarés au sein du groupe Covéa



Au-delà de la sélection des sinistres sécheresse se pose la question de la catégorie de sinistres à conserver. En effet lors de l'enregistrement d'un sinistre il est soit pris en charge, soit défini comme sans suite si l'assureur n'a pas à intervenir. Plus de 9% de la sinistralité concerne cette dernière catégorie. Avant d'utiliser ces données, il est nécessaire de s'assurer de la cohérence de ces sinistres avec la problématique du risque sécheresse. Si les sinistres correspondent à de fausses déclarations il est préférable de les éliminer directement de l'étude plutôt que de biaiser les conclusions résultant de l'analyse de ces données.

Afin de conclure sur ces sinistres, ils ont été croisés avec la présence d'argile sur le lieu du sinistre. Sachant que 90% d'entre eux sont compatibles avec une zone à fort risque sécheresse au sens du facteur de prédisposition, il s'agit vraisemblablement de sinistres causés par de la sécheresse pour lesquels l'arrêté de catastrophe naturelle n'a pas été obtenu ou pas sur la période de survenance du sinistre.

Il n'est en effet pas rare d'avoir une date de survenance déclarée 2 ans après en dehors du trimestre couvert par l'arrêté. C'est d'ailleurs ce que l'on peut facilement remarquer en analysant la vie du contrat après déclaration d'un premier sinistre sans suite. Ce dernier est régulièrement suivi d'autres déclarations de sinistres sécheresse pouvant déclencher la prise en charge des dommages par la compagnie d'assurance.

## 2.2.6. Particularité des données externes

L'utilisation de données externes provenant d'un achat de flux à un instant  $t$  pose la question de la continuité d'obtention des données.

Les données SWI présentées au point 2.2.3. proviennent de données achetées par le groupe Covéa auprès de Météo-France sur la période 2000 à 2019. Par choix, l'achat de ces données n'a pas été poursuivi par le groupe sur les années ultérieures. Différents moyens de contournement ont alors été étudiés afin d'assurer une continuité dans l'obtention de cet indicateur.

La gestion des SWI est désormais possible selon 4 indicateurs différents :

- **Le SWI opérationnel** correspond à la donnée achetée à Météo France jusqu'à fin 2019 soit les données sur la période 2000-2019 et qui ne sera pas actualisé. Cet indicateur tient compte des évolutions d'humidité observées en s'appuyant sur les données réelles du sol.

- **Le SWI uniforme** correspond à la donnée réellement utilisée par la commission pour l'acceptation des arrêtés de catastrophe naturelle. Contrairement au SWI opérationnel les spécificités du sol sont blanchies pour considérer le sol français comme une gigantesque plaine homogène. Cette donnée est disponible depuis 2000 et continue à être actualisée annuellement.
- **Les SWI alternatif opérationnel et uniforme** sont des indicateurs retraités par le service interne de Covéa en s'appuyant sur des données obtenues à partir de Copernicus<sup>13</sup> (fournisseur de données spatiales). L'objectif est de se rapprocher de la donnée SWI réelle avec ou sans les spécificités du sol. La profondeur d'historique étant moins conséquente, il est plus délicat de définir un atypisme du phénomène de la sécheresse à partir de ces éléments à date. Néanmoins ces données continueront à être enrichies.

D'après la figure 2.20, la proportion de communes ayant atteint le seuil critique au sens de la réglementation (soit un SSWI en-deçà de -2,06 ou au-delà de 2,06) d'après le SWI uniforme est surestimée de +5 points par rapport au SWI opérationnel. Les indicateurs alternatifs quant à eux affichent des résultats bien au-delà du fait d'un historique de données insuffisant. Ces 2 derniers indicateurs ne seront donc pas utilisés au cours des chapitres qui vont suivre.

Figure 2.20 : Proportion des communes au-delà du seuil critique

Type de SWI	2017	2018	2019	2020
Opérationnel	27,20%	35,82%	18,24%	
Uniforme	30,56%	43,49%	23,87%	21,97%
Alternatif Uniforme	58,51%	53,44%	28,93%	14,32%
Alternatif Opérationnel	51,77%	46,83%	34,70%	34,50%

La première application présentée dans ce mémoire s'appuiera sur le SWI opérationnel car c'était à date la seule donnée disponible. Dans la seconde application, la nécessité de mise à jour des données fera évoluer cet usage par l'utilisation des SWI uniforme.

*Le risque sécheresse apparaît comme spécifique à cause de la mouvance de la localisation des sinistres, de la méconnaissance de la sinistralité au moment de sa survenance réelle et d'une prise en charge par l'assureur dépendant des réglementations en vigueur. Ces trois éléments constituent donc différents facteurs à l'origine de l'aléa de la sinistralité.*

*L'analyse des différentes données exogènes à partir de l'origine de la sinistralité a permis de définir différentes variables permettant d'approcher d'un point de vue géographique ce phénomène. Ces diverses études préparatoires ayant été faites en amont de l'élaboration de ce mémoire, elles s'appuient sur des données figées à fin 2018. En effet les analyses ont été initiées fin 2020 et à cette date la sinistralité de l'année 2019 n'était que partiellement connue. Par la suite, les sinistres sur la période 2019 et 2020 ont pu être obtenus mais n'ont pas engendrés de mise à jour des points présentés dans ce chapitre.*

<sup>13</sup> Site Web [Accès aux données | Copernicus](#)

*En s'appuyant sur les résultats de ces études préparatoires ainsi que sur l'ensemble des données, deux applications vont permettre d'améliorer la maîtrise de ce risque. On s'intéressera dans un premier temps à la définition d'un zonier illustratif du risque d'un point de vue géographique. Dans un second temps la mise en place d'une méthode géographique permettra d'affiner le poids de la sinistralité à affecter par produit sur l'exercice le plus récent.*



## Chapitre 3 – Application au zonier

Quel que soit le risque climatique considéré, l'assuré n'a que peu de moyens préventifs à mettre en place et ne peut que subir sa sinistralité. La localisation géographique est donc un facteur majeur à l'origine de disparité du risque climatique en habitation.

Le premier point étudié concerne donc la **maitrise géographique de la sécheresse**. La révision du zonier technique doit permettre de conforter voire d'optimiser la connaissance actuelle des zones exposées au risque sécheresse afin d'améliorer la maitrise de ce risque dans le portefeuille assuré. L'identification des zones les plus à risque pourrait devenir un outil phare dans la relation assureur-assuré en permettant notamment de travailler autour des moyens de prévention afin de limiter l'impact d'un sinistre sécheresse.

### 3.1. Préambule à la modélisation

Avant toutes modélisations, il est nécessaire de traduire la problématique dans un cadre plus théorique via une formulation mathématique. Cette équation s'appuie sur l'indicateur à estimer ainsi que sur les données expliquant le phénomène.

De plus, l'analyse des variables est une étape clé qui influera la capacité du modèle à prédire correctement le risque. La qualité et la pertinence des données sélectionnées doivent donc être étudiées. Certaines données devront également être retraitées afin de respecter la typologie adéquate pour certains tests ainsi que pour clarifier la lecture des résultats.

En parallèle, les indicateurs de performances seront définis en accord avec l'objectif de la modélisation afin d'estimer la qualité des prédictions résultantes.

L'analyse et le retraitement des variables ont été implémentés à partir du logiciel SAS.

#### 3.1.1. Choix de la variable à modéliser

Contrairement à la modélisation d'un zonier par l'approche classique de la prime pure, la résolution du problème est ici abordée de manière différente pour mettre en avant non pas le risque réellement observé mais le **risque probable face au phénomène de la sécheresse**.

En effet, ce risque climatique est davantage corrélé à des éléments exogènes relevant plus des facteurs géologiques et météorologiques qu'à des critères de l'habitat considéré. Par ailleurs l'action de l'assureur n'intervenant qu'en cas de sinistralité validée par un arrêté de catastrophes naturelles sécheresse, la sinistralité enregistrée est plus que dépendante des règles d'obtention d'un arrêté de catastrophes naturelles qui sont extrêmement mouvantes dans le temps.

La finalité de ce chapitre sera donc d'identifier les communes les plus à risque face à la sécheresse en s'appuyant sur une probabilité de survenance communale. En décorrélant l'analyse de la sinistralité réelle et donc en s'affranchissant de la réglementation, on s'assure d'obtenir un zonier plus pertinent et plus stable face aux modulations réglementaires régulières.

D'un point de vue théorique, il s'agira d'expliquer la **variable binaire présence de sinistre Oui/Non** en utilisant non pas le résultat binaire mais bien la probabilité estimée pour la classification des communes.

### 3.1.2. Sélection des variables explicatives

Comme évoqué dans le chapitre précédent et en lien avec l'objectif de modéliser un indicateur à la commune, la grande majorité des variables explicatives intégrées au modèle seront donc des variables exogènes. Seuls quelques éléments de sinistralité et d'exposition s'appuient sur des informations contractuelles sur la période d'étude de 2000 à 2018. Tous ces indicateurs sont agrégés à la maille INSEE.

Pour synthétiser, les données utilisées dans le modèle peuvent être catégorisées en 4 rubriques (la liste exhaustive des données est disponible en [Annexe 1](#)) :

- Les **variables de risque** dépeignent la sinistralité réellement observée à la commune.
- Les **variables argile** précisent le facteur de prédisposition de la commune.
- Les **variables d'humidité** indiquent l'exposition de la commune face au facteur déclenchant.
- Les **variables exogènes** en provenance de l'INSEE viennent moduler les constats en approchant les typologies de construction.

### 3.1.3. Analyse et retraitement des variables explicatives

Contrairement à un cadre de modélisation standard, l'usage de données presque exclusivement exogènes limite grandement la problématique de qualité. Les différentes sources de variables utilisées étant soit des données officielles soit des données déjà retraitées, la pertinence de ces critères a déjà été analysée au préalable. De simples tests de cohérence sont ici suffisants pour valider ces éléments. A contrario il est fréquent de devoir retraiter voire supprimer des données aberrantes ou manquantes lors de l'utilisation de bases internes.

Pour poursuivre sur l'étude des variables explicatives et comprendre les interactions et dépendance entre ces dernières, une analyse de **corrélation** a été effectuée. En s'appuyant sur le calcul du **V de Cramér** il est aisé d'identifier les corrélations entre données afin d'éliminer les variables redondantes dont l'intégration pourrait biaiser les résultats du modèle.

Le V de Cramér permet de mesurer l'association entre 2 variables qualitatives ayant à minima 2 modalités. Cet indicateur basé sur le **test du  $\chi^2$  de Pearson** est plus usité que l'utilisation directe du test de Pearson car il permet de s'affranchir du biais causé par les grands échantillons.

Le test du  $\chi^2$  de Pearson s'appuie sur la statistique du  $\chi^2$  qui illustre l'écart entre la répartition réelle entre deux variables et la répartition théorique en cas d'indépendance complète de ces dernières. En prenant  $X$  et  $Y$  deux variables avec  $I$  et  $J$  modalités respective et  $n$  le nombre d'observations de l'échantillon, on note  $n_{ij}$  les effectifs observés et  $n_{ij}^*$  les effectifs théoriques obtenus si les variables étaient indépendantes avec  $i \in [1, I]$  et  $j \in [1, J]$ . A partir de ces éléments on peut calculer les effectifs suivants :

$$n_{i.} = \sum_{j=1}^J n_{ij}, \quad n_{.j} = \sum_{i=1}^I n_{ij}, \quad \text{et } n_{ij}^* = \frac{n_{i.} \times n_{.j}}{n}.$$

La valeur du  $\chi^2$  entre  $X$  et  $Y$  est alors définie par

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*}.$$

A partir de la valeur précédente, le  $V$  de Cramér est ainsi calculé :  $V = \sqrt{\frac{\chi^2}{n \times \min(I-1, J-1)}}$ .

Comme la régression logistique ainsi que toute la préparation des données ont été exécutées sous SAS, le  $V$  de Cramér peut être différent de la formule présentée ci-dessus dans le cas de variables n'ayant que 2 catégories. En effet, dans ce cas précis SAS calcule la valeur ci-dessous qui peut entraîner une valeur négative pour le  $V$  de Cramér :

$$V_{SAS} = \frac{n_{11} \times n_{22} - n_{12} \times n_{21}}{\sqrt{n_{1.} \times n_{2.} \times n_{.1} \times n_{.2}}}.$$

Par simplification, les analyses sont faites sur la valeur absolue du  $V$  de Cramér pour n'avoir à traiter que des valeurs positives. L'interprétation reste la même pour les valeurs négatives, plus le coefficient est proche de 0 plus l'indépendance est assurée et au contraire si le coefficient tend vers 1 ou -1 la relation de dépendance est proche.

La figure 3.1 donne l'interprétation du lien existant entre les variables à partir de la valeur absolue du  $V$  de Cramér.

Figure 3.1 : Calcul et interprétation du  $V$  de Cramér

Valeur absolue de $V$	Interprétation
Entre 0 et 0.05	Absence de liaison
Entre 0.05 et 0.1	Très faible
Entre 0.1 et 0.2	Faible
Entre 0.2 et 0.4	Modérée
Entre 0.4 et 0.8	Forte
Entre 0.8 et 1	Colinéarité

Source : Site sur le  $V$  de Cramér<sup>14</sup>

Cet indicateur ne permettant que l'analyse entre variables qualitatives, toutes les données continues (proportion d'argile et variables INSEE) ont été discrétisées de manière automatique en 5 classes comme visible sur l'exemple de la figure 3.2 :

- Isolement de la valeur 0
- Catégorisation par quartile des valeurs strictement positives

<sup>14</sup> <https://lemakistatheux.wordpress.com/2013/05/31/le-v-de-cramer/>

Figure 3.2 : Exemple de discrétisation des proportions d'argile

Classification	ALEA_FAIBLE	Classification	ALEA_MOYEN	Classification	ALEA_FORT
a. 0%	14 966	a. 0%	8 683	a. 0%	24 505
b. 0% - 14.67	4 979	b. 0% - 16.89%	6 549	b. 0% - 12.80%	2 594
c. 14.67% - 47.51%	4 979	c. 16.89% - 51.40%	6 550	c. 12.80% - 43.35%	2 594
d. 47.51% - 85.56%	4 978	d. 51.40% - 89.89%	6 551	d. 43.35% - 84.41%	2 594
e. 85.56% - 100%	4 980	e. 89.89% - 100%	6 549	e. 84.41% - 100%	2 595

Ces classes pourront d'ailleurs être utilisées dans les modélisations à la place des données brutes pour faciliter la lecture finale du modèle et pouvoir identifier quelques profils atypiques.

Certaines variables sont naturellement corrélées par construction comme les différents indicateurs argile et SWI présentés dans le chapitre 1. La préférence pour l'un ou l'autre de ces indicateurs sera actée par analyse des performances lors des modélisations. Aucune décision ne sera donc prise sur ces données suite à l'analyse de corrélation.

Pour le reste, une modalité a été supprimée pour chacun des indicateurs argiles et INSEE afin de s'assurer de n'avoir au maximum que des variables à colinéarité modérée. Par exemple la part des retraités dans la commune ainsi que la part des maisons en résidences principales construites avant 1919 ont été supprimés car l'information se déduit directement des autres catégories présentes.

Pour les variables présentant effectivement une corrélation élevée, l'une a été supprimée soit d'après sa forte interaction avec d'autres variables soit par pertinence avec l'objectif de la modélisation.

Ont ainsi été supprimées :

- La proportion des maisons en résidences principales au profit de conserver la part des maisons par commune,
- La densité de logement au profit de la densité de maison,
- L'exposition moyenne sur 19 ans au profit d'un indicateur simplifié sur la présence ou non de contrats durant cette même période.

Les matrices de corrélations initiale et après retraitement sont visibles en [Annexe 2](#).

Même si la suppression des corrélations est plus fortement préconisée pour les modèles classiques, supprimer ces effets peut également être bénéfique dans le cas de modèles d'apprentissage supervisé. Quel que soit le type de modèle testé, les variables intégrées seront celles sélectionnées après suppression des fortes corrélations.

### 3.1.4. Base de modélisation – base de test

Pour pouvoir comparer les performances de différentes méthodes il est nécessaire de s'appuyer sur une base exclue du calage afin de supprimer l'effet lié à l'apprentissage des modèles. Une partie de la base complète, soit certaines communes, est donc isolée dès le départ dans une **base de test** invariante qui permettra la comparaison de modèles entre eux. Le choix a été ici fait d'engendrer une répartition 70%-30% en faveur de la **base à modéliser**.

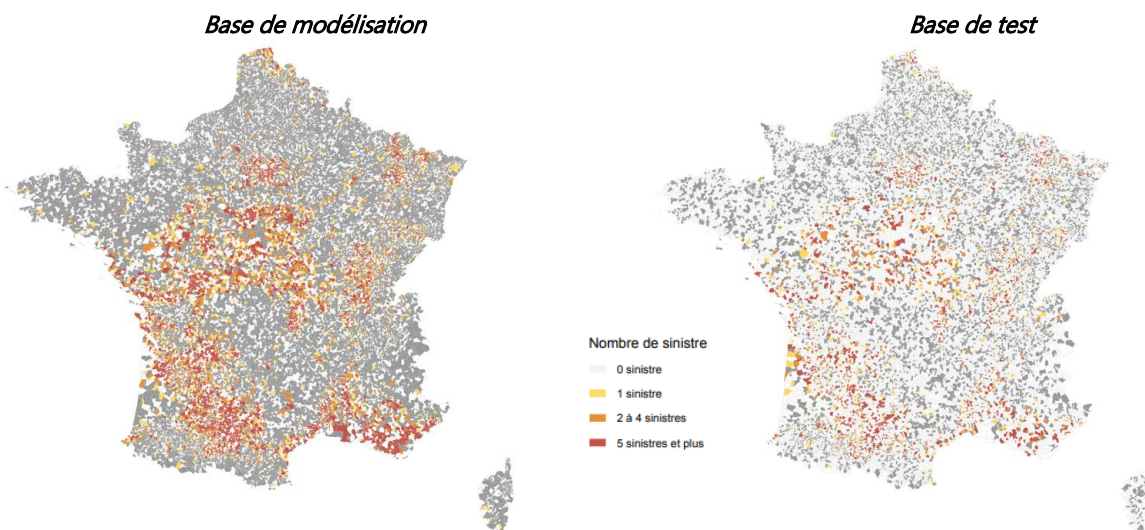
La particularité de la sécheresse réside dans sa sinistralité extrême faible. En effet, que l'on considère la marque MMA, le groupe Covéa voire l'ensemble des assureurs, la fréquence sécheresse reste très minime comparée à d'autres garanties couvertes par l'assurance habitation. Même si l'objectif de modélisation s'appuie sur un historique long de 19 ans, il n'empêche que la variable cible n'est absolument pas répartie de manière homogène : actuellement seules 25% des communes ont observé au moins un sinistre sur la période d'analyse.

L'échantillonnage des communes pour créer la base de test doit donc se faire sous contraintes afin de s'assurer de l'homogénéité des bases créées. Les contraintes retenues ont été appliquées sur les 3 variables suivantes :

- **Indicateur sinistre** pour conserver une homogénéité sur la typologie de commune,
- **Département** pour avoir une représentativité territoriale équilibrée,
- **Densité de maison** pour éviter le déséquilibre selon la structure communale.

La figure 3.3 ci-dessous illustre cartographiquement l'effet de cet échantillonnage sous contraintes. L'ensemble des communes de la France sont répartis entre la base de modélisation et la base de test. Les communes blanches correspondent aux communes non sélectionnées pour faire partie de l'une ou l'autre des bases. Visuellement les différentes zones sinistrées ressortent bien d'une carte à l'autre.

Figure 3.3 : Illustration de la base à modéliser et de la base de test après échantillonnage



Au-delà de pouvoir comparer les différents modèles entre eux il est également nécessaire de pouvoir juger de la qualité du modèle lors de sa création afin d'en améliorer si possible les performances au travers du retraitement de variables ou bien du calibrage de paramètres. Un retraitement identique à celui présenté ci-dessus a également été appliqué sur la base de modélisation afin de l'échantillonner sous les mêmes contraintes en une **base d'apprentissage** (70%) et une **base de validation** (30%).

Au final, la base complète est donc répartie de manière homogène dans les 3 bases suivantes :

- La base d'apprentissage avec 17 649 communes (soit 49% de la base totale) utilisée pour définir les modèles,
- La base de validation avec 7 095 communes (soit 21% de la base totale) visant à optimiser les performances du modèle,

- La base de test avec 10 138 communes (soit 30% de la base totale) pour la comparaison des différentes méthodes après finalisation des modèles.

Quel que soit le modèle testé dans la suite de ce chapitre il s'appuiera sur les bases d'apprentissage et de validation pour son optimisation. La base de test sera utilisée une fois les différents modèles définis afin de comparer leur performance sur un nouveau jeu de données.

### 3.1.5. Choix des indicateurs de performance

Comme l'objectif de la modélisation concerne une variable binaire, les indicateurs classiques d'analyse de performance s'appuient principalement sur la matrice de confusion résultante.

La figure 3.4 présente les 4 indicateurs usuels que l'on cherche à maximiser et dont l'intérêt peut varier selon la problématique à résoudre.

Figure 3.4 : Définition des indicateurs à partir de la matrice de confusion

		Prédictions	
		Score - 0	Score - 1
Modalités à prédire	Indic - 0	VN	FP
	Indic - 1	FN	VP

#### Accuracy ou Exactitude

Taux de prédiction correcte

$$Accuracy = \frac{VP + VN}{Total}$$

#### Recall ou Rappel

Indicateur des faux négatifs

$$Recall = \frac{VP}{VP + FN}$$

#### Precision ou Précision

Indicateur des faux positifs

$$Precision = \frac{VP}{VP + FP}$$

#### F1 score

Moyenne harmonique du Recall et de la Precision

$$F1\ score = \frac{2 \times Recall \times Precision}{Recall + Precision}$$

La qualité d'une prédiction repose directement sur sa faculté à prédire la bonne réponse la majorité du temps. L'Accuracy est donc le premier élément mis en avant lors de la modélisation. Cependant, plus les données à prédire sont déséquilibrées plus la valeur de l'Accuracy peut paraître trompeuse. Dans notre cas 76% des communes n'ont jamais été sinistrées contre 24% avec au moins 1 sinistre sur les 19 dernières années. En se plaçant dans la situation extrême d'un modèle qui ne prédirait que la non sinistralité pour toutes les communes, l'Accuracy serait alors de 76%. Sur un jeu de données équilibrées on pourrait considérer qu'il s'agit de performance acceptable pour autant ici l'intérêt du modèle serait nul car il ne répondrait en aucun cas au besoin d'identification des communes à risque.

Le Recall et la Precision sont deux mesures complémentaires plus ou moins mis en avant selon l'objectif de la modélisation qui vont apporter une information complémentaire sur la pertinence des prédictions. En appliquant le principe de prudence, un assureur va préférer surestimer la



sinistralité potentielle plutôt que la sous-estimer. Les performances du modèle seront donc validées en s'appuyant davantage sur la maximisation du Recall et sur l'évolution de l'Accuracy.

