INSTITUT DES ACTUAIRES

Lyon 1

ISFA

Mémoire présenté le :

pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA et l'admission à l'Institut des Actuaires

Par : Manuela TORRES GOMEZ

Titre : Modélisation historique et physique des Catastrophes Naturelles : réconciliation au travers d'un modèle hybride.

Confidentialité : ☐ NON ☒ OUI (Durée : ☐ 1 an ☒ 2 ans)

Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

Membre présents du jury de l'Institut des Actuaires signature

Membres présents du jury de l'ISFA

Entreprise :

Nom : Caisse Centrale de Réassurance

Signature :

Directeur de mémoire en entreprise :

Nom : COHIGNAC

Signature :

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 du candidat

Secrétariat

Bibliothèque :

Résumé

Mots clefs : catastrophes naturelles, théorie bayésienne, crédibilité, modélisation physique, modélisation par expérience, modélisation par exposition, modélisation hybride.
Afin de protéger les Français des conséquences financières d'un événement catastrophique d'origine naturelle, l'État Français a mis en place un mécanisme assurantiel appelé le Régime des Catastrophes Naturelles.

Ce régime repose sur un partenariat public-privé entre les assureurs et la Caisse Centrale de Réassurance ("CCR"), réassureur détenu à 100% par l'État. Il repose également sur la solidarité permettant une prise en charge des dommages matériels causés par des risques naturels à un coût supportable par tous. Ainsi, afin d'éviter toute anti-sélection, la couverture contre les catastrophes naturelles est rendue obligatoire au travers d'une surprime additionnelle aux contrats d'assurance dommages. La surprime est identique pour tous, quel que soit l'exposition aux catastrophes naturelles du risque assuré. Les montants en jeu pouvant être considérables, l'État permet à la CCR, grâce à sa garantie, d'offrir une protection illimitée (couvertures de réassurance illimitées) à ses clients assureurs.

La juste évaluation des risques est alors indispensable. Il s'agit non seulement de prendre en compte les événements passés, mais aussi d'anticiper les événements possibles. Pour ce faire, la CCR dispose de deux modèles. Un basé sur une approche purement historique de la sinistralité prise en charge par le Régime des Catastrophes Naturelles et l'autre, dit par exposition, basé sur une modélisation fine des phénomènes physiques considérés et des polices couvertes.

Ce mémoire a pour objectif de réconcilier les résultats de ces deux approches. L'idée étant de prendre en compte les résultats de l'approche historique pour les événements les plus fréquents et d'évoluer vers les résultats de l'approche par exposition pour les événements les plus rares. Nous présenterons ici deux approches : la première repose sur l'application de la théorie bayésienne et la deuxième est une combinaison convexe dont les poids dépendent de la période de retour. Une comparaison de ces deux méthodes permettra de choisir la distribution mélange qui sera utilisée lors de la création du Modèle Interne Partiel Hybride qui sera présenté par la suite.

Abstract

Key words : natural catastrophe, natural disaster, Bayesian theory, credibility, physical model, historical model, hybrid model.

In France, damage insurance policies are required to cover all losses caused by natural catastrophes since the legal framework "Régime CatNat" was created in 1982. This structure relies on the collaboration between private insurers and CCR from the public sector. Indeed, CCR's ownership is held by the State.

The idea behind "Régime CatNat" is to allow French people to cope with the cost of an extreme event due to a natural disaster by dividing the cost. Hence, the guarantee against natural catastrophes was made mandatory by adding a premium surcharge to all the damage insurance policies. In order to avoid only exposed policies paying a surcharge, the premium surcharge is equal for all the policyholders.

As the losses can be in the order of billions of euros, CCR's reinsurance contracts are State guaranteed which gives CCR the possibility of offering unlimited guarantees to the insurers. Thus, CCR is required to evaluate the commitment of all the parties to a high degree of accuracy.

To that end, CCR has developed two different models to assess its liabilities as well as those of its shareholder and clients. The first model is based on the historical data provided by the insurers while the second one is based on physical models. These physical models depend on the characteristics of the risk and the property at risk.

The purpose of this study is to create a hybrid distribution that would include information from two different sources : the historical data, and the output of the physical models. To that end, we will use two methods, the first one is an application of the Bayesian Theory, and the second one is a convex structure where the weigh of each source of information depend on the return period. We will compare these two methods and use the best one to create an hybrid partial internal model.

Remerciements

Je tiens à remercier toute l'équipe du Département Études Techniques Garantie d'État pour son accueil et la bonne ambiance au sein de l'équipe qui ont contribué au bon déroulement de l'alternance.

Je remercie plus particulièrement Thierry COHIGNAC, mon tuteur en entreprise, pour son investissement, ses conseils et son aide tout au long de mon alternance, ainsi que mon tuteur pédagogique, Nabil KAZI-TANI.

Je souhaite également remercier Geoffrey ECOTO pour son aide et ses conseils lors de l'utilisation des différents outils et le développement du modèle hybride.

Enfin, je remercie mon professeur de crédibilité de l'ISFA, Pierre THEROND pour son aide au niveau de la théorie de la crédibilité.

Table des matières

1 Introduction **7**

2 Régime d'indemnisation des catastrophes naturelles **9**
2.1 Principe du régime 10
2.2 Périls couverts 11
2.3 Biens garantis 13

3 Rôle de la CCR **14**
3.1 Département Etudes Techniques Garantie d'Etat 15
3.2 Couvertures proposées par la CCR pour les catastrophes naturelles 15
3.2.1 Définition d'une garantie *Quote-Part* 16
3.2.2 Définition d'une garantie *Stop-Loss* 16
3.2.3 Couverture des Catastrophes Naturelles 16

4 Modélisation physique des périls **19**
4.0.4 Modèle Inondation 21
4.0.5 Modèle Sécheresse 25
4.0.6 Modèle Séisme 28
4.0.7 Modèle Vents Cycloniques 32

5 Tarification des traités **34**
5.1 Tarification par expérience 34
5.2 Tarification par exposition 38

6 Objectif de l'étude **41**
6.1 Données historiques 42
6.2 Données sur l'exposition 42

6.3 Compatibilité des données . . . 43
6.4 Approches utilisées . . . 43
6.5 Critères basés sur l'expertise CCR . . . 44

7 Théorie bayésienne **45**
7.1 Généralités sur la théorie bayésienne . . . 45
7.1.1 Loi a priori . . . 46
7.1.2 Loi a posteriori . . . 46
7.1.3 Loi conjuguée . . . 46
7.2 Loi Log-Normale . . . 47
7.2.1 Espérance . . . 47
7.2.2 Paramètre de précision . . . 53
7.3 Loi Gamma . . . 57
7.3.1 Loi a priori . . . 57
7.3.2 Estimation des hyperparamètres a priori . . . 58
7.3.3 Modélisation a posteriori . . . 59

8 Crédibilisation au niveau de la charge du marché de la CCR **61**
8.1 Approche bayésienne . . . 61
8.1.1 Ajustement des lois de probabilités . . . 62
8.1.2 Application de la théorie bayésienne . . . 67
8.1.3 Crédibilisation du péril inondation . . . 69
8.1.4 Crédibilisation du péril sécheresse . . . 73
8.1.5 Crédibilisation de la sinistralité automobile . . . 75
8.1.6 Conclusion sur l'approche bayésienne . . . 78
8.2 Distribution hybride . . . 79
8.2.1 Présentation de l'approche . . . 79
8.2.2 Choix de la fonction quantile hybride . . . 80
8.2.3 Calcul des quantiles hybrides . . . 81
8.2.4 Crédibilisation hybride du péril inondation . . . 82
8.2.5 Crédibilisation hybride du péril sécheresse . . . 83
8.2.6 Crédibilisation hybride de la sinistralité automobile . . . 84
8.3 Comparaison des méthodes de crédibilisation . . . 85
8.3.1 Péril inondation . . . 85
8.3.2 Péril sécheresse . . . 86
8.3.3 Sinistralité automobile . . . 87

8.4 Conclusion sur la comparaison des méthodes de crédibilisation 88

9 Création d'un Modèle Interne Partiel Hybride 91

9.1 Mise en place des données 91
9.2 Description du modèle 92
9.3 Résultats au niveau de la CCR 94
9.3.1 Moyenne et écart-type 94
9.3.2 Niveaux de retour 95
9.3.3 Conclusion par rapport aux résultats CCR 97
9.4 Résultats au niveau d'une cédante 98

10 Conclusion générale et perspectives 101

11 Annexes 105

11.1 Annexe A : Tests statistiques 105
11.1.1 Test de Kolmogorov-Smirnov 106
11.1.2 Test de Cramér-von Mises 107
11.1.3 Exemples 107
11.2 Annexe B : Généralités sur la théorie bayésienne 110
11.2.1 Raisonnement proportionnel 110
11.2.2 Loi Normale-Normale 111

Chapitre 1

Introduction

La Caisse Centrale de Réassurance (ci-après "CCR") est une société anonyme détenue à 100% par l'État français. En tant que réassureur, la CCR intervient dans la plupart des branches et des marchés de la réassurance internationale comme par exemple l'incendie, les dommages aux biens, la vie et assurances de personnes, l'automobile, le transport maritime et aérien, entre autres.

Depuis l'instauration du régime d'indemnisation des catastrophes naturelles en 1982, elle intervient pour délivrer une couverture de réassurance, bénéficiant de la garantie de l'État, pour les risques de catastrophes naturelles en France, métropole et Outre-mer. Cette garantie est illimitée en montant. Ainsi, les assurés comme les assureurs bénéficient d'une sécurité maximale concernant l'indemnisation, dans l'hypothèse de la survenance de sinistres majeurs.

La CCR est chargée de concevoir, mettre en œuvre et gérer des instruments performants répondant à des besoins de couverture de risques exceptionnels, au service de ses clients et de son actionnaire. En particulier dans le cadre des catastrophes naturelles, la CCR a mis en place des modèles sur les principaux périls couverts par le régime d'indemnisation des catastrophes naturelles (inondations, y compris les submersions marines, sécheresses, séismes et vents cycloniques).

Ces modèles permettent principalement :

- d'estimer le coût d'un événement climatique pour CCR, le marché ou les assureurs, quelques jours ou quelques semaines après sa survenance (calcul des réserves) ;
- de mesurer l'exposition de l'État, de CCR et de ses clients aux événements possibles, même s'ils ne sont pas survenus ;
- d'analyser l'exposition des territoires aux risques naturels.

Dans ce contexte, la CCR a décidé de développer un Modèle Interne Partiel (ci-après "MIP") basé sur une modélisation à l'exposition.

Outre les utilisations ci-dessus du MIP, sa principale utilisation concerne le calcul du besoin en capital réglementaire sous solvabilité 2. Ainsi, le modèle est calibré de façon à ne pas sous-évaluer les événements extrêmes. Les événements fréquents ont donc tendance à être surestimés ou du moins à ne pas correspondre aux moyennes historiques observées depuis la création du régime.

La réconciliation des résultats du MIP et de l'historique du régime sera l'enjeu de ce mémoire.

Pour ce faire, nous allons considérer des modèles, que nous appellerons hybrides, prenant en compte l'historique mais également l'approche par exposition obtenue grâce au MIP (approche physique des périls modélisés).

A cette fin, deux approches vont être utilisées : une basée sur la théorie bayésienne de la crédibilité et l'autre basée sur un mélange convexe des distributions historiques et d'exposition dont les poids seront définis en fonction de la période de retour.

Les quatre premiers chapitres seront consacrés à expliquer plus en détail le principe du Régime des Catastrophes Naturelles, le rôle de la CCR, la modélisation physique des périls et la tarification des traités de réassurance soumis au régime. Dans le chapitre 5 nous expliquerons plus en détail le besoin de création d'une modélisation mélange à partir des modèles présentés dans les chapitres précédents. Ensuite, le chapitre 6 permettra d'introduire la théorie bayésienne et les notions nécessaires pour l'application envisagée. Enfin, les deux derniers chapitres seront dédiés à l'application de la théorie, dans un premier temps pour la création de distributions mélanges obtenues à partir des différentes méthodes, et dans un second temps, pour la création d'un Modèle Interne Partiel Hybride.

Chapitre 2

Régime d'indemnisation des catastrophes naturelles

Les catastrophes naturelles, telles que les inondations et les sécheresses, étaient autrefois exclues des contrats d'assurance. L'absence de couverture était due aux particularités des catastrophes naturelles, notamment :

- au risque de cumul, en effet ces phénomènes peuvent toucher un grand nombre de polices et les pertes cumulées des sinistres peuvent atteindre des montants de l'ordre de plusieurs milliards d'euro ;
- au risque d'anti-sélection, puisque seuls les risques exposés à des tels événements et localisés dans les zones à risque seraient amenés à se couvrir contre les catastrophes naturelles ;
- l'absence de statistiques fiables, due à la rareté des catastrophes naturelles, qui empêchait une modélisation crédible au sein des assureurs et des réassureurs.

La loi du 13 juillet 1982 rend l'inclusion de la couverture des dommages causés par les catastrophes naturelles obligatoire pour tous les contrats d'assurance couvrant des biens situés sur le territoire français. Cette loi créé alors le régime d'indemnisation des catastrophes naturelles (ci-après "régime CatNat") qui réside sur la solidarité et la responsabilité au travers de dispositifs de prévention.

2.1 Principe du régime

Pour qu'un événement soit indemnisé par le régime CatNat, un arrêté interministériel doit constater l'état de catastrophe naturelle pour la commune. Le préfet doit transmettre les demandes formulées par les maires des communes à la commission interministérielle. Pour que l'état de catastrophe naturelle soit reconnu, le dommage doit découler directement d'un phénomène naturel et doit avoir une intensité anormale.

Nous pouvons observer le nombre d'états de catastrophes naturelles reconnus entre 1982 et 2014 en fonction des communes sur la figure 2.1. Il est important de noter que pour certaines communes, le nombre de reconnaissances est supérieur à 25 états de catastrophe naturelle alors que d'autres communes n'ont jamais fait face à une catastrophe naturelle reconnue par la commission interministérielle. Toutefois, cela ne veut pas dire que les communes en question n'ont pas fait face à des événements catastrophiques non pris en charge par le régime des catastrophes naturelles.

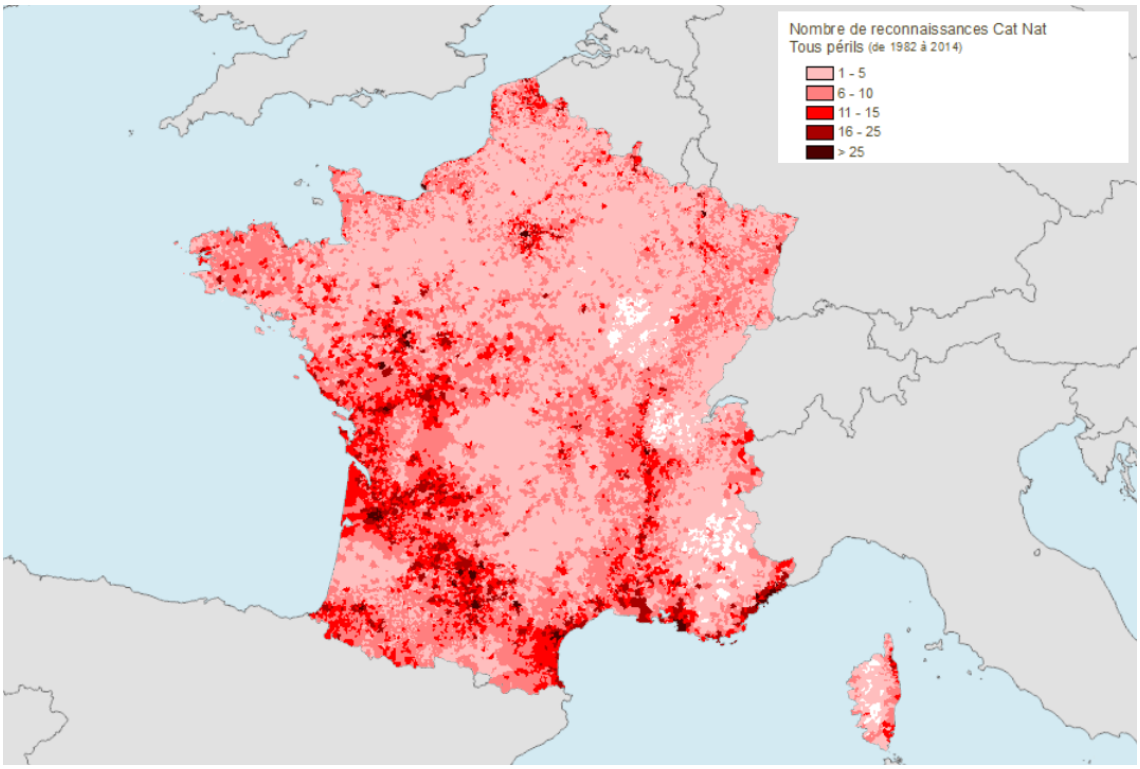

FIGURE 2.1 – Nombre de reconnaissances de catastrophes naturelles

2.2 Périls couverts

La commission interministérielle permet de définir si l'événement est reconnu en tant que catastrophe naturelle couverte par le régime. Les phénomènes assurés ne sont donc pas prédéfinis par une liste, mais l'historique permet d'établir une liste non exhaustive des périls couramment couverts par le régime d'indemnisation :

- inondations, il peut s'agir de ruissellements, de débordements, de remontées de nappe phréatique ou même de submersion marine ;
- coulées de boue ;
- séismes ;
- mouvements de terrain ;
- sécheresses géotechniques, il s'agit des mouvements de terrain dus à la sécheresse ;
- affaissements de terrain dus à des cavités souterraines et à des marnières ;
- raz de marrée ;
- avalanches ;
- cyclones de grande ampleur dans les départements et les collectivités d'outre-mer.