En complément de ces indicateurs, la qualité des modèles sera illustrée par l'aire sous la **courbe ROC** mettant en avant l'exactitude du modèle face à sa convergence et l'aire sous la **courbe Precision-Recall** illustrant l'équilibre entre faux positifs et faux négatifs. Les aires sous la courbe ROC et sous la courbe Precision-Recall vont venir illustrer la qualité du modèle selon l'axe de la prédiction. Elles sont très souvent utilisées comme critère de décision pour des données déséquilibrées car elles mettent en avant la qualité du modèle sur la prédiction du critère le moins représenté.

La courbe ROC permet d'identifier la discrimination du modèle en illustrant la qualité de prédiction sur la qualité sinistrée. L'**AUC ROC** donne la probabilité qu'une commune sinistrée ait un score plus élevée qu'une commune non sinistrée. Une valeur tendant vers 1 indique un modèle parfait, alors qu'une valeur aux alentours de 0,5 illustre un pouvoir prédictif nul.

La courbe Precision-Recall quant à elle évalue la performance globale d'un modèle. Pour un modèle parfait l'**AUC PR** serait à 1 alors que pour un modèle non prédictif l'aire atteindrait le taux de positif réel soit 24%.

## 3.2. Modèle classique - Régression logistique

La régression logistique est un modèle mathématique traduisant la problématique par une simple équation s'appuyant sur les variables explicatives.

L'un des avantages de cette méthode réside dans la facilité de lecture et de compréhension du modèle obtenu. Chaque prédiction pouvant être retrouvée par le calcul, le modèle est également facile à répliquer manuellement sur de nouvelles observations. De même, l'effet de chaque variable peut facilement s'identifier au travers des coefficients affectés à chacune de ces modalités. On peut donc une fois le modèle obtenu comparer l'impact des critères avec l'effet estimé a priori et identifier les profils les plus à risque.

La régression logistique a été travaillée à partir du logiciel SAS.

### 3.2.1. Théorie du modèle

En se basant sur notre problématique on note  $Y \in \{0, 1\}$  la variable binaire à prédire (0 : commune non sinistrée ; 1 : commune sinistrée) et  $X = \{X_1, \dots, X_J\}$  l'ensemble des  $J$  variables explicatives. On note  $n$  la taille de l'échantillon en identifiant une observation par  $\omega$ .

Chercher à prédire la sinistralité d'une commune à partir des variables explicatives revient mathématiquement à calculer la **probabilité conditionnelle**  $\pi(X) = P(Y = 1|X)$  soit la probabilité d'être sur une commune sinistrée d'après les caractéristiques de la commune.

L'estimation de  $\pi(X)$  s'obtient après quelques manipulations sur les probabilités conditionnelles visant à transformer le problème initial en une analyse des caractéristiques communes selon la présence ou non de sinistre. La démonstration en est faite ci-dessous.

En appliquant le théorème de Bayes, on a :

$$\pi(X) = P(Y = 1|X) = \frac{P(Y = 1) \times P(X|Y = 1)}{P(X)}.$$

De même, en appliquant conjointement la formule des probabilités totales ainsi que le théorème de Bayes, on obtient :

$$P(X) = P(X \cap Y = 0) + P(X \cap Y = 1)$$

$$P(X) = P(Y = 0) \times P(X|Y = 0) + P(Y = 1) \times P(X|Y = 1).$$

En réinjectant l'expression de  $P(X)$ , on obtient donc :

$$\pi(X) = P(Y = 1|X) = \frac{P(Y = 1) \times P(X|Y = 1)}{P(Y = 0) \times P(X|Y = 0) + P(Y = 1) \times P(X|Y = 1)}.$$

On peut de la même manière définir la probabilité pour la non sinistralité d'une commune par :

$$1 - \pi(X) = P(Y = 0|X) = \frac{P(Y = 0) \times P(X|Y = 0)}{P(Y = 0) \times P(X|Y = 0) + P(Y = 1) \times P(X|Y = 1)}.$$

On écrivant le ratio de ces deux probabilités, on obtient :

$$\frac{\pi(X)}{1 - \pi(X)} = \frac{P(Y = 1|X)}{P(Y = 0|X)} = \frac{P(Y = 1)}{P(Y = 0)} \times \frac{P(X|Y = 1)}{P(X|Y = 0)}.$$

En analysant chaque terme de la multiplication le premier terme est facilement calculable à partir des observations, seul le second terme est à estimer et fera donc l'objet de la modélisation.

La **régression logistique** cherche à estimer ce ratio en émettant l'hypothèse suivante :

$$\ln\left(\frac{\pi(X)}{1 - \pi(X)}\right) = b_0 + b_1X_1 + \dots + b_jX_j = bX \quad \text{avec } X = \{1, X_1, \dots, X_j\}.$$

Cette expression est appelé **logit**.

On parle alors de méthode semi-paramétrique car l'hypothèse du modèle s'appuie sur un rapport de distribution et non pas directement sur une distribution. L'hypothèse est ainsi moins restrictive et permet de couvrir un plus large panel de distributions.

Cette expression permet alors d'identifier simplement la **probabilité pour une commune d'être sinistrée** en calculant :

$$\pi(X) = P(Y = 1|X) = \frac{\exp(b_0 + b_1X_1 + \dots + b_jX_j)}{1 + \exp(b_0 + b_1X_1 + \dots + b_jX_j)}.$$

Les coefficients  $b_j$  sont obtenus au travers d'estimateur  $\hat{b}_j$  par la méthode du **maximum de vraisemblance**.

$$\text{Quand } n \rightarrow +\infty, \quad \ln\left(\frac{\pi(X)}{1-\pi(X)}\right) = bX \rightarrow \hat{b}X.$$

On suppose que toutes les observations sont indépendantes et identiquement distribuées. Pour chaque observation  $\omega$ , la probabilité conditionnelle de l'état de la commune d'après les critères explicatifs peut s'écrire :

$$P_b(Y(\omega)|X) = \begin{cases} \pi(\omega) & \text{si } Y(\omega) = 1 \\ 1 - \pi(\omega) & \text{si } Y(\omega) = 0 \end{cases}$$

$$P_b(Y(\omega)|X) = \pi(\omega)^{Y(\omega)} \times (1 - \pi(\omega))^{1-Y(\omega)}.$$

A partir de cette expression on peut donc définir la **vraisemblance** sur notre exemple :

$$L(b) = \prod_{\omega} P_b(Y(\omega)|X) = \prod_{\omega} \pi(\omega)^{Y(\omega)} \times (1 - \pi(\omega))^{1-Y(\omega)}.$$

Par passage au logarithme, on obtient l'expression de la **log-vraisemblance** :

$$\mathcal{L}(b) = \sum_{\omega} Y(\omega) \times \ln(\pi(\omega)) + (1 - Y(\omega)) \times \ln(1 - \pi(\omega)).$$

$\hat{b} = \{\hat{b}_1, \dots, \hat{b}_j\}$  correspondent aux coefficients qui maximisent la log-vraisemblance du modèle. Pour se faire, on cherche à résoudre :

$$\frac{\partial \mathcal{L}(\hat{b})}{\partial \hat{b}} = 0.$$

En pratique, des méthodes numériques itératives sont utilisées pour résoudre ce problème et définir les estimateurs. La méthode de **Newton-Raphson** en est une des plus connues et est classiquement utilisée lors de la régression logistique.

Une fois chaque coefficient estimé pour les différentes modalités de toutes les variables explicatives, la probabilité est calculée pour chaque observation. La régression logistique répondant à un objectif de classification, la finalité du modèle est de renvoyer la **prédiction binaire** pour chaque observation en complément de la probabilité résultante.

Ce scoring peut se faire soit à partir de la probabilité obtenue soit directement à partir du ratio  $\frac{\pi(X)}{1-\pi(X)}$  :

$$\begin{cases} \text{Si } \pi(X) > 0,5 \text{ ou } \frac{\pi(X)}{1-\pi(X)} > 1 \text{ alors } Y_{pred} = 1 \\ \text{Sinon } Y_{pred} = 0 \end{cases}.$$

### 3.2.2. Choix des variables explicatives

En traitant un peu plus tôt dans ce chapitre la corrélation des variables, un point a été laissé en suspens concernant le choix des variables les plus pertinentes pour expliquer le facteur de prédisposition ainsi que le facteur déclenchant. En effet, lors de la description des données plusieurs indicateurs ont été créés pour expliquer ces phénomènes.

Pour rappel, la variable argile peut être approchée par :

- L'usage de la part d'habitat pour chaque commune selon le risque argile (faible, moyen ou fort),
- 3 classifications du risque argile reposant soit sur une équation linéaire par pondération des aléas *iSOL6* soit sur la résultante de modélisation *CL\_1* et *CL\_2*.

La variation d'humidité quant à elle peut s'appuyer :

- Sur l'indicateur utilisée actuellement dans l'obtention d'un arrêté de catastrophes naturelles sécheresses *SSWI 3 mois*,
- Sur les indicateurs d'évolution du SWI sur 4 et 5 mois avec analyse de sa fréquence d'atteinte du seuil à -0,5 ou -0,4.

Pour identifier les variables optimales, un modèle de régression logistique a été lancé sur l'ensemble des combinaisons possibles. Sans effectuer de retraitements complémentaires ni d'action sur les données, les indicateurs optimaux ont été choisis en analysant les performances sur la base d'apprentissage et la base de validation. Nuancer les résultats obtenus directement suite au calage du modèle par les performances observées sur la base de validation permet d'effacer l'effet de sur-apprentissage qui peut intervenir sur la base d'apprentissage.

En s'appuyant sur les performances pures du modèle à partir du critère AIC, ainsi que sur des indicateurs de qualité comme l'aire sous la courbe ROC et l'aire sous la courbe Precision-Recall, le choix s'est porté sur l'utilisation des proportions de la commune par typologie d'argile pour le facteur de prédisposition et sur le nombre de fois où l'évolution du SWI sur 5 mois a été en-deçà du seuil à -0,4. Les résultats obtenus sur les différentes combinaisons de variables sont visibles en figure 3.5, la combinaison optimale y est illustrée en rouge.

Figure 3.5 : Performance de la régression logistique selon les variables testées

		Base apprentissage			Base validation		
		AIC	AUC ROC	AUC PR	AIC	AUC ROC	AUC PR
Proportion d'argile	INDIC_SWI_4M_S1	14 166,38	0,851	0,672	7 699,27	0,864	0,685
	INDIC_SWI_4M_S2	14 135,99	0,852	0,673	7 963,70	0,864	0,686
	INDIC_SWI_5M_S1	14 142,88	0,852	0,673	7 697,68	0,864	0,686
	INDIC_SWI_5M_S2	<b>14 107,11</b>	<b>0,853</b>	<b>0,674</b>	<b>7 682,05</b>	<b>0,865</b>	<b>0,687</b>
	INDIC_SSWI	14 166,74	0,851	0,675	7 705,39	0,864	0,684
Classif PRM	INDIC_SWI_4M_S1	14 555,95	0,840	0,658	7 897,15	0,854	0,672
	INDIC_SWI_4M_S2	14 507,71	0,841	0,659	7 886,81	0,854	0,673
	INDIC_SWI_5M_S1	14 531,04	0,841	0,660	7 894,86	0,854	0,673
	INDIC_SWI_5M_S2	14 482,07	0,842	0,661	7 878,04	0,855	0,673
	INDIC_SSWI	14 602,41	0,838	0,660	7 957,53	0,851	0,667
Classif interne 1	INDIC_SWI_4M_S1	14 340,04	0,846	0,666	7 816,88	0,858	0,677
	INDIC_SWI_4M_S2	14 300,30	0,847	0,668	7 809,47	0,858	0,678
	INDIC_SWI_5M_S1	14 316,35	0,847	0,668	7 816,57	0,857	0,677
	INDIC_SWI_5M_S2	14 275,22	0,848	0,669	7 802,04	0,858	0,678
	INDIC_SSWI	14 219,95	0,845	0,668	7 859,36	0,856	0,674
Classif interne 2	INDIC_SWI_4M_S1	14 507,88	0,841	0,660	7 965,88	0,851	0,665
	INDIC_SWI_4M_S2	14 559,40	0,843	0,661	7 955,42	0,851	0,667
	INDIC_SWI_5M_S1	14 479,90	0,842	0,662	7 960,23	0,851	0,666
	INDIC_SWI_5M_S2	14 433,30	0,843	0,663	7 942,96	0,852	0,667
	INDIC_SSWI	14 564,42	0,839	0,661	8 040,90	0,847	0,661

### 3.2.3. Sélection des variables explicatives

Une fois la sélection effectuée sur les critères expliquant les facteurs géologiques et météorologiques, il est nécessaire de s'intéresser à la validation de l'explicativité des différentes variables intégrées au modèle : données argiles, données humidités, données INSEE et données internes. L'[Annexe 3](#) reprend la liste des différentes données conservées dans la modélisation.

En s'appuyant sur les solutions algorithmiques classiques, la sélection finale des variables s'est faite à partir de la **méthode stepwise**. Cette méthode de sélection ascendante identifie les variables les plus pertinentes à intégrer dans le modèle à partir du résultat du **test de Wald**.

Le test de Wald donne pour chaque variable la probabilité qu'elle respecte l'une ou l'autre des hypothèses suivantes :

- **H0** : la variable n'influe pas le modèle
- **H1** : la variable influe le modèle

Le test calcule sous l'hypothèse H0 la contribution de la variable explicative sur la variable à expliquer. Plus la probabilité est élevée, plus le risque que la variable n'apporte que peu d'information au modèle augmente. Classiquement, avec une probabilité supérieure à 5% le risque de non contribution devient non négligeable et la variable est donc à supprimer du modèle.

L'algorithme fonctionne de manière itérative en ajoutant une à une les variables les plus significatives. La méthode stepwise contrairement à la méthode ascendante classique se laisse la possibilité de supprimer une variable précédemment intégrée dans la sélection si avec l'ajout d'une nouvelle elle s'avère ne plus être significative.

La part des maisons construites entre 1991-2005 et 2005-2014 sont 2 critères à supprimer du modèle au sens du test de Wald. Les résultats du test d'indépendance pour l'ensemble des variables explicatives sont présents en [Annexe 4](#).

### 3.2.4. Calage des coefficients et effets des variables

Après avoir validé la pertinence de chaque variable, de la même manière le test de Wald est également utilisé pour valider ou retraiter la pertinence des classes de chaque donnée.

Sur chacune d'elles sont définis deux hypothèses : **H0** :  $b_j = 0$  vs **H1** :  $b_j \neq 0$ .

On définit la statistique de test de Wald par :  $Wald = \frac{\hat{b}_j^2}{s_{b_j}^2}$  avec  $s_{b_j}$  erreur de  $\hat{b}_j$ .

La statistique suit un loi du  $\chi^2$  à  $J$  degré de liberté. La décision de rejeter H0 se fera dès lors que la p-value est inférieure ou égale à 5% ie  $P(\chi_{0.95}^2(J) \geq Wald) \leq 5\%$ .

Si une modalité apparait comme non pertinente, un regroupement est effectué soit auprès de la variable pivot si l'association est sensée soit avec une autre classe proche. Ce process se fait par itération successive jusqu'à ce que l'intérêt de chaque catégorie soit validé par le test.

Une fois le modèle final obtenu, il est aisé d'identifier l'effet de chaque variable à partir des coefficients simulés. La figure 3.6 prouve bien l'intérêt de l'usage des variables géographiques et météorologiques dans la modulation de la probabilité de base. Les variables INSEE viennent quant à elles nuancer également le résultat d'après les caractéristiques sous-jacentes des habitations.

Figure 3.6 : Effet sur la probabilité des différentes variables

Variable	Effet
Probabilité de base	8%
Part des communes avec risque argile fort	↗ - Positif
Part des communes avec moyen risque argile moyen	↗ - Positif
Part des communes sans risque argile	↘ - Négatif
Fréquence evol SWI sur 5 mois hors du seuil à -0,4	↗
Part des maisons par commune	↘
Densité de maisons par commune	↗ - Négatif
Part des CSP 1 - Agriculteur exploitant	↘ - Négatif
Part des CSP 2 - Artisan, Commerçant, Chef d'entreprise	↘ - Positif
Part des CSP 3 - Cadre, Profession intellectuelle supérieure	↘ - Positif
Part des CSP 4 - Profession intermédiaire	↘ - Négatif
Part des CSP 5 - Employé	↗ - Négatif
Part des CSP 6 - Ouvrier	↘ - Négatif
Part des CSP 8 - Autre sans activité professionnelle	↘ - Positif
Part des maisons résidences principales construites entre 1919 et 1945	↘ - Positif
Part des maisons résidences principales construites entre 1946 et 1970	↘ - Positif
Part des maisons résidences principales construites entre 1971 et 1990	↗ - Négatif
Part des maisons résidences principales construites après 2015	↗ - Négatif

Pour obtenir le modèle le plus cohérent possible la notion d'exposition a été intégrée à travers l'indication de présence ou non de contrat sur la commune pour nuancer l'explicativité obtenue sur des communes avec peu d'exposition. Cette notion n'ayant pour objectif que d'aider à calibrer le modèle, elle est par la suite blanchie pour ne pas avoir d'impact sur la valeur finale modélisée. C'est pourquoi l'effet de la variable n'est pas indiqué dans le tableau ci-dessus.

A partir des coefficients estimés variable par variable, modalité par modalité il a été possible de mettre en avant deux grands profils face à la sinistralité sécheresse :

- Le **profil le plus à risque** se trouve sur des communes où la part d'aléa moyen ou fort est extrêmement élevée et pour lesquelles les variations d'humidité sur 5 mois sont très fréquemment en deçà du seuil de -0.4. Ce profil correspond typiquement au phénomène physique de la sécheresse qui s'observe dans des zones argileuses suite à la rétractation du sol.
- Le **profil le moins risqué** quant à lui se trouve sur des communes faiblement exposées au facteur de prédisposition et de déclenchement.

Les variables INSEE modulent les constats en augmentant la probabilité des villes concentrant de nombreux logements y compris des appartements. Les logements sont d'ailleurs plus à risque dans les zones où la population est plus ouvrière et pour lesquelles les constructions ont été faites depuis les années 70. En extrapolant le profil, on retrouve une description des nouvelles banlieues péri-urbaines dont la qualité des matériaux utilisés entraîne une sensibilité plus accrue aux modifications volumique du sol.



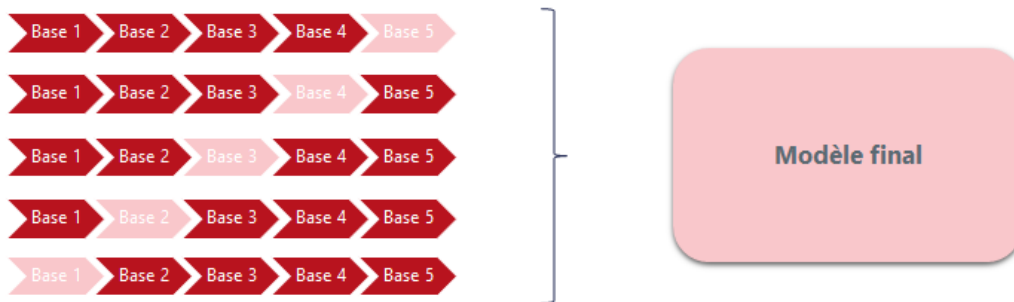
### 3.2.5. Utilisation de la validation croisée

La détermination de la base d'apprentissage et de la base de validation peut influencer plus ou moins fortement le résultat du modèle. En modifiant ne serait-ce que quelques observations, la réaction du modèle peut engendrer des conclusions différentes.

La **validation croisée** permet de lisser cet effet en s'appuyant sur toutes les données présentes. La base à modéliser se voit découpée en K bases (usuellement 5 ou 10 bases). Chaque base sert tour à tour de base de validation pendant que les autres sont utilisées pour caler le modèle. La modélisation est donc effectuée K fois et le modèle final est constitué d'une agrégation des K modèles. Ainsi chaque coefficient estimé correspond à la moyenne des coefficients obtenus sur chaque modèle. La probabilité résultante est donc lissée et stabilisée face à un changement d'échantillon.

La figure 3.7 illustre le fonctionnement de la validation croisée suite à un découpage en 5 classes.

Figure 3.7 : Validation croisée en 5 étapes



Cette méthode a été testée sur la régression logistique précédemment décrite en partant des variables sélectionnées et des modalités retraitées lors du premier modèle.

En zoomant sur l'estimation du coefficient sur un critère, la figure 3.8 décrit bien l'effet de convergence obtenu par agrégation du modèle. De même la probabilité résultante sur une commune converge elle aussi grâce à l'agrégation des modèles, la ville de Draguignan est donné comme exemple en figure 3.9.

Figure 3.8 : Convergence du coefficient sur un critère

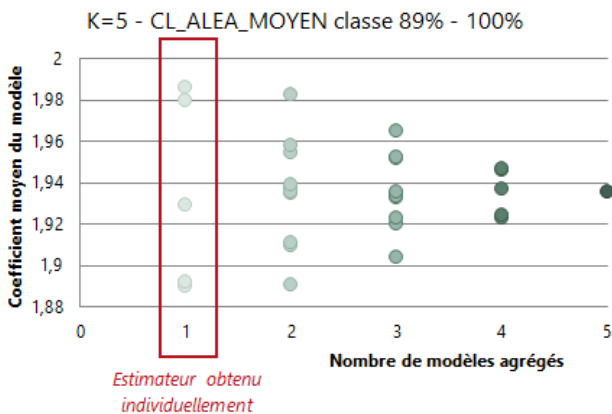
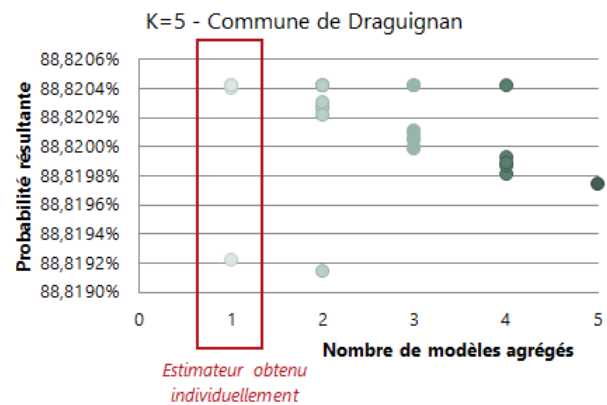


Figure 3.9 : Convergence de la probabilité sur Draguignan



Malgré la convergence visible des coefficients et des probabilités en utilisant la validation croisée, l'effet de convergence ne se voit qu'en zoomant à de nombreuses décimales. L'effet de la validation croisée dans cet exemple est donc assez limité car le modèle initial était déjà stable face à l'échantillonnage. La problématique avait effectivement été anticipée lors du découpage de la base à modéliser en intégrant des contraintes pour homogénéiser la structure des bases. Le visuel est par ailleurs biaisé par la trop forte similitude des modèles. En effet initialement 4 des 5 modèles pour la commune de Draguignan donnent une probabilité à 88,8204%. La combinaison progressive de ces 4 modèles entre eux engendre mécaniquement une probabilité résultante à 88,8204% biaisant ainsi l'effet de convergence.