Il est important de noter que, historiquement, les tempêtes ne sont pas prises en compte par le régime. Toutefois, les inondations causés par des fortes pluies suivant une tempête peuvent être indemnisées en tant qu'inondations.

La figure 2.2 présente la répartition de la sinistralité (en millions d'euro) des catastrophes naturelles prises en charge par le régime entre 1990 et 2010 pour l'ensemble du marché de l'assurance.

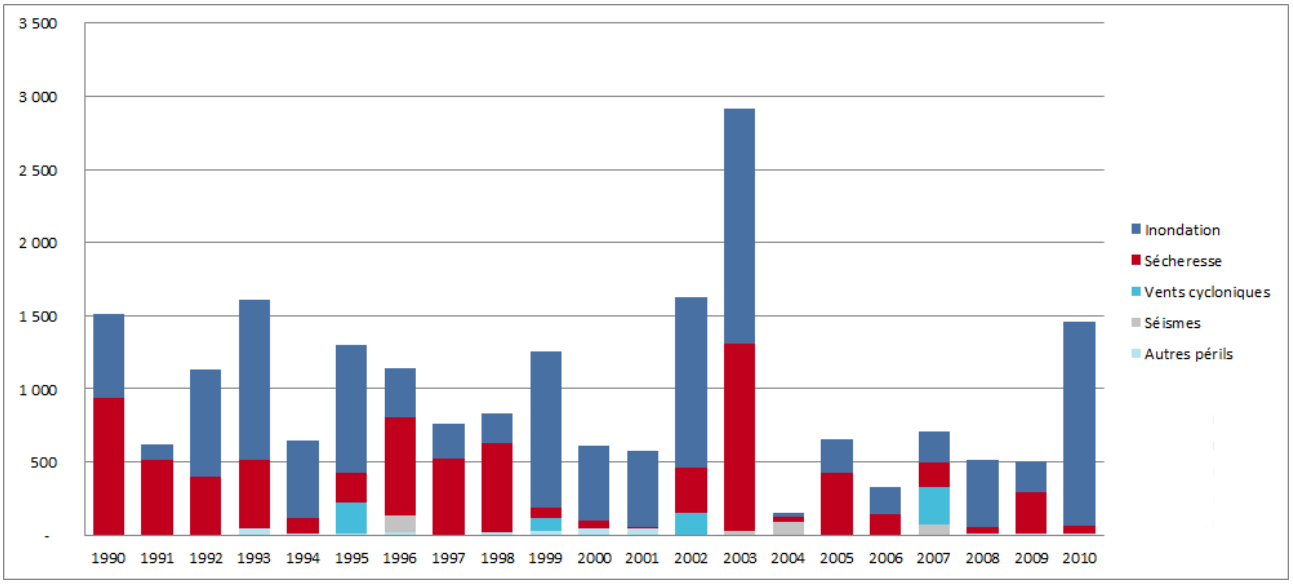

FIGURE 2.2 – Historique des périls pris en charge par le régime des catastrophes naturelles

La figure 2.3 présente la répartition moyenne par péril du coût des catastrophes naturelles pour le régime sur ses dix années. Nous remarquons que, d'après l'historique des catastrophes naturelles, il y a quatre périls prépondérants : les inondations, les sécheresses, les vents cycloniques et les séismes. Nous remarquons également que les périls inondation et sécheresse représentent en moyenne plus de 90% de la charge du régime des catastrophes naturelles. Toutefois, nous avons vu précédemment que les coûts des inondations et des sécheresses varient considérablement d'une année à l'autre.

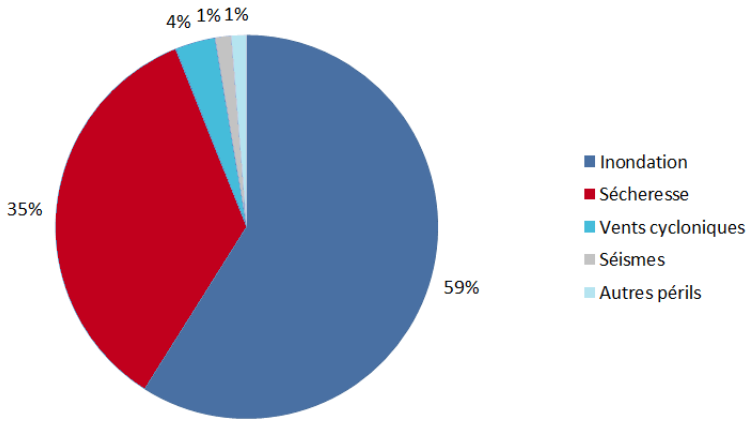

FIGURE 2.3 – Répartition de la charge du régime catastrophes naturelles par péril

2.3 Biens garantis

Nous avons vu précédemment que la couverture est obligatoire pour tous les contrats d'assurance couvrant des biens situés en France. Le régime couvre donc les biens immobiliers et non immobiliers assurés par des garanties incendie, accidents et risques divers tels que :

- les habitations et leur contenu ;
- les installations industrielles et commerciales et leur contenu ;
- les bâtiments appartenant aux collectivités locales et leur contenu ;
- les bâtiments agricoles et leur contenu ;
- les serres considérées en tant que bâtiment ou matériel ;
- les véhicules ;
- les frais de déblai, de démolition, de pompage et de nettoyage.

Peuvent aussi être indemnisés, les accessoires et équipements automobiles s'ils sont compris dans la garantie de base, les clôtures, murs de soutènement ou fondations s'ils sont couverts, ainsi que les forêts lorsqu'elles sont assurées par des contrats dommages aux biens.

Chapitre 3

Rôle de la CCR

La Caisse Centrale de Réassurance (ci-après "CCR") est une société détenue à 100% par l'État français. La particularité de cette entreprise est d'avoir la possibilité d'offrir, au travers de la Garantie d'État, des couvertures illimitées pour certains périls pour des sinistres ayant lieu sur le territoire français. On parle alors de réassurance publique. Les activités concernées par la Garantie d'État sont les suivantes :

- la réassurance des risques de catastrophes naturelles ;
- la réassurance des risques d'attentats et d'actes de terrorisme ;
- la réassurance des risques exceptionnels liés aux transports ;
- la réassurance de la responsabilité civile des exploitants de navires et installations nucléaires ;
- la réassurance du complément d'assurance-crédit public.

La CCR est la seule société de réassurance qui puisse proposer des couvertures illimitées. Ainsi, la société apporte aux cédantes une sécurité maximale et des avantages en termes de solvabilité.

Dans cette étude nous allons uniquement nous intéresser aux catastrophes naturelles. La figure 3.1 montre l'importance de la réassurance des catastrophes naturelles parmi les activités de la réassurance publique. En effet, cette activité représente plus de 90% du chiffre d'affaire de la réassurance publique.