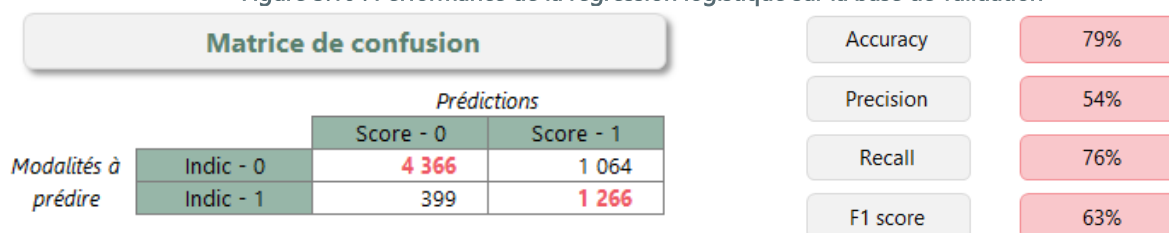
La régression logistique initiale sera donc préférée, le gain de précision face au temps de traitement ne justifiant pas de conserver un modèle plus complexe.

### 3.2.6. Performance du modèle

Les performances du modèle ont été analysées à partir de la base de validation utilisée uniquement pour définir le couple facteur de prédilection / facteur de déclenchement optimal et s'appuient sur les indicateurs précédemment présentés.

Comme l'illustre la figure 3.10 la catégorisation obtenue via la régression logistique n'est pas parfaite et présente quelques erreurs de classification. Cependant malgré 21% d'erreur d'estimation, le modèle est pertinent car il dépasse les performances du modèle estimant la non sinistralité pour toute commune (24% d'erreur). Par ailleurs en surestimant les communes risquées tout en minimisant l'identification à tort de communes sans risque, il respecte le besoin de prudence exprimé précédemment. Le taux d'erreur est en effet principalement porté par des faux positifs visible d'où une valeur du Recall importante.

Figure 3.10 : Performance de la régression logistique sur la base de validation



## 3.3. Modèle d'apprentissage supervisé – Forêt aléatoire

Pour modéliser une variable binaire un autre axe a été exploré, celui des **modèles d'apprentissages supervisés**. L'utilisation d'arbre et par extension de forêts aléatoires permet là encore d'obtenir la probabilité pour chaque commune d'être sinistrée ou non. Mais plutôt que de s'appuyer sur une équation aux paramètres fixes, les modèles d'apprentissages supervisés fonctionnent par itération pour améliorer continuellement leur approche face au problème donné.

Les données utilisées dans les modèles exposées par la suite sont les mêmes qui ont été intégrées initialement dans la régression logistique. Les modèles d'apprentissage supervisé n'exigeant pas

l'utilisation de typologie spécifique de données et les restitutions ne permettant pas un niveau de détail aussi fin que la régression logistique sur l'effet des variables, les données seront cette fois-ci utilisées avec leurs valeurs initiales continues.

Là encore, même si les performances des modèles seront associées au scoring binaire résultant, l'intérêt se portera davantage sur la probabilité pour répondre au besoin initial de zonage.

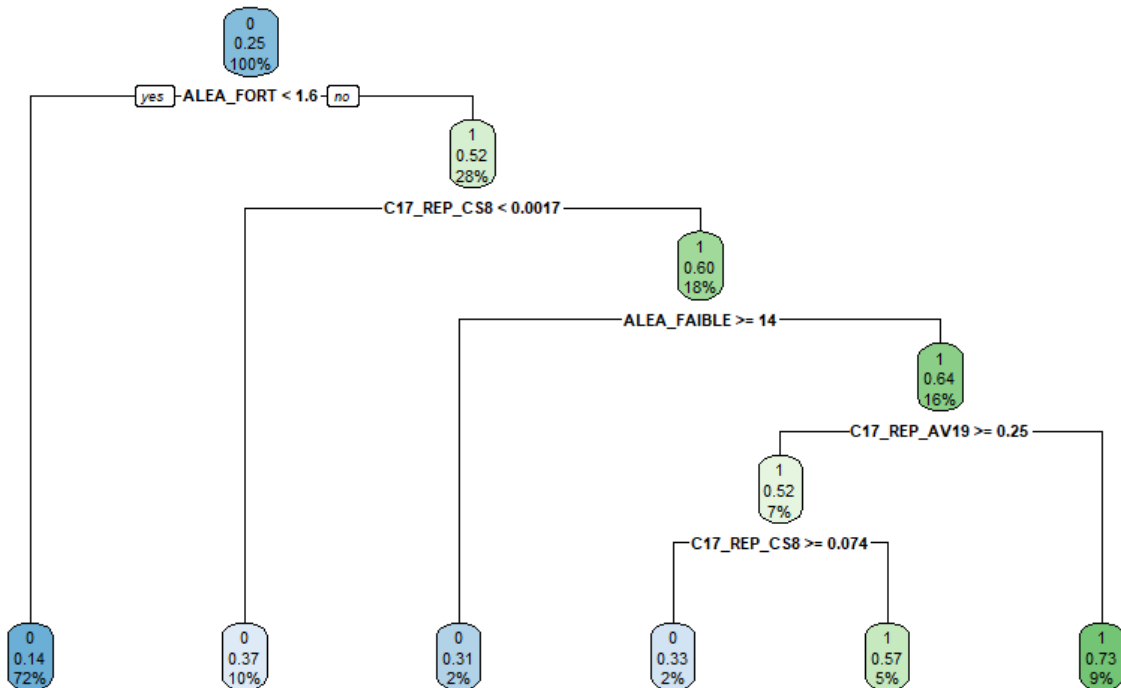
Tous les tests de machine learning ont été travaillés à partir du logiciel R.

### 3.3.1. Arbre seul

A la source de plusieurs méthodes de machine learning : l'**arbre**. C'est un objet rarement utilisé seul mais qui permet de visualiser la composition d'une forêt aléatoire. A chaque itération, l'arbre maximise la prédiction en ajoutant une nouvelle condition permettant d'approcher la valeur réelle à prédire.

La figure 3.11 montre les premiers nœuds d'un arbre identifiant la probabilité pour chaque commune d'être à risque ou non d'après ses caractéristiques.

Figure 3.11 : Arbre de décision pour prédire la sinistralité d'une commune



La force d'un arbre provient également de sa représentation, il est aisé de voir l'effet des variables qui modulent le phénomène à expliquer. Là encore le facteur de déclenchement ressort comme le critère principal à l'origine de la sinistralité. L'algorithme permettant d'atteindre le résultat est également facilement exploitable puisque son descriptif est donné dans les restitutions des logiciels et peut donc être répliqué sur de nouvelles données.

Même si l'arbre est un outil simple et visuel ce n'est pas forcément l'objet le plus robuste. En effet, dès que l'échantillon source est modifié l'arbre peut apparaître complètement différemment et donc être très instable en apparence. Pour essayer de pallier ce problème et d'obtenir une modélisation plus robuste plusieurs arbres sont agrégés ensemble. Cette **forêt** ainsi créée permet de gagner en stabilité et de limiter la variance des prédictions.

La création de forêts peut se faire d'après différentes méthodes qui diffèrent par la manière dont les arbres sont sélectionnés. Deux méthodes d'agrégation ont été testées pour identifier les communes à risques face à la sécheresse.

### 3.3.2. Forêt aléatoire par bagging

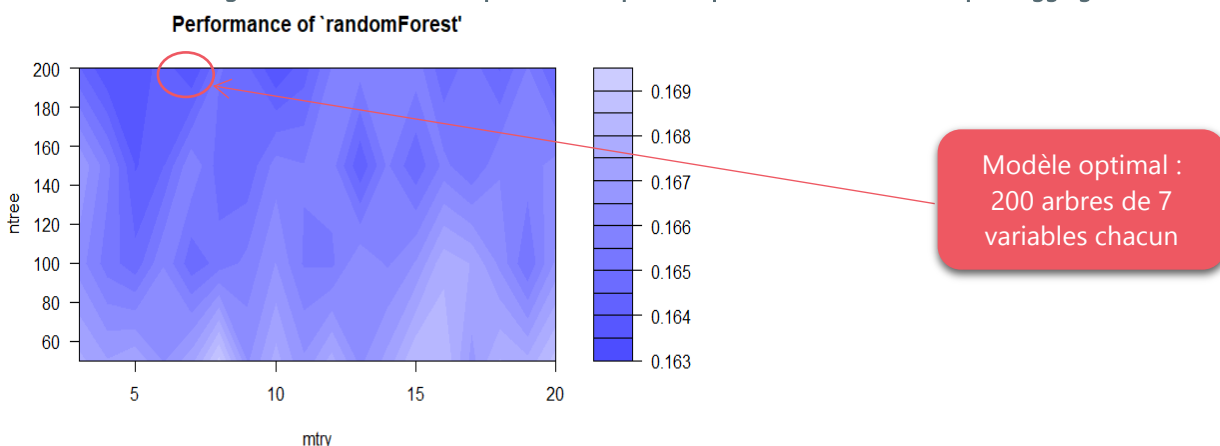
Dans le cadre d'une **forêt aléatoire** utilisant la méthode du **bagging**, chaque arbre se fait sur un échantillon aléatoire de la base d'apprentissage par tirage avec remise. Ces échantillons ne contiennent pas toutes les variables explicatives afin de limiter la colinéarité de chaque arbre. L'agrégation de ces différents objets permet d'obtenir un modèle plus robuste qu'un arbre unique.

Pour obtenir le modèle le plus performant, deux paramètres sont à ajuster :

- Le **nombre d'arbres** venant composer la forêt ;
- Le **nombre de variables explicatives** sélectionnées pour chaque arbre.

L'affinage de ces paramètres s'est fait via la méthode de validation croisée. En croisant différentes combinaisons faisant varier le nombre d'arbres (intervalle testé entre 50 et 200 arbres) avec le nombre de variables (nombre testé allant de 3 à 20), l'analyse de l'erreur d'estimation permet d'identifier la performance de chaque forêt. La figure 3.12 illustre graphiquement le résultat de la validation croisée qui estime les paramètres optimaux permettant de minimiser le taux d'erreur du modèle.

Figure 3.12 : Sélection des paramètres optimaux pour une forêt aléatoire par bagging



Le modèle retenu a donc été donc une forêt aléatoire par bagging composée de 200 arbres chacun constitué de 7 variables explicatives.

### 3.3.3. Forêt aléatoire par gradient boosting

Une **forêt aléatoire par gradient boosting** va permettre d'améliorer l'estimation de manière itérative en se basant cette fois-ci sur la connaissance obtenue des arbres précédents pour corriger les erreurs de l'arbre précédent.

Pour caler cette approche, différents paramètres influent les résultats :

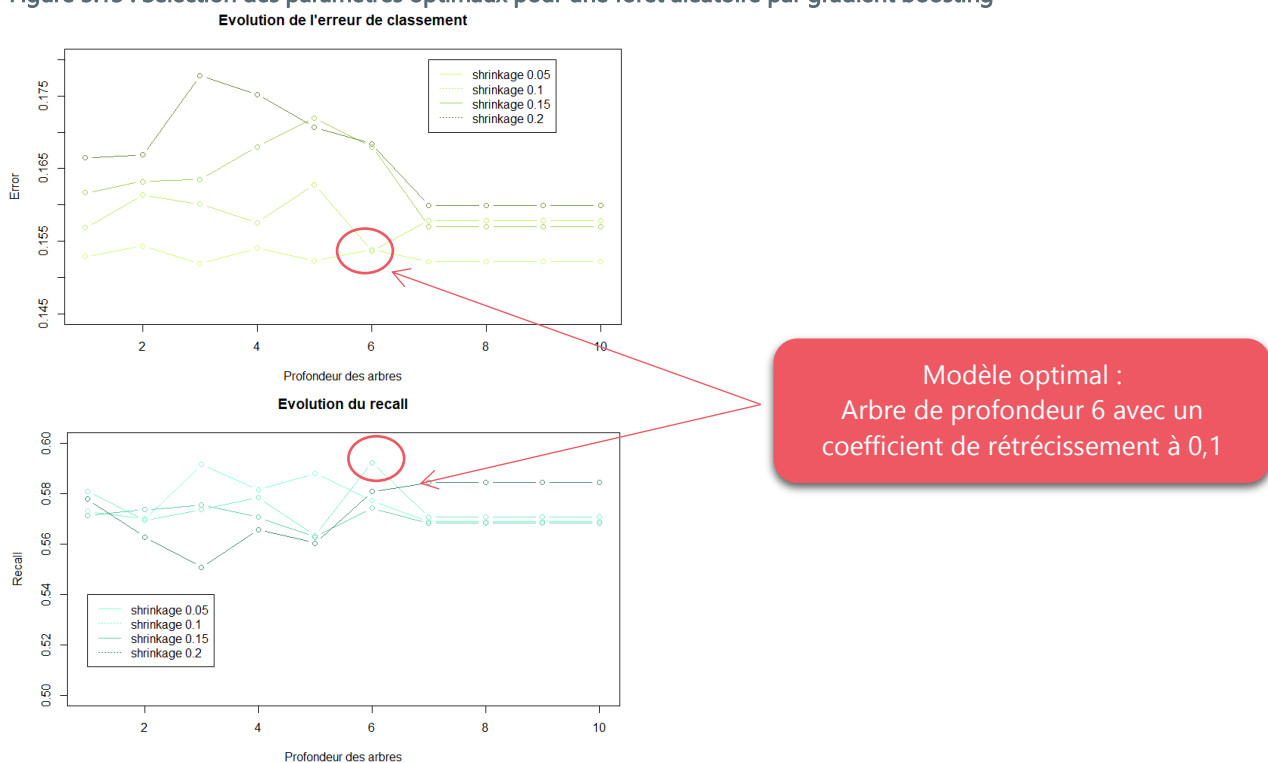
- Le **coefficient de rétrécissement** contrôlant la vitesse de convergence ;
- Le **sous-échantillonnage** à travers le nombre d'arbres utilisés ;
- La **profondeur des arbres** pour définir le nombre de nœuds possibles dans chaque arbre.

Le sous-échantillonnage a été fixé à 200 arbres, les deux autres paramètres ont été calés afin de maximiser les performances du modèle. Pour ce faire, plusieurs modèles ont été lancés en faisant varier le coefficient de rétrécissement ainsi que la profondeur des arbres. Le calage optimal a été obtenu après analyse de l'Accuracy et du Recall pour chaque combinaison de critères sur la base de validation.

Au-delà de rechercher le modèle le plus performant au global, la difficulté réside principalement dans la bonne estimation des communes à risque. En appuyant le choix du modèle sur la maximisation du Recall, la pertinence du modèle est dès lors axée vers le respect de la problématique.

La figure 2.13 illustre la sélection du modèle affichant le meilleur compromis entre les deux critères de décisions choisis.

Figure 3.13 : Sélection des paramètres optimaux pour une forêt aléatoire par gradient boosting



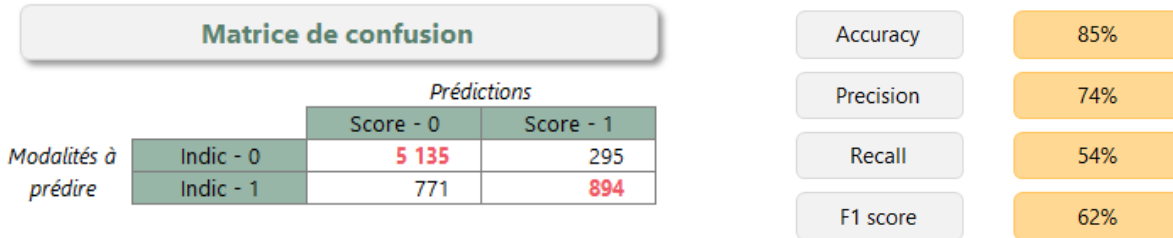
Le modèle sélectionné a donc été la modélisation de 200 arbres constitué de 6 nœuds avec un coefficient de rétrécissement à 0,1.

### 3.3.4. Performance des modèles

Comme pour la régression linéaire les performances sont étudiées sur la base de validation et s'appuient sur la matrice de corrélation ainsi que sur les indicateurs usuels définis à partir de cette dernière.

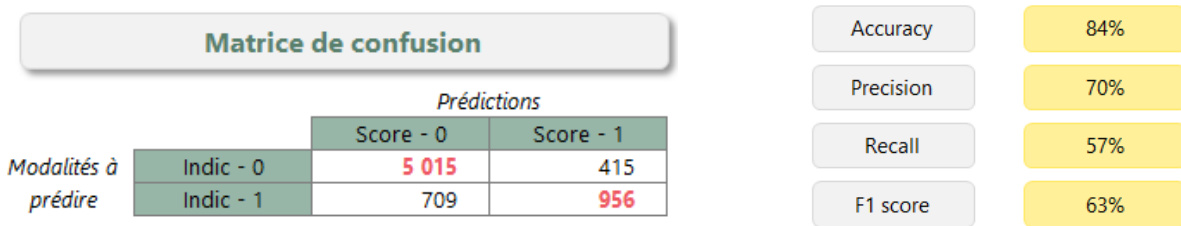
La figure 3.14 indique les performances de la classification obtenue à partir de la forêt aléatoire par bagging.

Figure 3.14 : Performance de la forêt aléatoire par bagging sur la base de validation



La figure 3.15 indique les performances de la classification obtenue à partir de la forêt aléatoire par gradient boosting.

Figure 3.15 : Performance de la forêt aléatoire par gradient boosting sur la base de validation



Ces deux méthodes d'apprentissages supervisés présentent des disparités en terme de performance par rapport à la régression logistique. L'Accuracy est bien meilleure avec ces modèles en limitant le taux d'erreur à 15% là où il dépassait les 20% avec un modèle classique. Dans la résolution d'un problème standard ces modèles seraient donc préférés. Cependant notre problématique ici réside également dans la qualité de l'estimation et notamment dans la prudence des résultats face à la commune à risque.

La régression logistique a tendance malgré son taux d'erreur plus important à orienter ses performances vers une surestimation des faux positifs avec peu de faux négatifs. Autrement dit, peu de communes sinistrées sont oubliées du modèle par contre l'identification à tort de certaines en tant que sinistrées est importante. Les deux modèles de machine learning affichent quant à eux un schéma inversé : le modèle se trompe peu dans l'estimation des communes non risqués par contre il a plus de mal à estimer les communes risqués. En se trompant sur plus de 700 communes à risque contre 400, c'est la modélisation par régression logistique qui apparaît comme étant la plus satisfaisante au regard du besoin sur les données de la base de validation.



### 3.4. Comparaison des modèles

Les 3 modèles ainsi créés, l'objectif est de déterminer le modèle qui répondra le plus favorablement à la catégorisation des communes face au risque de la sécheresse. L'analyse des performances des modèles pris individuellement a déjà orienté le choix vers la régression logistique, mais la décision finale ne peut se prendre qu'en analysant les différentes méthodes sur une base permettant d'analyser la qualité de la prédiction face à l'arrivée d'un nouveau risque soit la base de test initialement isolé.

Puis, au-delà même de la qualité théorique du modèle il faut également pouvoir juger de la pertinence des résultats face à la réalité.

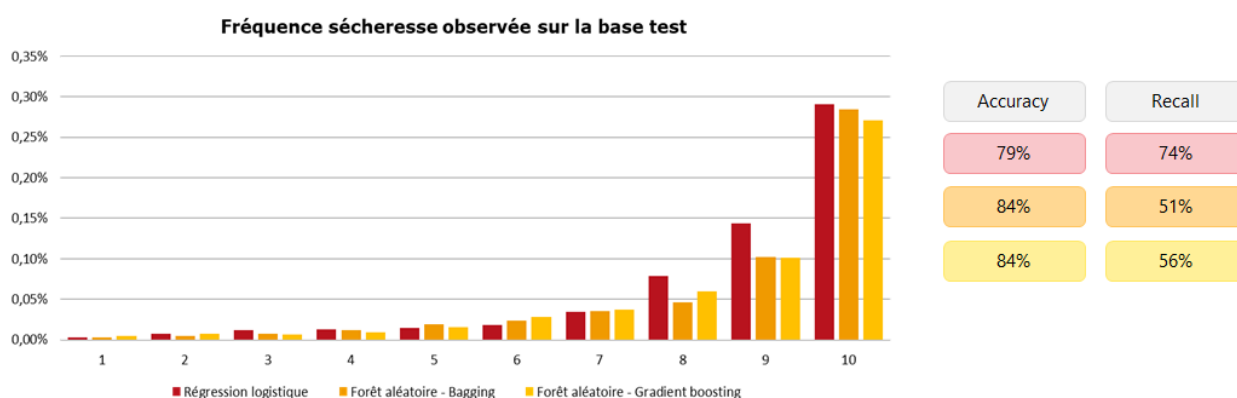
#### 3.4.1. Performance sur la base de test

Pour s'assurer d'une comparaison optimale entre les modèles, l'analyse des performances se fait sur la base de test exclue dès le départ des données utilisées dans le cadre de la modélisation. C'est en effet la seule base qui n'a jamais été utilisée ne serait-ce que pour optimiser les paramètres du modèle. Les performances observées sur cette base sont donc complètement décorrélées de la modélisation initiale.

En plus du calcul de l'Accuracy et du Recall qui illustrent les performances de l'estimation, une analyse de pertinence vient compléter l'analyse des résultats. En effet, le but de cette étude est de permettre d'estimer les communes à risque. Regarder la fréquence est donc un indicateur cohérent pour illustrer la logique des estimations même si certaines communes peuvent paraître et être risquées sans pour autant n'avoir jamais enregistré de sinistres dans le portefeuille Covéa.

La figure 3.16 présente l'évolution de la fréquence après découpage par décile de la probabilité obtenue par commune sur les différents modèles testés, ainsi que l'Accuracy et le Recall sur ces mêmes modèles.

Figure 3.16 : Qualité des modèles sur la base de test



Le phénomène observé sur la base de validation est à nouveau perçu ici : les forêts aléatoires sont plus qualitatives dans leur globalité mais ont du mal à estimer correctement les communes sinistrées.

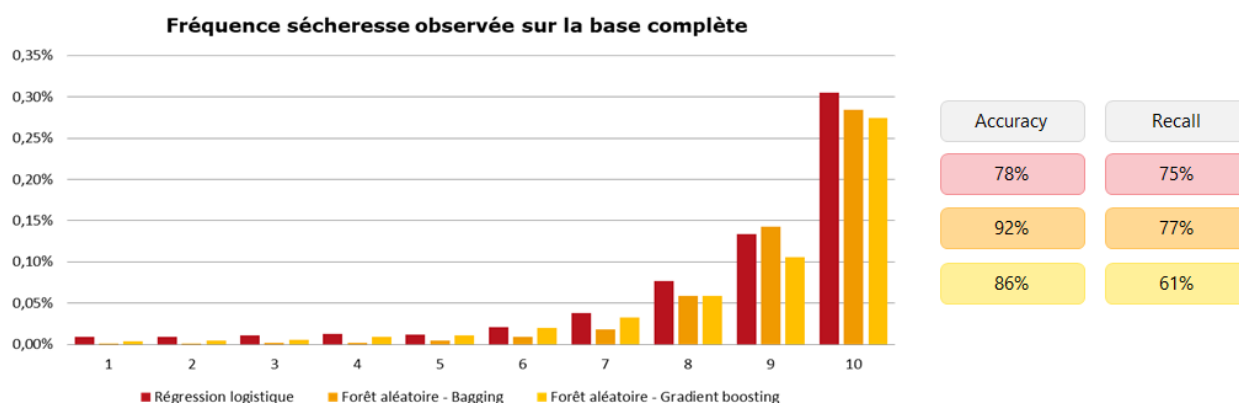
En terme de fréquence, chaque modèle prouve bien sa qualité en affichant une forte segmentation de la fréquence par zone. Néanmoins, la régression logistique apparaît comme plus pertinente avec une segmentation plus régulière. Comme précédemment, c'est donc la régression logistique qui se détache des autres modèles en répondant plus précisément à l'estimation des communes risqués face à la sécheresse.