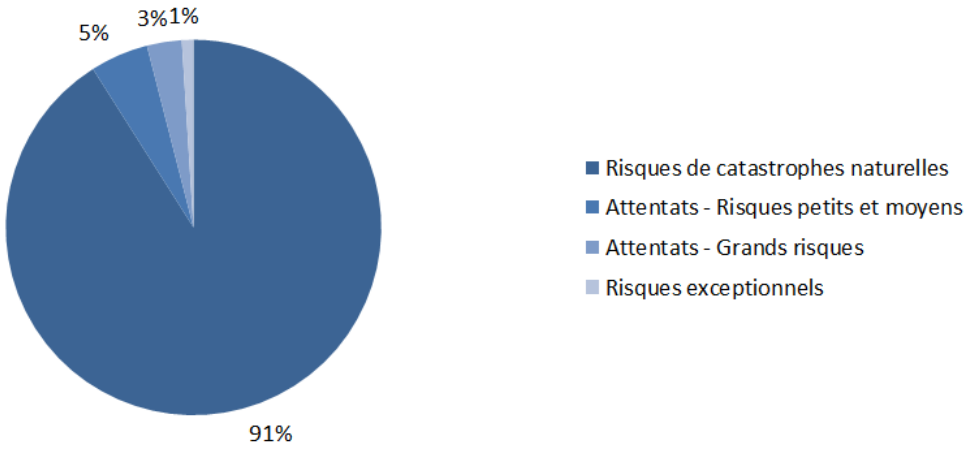

FIGURE 3.1 – Répartition du chiffre d'affaires par activité

3.1 Département Etudes Techniques Garantie d'Etat

Le département "Études Techniques Garantie d'État" au sein de la direction "Catastrophes Naturelles, Fonds Publics & Etudes Techniques" s'occupe de l'ensemble des études menées sur les différents périls couverts par la Garantie d'État. Ce département est composé de deux pôles :

- le pôle Recherche & Développement et services :
  Ce pôle développe des modèles physiques pour les différents périls couverts, ainsi que de l'offre de services proposés aux cédantes et aux partenaires ;

- le pôle Conformité des modèles, Actuariat et Information scientifique :
  Ce pôle s'intéresse aux études actuarielles pour les activités avec Garantie d'État. Ces études sont principalement liées à l'équilibre du régime Catastrophes Naturelles dont la CCR est l'un des piliers.

3.2 Couvertures proposées par la CCR pour les catastrophes naturelles

Dans le cadre de la garantie d'État, la CCR propose une couverture illimitée composée de deux garanties : une garantie *Quote-Part* et une garantie *Stop-Loss*.

3.2.1 Définition d'une garantie *Quote-Part*

Un *Quote-Part* (ci-après "QP") est une garantie qui permet à l'assureur, dénommé cédante, de céder une partie du risque au réassureur. La cédante s'engage ainsi à céder une part des primes encaissées au réassureur et en contrepartie le réassureur s'engage à prendre en charge la même part des sinistres. La part est appelée le taux de cession. Si on note t le taux de cession et X_i les sinistres de la cédante sur l'année. Alors, le coût à la charge du quote-part du réassureur est le suivant :

$$\text{Charge QP Réassureur} = t \times \sum X_i$$

La part des sinistres conservée par la cédante, aussi connue sous le nom de rétention, s'écrit :

$$\text{Rétention} = (1-t) \times \sum X_i$$

3.2.2 Définition d'une garantie *Stop-Loss*

Un excédant de perte annuelle, aussi connu sous le nom de *Stop-Loss* (ci-après "SL"), est une garantie qui permet de protéger le ratio Sinistres/Primes de la cédante. Le traité fait mention d'une priorité et d'une portée. Le réassureur s'engage à payer la charge annuelle de la cédante dépassant la priorité. Néanmoins, le coût pour le réassureur est généralement limité à la portée. La cédante absorbe la charge annuelle dépassant la somme de la priorité et de la portée, appelée plafond. Si on note S la charge annuelle de la cédante, nous obtenons :

$$\text{Charge SL Réassureur} = \min\left\{\max\left\{S - \text{ priorité }, 0\right\}, \text{ portée}\right\}$$

3.2.3 Couverture des Catastrophes Naturelles

Dans le cas de la couverture proposée par la CCR pour les catastrophes naturelles, le *Quote-Part* agit sur la charge annuelle de la cédante, et le *Stop-Loss* agit ensuite sur la rétention de la cédante. Par ailleurs, comme la couverture SL est dite illimitée, le traité a une portée infinie, ainsi le coût à la charge la CCR au niveau du SL n'est pas plafonné. De cette manière, la garantie de l'État permet à la cédante de plafonner sa perte annuelle à la priorité du SL.

Après l'application du QP nous obtenons donc :

$$\text{Charge QP CCR} = t \times \sum X_i \qquad \text{et} \qquad \text{Rétention} = (1-t) \times \sum X_i$$

L'application du SL nous donne ensuite :

$$\text{Charge SL CCR} = \max\{\text{ Rétention} - \text{priorité}\ ,\ 0\ \} \qquad \text{et} \qquad \text{Charge Cédante} = \min\{\text{ Rétention}\ ,\ \text{priorité}\ \}$$

La figure 3.2 permet de mieux comprendre l'application des deux garanties. Dans un premier temps nous appliquons le quote-part à la sinistralité annuelle de la cédante. Ensuite nous appliquons l'excédant de perte annuelle à la rétention de la cédante.

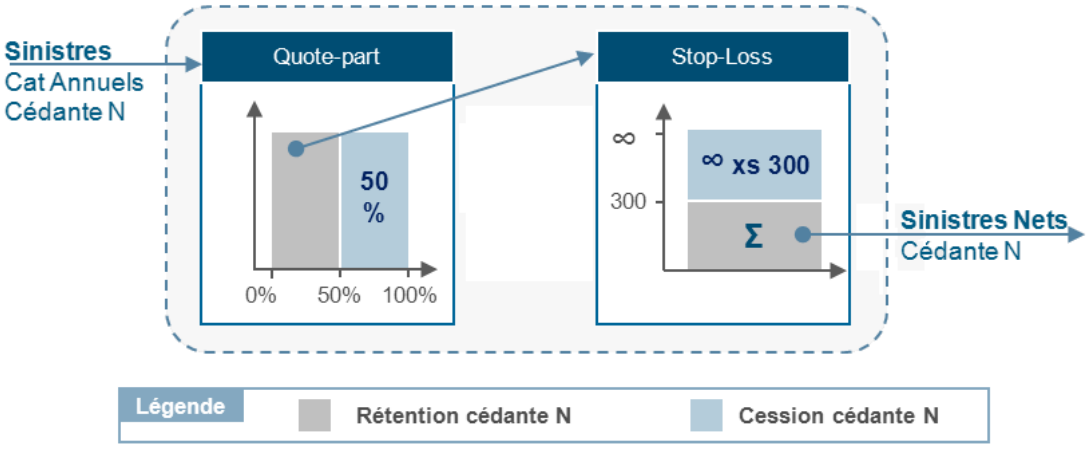

FIGURE 3.2 – Application des traités

Prenons par exemple d'une cédante, que nous appellerons la Cédante N, avec un quote-part ayant un taux de cession de 50% et un stop-loss ayant une priorité de 100 M€. La figure 3.3 illustre cet exemple. Nous remarquons que le coût annuel net de la cédante est bien plafonné.

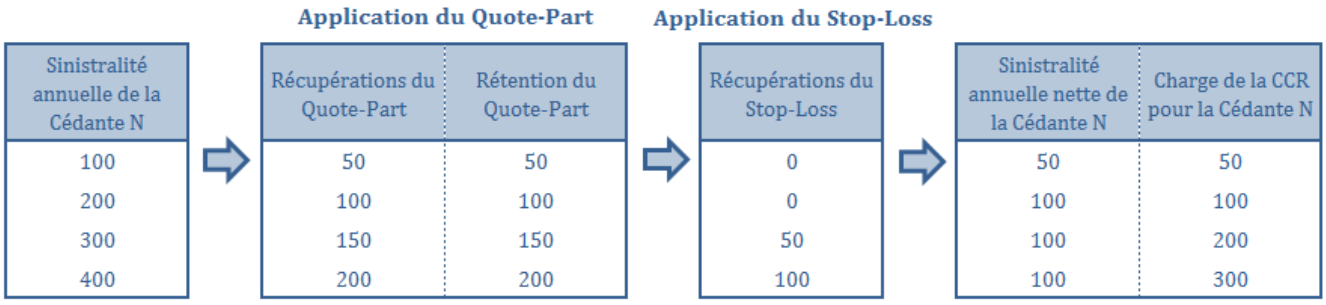

FIGURE 3.3 – Exemple d'application des traités

Toutefois, la CCR ne conserve pas la totalité du risque auquel elle est exposée via les traités avec garantie d'État. En effet, la perte annuelle de la CCR est couverte par l'État de telle manière que l'État intervient au-delà d'un montant appelé Seuil d'Intervention de l'Etat. Dans le cadre des catastrophes naturelles, le seuil est calculé en fonction des provisions et de la réserve spéciale constituée pour ces périls.

Chapitre 4

Modélisation physique des périls

La CCR a développé des modèles physiques permettant de modéliser les différents périls pris en charge par le régime des catastrophes naturelles. Le but de ces modèles est d'évaluer l'exposition des différentes cédantes et de la CCR en fonction de la localisation des polices et de l'intensité du phénomène. Nous avons vu précédemment que, du point de vue de la charge annuelle du régime des catastrophes naturelles, les quatre principaux périls pris en charge par le régime sont : les inondations, les sécheresses, les séismes et les vents cycloniques. Des modèles physiques ont donc été développés pour ces quatre périls.

Les modèles physiques sont développés soit en interne, soit par des agences externes à la CCR. Ces modèles se décomposent généralement en quatre modules :

- le module d'aléa permet de générer un catalogue d'événements associés à leur probabilité de survenance et à l'intensité physique du phénomène ;
- le module de vulnérabilité a pour but de recenser l'ensemble des polices et les données relatives à la leur localisation, leurs caractéristiques et les valeurs assurées ;
- le module de dommages évalue le taux de destruction à partir des données provenant des deux modules précédents.

A ces modèles physiques vient s'ajouter un module financier qui permet d'appliquer les conditions contractuelles des traités d'assurance et de réassurance.

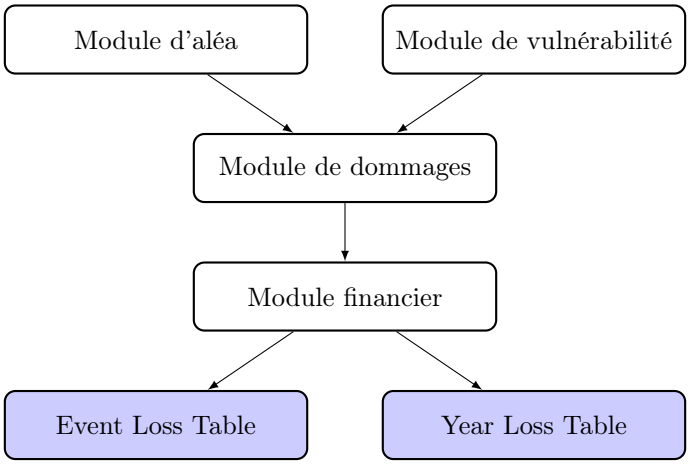

FIGURE 4.1 – Modélisation physique des catastrophes naturelles

La figure 4.1 montre le schéma suivi dans le cadre de la modélisation physique des catastrophes naturelles. Nous remarquons que la sortie des modèles peut être : soit une *Event Loss Table*, soit une *Year Loss Table*.

Une *Event Loss Table* est un catalogue d'événements auxquels on associe une fréquence et les paramètres d'une loi permettant d'estimer le coût du sinistre. Il est ainsi possible de tirer au sort un événement à partir de sa fréquence et ensuite de modéliser son coût à partir d'une loi définie par les paramètres de l'événement. Nous remarquons que les *Event Loss Tables* ne permettent pas de prendre en compte la corrélation entre les différents événements. Ainsi, il n'est pas possible de modéliser certaines caractéristiques des catastrophes naturelles telles que les répliques des tremblements de terre ou encore des inondations qui s'enchaînent tout au long d'un fleuve et de ses ramifications.

Les *Year Loss Tables* permettent quant à elles de modéliser cette corrélation. Il s'agit d'un catalogue d'années associées à plusieurs événements qui sont corrélés. Il est ainsi possible de tirer au sort des événements corrélés tels qu'un tremblement de terre et ses répliques. Le coût de chaque événement est ensuite modélisé à l'aide d'une loi prédéfinie et des paramètres renseignés dans la *Year Loss Table*.

Prenons l'exemple d'un péril pour lequel le coût d'un scénario i est modélisé à partir d'une loi Log-Normale de paramètres (μ_i, σ_i). Les figures 4.1 et 4.2 permettent de mieux comprendre la différence entre le contenu de chacune de ces deux tables.

Id évènement	Fréquence	Espérance	Variance
00001	f_1	μ_1	σ_1
00002	f_2	μ_2	σ_2
00003	f_3	μ_3	σ_3
⋮	⋮	⋮	⋮

TABLE 4.1 – *Event Loss Table*

Année	Id évènement	Fréquence	Espérance	Variance
00001	00001	f_1	μ_1	σ_1
00001	00002	f_2	μ_2	σ_2
00001	00003	f_3	μ_3	σ_3
00002	00004	f_4	μ_4	σ_4
00002	00005	f_5	μ_5	σ_5
00003	00006	f_6	μ_6	σ_6
⋮	⋮	⋮	⋮	⋮

TABLE 4.2 – *Year Loss Table*

La CCR a opté pour des *Year Loss Tables* afin de prendre en compte la corrélation entre les événements. Par la suite nous allons décrire les modèles développés pour les différents périls.

4.0.4 Modèle Inondation

Le modèle inondation a été principalement développé en interne. Il faut premièrement distinguer les inondations prises en charge par le régime d'indemnisation des catastrophes naturelles. Il n'est pas possible d'établir une liste exhaustive, cependant, les phénomènes généralement couverts sont les suivants :

- les débordements des cours d'eau ;
- les ruissellements causés par des précipitations intenses ;
- les remontées des nappes phréatiques ;
- les submersions marines.

Pour que l'état de catastrophe naturelle soit reconnu pour une commune, il faut que l'aléa dépasse un certain seuil, qui a été établi par la jurisprudence. Dans ce cas, l'événement est pris en charge

par le régime.

Module d'aléa

Le module aléa du modèle inondation permet de simuler le débordement de tous les cours d'eau ainsi que le ruissellement. Pour cela, le module dispose d'un générateur de débits et d'un générateur de pluies.

La simulation des débits s'appuie sur la théorie des copules. Dans un premier temps, nous modélisons la loi marginale du débit de chaque station hydrométrique par une loi que nous choisissons parmis les distributions suivantes : Generalized Extreme Value, Weibull, Log-Normale et Gamma. Cette loi est ajustée aux débits maximaux mensuels qui sont extraits de la banque Hydro. La dépendance entre les différentes stations et les mois est ensuite modélisée par une copule Gaussienne. Ce générateur permet ainsi de construire un catalogue d'événements fictifs.

Le générateur des pluies a été, quant-à-lui, développé au sein de l'Institut National de Recherche en Sciences et Technologies pour l'Environnement et l'Agriculture. Il se base sur des données issues de Météo France qui correspondent à l'historique des pluies horaires. Ce générateur permet d'obtenir un catalogue d'événements de précipitations fictives qui sont ensuite utilisés pour estimer le débit des cours d'eau. Nous appliquons ensuite le modèle des débits décrit ci-dessus pour obtenir un catalogue de débordements.

Nous pouvons observer un exemple d'aléa modélisé sur la ville de Paris sur la figure 4.2. Sur cette dernière figurent l'aléa de débordement sur l'image à gauche et la superposition du débordement et du ruissellement sur l'image à droite.