### 3.4.2. Performance sur la base complète

Après avoir comparé les modèles entre eux sur la base de test il est intéressant de regarder ces mêmes résultats sur la base complète pour analyser le comportement sur un volume de données plus conséquent et représentatif de l'intégralité du territoire de la France métropolitaine.

La figure 3.17 présente ces différents éléments après application des 3 modèles sur l'ensemble des communes.

Figure 3.17 : Qualité des modèles sur la base complète



Les performances sur la base complète illustrent explicitement l'effet de sur-apprentissage sur la forêt aléatoire par bagging. En effet, que ce soit en terme d'Accuracy ou de Recall le modèle apparaît comme étant le plus pertinent. Cependant comme nous avons pu le voir précédemment, ces résultats sont à nuancer avec ceux obtenus sur la base de validation et la base de test. Le modèle est donc instable face aux changements d'échantillonnage car il affiche des estimations quasiment parfaite après calage sur la base d'apprentissage mais n'est pas capable de répliquer correctement ces informations sur une nouvelle base.

Les deux typologies de modèles testés n'ont pas vocation à répondre identiquement à une même question. La régression logistique va porter un point d'intérêt tout particulier à identifier le modèle le plus simple permettant d'expliquer au mieux les observations. Les modèles d'apprentissages supervisés cherchent quant à eux à maximiser la performance quel que soit la complexité du modèle résultant. Orienter le calage vers la performance conduit le plus souvent à du sur-apprentissage ce qui entraîne des performances dégradées dès lors que l'échantillon est modifié. Pour autant ces 2 axes pourraient s'équilibrer sur un problème de classification plus homogène.

Cet exemple illustre bien l'intérêt d'évaluer la qualité d'une modélisation sur une base invariante exclue du calage afin d'observer les réelles aptitudes du modèle. La régression logistique restera donc la modélisation conservée pour répondre au besoin.

Par ailleurs contrairement à la régression logistique, l'utilisation de forêts aléatoires induit une intégration de modèle à l'aveugle. En effet, même si l'arbre est un objet visuel qui permet de comprendre l'effet de chaque variable il n'en est pas de même quand on commence à les agréger. Les modèles d'apprentissages supervisés fonctionnent selon un principe de boîte noire qui ne permet pas de connaître l'algorithme final utilisé pour obtenir la valeur prédite. Il est donc plus compliqué de comparer l'impact des critères dans le modèle avec des hypothèses a priori. Cette validation pratique reste néanmoins souvent nécessaire pour se conforter dans l'utilisation du modèle et faire adhérer les utilisateurs.

### 3.5. Comparaison du zonage

Le nouveau zonier ainsi défini doit permettre d'améliorer la maîtrise du risque sécheresse par une meilleure connaissance des zones à risques et donc une meilleure estimation de la localisation de la sinistralité probable.

#### 3.5.1. Classification par la méthode de Jenks

Toutes les versions du zonier ont été obtenues en classifiant les communes par la méthode de Jenks. La **classification par seuil naturels (Jenks)** permet de classifier les données en optimisant le regroupement des observations semblables dans des classes très différentes. D'un point de vue théorique, cette méthode consiste à maximiser la variance intra-classe tout en minimisant la variance interclasse.

En notant  $K$  le nombre de classe,  $n$  le nombre d'observation à classifier,  $n_k$  le nombre d'observations dans la classe  $k$  et  $y_i^k$  la valeur de l'observation  $i$  appartenant à la classe  $k$ ,  $\bar{y}$  la moyenne totale et  $\bar{y}^k$  la moyenne par classe sont définies comme suit :

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad \text{et} \quad \bar{y}^k = \frac{1}{n_k} \sum_{i=1}^{n_k} y_i^k .$$

A partir de ces notions, les variances inter et intra-classe peuvent être caractérisées par :

$$\text{Variance interclasse} = \frac{1}{n} \sum_{k=1}^K n_k \times (\bar{y}^k - \bar{y})^2 ,$$

$$\text{Variance intra-classe} = \frac{1}{n} \sum_{k=1}^K \sum_{i=1}^{n_k} (y_i - \bar{y}^k)^2 .$$

L'algorithme de classification fonctionne de manière itérative en s'appuyant sur les observations et le nombre de classes souhaitées (notée  $k$  ici) :

- Initialisation du process :  $k - 1$  valeurs aléatoires sont calculées afin de définir les bornes des  $k$  classes. Les observations sont réparties dans les différentes classes à partir des limites de valeurs ainsi définies.

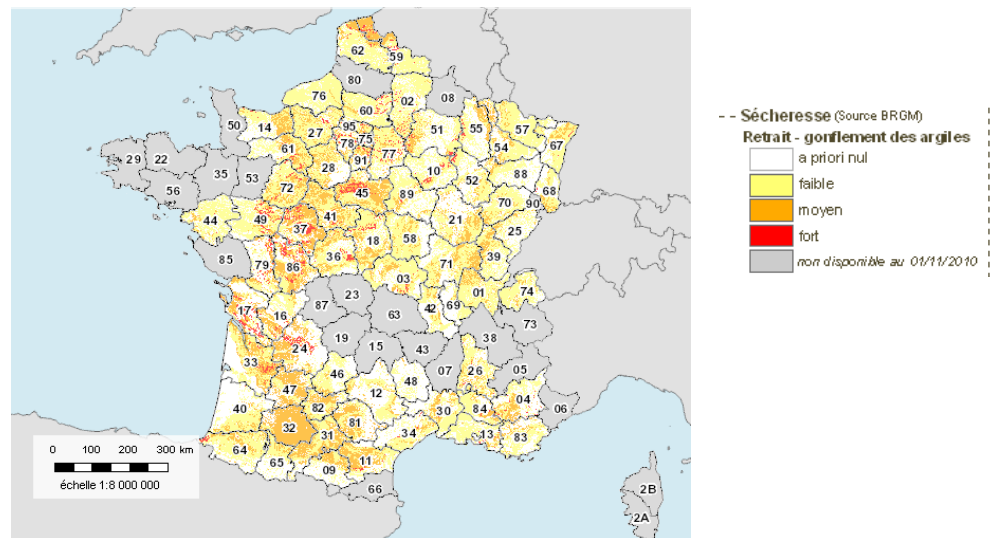
- Calcul des indicateurs de performance : La valeur moyenne de chaque classe est calculée ainsi que l'écart quadratique de chaque individu à sa moyenne  $(y_i - \bar{y}^k)^2$ . La somme de ces écarts est également calculée pour chaque classe et au global.
- Modification des affectations : Les individus présentant l'écart quadratique le plus important sont transférés dans la classe adjacente dont les bornes sont ajustées en conséquence.
- Le calcul des indicateurs de performance ainsi que la modification des affectations sont réitérés jusqu'à atteindre les valeurs optimales pour la variance inter et intra-classe.

### 3.5.2. Comparaison avec le zonier actuel

Le zonier sécheresse actuellement intégré au produit habitation date de 2011. Il avait déjà été travaillé en s'appuyant sur le facteur de prédisposition connu à date à partir d'une carte du retrait et gonflement des argiles publiée par la CCR en 2010. C'est cette même carte qui a été revue récemment par la CCR pour tenir compte des grosses sécheresses enregistrées depuis.

Comme l'illustre la figure 3.18, une vingtaine de départements ne présentaient alors aucune information. Avec les actualisations successives de cette carte, les données ont pu être enrichies sur l'ensemble du territoire.

Figure 3.18 : Carte du retrait et gonflement des argiles en 2010



Source : CCR

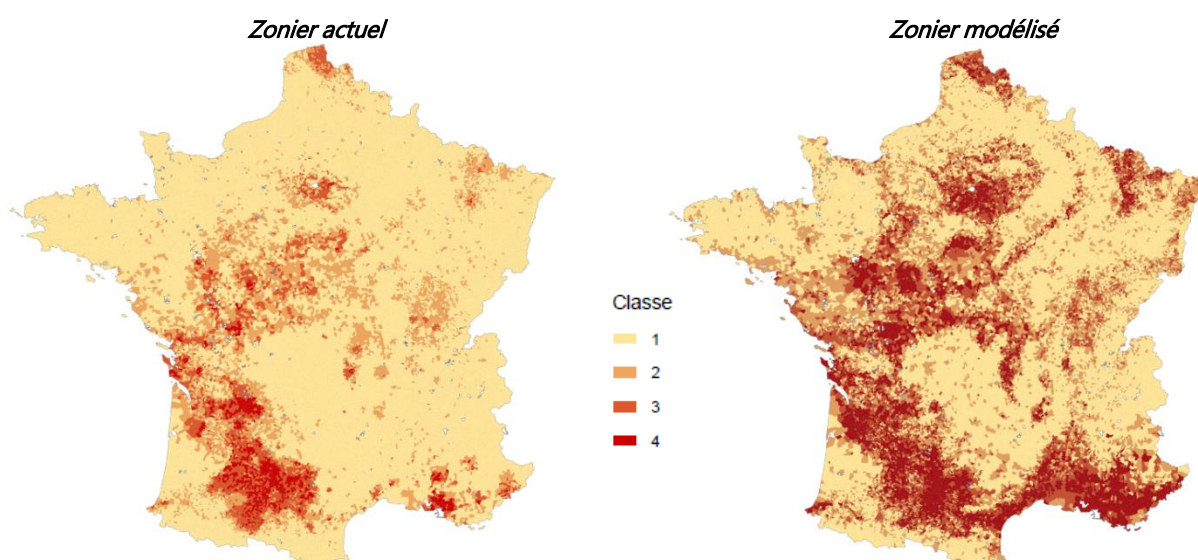
Cette donnée argile avait été approchée pour intégration dans le zonier en s'appuyant sur la liste des arrêtés de catastrophes naturelles obtenus pour de la sécheresse et en croisant cette information avec la taille de la commune d'après sa superficie et sa population, ainsi que son pouvoir politique d'après sa caractérisation administrative (chef-lieu de canton, d'arrondissement, de département ou de région).

La révision de notre zonier s'inscrit donc dans une démarche d'actualisation en s'appuyant sur des données réelles et révisées des sécheresses observables ces 10 dernières années, maximisant ainsi la connaissance du risque.

Le zonier actuel étant discrétisé en 4 classes, une proposition du nouveau zonier est donc également découpée en 4 classes par la méthode de Jenks afin de pouvoir effectuer des comparaisons entre les deux.

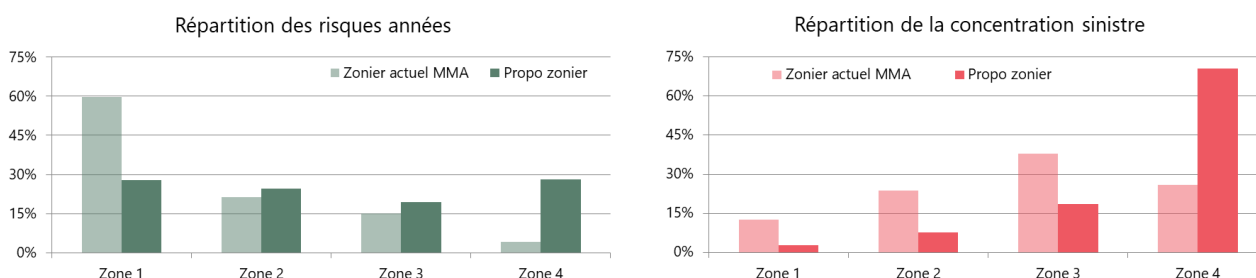
Comme on peut le constater sur la figure 3.19, le nouveau zonier proposé amplifie les zones à risque en respectant cependant leur identification avec le zonier actuel. Auparavant 76% des communes sans risque (zone 1) concentraient 60% du portefeuille de la marque. Avec la proposition du nouveau zonier ce ne sont désormais plus que 52% des communes qui vont abriter moins d'un tiers du portefeuille en zone sans risque.

Figure 3.19 : Comparaison avec le zonier sécheresse actuel



Afin de valider la pertinence du nouveau modèle, certaines répartitions ont été calculée afin d'identifier les effets du nouveau zonier. La figure 3.20 présente les résultats de ces tests sur l'exposition et la sinistralité.

Figure 3.20 : Impact du nouveau zonier sur le portefeuille



Le zonier actuel étant extrêmement segmentant, très peu de communes et donc très peu de contrats étaient considérés à risque accru (zone 4). Le résultat de la modélisation oriente désormais davantage à une plus forte prudence en allant presque jusqu'à équilibrer l'exposition dans chaque classe.

Par ailleurs, la classification des communes est validée par l'analyse de la concentration sinistre. Même si augmenter l'exposition dans la zone 4 conduit forcément à faire baisser la fréquence sur



cette classe (0,7% avec l'ancien zonier contre 0,3% avec le nouveau), les communes intégrées concentrent à elles seules 75% de la sinistralité.

Le nouveau zonier permet donc une meilleure identification des zones à risque d'après la sinistralité observée avec une prudence complémentaire par rapport au risque prédit par le zonier actuel.

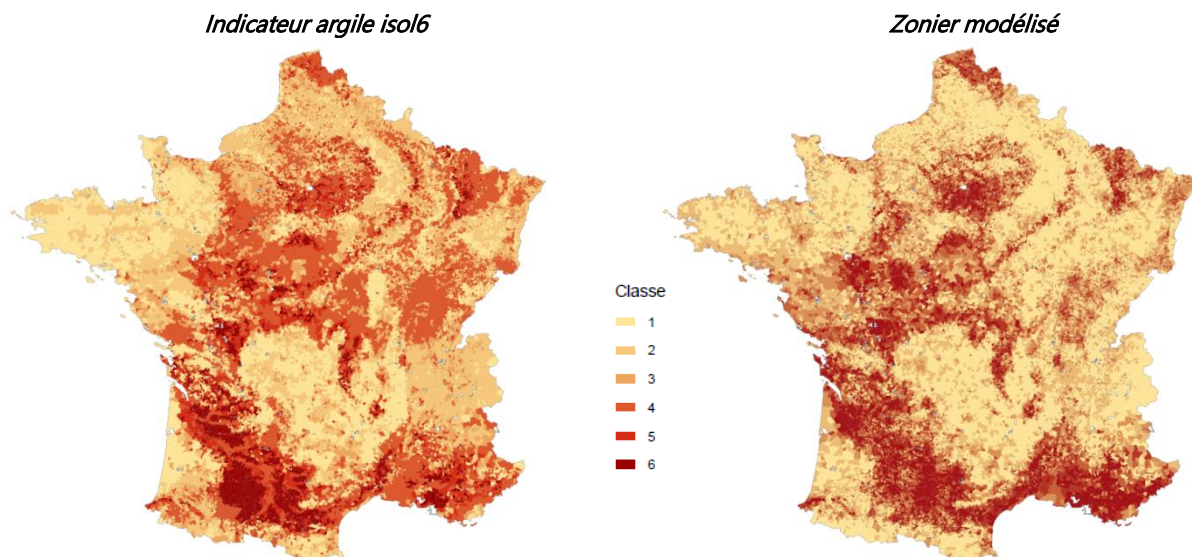
### 3.5.3. Comparaison avec la carte argile

Parmi toutes les variables intégrées au modèle, la variable argile en tant que facteur de prédisposition se doit d'être fortement liée à l'estimation des communes à risque. En effet, de par la construction de la carte RGA, l'absence d'argile coïncide dans 95% des cas avec une absence de sinistralité. Comme cet indicateur fait également partie des critères d'acceptation d'un arrêté de Catastrophes Naturelles sécheresse, il y a peu de chance d'observer de nombreux sinistres en dehors de ces zones.

Sachant que dans la modélisation l'effet argile apparaissait également comme le critère le plus explicatif de la sinistralité d'une commune, comparer le résultat de la modélisation avec la carte argile permet de s'assurer de la cohérence et de l'impact de cet effet dans la restitution finale.

En se plaçant sur le même nombre de catégorie on observe bien figure 3.21 le lien fort entre l'argile et le zonier obtenu. Les autres variables intégrées dans le modèle sont venues nuancer le constat lié à l'impact géologique seul.

Figure 3.21 : Comparaison avec l'indicateur Argile en 6 classes



La classification des communes n'étant pas basée sur la même méthode, il est logique d'observer des disparités entre les deux cartes. Il y a cependant des tendances très similaires mettant en avant les mêmes zones géographiques. Par ailleurs, sur 90% des communes les deux estimations sont identiques à 1 classe près. La variable argile a donc bien toute son importance dans la révision de ce zonier.

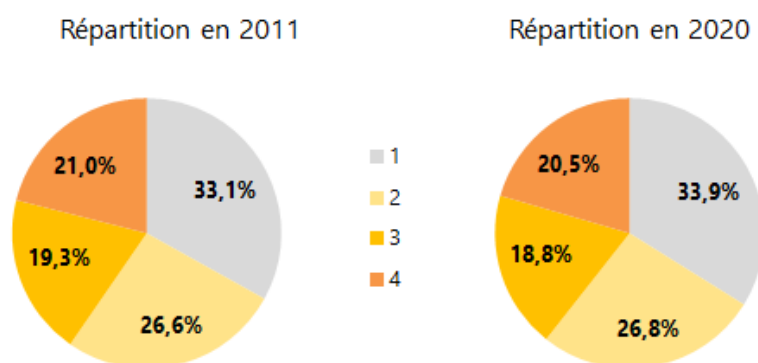


### 3.5.4. Illustration du nouveau zonier face à l'exposition

Le nouveau zonier défini dans cette étude permet d'apporter une meilleure connaissance du risque pris par l'entreprise au sens de la sinistralité sécheresse. Comme évoqué précédemment, la sinistralité sécheresse touche de manière presque exclusive les propriétaires de maisons. En représentant plus de la moitié du portefeuille, il est intéressant d'observer l'évolution du risque sécheresse dans l'exposition MMA.

En reprenant l'analyse du portefeuille historique de la marque présentée dans le 1<sup>er</sup> chapitre, il était clairement apparu que les modifications observées entre 2011 et 2020 n'avaient que peu transformées la répartition des risques d'un point de vue géographique. On observe effectivement ce même phénomène quand on croise ces expositions avec le nouveau zonier technique. La figure 3.22 illustre la répartition de l'exposition en 2011 et 2020 d'après la cartographie du risque sécheresse.

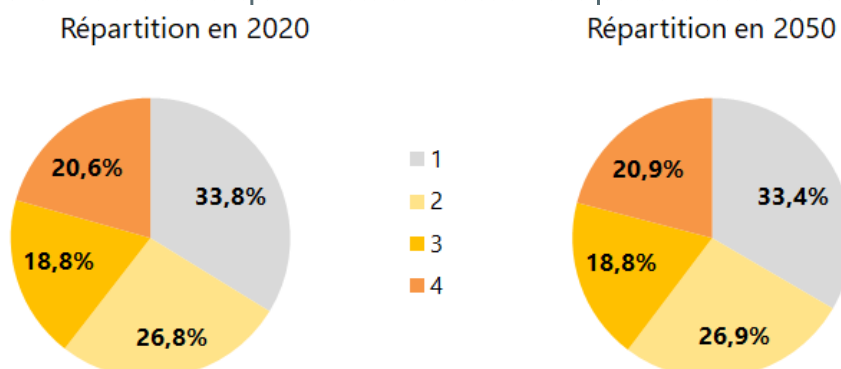
Figure 3.22 : Répartition des propriétaires de maisons selon le risque sécheresse entre 2011 et 2020



L'évolution du portefeuille en 10 ans n'a eu un impact que très limité sur l'exposition des risques au sens de la classification du risque sécheresse du zonier. A noter que la seule distorsion observable a permis de limiter de 0,5 pts le poids des contrats assurés en risque le plus accru.

En s'appuyant cette fois-ci sur la projection du portefeuille MMA en 2050 tenant compte de la migration de la population, le même exercice permet d'identifier une trajectoire possible que pourrait prendre l'exposition face à la sécheresse. Comme visible sur la figure 3.23, la migration de la population n'aurait que peu d'impact sur la répartition des habitations face au risque de la sécheresse. En croisant cette nouvelle structure de portefeuille au nouveau zonier seuls 0,4% des contrats s'avèreraient être en zone plus risquée.

Figure 3.23 : Evolution de la répartition des contrats selon le risque sécheresse à horizon 2050



Aux regards du passé, du présent et d'un futur probable, le portefeuille de MMA ne se modifie que de peu face au risque de la sécheresse. La vocation d'un tel exercice permet de mieux maîtriser le risque pris par l'assureur vis-à-vis de la sécheresse ainsi que son évolution. Cette analyse à vocation à être actualisée de manière régulière afin de pouvoir identifier si à un moment donné l'évolution de la souscription et de la vie des contrats engendre une hausse du risque sécheresse porté par la marque.

*Les différents travaux menés dans le cadre de cette modélisation ont permis tout d'abord de s'acculturer à un phénomène spécifique et surtout de travailler à partir de données externes extrêmement ciblées sur un risque climatique prenant de plus en plus d'ampleur dans le monde de l'assurance IARD.*

*Cette application d'un point de vue géographique a permis de retravailler le zonier actuel avec le souhait d'améliorer la maîtrise et la connaissance du risque sécheresse sur le portefeuille MMA. Le nouveau zonier défini dans ce chapitre intègre de nouvelles zones de risque méconnues jusqu'alors dans la représentation cartographique de risque sécheresse. En se décorrélant des évolutions réglementaires, le zonier finalement obtenu permet de rester le plus stable et le plus explicatif possible du risque face aux différentes évolutions impactant la prise en charge ou non de tel ou tel autre sinistre.*

*En s'appuyant sur cette information il est également du rôle de l'assureur d'apporter une information sur la prévention auprès de l'assuré en identifiant le risque encouru à partir de l'adresse de l'habitation à assurer. Le partage des bonnes pratiques telles que l'élagage régulier des arbres alentours, la vérification de l'état des canalisations et évacuations, l'éloignement au maximum de la végétation et des puits ; permet à la fois une meilleure maîtrise du risque pour l'assureur mais surtout d'éviter tant que faire se peut, un sinistre aux conséquences parfois dévastatrice pour l'assuré. La gestion de la sinistralité sécheresse deviendrait ainsi plus résiliente.*

*Néanmoins, cette étude ayant été la première à être menée dans le cadre de ce mémoire elle s'est appuyée sur un historique sinistre figée à fin 2018. Ne pas avoir intégré la sinistralité des derniers exercices y compris l'année 2022 pourrait engendrer une non identification de certaines zones dans le Nord-Est de la France par exemple. Par ailleurs, l'indicateur de la présence d'argile d'en le sol obtenu à partir de la carte RGA risque de devenir incomplet assez rapidement puisque le BRGM a annoncé une mise à jour à venir de cette carte. Sans pour autant refaire l'entièreté de l'étude avec des données actualisées, il sera certainement nécessaire d'identifier les changements majeurs concernant la base RGA utilisée et la nouvelle base fournie afin de modifier les communes pour lesquelles la présence d'argile et donc le potentiel risque de sinistralité aurait été oublié.*

*Malgré l'identification des zones à risques, la sécheresse est fortement instable d'une année à une autre. Avoir une exposition constante au cours du temps n'est donc pas gage d'une stabilité annuelle de la sinistralité. Afin d'apporter une information complémentaire on souhaite pouvoir définir le poids de la sinistralité sur l'exercice le plus récent vis-à-vis de notre portefeuille. En s'appuyant là encore sur une analyse géographique, on souhaite définir une approche nous permettant d'affiner la connaissance de l'exposition année après année face à cette sinistralité.*

## Chapitre 4 – Application au provisionnement

Le second point traité dans le cadre de ce mémoire concerne la **maitrise de la charge associée aux sinistres sécheresses**. En effet, quelle que soit la garantie étudiée, le **cycle de production inversé** typique du secteur de l'assurance nécessite d'estimer et de prévoir correctement la charge globale des sinistres. Ceci dans le but d'être en mesure de faire face aux sinistres exercice par exercice même si le coût connu à un instant  $t$  sera réévalué par la suite.

La vision **économique** d'un sinistre aussi appelée **ultime** permet d'intégrer une vision prudente du sinistre en estimant son coût final via des techniques de liquidation à partir de la sinistralité observée.

La connaissance de la vision ultime de la sinistralité à un niveau plus détaillé permet de prendre les décisions adéquates lors des revalorisations tarifaires dans le respect de la rentabilité et de la pertinence des produits commercialisés. Il est donc nécessaire d'obtenir une vision précise de la sinistralité sur l'ensemble des garanties couvertes par le produit d'assurance.

En s'appuyant sur l'estimation de la sinistralité faite sur l'ensemble de la branche habitation, l'objectif de cette dernière application est de définir une méthode applicable à la sinistralité sécheresse afin d'être en mesure à partir du portefeuille assuré de définir produit par produit la charge associée à ce risque sur le dernier exercice malgré une déclaration des sinistres incomplète.

### 4.1. Process actuel

L'estimation de la charge sinistre par garantie sur l'intégralité d'un périmètre est calculée lors de l'établissement des business plans. Elle fait ainsi partie intégrante de la définition de la stratégie à adopter à plus ou moins court terme afin de respecter les objectifs fixés par l'entreprise. Pour conserver la compétitivité d'un périmètre il est nécessaire de décliner la stratégie générale sous forme de mesures segmentées. Ces dernières ne peuvent être définies que par une connaissance accrue de la sinistralité à un niveau de détail plus fin de manière à mettre en place des mesures pertinentes.

L'enveloppe globale de la charge pour les exercices achevés est calculée par le service de pilotage économique afin de fournir un coût du risque pour l'ensemble de la branche habitation. Au service de tarification par la suite d'être en mesure de segmenter correctement cette charge pour la démultiplier sur les différents produits et selon différents critères de risques.

Actuellement la sinistralité sécheresse suit un process classique au même titre qu'un sinistre incendie ou dégât des eaux. Pourtant les fortes spécificités géographiques de ce risque ainsi que la méconnaissance de sa sinistralité du fait d'un délai de déclaration long rendent compliquées l'usage de telles méthodes.

### 4.1.1. Utilisation des triangles de liquidation

Afin de répartir le coût économique par produit et aux critères il est d'usage de s'appuyer sur des **triangles de liquidation**. L'usage de triangles permet d'estimer la vision liquidée d'une charge sinistre récente d'après la réalité des années passées. En se basant sur la sinistralité passée des 10 dernières années en vision cumulée, l'objectif est d'identifier sur les exercices les plus récents la charge ultime et donc le montant à provisionner pour tenir les engagements auprès des assurés. L'exemple proposé en figure 4.1 illustre les données sinistres observées sur la garantie dégât des eaux. Pour chaque exercice la charge est connue lors de l'année de survenance du sinistre puis année après année après révision de l'estimation initiale, le montant final n'étant connu qu'au moment de la clôture de l'ensemble des sinistres.

Figure 4.1 : Charge sinistre observée cumulées sur la période 2011-2020

Montant d'Evaluation en K€										
	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
<b>N</b>	176 405	194 881	218 439	202 885	188 183	215 095	194 426	209 749	191 795	173 341
<b>N+1</b>	200 911	207 610	230 602	217 439	197 892	223 823	219 106	246 200	222 181	?
<b>N+2</b>	201 242	207 806	235 335	218 302	195 026	221 725	220 061	245 185	?	?
<b>N+3</b>	201 338	208 120	237 534	213 586	193 314	221 333	221 935	?	?	?
<b>N+4</b>	201 021	211 867	237 366	214 876	192 473	224 029	?	?	?	?
<b>N+5</b>	201 158	211 301	238 152	214 301	193 194	?	?	?	?	?
<b>N+6</b>	200 766	210 781	239 049	215 614	?	?	?	?	?	?
<b>N+7</b>	200 872	212 459	237 048	?	?	?	?	?	?	?
<b>N+8</b>	200 762	212 484	?	?	?	?	?	?	?	?
<b>N+9</b>	201 086	?	?	?	?	?	?	?	?	?

Comme expliqué précédemment, l'intérêt pour le service de tarification n'est pas d'estimer le coût exact à l'ultime mais plutôt de pouvoir en connaître la répartition parmi l'ensemble des différents produits du périmètre et des risques en portefeuille.

La méthode de liquidation utilisée correspond à la méthode de **chain ladder classique** soit un modèle déterministe. Un tel modèle ne nécessite aucune hypothèse sur la distribution des coûts et donc son estimation ne repose sur aucune loi statistique. L'intérêt de ce modèle réside donc principalement dans sa simplicité de mise en place sans condition préalable ainsi que dans sa clarté d'estimation.

En ne supposant aucune distribution spécifique sur les données, tous les sinistres sont considérés comme indépendants et l'estimation de la vision liquidée s'obtient par extrapolation des évolutions constatées sur les années précédentes.

En notant  $i$  la colonne correspondant à l'année d'exercice,  $j$  la ligne correspondant à l'année d'actualisation de la charge et  $n$  le nombre d'exercice présent dans le tableau, le **coefficient de passage** à  $j + 1$  est calculée de la manière suivante :

$$\alpha_{j+1} = \times \frac{\sum_{i=1}^{n-j-1} Charge_{i,j+1}}{\sum_{i=1}^{n-j-1} Charge_{i,j}}, \text{ pour } j \in [0, n - 1].$$

Ce dernier ne s'appuie que sur les données sinistres observées et permet d'estimer la charge au bout d'un an l'en actualisant comme suit :

$$Charge_{i,j+1} = \alpha_{j+1} \times Charge_{i,j}.$$

En reprenant l'exemple ci-dessus et en appliquant la formule d'estimation de la charge liquidée, le triangle de liquidation peut ainsi se remplir pour connaître la vision ultime de la sinistralité sur chaque exercice. La figure 4.2 présente donc le triangle complété toujours sur l'exemple précédent de la garantie dégâts des eaux.

Figure 4.2 : Triangle de liquidation finalisé sur la période 2011-2020

Montant d'Evaluation en K€											Coefficient de passage
	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	
<b>N</b>	176 405	194 881	218 439	202 885	188 183	215 095	194 426	209 749	191 795	173 341	
<b>N+1</b>	200 911	207 610	230 602	217 439	197 892	223 823	219 106	246 200	222 181	190 164	1,097
<b>N+2</b>	201 242	207 806	235 335	218 302	195 026	221 725	220 061	245 185	222 321	190 284	1,001
<b>N+3</b>	201 338	208 120	237 534	213 586	193 314	221 333	221 935	244 803	221 975	189 987	0,998
<b>N+4</b>	201 021	211 867	237 366	214 876	192 473	224 029	223 050	246 033	223 090	190 942	1,005
<b>N+5</b>	201 158	211 301	238 152	214 301	193 194	224 136	223 156	246 150	223 196	191 033	1,000
<b>N+6</b>	200 766	210 781	239 049	215 614	193 483	224 472	223 491	246 519	223 531	191 319	1,002
<b>N+7</b>	200 872	212 459	237 048	215 542	193 419	224 397	223 417	246 437	223 457	191 255	1,000
<b>N+8</b>	200 762	212 484	236 999	215 498	193 379	224 351	223 371	246 387	223 411	191 216	1,000
<b>N+9</b>	201 086	212 827	237 382	215 846	193 691	224 713	223 732	246 784	223 771	191 525	1,002

Sont ainsi traités les montants d'évaluation ainsi que les recours pour actualiser le restant à charge de l'assureur par garantie et pour tous les produits. Une fois chacun de ces éléments obtenus, l'agrégation de la charge totale sur les produits de la branche permet de définir le poids de chacun d'entre eux. C'est à partir de l'application de ces poids sur la sinistralité globale fournie sur l'ensemble du périmètre habitation que l'on peut définir la charge sinistre à l'ultime produit par produit.

#### 4.1.2. Spécificité de la sécheresse

L'utilisation des triangles de liquidation a tout son sens sur une garantie homogène. Cependant lors de son application à la sinistralité sécheresse le coefficient de passage calculé entre l'année 1 et 2 apparait comme exorbitant à cause de la connaissance tardive des sinistres. Ce phénomène est illustré sur la figure 4.3 d'après l'exemple de la sinistralité sécheresse en habitation entre 2011 et 2020.

Figure 4.3 : Triangle de liquidation sur la sinistralité sécheresse

Montant d'Evaluation en K€											Coefficient de passage
	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	
<b>N</b>	15	66	16	32	16	33	17	102	100	111	
<b>N+1</b>	10 316	2 051	18	163	588	7 855	10 722	21 556	11 032	17 928	162,153
<b>N+2</b>	13 649	3 002	21	408	1 506	9 112	11 919	15 258	11 365	18 469	1,030
<b>N+3</b>	14 502	2 866	21	318	1 454	11 041	13 786	16 941	12 619	20 506	1,110
<b>N+4</b>	14 919	3 591	26	320	1 416	12 755	15 075	18 526	13 799	22 424	1,094
<b>N+5</b>	15 842	3 749	25	352	1 506	13 512	15 970	19 625	14 618	23 755	1,059
<b>N+6</b>	15 971	3 796	26	409	1 524	13 670	16 156	19 854	14 788	24 032	1,012
<b>N+7</b>	16 344	3 926	85	421	1 567	14 058	16 615	20 418	15 208	24 715	1,028
<b>N+8</b>	16 355	3 933	85	421	1 568	14 071	16 631	20 437	15 223	24 738	1,001
<b>N+9</b>	16 815	4 043	88	433	1 612	14 467	17 098	21 012	15 651	25 434	1,028

Sur la base d'un historique de 10 ans le premier coefficient pour passer de l'année 1 à 2 engendre une multiplication de la charge initiale par 162 du fait d'une charge encore inconnue à date.

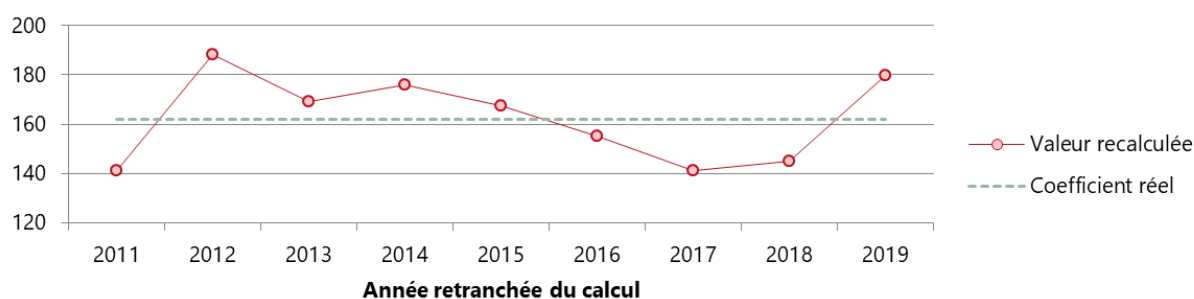
Dans une vision encore plus extrême la charge sécheresse pourrait se retrouver complètement blanchie sur la dernière année tant qu'aucune sinistralité n'est déclarée en année N. Le calcul de la

charge se faisant une fois le dernier exercice achevé, la sinistralité est vue à fin février N+1. Les sinistres sécheresses ne sont alors que très peu voire encore pas déclarés puisqu'ils s'appuient sur des arrêtés de catastrophe naturelle non encore émis. De plus avec l'allongement du délai de prescription suite à la dernière réforme appliquée sur la sécheresse, ce phénomène pourrait s'accroître.

En outre l'évolution du coefficient de passage entre l'année 1 et 2 est extrêmement volatile. L'hétérogénéité des exercices précédents influe fortement sur la valeur de son coefficient. En supprimant itérativement une année dans son calcul, on observe bien une forte variabilité du coefficient de passage. La figure 4.4 présente l'évolution de ce coefficient selon l'année exclue du calcul.

Figure 4.4 : Evolution du coefficient de passage sur la première année

#### Evolution du coefficient de passage entre année 1 et 2



Au-delà de l'aspect visuel de la volatilité, il est également possible d'illustrer ce phénomène en s'appuyant sur un indicateur mathématique de dispersion. La métrique usuellement appliquée est l'**écart type** qui calcule à partir d'un échantillon la moyenne quadratique des écarts par rapport à la moyenne. Cependant, pour pouvoir comparer deux écarts types calculés sur des variables différentes et ainsi statuer sur l'homogénéité ou l'hétérogénéité des variables, il est nécessaire de s'appuyer sur un indicateur insensible aux ordres de grandeur. C'est pour répondre à cette problématique que le **coefficient de variation** a été créé et est défini comme suit :

$$\text{Coefficient de variation} = \frac{\text{Ecart type}}{\text{Moyenne}}$$

En calculant comme précédemment les différents coefficients de passage entre l'année 1 et 2 par retrait successif d'un an sur l'exemple de la garantie dégâts des eaux il est ainsi possible de comparer ces différentes métriques. D'après les résultats exposés en figure 4.5 les coefficients de passage de la garantie dégâts des eaux sont extrêmement stables par rapport à la sécheresse avec un coefficient de variation 18 fois moins important. Sur ces exemples, les données étant différentes avec des ordres de grandeur variables, statuer directement sur l'écart type aurait pu biaiser la conclusion.

Figure 4.5 : Métrique de dispersion sur le coefficient de passage entre l'année 1 et 2

	Ecart type	Moyenne	Coefficient de variation
Dégâts des eaux	0,61%	1,10	0,56%
Sécheresse	1656,40%	162,57	10,19%



L'utilisation des triangles de liquidation est donc pertinente pour des données homogènes mais ce n'est pas le cas pour la sécheresse.

Afin de palier ce phénomène, la sécheresse est actuellement intégrée à l'ensemble de la charge des catastrophes naturelles incluant entre autre la partie inondation. Le fait de cumuler plusieurs événements rend certain la présence d'un coût sinistre non nul dès la première année et permet de lisser la charge sinistre globale finale. En analysant le triangle de liquidation présenté en figure 4.6, l'effet d'incohérence lié à la sécheresse et quasiment invisible du fait d'un risque ne représentant qu'une part minime du coût en année N parmi la totalité de la charge des catastrophes naturelles.

Figure 4.6 : Triangle de liquidation sur la sinistralité catastrophes naturelles

Montant d'Evaluation en K€											Coefficient de passage
	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	
N	4 851	995	5 384	11 475	10 103	29 589	266	16 790	9 839	6 161	
N+1	16 613	3 366	6 168	13 496	10 717	37 592	11 120	39 365	25 076	11 283	1,831
N+2	19 809	4 378	6 601	13 229	10 935	37 272	12 257	33 242	24 946	11 224	0,995
N+3	20 818	4 395	6 649	12 948	10 677	39 786	14 185	34 825	26 135	11 759	1,048
N+4	21 249	5 153	6 708	13 247	10 988	41 728	14 750	36 214	27 177	12 228	1,040
N+5	22 187	5 357	7 169	13 969	11 074	43 483	15 371	37 737	28 320	12 742	1,042
N+6	22 275	5 443	7 592	13 701	11 149	43 777	15 475	37 993	28 511	12 829	1,007
N+7	22 663	5 584	7 633	13 921	11 329	44 483	15 724	38 605	28 971	13 035	1,016
N+8	22 711	5 590	7 647	13 948	11 350	44 567	15 754	38 678	29 025	13 060	1,002
N+9	23 172	5 703	7 802	14 231	11 581	45 472	16 074	39 463	29 615	13 325	1,020

Bien que l'utilisation du triangle de liquidation sur l'ensemble des risques de catastrophes naturelles semble s'appuyer sur des données plus stables, le passage de l'année 1 à 2 reste bien plus volatile que pour toutes autres garanties classiques. En appliquant la méthode présentée ci-dessus pour illustrer de la stabilité de ce coefficient par blanchiment successif d'une année, l'agrégation de l'ensemble des sinistres de catastrophes naturelles permet simplement de diminuer le coefficient de variation de 3 points. Ce dernier restant là encore bien au-delà d'une garantie comme le dégât des eaux.

Pour améliorer la qualité de notre estimation et être en capacité d'identifier la sinistralité sécheresse à l'ultime à affecter produit par produit, il est nécessaire de s'appuyer sur une autre méthode d'estimation. La partie catastrophes naturelles hors sécheresse restera quant à elle estimée d'après les triangles de liquidation car n'étant pas concernée par la déclaration tardive des sinistres, la méthode actuelle est bien adaptée.

### 4.1.3. Choix d'une méthode d'estimation géographique

Diverses méthodes de provisionnement peuvent être utilisées en alternative à la méthode de chain ladder classique. Elles ont pour objectif d'améliorer la stabilité de l'estimation de la charge à l'ultime en intégrant des hypothèses complémentaires au simple usage d'une projection liée à l'historique.

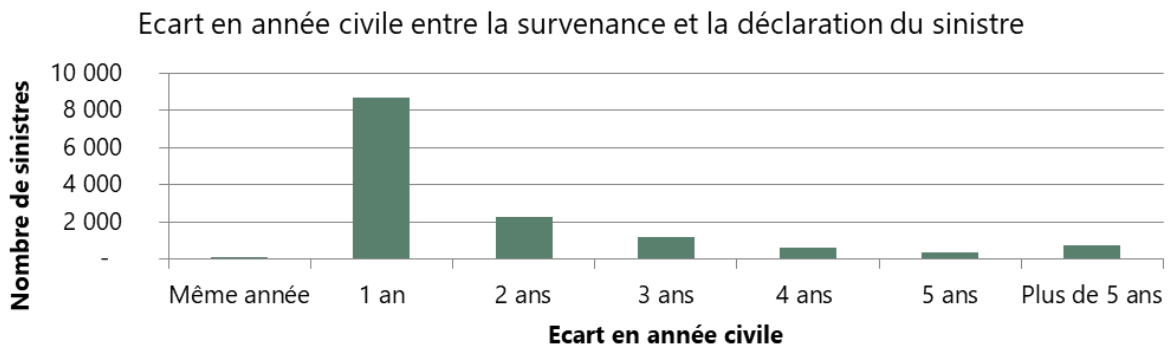
La méthode de **Bornhuetter-Ferguson** permet d'intégrer une notion d'expertise dans l'estimation de la sinistralité à l'ultime en intégrant dans l'estimation de la charge une pondération liée à l'atteinte d'un S/C. Contrairement à la méthode de **Loss Ratio simple**, la charge à l'ultime ne dépend pas que du S/C à dire d'expert mais tient également compte de la liquidation de l'historique observé.

Concernant la méthode de Bornhuetter-Ferguson il est impossible de définir en amont un S/C cible. En effet au même titre que l'ensemble des événements de catastrophes naturelles, les différents contrats IARD ne disposent pas d'une garantie liée afin d'en équilibrer le résultat. Ces sinistres sont assurés majoritairement par le système de solidarité mis en place par l'état puis par les réassureurs. De même, comme l'ont montré les différentes études préalables, la sinistralité sécheresse est un événement imprévisible d'une année sur l'autre. Il n'y a donc aucun sens à chercher à provisionner le dernier exercice en s'appuyant sur une hypothèse d'atteinte d'un S/C cible puisque que ce dernier n'est pas maîtrisable.

On peut également s'intéresser à la méthode **Fréquence-Coût** qui identifie la charge à l'ultime en se basant sur une approche classique en tarification : l'analyse séparée du nombre et du montant. L'estimation dissociée de la fréquence ainsi que du coût permet d'intégrer dans la prévision des phénomènes influant sur les sinistres de manière bien plus précise que dans les méthodes de provisionnement classiques. La force de cette méthode réside dans la prise en compte des déclarations tardives des sinistres. Par exemple, en s'intéressant à l'ajustement d'une loi selon le délai de déclaration on peut ainsi moduler la connaissance de la sinistralité à date en identifiant la part des sinistres déclarés après l'achèvement de l'exercice comptable.

Comme évoqué tout au long de ce mémoire, l'une des particularités de la sinistralité sécheresse réside dans la déclaration très tardive des sinistres. La figure 4.7 illustre la répartition du délai par exercice civil entre la date de survenance et la date de déclaration des sinistres.

Figure 4.7 : Ecart en exercice civil entre la déclaration et la survenance du sinistre



En considérant un historique des sinistres entre 2000 et 2020, moins de 1% des sinistres ont été déclarés durant l'année de survenance. Il est donc complexe avec si peu d'information à la clôture de l'exercice de définir la tendance des sinistres à venir. On observe une déclaration pour près de 60% des sinistres entre 1 an et 1 an et demi après la survenance de l'événement sécheresse

Notons par ailleurs que le but de l'exercice est bien de définir ici une méthode qui sera appliquée sur l'ensemble des produits formant la branche habitation. MMA a fait le choix de laisser coexister plusieurs versions de produits en conservant certaines offres historiques dans son portefeuille. Certains produits sont donc très limités en terme d'exposition. Il n'est donc pas rare que sur l'un ou l'autre le nombre de sinistre déclaré avant la fin de l'exercice soit à 0. Dans ce cas, aucune des méthodes évoquées n'est en capacité à partir d'une sinistralité initiale à 0 d'estimer un coût à venir. Dans l'optique de définir une méthode de provisionnement sur le dernier exercice, il est difficile voire impossible de s'appuyer sur les alternatives citées ci-dessus.

La solution préconisée est donc de s'intéresser à une approche géographique qui s'appuiera sur la détermination des communes à risque ainsi que sur le portefeuille assuré sur ces communes. Une telle méthode à l'avantage de ne pas s'appuyer sur la déclaration de sinistres et permet également de tenir compte de la réglementation en vigueur. En ne se référant pas à l'historique, elle restera toujours utilisable même si les règles d'acceptation des arrêtés sont modifiées à partir de l'année étudiée.