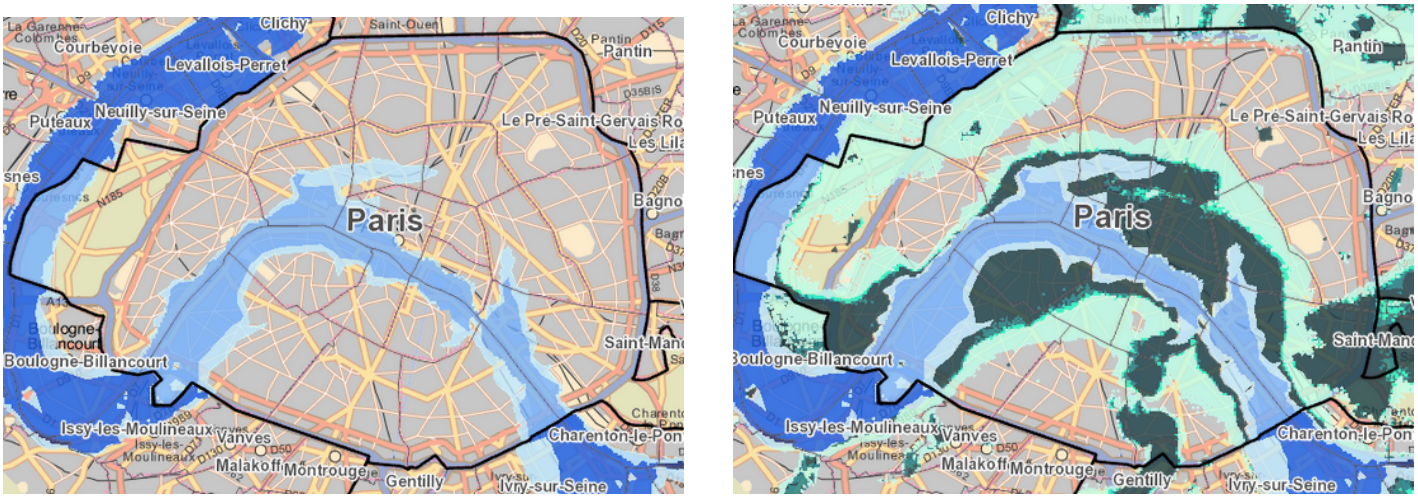

FIGURE 4.2 – Exemple d'aléa d'inondation modélisé sur Paris

Module de vulnérabilité

Afin de modéliser l'exposition de l'ensemble des risques de chaque cédante par rapport à leur localisation, il est nécessaire de recueillir au-près de chacune d'entre elles des données suffisamment précises sur le portefeuille. En effet, il est nécessaire de connaître la localisation des polices mais aussi la nature du risque, l'usage du risque, ainsi que sa valeur assurée. En fonction des données recueillies il est possible de géocoder les risques à l'adresse ou à la commune. Pour cela il faut associer une latitude et une longitude à chaque police. Il est ensuite possible de superposer les zones d'aléa qui ont été simulées par le module d'aléa et les polices géocodées.

Module de dommages

Ce module permet principalement d'estimer la probabilité pour chaque police d'être touchée par le sinistre et le taux de destruction pour chaque police touchée. La probabilité de sinistre pour chaque police est estimée à partir de :

- la probabilité de demande d'état de catastrophe naturelle de la commune en connaissant l'aléa $\mathbb{P}(D_C|\text{Aléa})$;
- la probabilité de reconnaissance de cet état pour la commune si la commune a fait la demande $\mathbb{P}(R_C|D_C)$;
- la probabilité de sinistre pour la police i si l'état de catastrophe naturelle est reconnu pour la commune $\mathbb{P}(S_i|R_C)$.

Les probabilités de demande et de reconnaissance de l'état de catastrophe naturelle des communes sont estimées à partir des données provenant du module d'aléa et de l'historique des phénomènes reconnus en tant que catastrophes naturelles. Le module d'aléa permet de prendre en compte l'intensité du phénomène, c'est-à-dire la hauteur maximale de débordement et le débit maximal de ruissellement sur l'ensemble des risques de la commune. L'historique permet de définir si l'intensité simulée serait prise en charge par le régime d'après l'historique des reconnaissances associées à de telles intensités.

Ensuite, la probabilité de sinistre pour la police i si la commune a la reconnaissance d'état de catastrophe naturelle, est calculée en fonction de l'intensité du phénomène et de la vulnérabilité, c'est-à-dire des caractéristiques des polices.

Après avoir estimé ces probabilités, nous pouvons estimer la probabilité de sinistre :

$$\mathbb{P}(S_i) = \mathbb{P}(D_C|\text{Aléa}) \times \mathbb{P}(R_C|D_C) \times \mathbb{P}(S_i|R_C)$$

Le coût du sinistre est estimé à partir du taux de destruction, ce dernier est estimé en fonction de la hauteur de débordement pour les polices situées dans la zone de débordement et en fonction du débit de ruissellement pour les polices situées dans la zone de ruissellement. Nous multiplions ensuite le taux de destruction t_i par la valeur assurée VA_i pour obtenir le coût C_i pour chaque police :

$$C_i = VA_i \times t_i$$

Nous pouvons observer le coût associée aux inondations entre 1995 et 2012 sur la figure 4.3. La sinistralité prise en compte correspond à la sinistralité à la charge du régime des catastrophes naturelles. Cette figure permet de remarquer l'importance d'avoir la localisation au moins par commune des différentes polices. En effet, le coût de tels événements varie de manière importante d'une commune à l'autre.

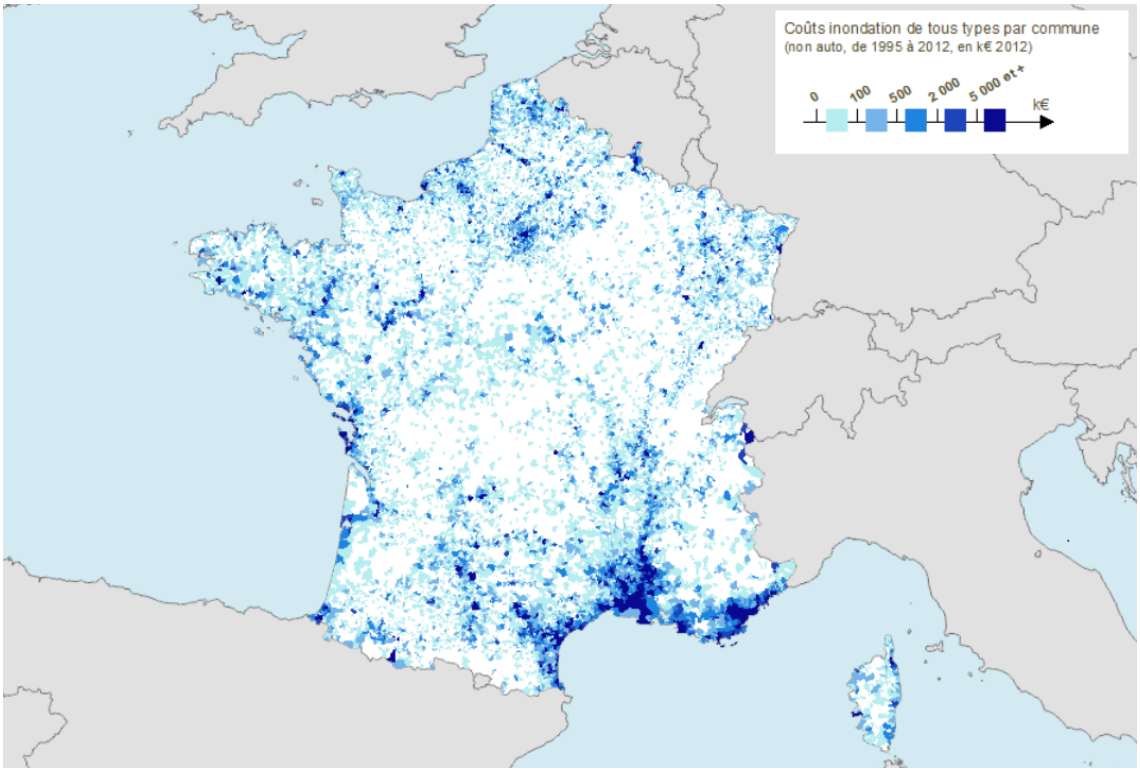

FIGURE 4.3 – Coût moyen des inondations pour les différentes communes

4.0.5 Modèle Sécheresse

Du point de vue du régime de catastrophes naturelles, les phénomènes généralement pris en charge en cas de sécheresse sont les phénomènes de retrait et de gonflement des sols argileux et marneux dus à la déshydratation et à la ré-hydratation de ces derniers. Ces phénomènes sont appelés sécheresses RGA pour Retrait-Gonflement des Argiles. Le régime ne prend pas en compte les sécheresses agricoles.

Module d'aléa

L'aléa du modèle sécheresse dépend de deux valeurs, l'indice d'humidité du sol aussi appelé le *Soil Wetness Index* (ci-après "SWI") et d'un indice développé par le Bureau de Recherches Géologiques et Minières que nous appellerons par la suite l'aléa BRGM. A partie de ces deux paramètres, il est possible de générer des sinistres fictifs.

En ce qui concerne l'indice d'humidité du sol, le territoire est découpé en mailles. Une série temporelle permet ensuite de modéliser un SWI tous les 10 jours pour chaque maille. Cette méthode assure la concordance entre les données historiques et les SWI simulés ainsi que la cohérence spatial du phénomène. Après ajustement d'une série temporelle à chaque maille il est possible de simuler des SWI fictifs pour chacune d'entre elles.

Par ailleurs, l'aléa BRGM divise également le territoire français en 4 zones en fonction de la probabilité de survenance d'une sécheresse RGA. Cet indice peut prendre des valeurs parmi les suivantes :

- a priori nul ;
- faible ;
- moyen ;
- fort.

A partir de ces données, le module d'aléa génère des événements caractérisés en tant que phénomènes physiques. Des indicateurs géologiques et météorologiques permettent ensuite de déterminer la probabilité de demande de l'état de catastrophe naturelle de la commune ainsi que la probabilité de reconnaissance de cet état. L'indicateur géologique est calculé pour chaque police et chaque commune en fonction de l'aléa BRGM et les indicateurs météorologiques sont calculés par commune à partir des indices SWI.

Module de vulnérabilité

De la même manière que pour le péril inondation, il est nécessaire de géocoder les polices afin de pouvoir quantifier l'impact d'un événement sur les différentes polices. Les risques sont géocodés le plus précisément possible, soit à l'adresse exacte, soit au centre de la rue, soit centre de la commune. Le géocodage permettra par la suite de superposer les risques et les zones touchées par les sécheresses afin d'estimer un coût pour chaque police.

Module de dommages

A partir des données provenant des modules d'aléa et de vulnérabilité, le module de dommages évalue le coût des événements fictifs et leurs probabilités. Le module d'aléa renvoie des informations qui concernent les zones sinistrées et le module de vulnérabilité permet d'identifier les polices situées dans ces zones. Comme pour le péril inondation, la probabilité de sinistre $\mathbb{P}(S_i)$ peut être estimée comme suit :

$$\mathbb{P}(S_i) = \mathbb{P}(D_C|\text{Aléa}) \times \mathbb{P}(R_C|D_C) \times \mathbb{P}(S_i|R_C)$$

Nous rappelons les notations suivantes :

- $\mathbb{P}(D_C|\text{Aléa})$ est la probabilité de demande d'état de catastrophe naturelle de la commune en connaissance de l'aléa ;
- $\mathbb{P}(R_C|D_C)$ est la probabilité de reconnaissance de l'état de catastrophe naturelle si la commune fait la demande ;
- $\mathbb{P}(S_i|R_C)$ est la probabilité de sinistre pour la police i si l'état de catastrophe naturelle est reconnu pour la commune.

La probabilité de demande de l'état de catastrophe naturelle par une commune est estimée à partir des indicateurs météorologiques et géologiques au niveau de la commune mais aussi de l'historique des demandes. La probabilité de reconnaissance de l'état de catastrophe naturelle après la demande dépend, quant-à-elle, de certains critères d'éligibilité.

Enfin, la probabilité de sinistres en cas de reconnaissance de l'état de catastrophe naturelle est également estimée à partir des indicateurs météorologiques au niveau de la commune et des indicateurs géologiques au niveau du risque. Un modèle a été développé pour chacune des trois zones d'aléa BRGM à risque, à savoir : aléa faible, moyen et fort. Les zones à aléa nul sont exclues de ces estimations.

En outre, nous pouvons estimer le taux de destruction en fonction de l'intensité du phénomène à l'aide des indicateurs météorologiques et géologiques. Le coût C_i pour le risque i sachant qu'il a été sinistré est ensuite calculé à partir du taux de destruction t_i et de la valeur assurée VA_i.

$$C_i = VA_i \times t_i$$

De manière similaire aux inondations, nous pouvons observer le coût associée aux sécheresses entre 1995 et 2012 sur la figure 4.4. Encore une fois, la sinistralité prise en compte correspond à la sinistralité à la charge du régime des catastrophes naturelles. Nous notons à nouveau l'importance d'avoir le géocodage des polices par commune minimum.

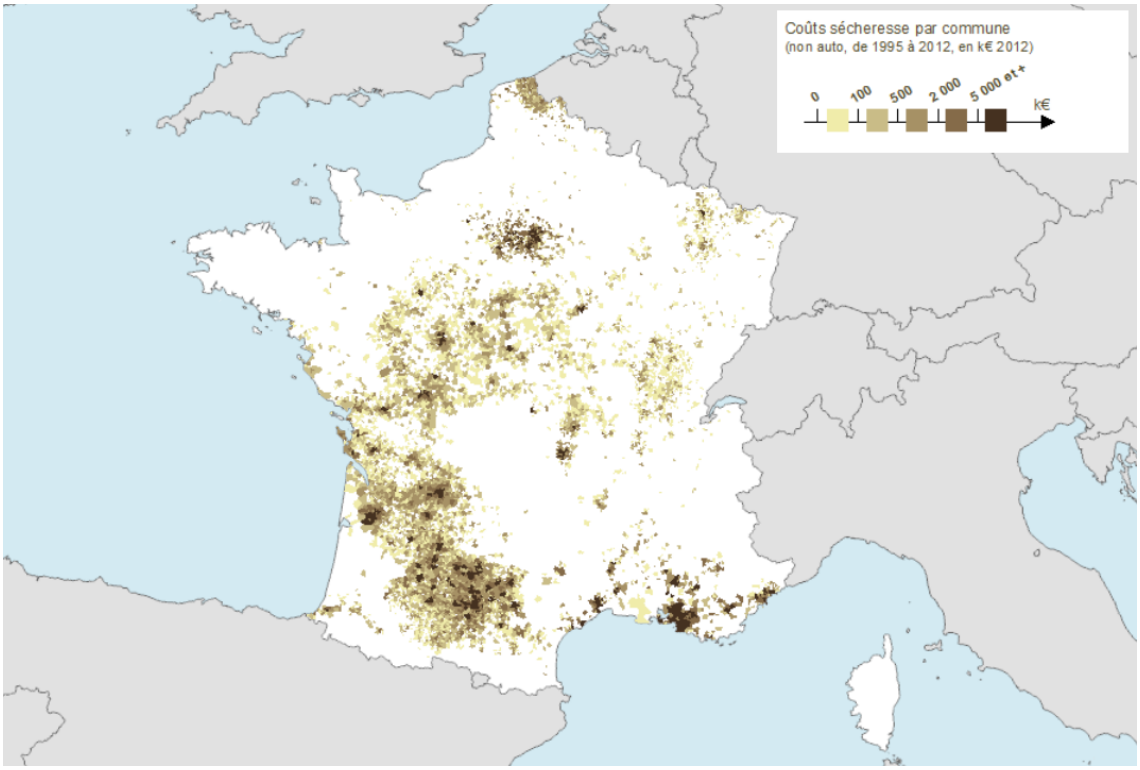

FIGURE 4.4 – Coût moyen des sécheresses pour les différentes communes

4.0.6 Modèle Séisme

Le modèle physique pour le péril séisme a été développé par la société RMS. Cette dernière transmet à la CCR une *Year Loss Table* (ci-après "YLT") permettant de modéliser le coût des événements. Les explications ci-dessous sont issues des explications et de la documentation transmise par la société en question. Le logiciel développé par la société RMS est appelé RiskLink.

Module d'aléa

Le module d'aléa du péril séisme est constitué de deux parties. La première partie est la création d'un catalogue d'événements sismiques caractérisés par la magnitude, la localisation et la période de retour :

- la localisation est modélisée à partir d'événements historiques, de failles réelles et de failles dites synthétiques. Les failles synthétiques permettent de modéliser les événements qui sont

associés à des failles qui ne sont pas encore connues ;

- la période de retour est définie en fonction d'études approfondies géologiques pour les failles réelle et les failles synthétiques. Dans le cas des événements historiques, les périodes de retour sont suffisamment importantes pour que la probabilité d'avoir un même événement deux fois dans le catalogue soit négligeable ;

- la magnitude est quantifiée selon l'échelle de magnitude du moment. Une limite inférieure est définie comme la magnitude minimale causant des dommages importants. La limite supérieure dépend de l'impact que peut avoir une faille. Dans le cas des failles réelles, cela dépend de la taille de la faille, et dans le cas des failles synthétiques cela dépend de l'historique des tremblements de terre et de la région.

Nous pouvons observer l'aléa des séismes en France d'après le Ministère de l'Écologie, du Développement Durable et de l'Énergie sur la figure 4.5.