Ce type de modélisation basée sur l'approche géographique à plus large échelle est d'ailleurs largement usité par les différentes compagnies d'assurance via les solutions proposées par les modèles de Moody's RMS. Leader mondial de la modélisation des catastrophes, cette entreprise propose notamment aux assureurs divers modèles apportant une information clé dans la meilleure maîtrise de la sinistralité climatique.

## 4.2. Identification des communes à risque

La vision économique d'un sinistre s'appuie sur les spécificités annuelles de la sinistralité (emplacement, caractéristiques du contrat ...) avec comme objectif l'identification du coût final de ces sinistres. Il faut donc réussir à identifier sur les exercices les plus récents le portefeuille en risque d'être touché par de la sécheresse. Ce sont les caractéristiques de ce même portefeuille qui permettront de répartir la charge sinistre économique par produit.

Pour définir le portefeuille à risque, la première étape est donc de définir annuellement une **cartographie des communes concernées par le risque sécheresse**.

### 4.2.1. Application des règles des arrêtés de catastrophe naturelle

Les communes risquant de déclarer de la sinistralité sécheresse auprès de leur assureur correspondent aux communes qui sont susceptibles d'obtenir un arrêté de catastrophe naturelle pour le risque sécheresse. L'utilisation des indicateurs actuellement intégrés dans l'acceptation des arrêtés seront donc utilisés pour permettre d'estimer les communes potentiellement sinistrées.

Les règles appliquées par la commission depuis 2019 avec effet rétroactif sur les arrêtés déclarés à partir de 2018 sont les suivantes :

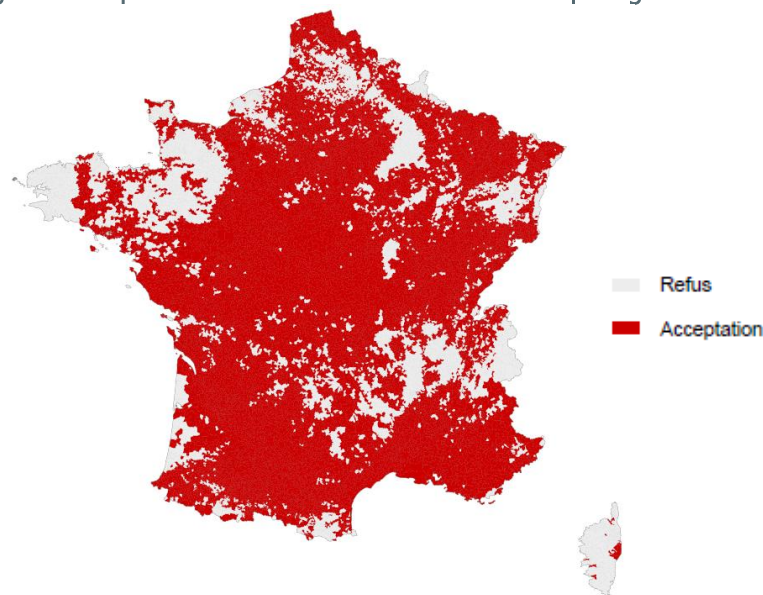
- Présence d'argile en aléa moyen ou fort sur à minima 3% de la commune ;
- Identification d'une période de sécheresse atypique trimestre par trimestre vis-à-vis d'un historique de 50 ans en s'appuyant sur le SSWI standardisé 3 mois.

Ces règles n'ont d'ailleurs que très peu évoluées depuis la modification appliquée en 2016 à effet rétroactif des sécheresses de 2015. En effet, seule la présence d'argile se devait d'atteindre un taux de présence avérée de 10% et non de 3%. Pour le reste les SSWI étaient déjà appliqués pour identifier la période de sécheresse. Par simplicité, la règle de détection de présence d'argile dès 3% sera appliquée sur l'intégralité de la sinistralité, donc depuis 2015. Il pourra donc en résulter une légère surestimation des communes concernées sur la période 2015-2017.

Pour simplifier et identifier plus largement le portefeuille sur les communes à risque, la commune sera considérée à risque sur l'année complète et non pas uniquement sur le trimestre où l'arrêté pourrait être obtenu. Ainsi on s'intéressera à tous les contrats présents au cours de l'année sans pour autant mettre en lien la présence du contrat avec le trimestre atypique d'un point de vue climatique.

L'**effet argile** étant constant au cours du temps, la cartographie des communes au sens du risque argile sera identique année après année. Elle a été établie en s'appuyant sur les proportions d'argiles avec aléa moyen ou fort calculées à la commune, conformément à ce qui est défini dans la réglementation. La figure 4.8 représente les communes à risque au sens du facteur de prédisposition.

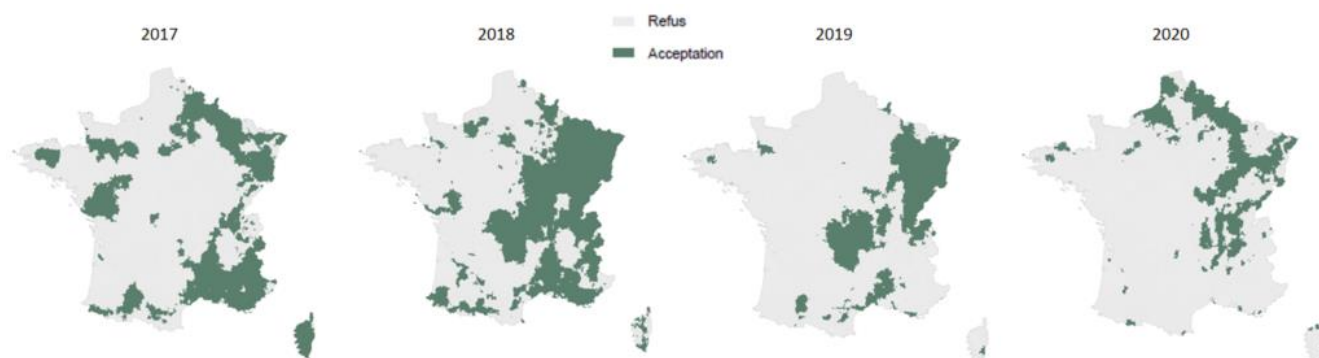
Figure 4.8 : Représentation des communes au sens du risque argile



Le seuil étant fixé à 3% il est logique d'observer 77% des communes concernées par cet événement. Si le seuil retenu avait été le 10% utilisé sur les sécheresses observées entre 2015 et 2017, la carte aurait été sensiblement identique en mettant en avant 72% de communes à risque. Ce peu d'écart confirme donc bien le choix de n'appliquer qu'une unique règle sur l'historique étudié.

Pour identifier l'**effet atypisme de l'humidité** l'indicateur utilisé s'appuie sur le SWI uniforme. Ce critère présenté dans le premier chapitre est la donnée appliquée par la commission pour déterminer les arrêtés, il est donc plus logique de s'appuyer sur cette même variable. Comme ce phénomène dépend d'éléments observés à une date donnée, la cartographie des communes sera différentes année après année. Même si les arrêtés ne couvrent qu'un trimestre, par simplification une seule cartographie par année a été faite en s'appuyant sur l'ensemble des communes qui ont présenté sur l'année au moins un trimestre atypique au sens du SWI. En regardant sur la période 2017-2020 au travers de la figure 4.9, la diversité géographique du phénomène est bien visible.

Figure 4.9 : Illustration du facteur déclenchant sur la période 2017-2020



En cumulant ces deux règles réglementaires et en croisant avec la sinistralité enregistrée, il n'y a que peu de sinistres déclarés qui coïncident avec notre estimation théorique d'arrêtés de catastrophes naturelles pour risque sécheresse des communes. Comme l'illustre la figure 4.10 seuls 43% des sinistres sur la période 2017-2020 répondent aux règles strictes des indicateurs appliqués sur les communes. Ce taux est d'ailleurs variable d'années en années et s'amenuise sur les années les plus récentes du fait d'une connaissance encore incomplète de ces sinistres.

Figure 4.10 : Répartition des sinistres selon l'estimation théorique des arrêtés

Arrêtés	2017	2018	2019	2020
Sans	49,9%	39,1%	64,4%	80,3%
Avec	50,1%	60,9%	35,6%	19,7%

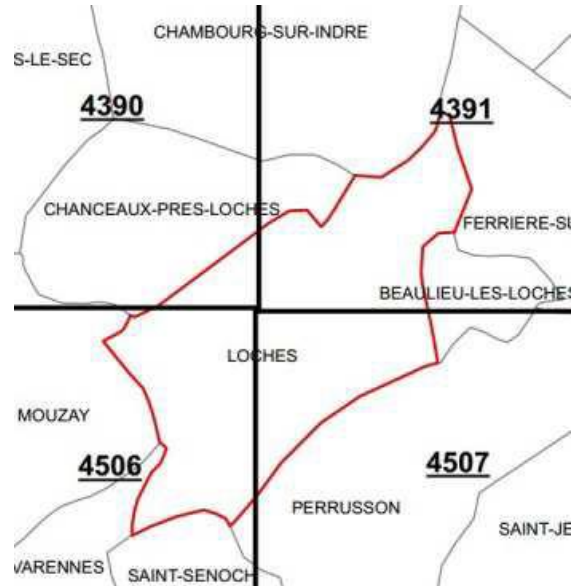
Il est donc clair au vu de ces résultats que l'identification des communes pouvant obtenir un arrêté de catastrophe naturelle pour la sécheresse ne peut uniquement s'appuyer sur les critères théoriques d'argile et d'atypisme pour l'humidité. Sachant que la sinistralité n'est prise en charge qu'en présence d'une acceptation gouvernementale, un ou plusieurs facteurs externes influent sur l'obtention de cet accord.

Il est en effet clairement établi par le gouvernement qu'une dérogation à ces critères techniques peut être obtenue. En se référant à la circulaire transmise aux préfets de département le 10 mai 2019 exposant la « Révision des critères permettant de caractériser l'intensité des épisodes de sécheresse-réhydratation des sols à l'origine de mouvements de terrain différentiels. » on y trouve la mention suivante : « En conséquence, si les zones d'aléa prises en compte recouvrent une surface inférieure à 3% de la superficie du territoire communal, il est demandé à la commune de fournir une étude de sol complémentaire. Toute étude géotechnique disponible en mairie peut être utilement mobilisée (études établies à l'occasion de la réalisation de travaux publics, etc.). Il peut s'agir également d'études géotechniques dont disposent les particuliers ou les entreprises sinistrés. » (extrait de l'Article 3 de la circulaire).

Par ailleurs, le maillage utilisé lors de l'identification du facteur d'humidité est en faveur des communes. La donnée n'étant pas calculée à la commune mais sur un maillage normé par carré de 64km<sup>2</sup>. Toutes les communes présentant une part même minime de leur parcelle dans le carré concerné par une humidité atypique sont potentiellement éligible à l'obtention d'un arrêté de catastrophes naturelles sécheresse. En s'appuyant sur l'exemple de la figure 4.11, si le carré 4390

valide le critère météorologique la commune de Loche est alors éligible à l'obtention d'un arrêté même si les carrés 4391, 4506 et 4507 ne respectent pas le critère.

Figure 4.11 : Illustration du maillage de calcul du facteur météorologique



Source : Article 3 de la circulaire « Révision des critères permettant de caractériser l'intensité des épisodes de sécheresse-réhydratation des sols à l'origine de mouvements de terrain différentiels. »

La donnée telle que nous avons pu la travailler dans ce mémoire a été rattachée à la commune afin de refléter l'évolution réelle de l'humidité observée sur cette commune. La valeur ainsi obtenue est plus en adéquation avec le phénomène déclenchant le retrait et le gonflement de l'argile mais ne permet pas d'illustrer l'effet favorable du traitement effectué lors des demandes d'obtention des arrêtés de catastrophes naturelles sécheresse.

L'ensemble de ces dérogations expliquent donc que l'usage direct des critères techniques ne permettra jamais d'être en adéquation avec les décisions prises par le gouvernement. Il est donc nécessaire d'estimer les communes pouvant obtenir un arrêté.

#### 4.2.2. Modélisation des communes à risque

Pour améliorer la détermination des communes à risque, une modélisation a été testée afin d'intégrer dans l'identification de ces communes quelques données exogènes susceptibles d'intervenir dans l'acceptation de l'arrêté.

La modélisation de la **présence d'arrêté sur la commune** sera obtenue en s'appuyant sur la base **GASPAR**. Cette base de données officielle recense les diverses informations relatives aux risques naturels ou technologiques. De multiples données sont disponibles dans cette base dont la liste est présentée en figure 4.12. Dans le cadre de cette étude seules les « Reconnaissances de l'état de catastrophes naturelles » ayant pour objet la sécheresse seront étudiées.



Figure 4.12 : Présentation des données disponibles dans la base GASPAR

Ce jeu de données correspond aux procédures administratives relatives aux risques (BD GASPAR), mises à jour directement par les services instructeurs départementaux, réunissent des informations sur les documents d'information préventive ou à portée réglementaire :

- les plans de prévention des risques naturels et assimilés et technologiques ;
- les procédures de type "Reconnaissance de l'état de catastrophes naturelles", actualisés dans les 30 jours après leur parution au Journal Officiel ;
- les documents d'information préventive :
  - dossier de Transmission d'Information au Maire (TIM) ;
  - document d'Information Communal des populations sur les Risques Majeurs (DICRIM) ;
  - Plan Communal de Sauvegarde (PCS) ;
  - Atlas des Zones Inondables (AZI).

Source : Site [georisques.gouv.fr](https://www.georisques.gouv.fr)<sup>15</sup>

La modélisation ayant pour but la prédiction des arrêtés futurs, il est plus pertinent de ne s'appuyer que sur les 2 dernières réglementations en vigueur car assez semblables. La période intégrée dans le modèle correspondra donc aux arrêtés obtenus entre 2015 et 2020.

Afin de générer une équation simple, peu de données compléteront les éléments à l'origine de la survenance de la sécheresse. La prédiction de l'arrêté s'appuiera donc sur :

- L'effet prédisposition déjà identifié au travers de la cartographie des communes ayant une part de risque argile au-delà de 3% ;
- L'effet déclencheur d'après la valeur absolue maximale du SSWI 3 mois atteint sur l'année ;
- L'effet administratif en intégrant la densité de maison et la population sur la commune.

La modélisation ne s'appuie pas directement sur l'identification de l'atypisme lié à l'humidité afin de conserver une marge d'erreur sur la valeur de l'indicateur. Par ailleurs, l'intégration des données INSEE permettra de moduler également l'effet strict des règles dans l'obtention de l'arrêté. En effet il est probable qu'une commune ayant peu d'habitants voire peu de maisons ne fera pas les démarches pour obtenir une reconnaissance d'état de catastrophes naturelles. A contrario, une commune fortement peuplée pourrait obtenir une dérogation sur l'un ou l'autre des critères techniques de par son exposition face au risque.

Comme pour la modélisation du zonier technique les variables quantitatives ont été transformées en variables qualitatives pour en simplifier la lecture. Les modalités ont également été retravaillées afin de respecter le test de Wald et de ne conserver que des variables aux facteurs les plus pertinents possibles.

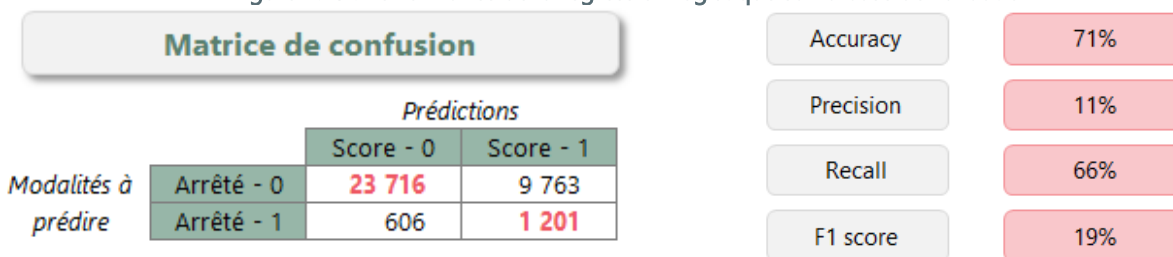
Le modèle a été calibré sur l'ensemble des arrêtés observés sur la période 2015-2019 afin de conserver l'année 2020 comme base de validation.

La figure 4.13 présente les performances du modèle obtenu sur l'année 2020. L'erreur d'estimation reste conséquente avec 29% de mauvaises affectations principalement portée par les faux positifs. Cette anomalie est facilement explicable par le facteur humain à l'origine de la décision. Les arrêtés répondent à un besoin émis par les collectivités, en l'absence de demande il n'y a donc pas de décision rendu même si d'un point de vue purement géographique la commune le mériterait.

---

<sup>15</sup> <https://www.georisques.gouv.fr/donnees/bases-de-donnees/base-gaspar>

Figure 4.13 : Performance de la régression logistique sur la base de validation



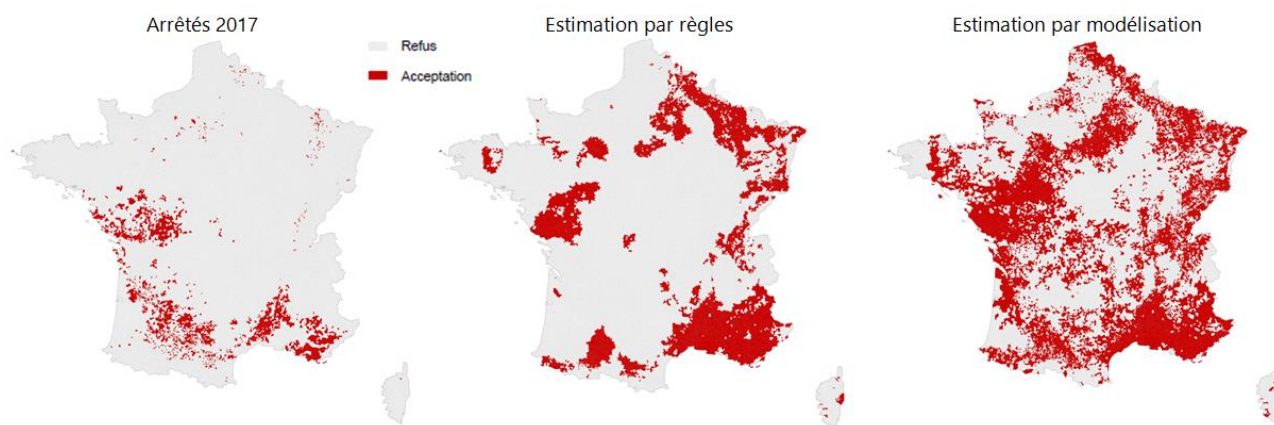
Pour autant cette prédication est nettement meilleure qu'en appliquant strictement les règles puisque désormais 81% des sinistres coïncident avec des communes estimées comme pouvant avoir un arrêté pour cause de sécheresse. Comme le montre la figure 4.14 ce taux est là encore un peu moins performant sur les exercices récents pour cause de sinistralité incomplète à date.

Figure 4.14 : Répartition des sinistres selon l'estimation modélisée des arrêtés

Arrêtés	2017	2018	2019	2020
Sans	14,9%	6,9%	32,9%	30,9%
Avec	85,1%	93,1%	67,1%	69,1%

Malgré les performances contestables de la modélisation, cette estimation reste plus pertinente et permet ainsi de définir une cartographie annuelle des communes risquant d'obtenir un arrêté de catastrophe naturelle. La figure 4.15 illustre sur l'année 2017 les deux méthodes d'estimations par rapport à la réalité des arrêtés actés à date.

Figure 4.15 : Illustration de l'estimation des arrêtés sur 2017



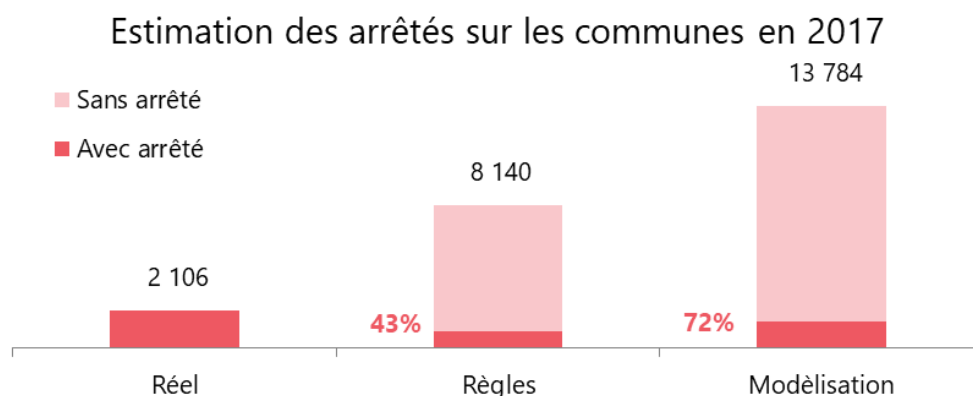
De par la construction de notre cartographie des risques il est légitime d'engendrer une surestimation du nombre de commune à risque. Il ne s'agit que d'une pré-estimation géographique, les demandes réelles d'arrêtés doivent donc principalement se trouver parmi la liste de ces communes mais toutes ne demanderont pas en arrêtés pour cause de sécheresse.

Dans cette exemple la modélisation a permis d'extrapoler les communes légitimes au sens des critères techniques pour intégrer les communes ayant réellement obtenu un arrêté sur toute la partie aquitaine de la France.

Le graphique présenté en figure 4.16 met en avant les correspondances entre les arrêtés sur l'année 2017 et les communes estimées selon les deux méthodes. Quelle que soit la technique

appliquée, le nombre de commune à risque est largement surestimé par rapport aux réels arrêtés. La modélisation affiche certes une multiplication par 6,5 dans l'estimation des communes à risque pour autant 72% des communes ayant réellement eu un arrêté de catastrophe naturelle sont présentes dans son estimation contre 43% uniquement en s'appuyant sur les règles strictes.

Figure 4.16 : Correspondance arrêtés et estimations sur les communes sur 2017



Même si la modélisation maximise fortement le nombre de communes par rapport au réel, cette méthode sera préférée de par sa pertinence dans l'estimation des communes ayant eu des arrêtés et ayant été sinistrées. L'estimation reste bien évidemment perfectible et ne représente qu'une approche de la réalité. Toute la difficulté résidant dans la connaissance imparfaite des arrêtés pouvant être obtenue année après année puisque toutes les communes légitimes n'en font pas forcément la demande.

C'est donc à partir de cette cartographie des arrêtés probables que sera défini le portefeuille à risque permettant ainsi de répartir la charge économique selon ses caractéristiques.

### 4.3. Identification du portefeuille à risque

Une fois les communes définies, le portefeuille à risque est estimé en s'appuyant sur les contrats en vigueur à fin d'année dont les adresses de risques se trouvent sur ces communes. La connaissance des critères et du produit du contrat permettront par la suite de répartir la charge sinistre selon le poids de chacun de ces éléments dans le portefeuille complet.

#### 4.3.1. Périmètre produits sélectionnés

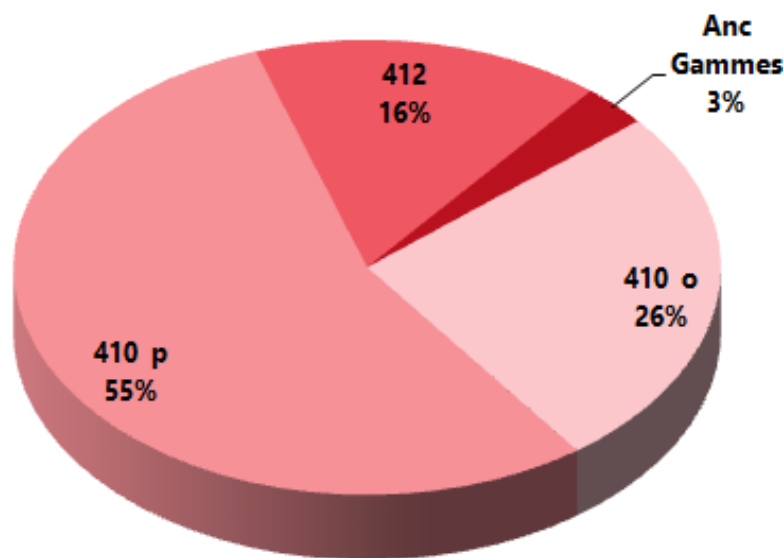
Ces contrats sont sélectionnés sur l'ensemble des produits du périmètre habitation concerné par le risque sécheresse soit l'intégralité des contrats couvrants les propriétaires. Comme évoqué au préalable, seule cette population pourra faire jouer son assurance dans le cas d'un sinistre sécheresse.

La connaissance des caractéristiques de ces contrats permettra d'alimenter le découpage de la charge sinistre selon le produit. Pour rappel, actuellement sur le portefeuille habitation de la marque MMA plusieurs produits coexistent du fait de nombreuses offres historiques.

Les différents produits suivis sont les suivants dont leur poids est représenté sur la figure 4.17 :

- Les contrats 410 o sont des contrats habitation permettant l'assurance des résidences principales et secondaires. Suite à une évolution d'offre ce contrat est désormais clos à la vente depuis 2020 mais contient encore une part importante du portefeuille dont 70% des maisons pour les propriétaires occupants.
- Les contrats 410 p concernent l'assurance active des résidences principales et secondaires et son portefeuille détenait 30% des maisons pour les propriétaires occupants à fin 2021.
- Les contrats 412 permettent l'assurance des propriétaires non occupants d'un bâtiment d'habitation.
- Les contrats Anciennes Gammes correspondant aux anciens contrats hérités des différentes fusions et de l'histoire de la marque sur l'intégralité des assurances touchant à l'habitat.

Figure 4.17 : Répartition des contrats en portefeuille à fin 2021 selon le produit



#### 4.3.2. Définition du critère de répartition

Une fois le portefeuille connu, l'objectif est de répartir la charge sinistre à partir de ces contrats risqués du point de vu de la sécheresse. Plusieurs clés de répartition peuvent être utilisées afin de définir la charge à affecter sur chaque périmètre.

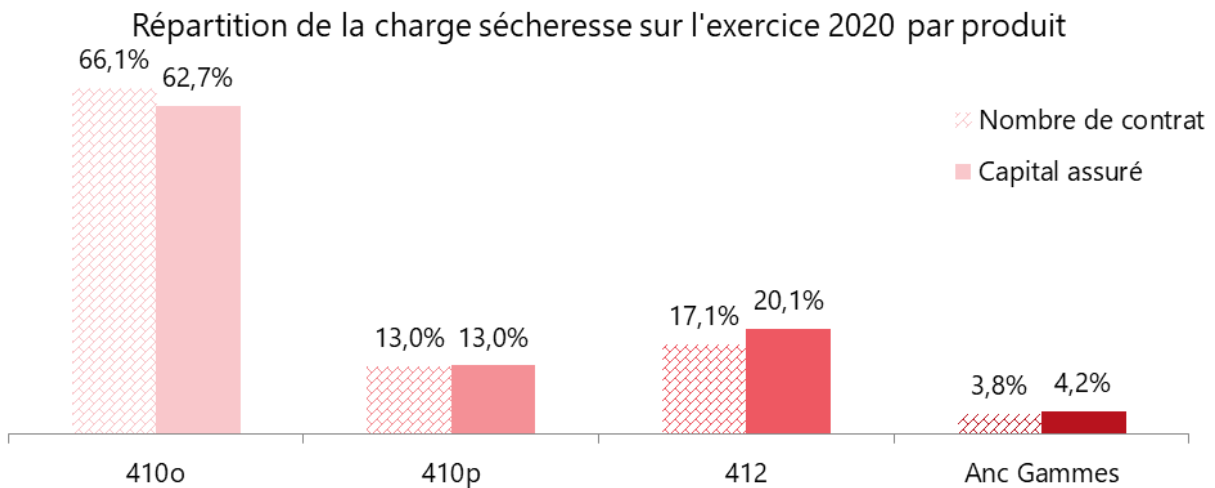
Une première solution est de s'appuyer sur le dénombrement des contrats par produit. Le **risque année** qui tient compte du temps de présence du contrat sur l'année pourrait être utilisé. Pour autant, la sinistralité sécheresse n'étant absolument pas lié au comportement client il est illusoire de penser qu'un contrat peu présent est moins risqué. Par ailleurs avec l'allongement du délai de prescription, un assureur pourra se retrouver fréquemment à prendre en charge des sinistres de contrat dont il n'est plus détenteur à la date de déclaration de ce dernier. S'appuyer sur le **nombre de contrat** est donc plus pertinent pour dénombrer les risques.

Une deuxième possibilité est de s'intéresser au **coût immobilier du bien** en s'appuyant sur le capital assuré. On peut ainsi mieux appréhender la typologie du risque qui est assuré et tenir compte du coût probable du sinistre sécheresse. Le capital immobilier n'est pas une donnée connue pour les

contrats, cependant dans le cadre de Solvabilité 2 une estimation est faite de ce capital pour chaque bien afin d'avoir une meilleure connaissance de nos portefeuilles. Ce capital est calculé en multipliant la superficie des habitations assurées par le prix du m<sup>2</sup> d'après les valeurs vénales immobilières et foncières publiées annuellement aux éditions Callon<sup>16</sup>. L'usage de ces données permet de définir une valeur cohérente avec le marché en différenciant le prix par département et par logement. A titre indicatif une maison se voyait attribuer un prix du m<sup>2</sup> compris entre 1300 €/m<sup>2</sup> et 1700 €/m<sup>2</sup> en 2020.

L'utilisation du capital immobilier plutôt que le nombre de contrats permet de tenir compte des différences entre les contrats assurés. Comme on peut le voir sur la figure 4.18 le capital assuré va venir maximiser le poids à affecter sur les anciennes gammes et l'assurance des propriétaires non occupant (412) par rapport à leur nombre réel.

Figure 4.18 : Décomposition de la charge sécheresse par produit sur l'année 2020



Les vieux contrats couvrent souvent des risques complexes et grands qui ne feraient plus partie du périmètre d'acceptation selon les bornes actuelles de nos contrats. Il est donc essentiel de ne pas sous-estimer ces risques. Il est d'ailleurs plus pertinent de s'appuyer sur un montant dès lors qu'il s'agit de répartir une charge et non une fréquence. C'est donc l'utilisation du capital assuré qui sera retenu comme clé de répartition dans la suite de cet exposé.

#### 4.4. Comparaison avec le process actuel

Pour pouvoir comparer les résultats du nouveau process défini ci-dessus avec la méthodologie actuellement utilisée il est nécessaire de se ramener à périmètre équivalent et donc d'intégrer en complément du retraitement de la sécheresse la partie autres événements catastrophes naturelles composée notamment des sinistres inondations. L'estimation à l'ultime de cette garantie n'a pas été remise en cause dans le cadre de cet exposé, elle reste donc déterminée à partir des triangles de liquidation par chain ladder classique.

<sup>16</sup> Cotes annuelles des valeurs vénales immobilières et foncières

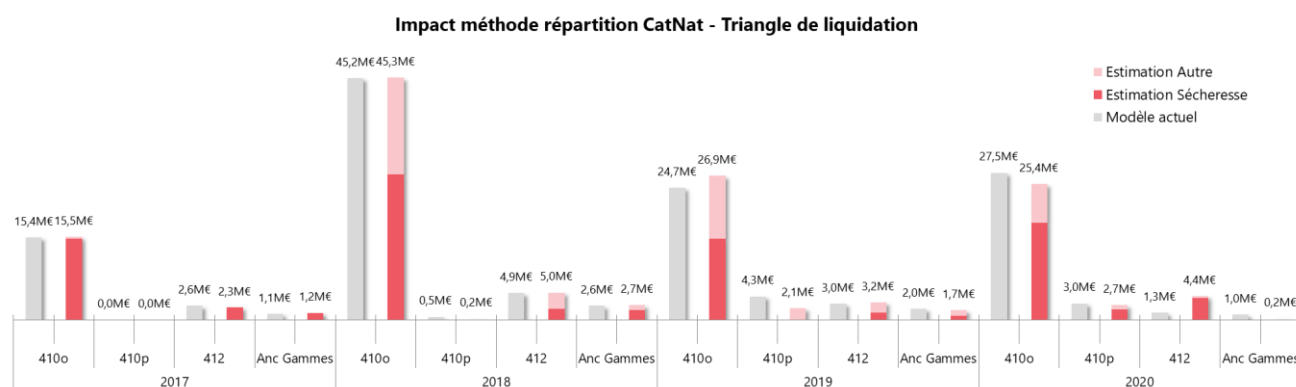
#### 4.4.1. Charge économique à partir des triangles de liquidation

En conservant la méthode des triangles de liquidation, la seule modification par rapport à l'existant est donc de dissocier la partie sécheresse des autres risques de catastrophes naturelles et de définir un triangle de liquidation par événement dissocié pour chacun des 4 produits présentés ci-dessus.

La répartition de la charge économique globale se fait donc en s'appuyant sur la part des coûts sinistres à l'ultime du produit parmi le coût global de l'ensemble de la branche pour chacun des deux groupes de sinistres.

La figure 4.19 illustre l'effet de l'utilisation des triangles de liquidation en différenciant les produits et la garantie sur la période 2017-2020.

Figure 4.19 : Evolution de la répartition de la charge économique sur la période 2017-2020



Sur la période 2017-2018, il n'y a que peu d'écart entre les méthodes malgré le découpage des événements assurés en catastrophes naturelles. Par contre, sur les exercices les plus récents les variations sont bien plus conséquentes car la répartition s'appuie sur des triangles instables dont le premier coefficient est souvent exorbitant concernant la sécheresse. A titre d'exemple, sur l'année 2020 la charge estimée sur le produit majoritaire en portefeuille (410 o) serait tout événement confondu de 25,4M€ dont 18,1M€ imputée à la sécheresse, contre 27,5M€ avec la méthode initiale. La dissociation des événements engendre donc une répartition légèrement différente par produit.

Par ailleurs, pour les portefeuilles les moins exposés, il n'est pas rare que le risque sécheresse soit complètement élué à tort la première année. C'est le cas dans l'exemple ci-dessus sur le périmètre des anciennes gammes pour lequel aucun sinistre sécheresse n'a été enregistré sur l'exercice 2020 avant la fin de l'année civile. En se plaçant en vision fin juin 2022, près de 60 000 € de sinistres ont pourtant été déclarés. Anticiper une charge sans déclaration initiale n'est cependant pas possible à partir de triangle de liquidation.



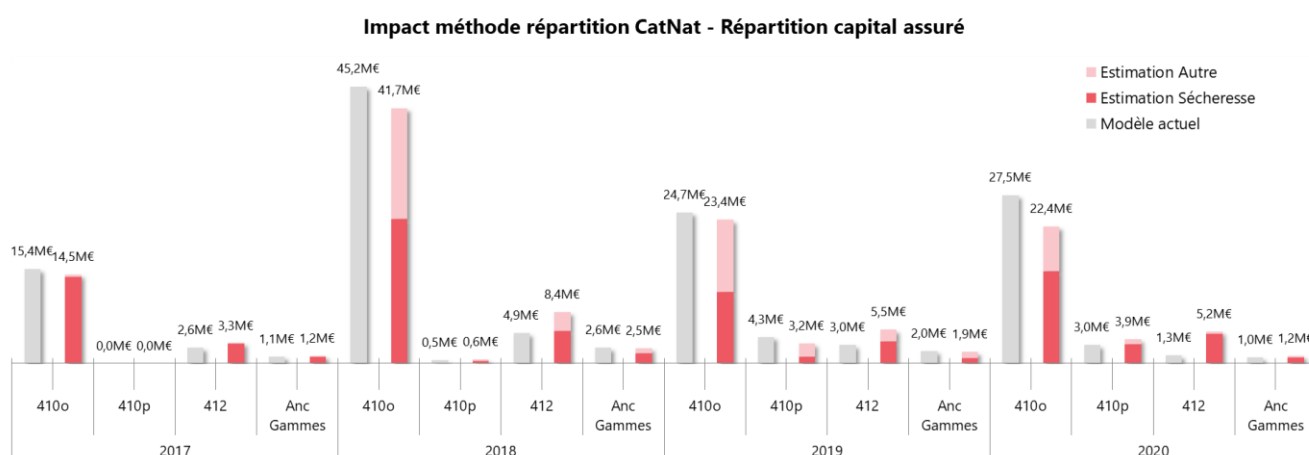
#### 4.4.2. Charge économique à partir de la modélisation

La modélisation n'a eu pour objectif que de retraiter différemment la gestion de la sinistralité sécheresse sur les derniers exercices. La part de la charge attribuée aux autres événements catastrophes naturelles conserve bien la méthode usuelle des triangles de liquidation.

Le même exercice de répartition par produit est effectué en s'appuyant cette fois-ci sur la modélisation des communes et sur la répartition de leur portefeuille d'après le coût immobilier des biens assurés. Après avoir identifié sur la période 2017 à 2020 les communes à risque, sélectionné le portefeuille présent exercice par exercice sur ces communes et sommé par produit les capitaux assurés, on a pu obtenir la part des capitaux risqués face à la sécheresse pour chacun des exercices. Le poids de ces capitaux assurés par produit parmi l'ensemble des capitaux des contrats en portefeuille sur les communes à risque a permis de répartir par produit la charge sinistre sécheresse ultime fournie sur l'ensemble de la branche habitation.

La figure 4.20 met en avant des disparités par rapport au découpage par triangle de liquidation sur la charge totale des catastrophes naturelles.

Figure 4.20 : Evolution de la répartition de la charge économique sur la période 2017-2020



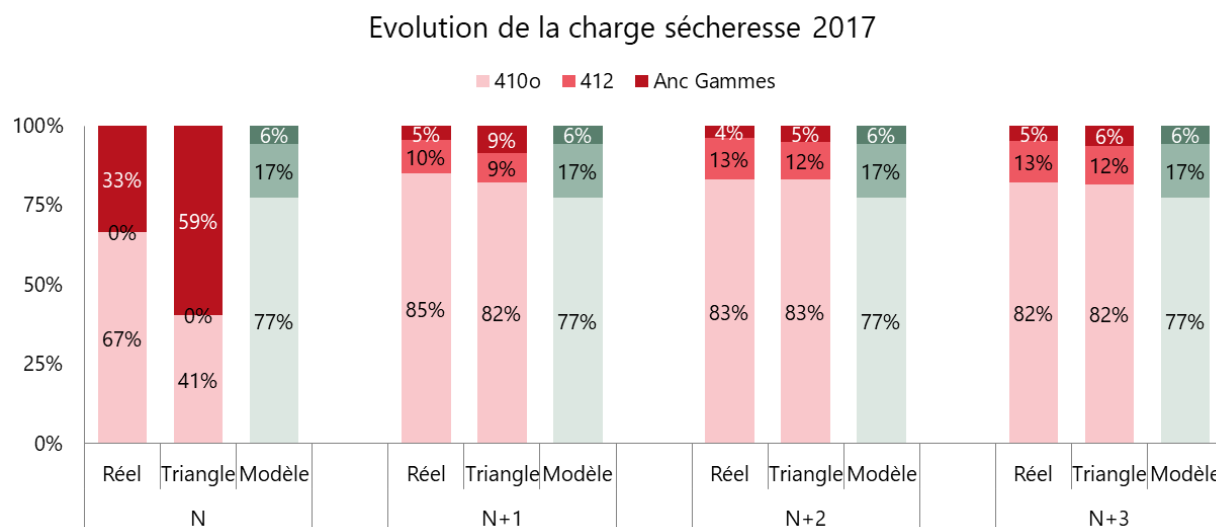
Le traitement différencié de la sécheresse permet avec cette méthode d'estimer un coût même sur un périmètre peu exposé avec une sinistralité peu ou pas encore connue. Ce sont donc les périmètres des Anciennes Gammes et de la 410 p qui se retrouvent ainsi plus valorisés que lors de l'utilisation seule des triangles de liquidation sans distinction de la garantie sécheresse. En se focalisant sur l'année 2020 la charge estimée sur le produit 410 o n'est alors plus que de 22,4M€ dont 15,0M€ en sécheresse, contre 27,5M€ avec la méthode actuelle. Le poids de la sinistralité sécheresse est moindre sur ce périmètre par rapport à l'usage des triangles de liquidation différencié par événement ce qui permet d'améliorer son estimation même sur des produits avec peu d'exposition.

#### 4.4.3. Pertinence de la modélisation

En regardant année après année l'évolution de la connaissance sinistre sur un exercice ainsi que l'évolution de son triangle de liquidation, il est flagrant que l'usage de la modélisation n'a de sens

que pour l'exercice le plus récent. La figure 4.21 illustre l'évolution de la sinistralité sécheresse 2017 selon le recul annuel, N correspondant à fin 2017.

Figure 4.21 : Evolution de la répartition de la charge sur l'année 2017 selon le recul



Malgré l'usage de coefficient de passage, le triangle de liquidation reste fortement orienté par la sinistralité réelle d'où une forte similarité avec la répartition de la sinistralité réelle. La modélisation quant à elle est complètement détachée de la date de vision et reste donc stable quelle que soit le recul adopté.

Au bout de 2 ans la sinistralité se stabilise permettant ainsi d'avoir une vision réelle en phase avec la sinistralité ultime. Pour autant, sur la première année seule la modélisation permet d'obtenir une vision correcte de la sinistralité à l'ultime en respectant les ordres de grandeur de la vision finale et surtout en n'omettant pas un produit dont la sinistralité n'a pas encore eu le temps de se déclarer comme c'est le cas pour le produit 412. Ainsi, dès la première année on aurait pu être en capacité d'estimer la charge sur ce produit à 3,2M€ pour une charge réelle stabilisée à 2,5M€ avec l'usage de la méthode géographique contre 0 en se basant uniquement sur une liquidation de la sinistralité connue à date.

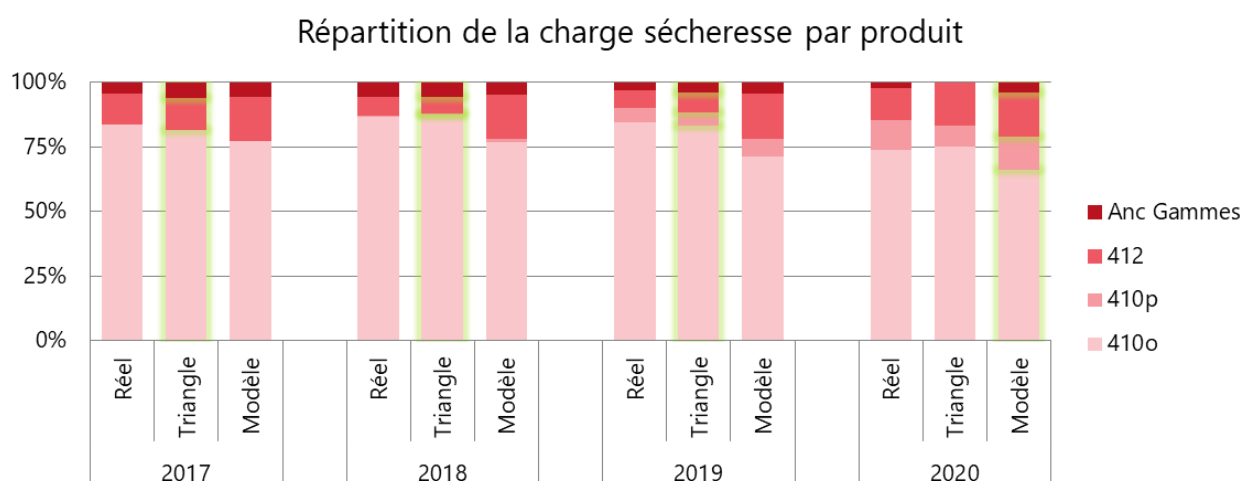
Il apparait donc comme raisonnable de s'appuyer sur le résultat de la modélisation à minima pour le dernier exercice, les autres pouvant être estimés classiquement par l'usage des triangles de liquidation. On aurait ainsi pu estimer la répartition des 19M€ de charge en 1,1M€ sur les Anciennes Gammes, 3,2M€ sur la 412 et 14,7M€ sur la 410 o ; en cohérence avec une version réelle stabilisée à 1,0M€ sur les Anciennes Gammes, 2,5M€ sur la 412 et 15,5M€ sur la 410 o.

La figure 4.22 présente le choix de la répartition qui aurait pu être fait sur la période 2017-2020 d'après les éléments présents à fin 2020. La sinistralité réelle est quant à elle actualisée à fin juin 2022 afin d'illustrer la pertinence des triangles et de la modélisation selon la période.

La répartition sélectionnée année après année est encadrée en vert sur le graphique soit :

- L'utilisation des triangles de liquidation dès lors que le recul sur la sinistralité est suffisant,
- L'application du modèle pour pallier les erreurs d'estimation sur le ou les exercices les plus récents.

Figure 4.22 : Répartition de la charge sécheresse sur la période 2017-2020 selon la méthode d'estimation



Sur cet exemple la modélisation aurait permis d'identifier la présence de sinistres sur les contrats Anciennes Gammes là où la sinistralité à date ne permettait pas d'en tenir compte dans l'élaboration des triangles de liquidation. Pour l'année 2020, sur les produits les moins exposés (410 p et Anciennes Gammes) la liquidation de la charge historique aurait estimé une charge à 2,6M€. En s'appuyant sur l'approche géographique le coût s'élèverait à 4,3M€ sur la 410 p et 1,3M€ sur les Anciennes Gammes. En s'appuyant sur la sinistralité réelle avec un recul d'un an et demi la sinistralité enregistrée se situe à 3,6M€ sur la 410 p et 0,7M€ sur les Anciennes Gammes. L'approche géographique a donc bien tout son sens l'exercice pour lequel la sinistralité n'est que peu connue en proposant une estimation plus cohérente que la liquidation simple.

Un tel système sera donc à répliquer annuellement pour identifier les communes à risque ainsi que le portefeuille exposé sur le dernier exercice analysé.

*Avoir une vision précise de la sinistralité sécheresse pour les années les plus récentes est un exercice difficile qui s'apparente plus à de la prédiction qu'à de la liquidation. La méthode proposée au cours de ce chapitre a permis d'améliorer la répartition actuelle de la sinistralité en intégrant notamment la possibilité de distinguer la charge sécheresse du reste des catastrophes naturelles. En affinant la répartition de la charge on permet ainsi d'obtenir une vision consolidée plus juste produit par produit de la sinistralité globale des contrats en portefeuille.*

*Sans pour autant obtenir une estimation identique à la sinistralité réelle, l'approche géographique apporte une solution à la problématique mise en avant par les triangles de liquidation et permet notamment de ne pas négliger la présence d'une charge sur des produits à faible exposition.*

*La répartition de la charge par usage de l'identification d'un portefeuille sinistré n'a de sens que pour les exercices les plus récents. Dès lors que 2 ou 3 années sont écoulées, les coefficients de liquidation sont suffisamment stables pour permettre une estimation pertinente de la charge économique par triangle de liquidation puisque une grande partie des sinistres ont pu être déclarés par les assurés. Pour autant, l'allongement du délai de prescription dans le cadre de la déclaration des sinistres sécheresses risque d'amplifier le phénomène visible sur le dernier exercice. L'usage d'une méthode d'approximation géographique comme défini dans ce chapitre devra peut-*

*être être généralisée sur plus d'exercices afin de pallier la méconnaissance des sinistres sur une plus longue période.*

*Enfin, la méthode d'approche proposée dans cette partie s'appuie sur les indicateurs en vigueur dans l'obtention du système d'arrêtés de catastrophes naturelles malgré quelques ajustements. Pour conserver cette pratique il faudra être en veille constante sur les évolutions réglementaires afin de s'assurer de répliquer le système des arrêtés effectifs. La méthode présentée n'est donc pas pérenne en l'état.*

## Conclusion

Les différentes études présentées dans ce mémoire ont permis de conforter le besoin de maîtrise du risque sécheresse face aux changements climatiques observés. L'assurance habitation étant de plus en plus impactées par les risques dits de catastrophes naturelles, il est légitime de chercher à mieux en maîtriser l'origine ainsi que les impacts.

Bien que la sinistralité sécheresse soit un phénomène rare, il est malheureusement de plus en plus fréquent ces dernières années, et est considéré comme un risque à forte croissance dans l'avenir. L'année 2022 apparaît d'ailleurs comme la sécheresse la plus importante enregistrée depuis les années 2000. La CCR estime que la charge pour le marché français relative à cet épisode serait aux alentours de 2,5Md€ soit deux fois plus que le coût de la sécheresse de 2023. Le groupe Covéa se retrouve impacté à même hauteur de la marché avec une charge estimée à 490M€ là encore doublant le coût observé en 2003. Le réchauffement climatique est l'une des principales raisons de cette augmentation. Après avoir examiné les différentes études d'impact, il est apparu que la stabilité du portefeuille MMA pourrait aider à limiter les perturbations causées par les événements climatiques extrêmes, tels que les sécheresses. Cependant, même si cette stabilité est bénéfique, elle ne serait pas suffisante pour protéger le portefeuille dans un scénario alarmiste lié au réchauffement climatique.

L'analyse des différents facteurs à l'origine du sinistre sécheresse ainsi que de sa prise en charge par l'assureur a été un premier point d'apprentissage. En effet, au-delà même de l'aléa de survenance de l'événement il faut tenir compte d'un aléa temporel et d'un aléa réglementaire. Contrairement aux sinistres classiques couverts par l'assurance habitation, les dégâts matériels causés par la sécheresse ne sont pas visibles au moment de sa survenance mais plusieurs années après. De plus comme tous sinistres de catastrophes naturelles, sa prise en charge dépend de la reconnaissance de l'événement comme catastrophes naturelles par un arrêté. Pour ne pas pénaliser les assurés, les critères d'acceptation et les indicateurs utilisés sont régulièrement revus afin d'indemniser la majorité des sinistres. Même si les indicateurs permettant l'acceptation restent toujours explicatifs du phénomène au sens géologique et climatique, la manière dont ils sont calculés impacte fortement la délivrance ou non de l'arrêté. Le dernier exemple en date concerne la mise en application du PPL Baudu avec un allongement du délai de prescription de 2 à 5 ans ce qui aura pour conséquence d'augmenter considérablement le nombre de sinistres et donc la charge de chaque exercice. Ces différents éléments mettent donc en avant une incertitude forte concernant la prise en charge de la sinistralité par l'assureur.

Suite aux études préalables effectuées sur les différentes variables identifiées comme explicatives du risque sécheresse, deux analyses ont permis d'utiliser ces données afin d'améliorer la connaissance et la maîtrise du risque face au portefeuille assuré.

La première application via la révision du zonier existant a permis d'améliorer l'identification des zones potentiellement à risque en fournissant une cartographie stable et prudente des communes vis-à-vis de la sinistralité sécheresse. En choisissant d'établir un zonier non corrélé aux derniers critères d'acceptation d'un arrêté pour sécheresse on s'assure d'expliquer l'intégralité des sinistres historiques et à venir. Afin d'obtenir un zonier prudent, le choix a été fait d'accepter de perdre en exactitude générale lors de la modélisation au profit d'une bonne estimation des communes sinistrées. C'est d'ailleurs à cause de ce choix que la régression logistique a été préférée par rapport aux deux modèles de machine learning testés. Le nouveau zonier est principalement expliqué par

l'argile, à l'origine de la sinistralité, soit par la cartographie du RGA mise en place par le BRGM. Cependant, la MRN vient de communiquer sur une mise à jour de cette carte courant 2023. Il serait donc nécessaire de s'intéresser aux évolutions majeures de l'indicateur pour mettre à jour les communes pour lesquelles le risque paraîtrait sous-estimé au regard de la mise à jour du risque argile. C'est avant tout un outil essentiel qui permet à l'assureur de maîtriser son exposition face aux différents risques assurés en identifiant toute potentielle distorsion de l'équilibre actuel. Cette nouvelle cartographie du risque va également permettre à l'assureur d'intégrer dans son devoir de conseil un volet préventif en axant le partage de mesures préventives auprès des clients concerné par un risque élevé. Au vu de l'ampleur que prend la sinistralité sécheresse on pourrait également imaginer aller plus loin en proposant une aide financière afin d'accompagner les clients qui accepteraient de consolider leurs fondations afin de limiter la sinistralité à venir.

La seconde application s'est intéressée à définir une méthode alternative au triangle de liquidation pour mieux estimer la sinistralité sur le dernier exercice aux regards du portefeuille exposé à cette sinistralité. Là encore, la proposition d'une approche géographique a permis par la réplication du système d'arrêtés de catastrophes naturelles d'identifier le portefeuille à risque sur l'ensemble de produits de la branche habitation avant même la déclaration exhaustive des sinistres. Les résultats obtenus ont bien montré toute la difficulté d'estimer à l'avance les communes à risque. En effet, parmi elles une partie ne fera pas la demande d'arrêt. La modélisation engendre donc une large surestimation des communes qui obtiendront réellement un arrêt. Cette méthode a cependant permis d'observer une meilleure estimation de la charge sinistre comparée aux triangles de liquidation sur la première année. Cette méthode s'appuyant sur les indicateurs actuellement en vigueur dans l'obtention du système d'arrêtés de catastrophes naturelles, elle nécessite d'être en veille constante sur les évolutions réglementaires afin d'adapter le choix des indicateurs à ceux étudiés par la commission interministérielle.

*Pour conclure, le risque sécheresse est un risque qui prend de plus en plus d'ampleur dans l'assurance habitation. Les différents projets réglementaires en cours illustrent bien toute la complexité de maintenir les indemnisations avec le système actuel. La prise en charge de la moitié de la sinistralité par le fonds de catastrophes naturelles n'est plus une solution viable en conservant la taxe CatNat à 12%, les compagnies de réassurance réévaluent année après année leurs traités afin d'être en mesure de tenir leurs engagements auprès des assureurs et le changement climatique n'augure pas d'une réduction de la sinistralité dans les années à venir.*



## Bibliographie

### Documentation sur le fonctionnement du système d'arrêtés de catastrophes naturelles

<https://www.service-public.fr/particuliers/vosdroits/F3076> (Lien entre assurance et catastrophe naturelle avec descriptif du process d'indemnisation), site consulté le 2 novembre 2021

<https://www.georisques.gouv.fr/donnees/bases-de-donnees/base-gaspar> (Base de données gaspar recensant l'ensemble des procédures administratives concernant les risques de catastrophes naturelles), site consulté le 4 août 2022

### Documentation sur la sinistralité sécheresse

MARCA A. (2020) *Réalisation de zoniers en assurance habitation pour les risques climatiques*, Mémoire d'Actuariat, ISUP

CHARPENTIER A., MOLLY J., HANI A. (2021) *Predicting Drought and Subsidence Risks in France*, UQAM, EURIA, Willis Re

<https://notre-environnement.gouv.fr/donnees-et-ressources/ressources/graphiques/article/graphique-27> (Illustration de la sinistralité sécheresse sur la période 1988-2019), Site consulté le 2 novembre 2021

<https://www.georisques.gouv.fr/articles-risques/retrait-gonflement-des-argiles/exposition-du-territoire-au-phenomene> (Carte du RGA réalisée par le BRGM), site consulté le 25 août 2021

France Assureurs, *Impact du changement climatique sur l'assurance à l'horizon 2050*, Octobre 2021

MRN, *Sécheresse Géotechnique - De la connaissance de l'aléa à l'analyse de l'endommagement du bâti*, Décembre 2018

<https://www.toutsurlebeton.fr/mise-en-oeuvre/quelle-profondeur-pour-vos-fondations/> (Explication de la qualité des fondations requises dans la construction des bâtiments), site consulté le 3 novembre 2021

Circulaire du Ministère de l'intérieur, *Révision des critères permettant de caractériser l'intensité des épisodes de sécheresse-réhydratation des sols à l'origine de mouvements de terrain différentiels*, 10 mai 2019

*La Cotes annuelles des valeurs vénales immobilières et foncières*, Edition Callon

## Documentation sur la régression logistique

[https://eric.univ-lyon2.fr/~ricco/cours/slides/regression\\_logistique.pdf](https://eric.univ-lyon2.fr/~ricco/cours/slides/regression_logistique.pdf) (Cours sur les fondamentaux de la régression logistique), site consulté le 13 mars 2022

[https://fr.wikipedia.org/wiki/Maximum\\_de\\_vraisemblance](https://fr.wikipedia.org/wiki/Maximum_de_vraisemblance) (Calcul des estimateurs par maximum de vraisemblance), site consulté le 13 mars 2022

<https://help.sap.com/viewer/00f68c2e08b941f081002fd3691d86a7/release/fr-FR/235c79933a7b4f398369e23a04520a3e.html> (Création et interprétation de la courbe ROC et de l'aire sous la courbe) site consulté le 10 septembre 2021

<https://kobia.fr/classification-metrics-auc-pr/> (Création et interprétation de la courbe PR et de l'aire sous la courbe) site consulté le 10 septembre 2021

<https://lemakistatheux.wordpress.com/2013/05/31/le-v-de-cramer/> (Interprétation du V de Cramér), site consulté le 10 septembre 2021

[https://documentation.sas.com/doc/en/pgmsascdc/9.4\\_3.5/procstat/procstat\\_freq\\_details08.htm](https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.5/procstat/procstat_freq_details08.htm) (Implémentation du V de Cramér dans SAS), site consulté le 10 septembre 2021

## Documentation sur les modèles de machine learning

<https://www.lovelyanalytics.com/2016/08/18/un-arbre-de-decision-avec-r/> (Création d'un arbre de décision sous R), site consulté le 15 octobre 2021

<https://lesforetsaleatoiresavecrob.in.genuer.fr/> (Introduction aux forêts aléatoires sous R), site consulté le 15 octobre 2021

<http://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-m-app-agreg.pdf> (Modèles d'agrégation en forêts aléatoires), site consulté le 15 octobre 2021

MUELLER JP, MASSARON L. (2019) *Le machine learning pour les nuls*, Collection Pours les Nuls

## Documentation sur la discrétisation par la méthode de Jenks

<https://www.spatialanalysisonline.com/HTML/index.html> (Définition de l'algorithme de Jenks), site consulté le 3 octobre 2021

<https://pro.arcgis.com/fr/pro-app/latest/help/mapping/layer-properties/data-classification-methods.htm> (Illustration de la méthode de Jenks), site consulté le 3 octobre 2021

## Documentation sur les triangles de liquidation

KHORDJ M, HAFFAR A, TEULON F (2017). *Provisionnement et mesure de risque en assurance dommage dans le cadre de Solvabilité II*. Gestion 2000, Volume 34, pages 137-168

BARON T. (2011) *Mesure de l'Incertitude sur les Provisions en Assurance Non Vie - Approche Bornhuetter-Ferguson*, Mémoire d'Actuariat, ENSAE

LALOUM L. (2021) *Approche hybride au calcul de la provision IBNR : méthode fréquence / coût pour l'estimation des sinistres tardifs*, Mémoire d'Actuariat, ISUP

## Scénarii de migration de la population (INSEE)

Projections de population - Insee Résultats N°187, *Projections de population 2013-2070 pour la France*, paru le 03/11/2016

[Projections de population 2013-2050 pour les départements et les régions | Insee](#) (Démultiplication de la projection de la population française à horizon 2050 par département), site consulté le 16 juin 2022

## Projection de l'impact du changement climatique sur la sinistralité sécheresse (CCR)

CCR, *Modélisation de l'impact du changement climatique sur les dommages assurés dans le cadre du régime Catastrophes Naturelles*, Décembre 2015

CCR et Météo France, *Conséquences du changement climatique sur le coût des catastrophes naturelles en France à horizon 2050*, Septembre 2018



## Annexes

## Annexe 1 : Liste des variables intégrées à la modélisation du zonier technique

Catégories	Champs	Sources	Commentaire
Risque	INSEE	données internes	Code INSEE
	RA_19ANS	données internes	Risque année moyen sur la période d'analyse (19 ans)
	INDIC_RA	données internes	Présence d'exposition sur la commune
	NB_SIN	données internes + MAAF et GMF	Nombre de sinistre sur la commune
	INDIC_SIN	données internes + MAAF et GMF	Booléen sur la présence de sinistre ou non sur la commune
Argile	ALEA_FAIBLE	données PRM <sup>17</sup> - Sinistralité marché	Pourcentage commune avec risque argile faible
	ALEA_MOYEN	données PRM - Sinistralité marché	Pourcentage commune avec risque argile moyen
	ALEA_FORT	données PRM - Sinistralité marché	Pourcentage commune avec risque argile fort
	SANS_ALEA	données PRM - Sinistralité marché	Pourcentage commune sans risque argile
	CLASSE	données PRM - Sinistralité marché	Indicateur iSOL6 calculé par PRM
	CL_1	modèle interne	Classe résultante de la régression logistique
	CL_2	modèle interne	Classe résultante de la régression logistique (passage au log)
Humidité	FREQ_INDIC_4M_S1	modèle interne	Fréquence evol SWI sur 4 mois hors du seuil à -0,5 (analyse de 2001 à 2019)
	FREQ_INDIC_4M_S2	modèle interne	Fréquence evol SWI sur 4 mois hors du seuil à -0,4 (analyse de 2001 à 2019)
	CL_FREQ_INDIC_4M_S1	modèle interne	Classification selon la fréquence atteinte sur la commune
	CL_FREQ_INDIC_4M_S2	modèle interne	Classification selon la fréquence atteinte sur la commune
	FREQ_INDIC_5M_S1	modèle interne	Fréquence evol SWI sur 5 mois hors du seuil à -0,5 (analyse de 2001 à 2019)
	FREQ_INDIC_5M_S2	modèle interne	Fréquence evol SWI sur 5 mois hors du seuil à -0,4 (analyse de 2001 à 2019)
	CL_FREQ_INDIC_5M_S1	modèle interne	Classification selon la fréquence atteinte sur la commune
	CL_FREQ_INDIC_5M_S2	modèle interne	Classification selon la fréquence atteinte sur la commune
	FREQ_INDIC_SSWI	donnée PRM	Fréquence SSWI_3M en dehors de -2,06/2,06 (analyse de 2001 à 2019)
	CL_FREQ_INDIC_SSWI	modèle interne	Classification selon la fréquence atteinte sur la commune

<sup>17</sup> PRM : Service Pôle Risques Majeurs de Covéa

Catégories	Champs	Sources	Commentaire
Exogène	C17_REP_CS1	INSEE - Ménage	CSP 1 - Agriculteur exploitant
	C17_REP_CS2	INSEE - Ménage	CSP 2 - Artisan, Commerçant, Chef d'entreprise
	C17_REP_CS3	INSEE - Ménage	CSP 3 - Cadre, Profession intellectuelle supérieure
	C17_REP_CS4	INSEE - Ménage	CSP 4 - Profession intermédiaire
	C17_REP_CS5	INSEE - Ménage	CSP 5 - Employé
	C17_REP_CS6	INSEE - Ménage	CSP 6 - Ouvrier
	C17_REP_CS7	INSEE - Ménage	CSP 7 - Retraité
	C17_REP_CS8	INSEE - Ménage	CSP 8 - Autre sans activité professionnelle
	C17_REP_MAI	INSEE - Logement	Proportion des maisons au code INSEE
	C17_REP_RP_MAI	INSEE - Logement	Proportion des maisons Résidences Principales
	C17_REP_AV19	INSEE - Logement	Proportion des maisons RP construites avant 1919
	C17_REP_19_45	INSEE - Logement	Proportion des maisons RP construites entre 1919 et 1945
	C17_REP_46_70	INSEE - Logement	Proportion des maisons RP construites entre 1946 et 1970
	C17_REP_71_90	INSEE - Logement	Proportion des maisons RP construites entre 1971 et 1990
	C17_REP_91_05	INSEE - Logement	Proportion des maisons RP construites entre 1991 et 2005
	C17_REP_05_14	INSEE - Logement	Proportion des maisons RP construites entre 2006 et 2014
	C17_REP_AP_15	INSEE - Logement	Proportion des maisons RP construites après 2015
	DENS_LOG	Superficie iris - PRM	Densité de logement par code INSEE
	DENS_MAI	Superficie iris - PRM	Densité de maisons par code INSEE

Toutes ces variables ont servis à la fois à la régression logistique ainsi qu'aux modèles de machine learning. Dans le dernier cas, la variable initiale non classifiée a été directement utilisée.



# Annexe 2 : Matrice de corrélation sur l'ensemble des variables

Le premier tableau illustre la valeur absolue du V de Craméri obtenue en croisement de l'ensemble des variables explicatives.

Table of absolute Cramer's V values for 30 variables. The table is symmetric with 1.00 on the diagonal. Notable high correlations include CL\_FREQ\_INDIC\_5M\_S1 (0.63) and CL\_FREQ\_INDIC\_5M\_S2 (0.66). The table uses a color scale from light yellow to dark red.

## Corrélation naturelle par construction des indicateurs :

- Ensemble des variables expliquant le **facteur de prédisposition**
- Ensemble des variables expliquant le **facteur déclenchant**

Le second tableau présente la matrice de corrélation sur les variables sélectionnées après suppression des données trop fortement corrélées des liens du V de Cramér.

Table of selected variables showing correlations. The variables are CL\_ALEA\_MOYEN, CL\_ALEA\_FORT, CL\_SANS\_ALEA, and 19 indicators from CL\_C17 to INDIC\_RA. The table shows various correlation values, with some cells highlighted in red to indicate higher correlations.

### Annexe 3 : Liste des variables retenues dans la modélisation par régression logistique

La liste ci-dessous présente l'ensemble des variables qui ont été conservés après avoir défini les critères optimaux concernant le facteur déclenchant et le facteur de prédisposition.

Catégories	Champs	Commentaire
<b>Variable cible</b>	INDIC_SIN	Booléen sur la présence de sinistre ou non sur la commune
<b>Variable explicative Risque</b>	INDIC_RA	Présence d'exposition sur la commune
<b>Variable explicative Argile</b>	ALEA_FAIBLE	Pourcentage commune avec risque argile faible
	ALEA_MOYEN	Pourcentage commune avec risque argile moyen
	ALEA_FORT	Pourcentage commune avec risque argile fort
<b>Variable explicative Humidité</b>	CL_FREQ_INDIC_5M_S2	Classification selon la fréquence de l'évolution sur SWI sur 5 mois en dehors du seuil à -0,4
<b>Variable explicative Exogène</b>	C17_REP_CS1	CSP 1 - Agriculteur exploitant
	C17_REP_CS2	CSP 2 - Artisan, Commerçant, Chef d'entreprise
	C17_REP_CS3	CSP 3 - Cadre, Profession intellectuelle supérieure
	C17_REP_CS4	CSP 4 - Profession intermédiaire
	C17_REP_CS5	CSP 5 - Employé
	C17_REP_CS6	CSP 6 - Ouvrier
	C17_REP_CS8	CSP 8 - Autre sans activité professionnelle
	C17_REP_MAI	Proportion des maisons au code INSEE
	C17_REP_19_45	Proportion des maisons RP construites entre 1919 et 1945
	C17_REP_46_70	Proportion des maisons RP construites entre 1946 et 1970
	C17_REP_71_90	Proportion des maisons RP construites entre 1971 et 1990
	C17_REP_91_05	Proportion des maisons RP construites entre 1991 et 2005
	C17_REP_05_14	Proportion des maisons RP construites entre 2006 et 2014
	C17_REP_AP_15	Proportion des maisons RP construites après 2015
	DENS_MAI	Densité de maisons par code INSEE

#### Annexe 4 : Visualisation du test d'indépendance du $\chi^2$ des variables explicatives

Le tableau ci-dessous présente le résultat du test d'indépendance du  $\chi^2$  sur les variables explicatives ayant engendré la suppression de la part des maisons en résidences principales construites entre 1991 et 2014.

Analyse des effets Type 3			
Effet	DDL	Khi-2 de Wald	Pr > khi-2
Part des communes avec risque argile fort	4	1 935,11	<,0001
Part des communes avec moyen risque argile moyen	4	659,31	<,0001
Fréquence evol SWI sur 5 mois hors du seuil à -0,4	7	110,45	<,0001
Part des CSP 1 - Agriculteur exploitant	4	99,97	<,0001
Part des maisons résidences principales construites entre 1919 et 1945	4	69,11	<,0001
Part des CSP 8 - Autre sans activité professionnelle	4	61,21	<,0001
Part des CSP 2 - Artisan, Commerçant, Chef d'entreprise	4	58,61	<,0001
Part des CSP 6 - Ouvrier	4	49,63	<,0001
Part des communes sans risque argile	3	46,39	<,0001
Part des CSP 4 - Profession intermédiaire	4	37,50	<,0001
Part des CSP 3 - Cadre, Profession intellectuelle supérieure	4	29,81	<,0001
Part des maisons résidences principales construites entre 1971 et 1990	4	26,76	<,0001
Part des CSP 5 - Employé	4	22,19	0,0002
Densité de maisons par commune	3	21,45	<,0001
Part des maisons par commune	3	20,11	0,0002
Part des maisons résidences principales construites entre 1946 et 1970	4	18,75	0,0009
Part des maisons résidences principales construites après 2015	4	14,08	0,007
Présence d'exposition	1	5,36	0,0206
<b>Part des maisons résidences principales construites entre 2006 et 2014</b>	<b>4</b>	<b>3,12</b>	<b>0,5373</b>
<b>Part des maisons résidences principales construites entre 1991 et 2005</b>	<b>4</b>	<b>1,63</b>	<b>0,8037</b>