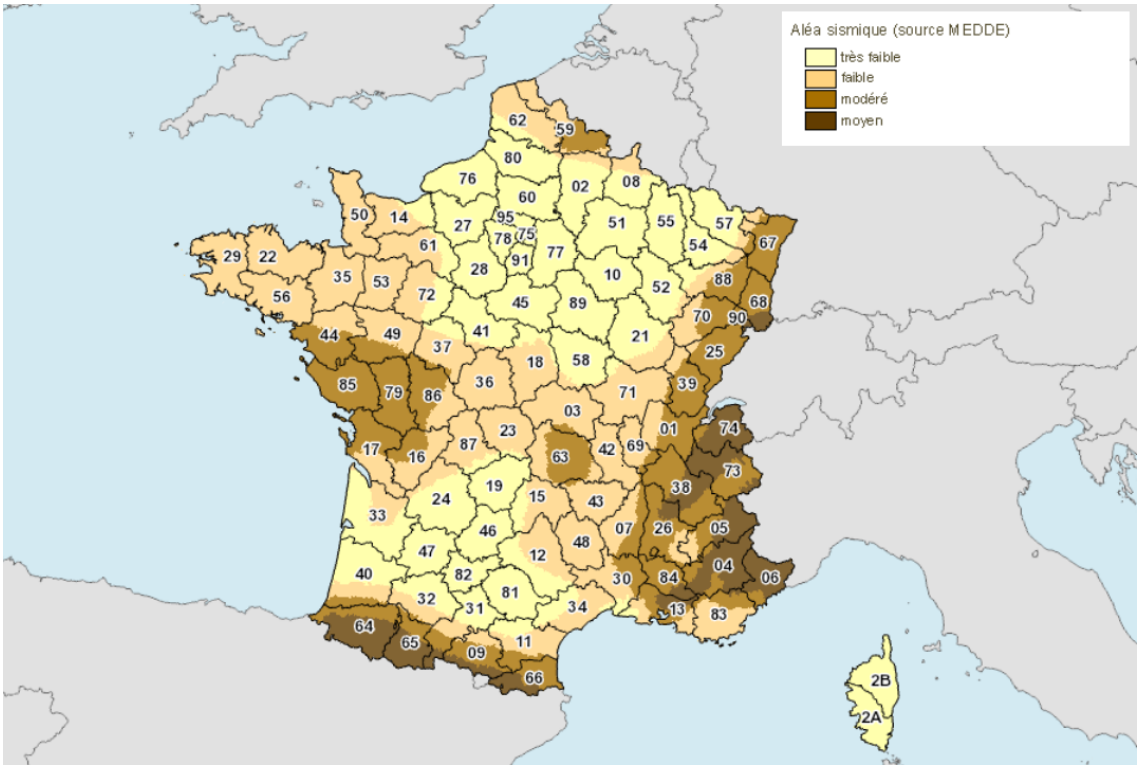

FIGURE 4.5 – Aléa des séismes

La deuxième partie de ce module consiste à modéliser l'impact de chaque événement en termes de mouvement du sol dans le but de pouvoir déterminer un taux de destruction. Le logiciel RiskLink permet de modéliser l'impact physique du phénomène pour chaque événement du catalogue décrit précédemment. Cette modélisation se fait en deux étapes :

- le phénomène principal ayant un impact sur les constructions est le tremblement du sol dû à l'atténuation des ondes. Ce dernier dépend de la distance par rapport à la source appelée le foyer ;

- un deuxième phénomène peut se superposer à l'atténuation des ondes, il s'agit de l'amplification ponctuelle des ondes due aux caractéristiques superficielles du sol.

Module de vulnérabilité

Comme pour les modèles précédents, il est nécessaire de localiser les polices qui se trouvent dans la zone touchée par la catastrophe afin de quantifier l'impact de l'événement. Le module de vulnérabilité permet encore une fois de géocoder les polices le plus précisément possible et de recueillir un maximum d'informations sur le risque couvert.

Dans le cas des séismes, des informations plus détaillées concernant les caractéristiques du bâtiment sont nécessaires. En effet, l'impact d'un événement dépend également de la résistance des constructions. Il faut donc tenir compte de nombreux paramètres tels que :

- le type d'occupation ;
- la hauteur des constructions ;
- les matériaux utilisés dans les constructions ;
- la date de construction des bâtiments ;
- l'application de règles para-sismiques ;
- l'état des constructions.

Module de dommages

Le module dommages de ce péril permet de modéliser le taux de destruction sous la forme de courbes. Chaque courbe de taux de destruction est définie à partir du taux moyen et de l'incertitude, et elle est calibrée en fonction de l'historique et d'études d'ingénierie.

Ce module permet d'estimer le taux d'endommagement pour chaque police en fonction de la courbe de taux et des données provenant du module d'aléa et du module de vulnérabilité. L'événement défini la courbe qui sera utilisée en fonction de la magnitude. Les spécificités des polices, telles que les caractéristiques des constructions décrites précédemment, vont ensuite déterminer le taux d'endommagement de chacune d'entre elles.

De cette manière, le taux de destruction prend en compte les données du module d'aléa qui concernent l'événement et varie en fonction des données du module de vulnérabilité qui concernent les caractéristiques des polices.

4.0.7 Modèle Vents Cycloniques

Le régime des catastrophes naturelles ne prend en charge que les événements cycloniques ayant lieu dans les Départements d'Outre-Mer (ci-après "DOM"). Cette couverture concerne tous les sinistres causés par les vents cycloniques, que ce soit directement ou indirectement via des inondations, des submersions marines, des crues, des glissements de terrain, ou même des coulées de boue. Deux modèles ont été créés afin de modéliser le péril aux Antilles et à la Réunion. L'activité cyclonique aux Antilles est modélisée à l'aide du logiciel RiskLink. Le modèle utilisé pour l'île de la Réunion a été développé en interne.

Module d'aléa

Le module d'aléa utilisé pour les vents cycloniques sur les Antilles a été développé par la société RMS. Il s'agit d'un catalogue d'événements qui peuvent être historiques ou fictifs qui sont caractérisés par :

- la trajectoire du cyclone ;
- des indices météorologiques sur toute la trajectoire ;
- la fréquence annuelle de survenance.

La fréquence est estimée à partir de la fréquence historique des événements et prend également en compte l'avis des experts quant à l'évolution des cyclones.

Le module aléa pour l'activité cyclonique sur l'île de la Réunion renvoie aussi un catalogue d'événements. Pour cela, le module a été développé selon trois axes qui sont supposés indépendants :

- les vents ;
- le ruissellement suite à des fortes pluies ;
- la submersion marine.

Module de vulnérabilité

Dans le cas des Antilles, le module de vulnérabilité permet de recueillir des caractéristiques des constructions nécessaires pour l'estimation du coût des dommages, telles que :

- le type de construction ;
- le type d'occupation ;
- la hauteur des constructions ;

- la date de construction des bâtiments ;
- la superficie des constructions.

Pour l'activité cyclonique sur l'île de la Réunion, le module de vulnérabilité permet de construire un portefeuille "marché" sur l'île à partir des données de l'Institut National de la Statistique et des Études Économiques (ci-après "INSEE"). Le portefeuille est géocodé à la commune et est caractérisé par le nombre de polices pour chaque type de bâti, les valeurs assurées et la prime correspondant au Régime CatNat.

Module de dommages

Dans le cas des Antilles, le module de dommages estime le taux d'endommagement moyen et son coefficient de variation pour chaque événement du catalogue fourni par le module d'aléa, et ce pour les structures, le contenu et les pertes d'exploitation. Un coefficient de redressement est ensuite appliqué au taux d'endommagement moyen en fonction des spécificités des constructions qui sont fournies par le module de vulnérabilité.

Dans le cas de l'île de la Réunion, l'endommagement est mesuré en fonction des matériaux et de leur vulnérabilité aux aléas. Pour cela, le module de dommages fait appel à des courbes d'endommagement en fonction de l'intensité de l'aléa qui est modélisée par le module d'aléa et qui permettent d'estimer le taux d'endommagement sur chaque police du portefeuille créé par le module de vulnérabilité.

Chapitre 5

Tarification des traités

Nous avons vu précédemment que la prime associée au quote-part correspond à une portion des primes de la cédante. Cependant, la prime de réassurance associée à l'excédant de perte annuelle est quant-à-elle calculée via une méthode de tarification. Pour cela, la CCR a développé deux approches de tarification : une tarification par expérience qui repose sur l'historique de la sinistralité des cédantes, et une tarification par exposition qui repose sur la modélisation physique des périls. Les deux méthodes se basent sur une approche probabiliste.

5.1 Tarification par expérience

La tarification par expérience se fait à partir de l'historique de la sinistralité annuelle de la cédante. Celle-ci se base sur une méthode probabiliste qui consiste à ajuster une loi de probabilités aux rapports Sinistres/Primes (ci-après "S/P").

L'utilisation du rapport S/P permet de supprimer les évolutions qui sont liées à l'inflation ainsi qu'aux variations de la taille du portefeuille des cédantes. De cette manière, les augmentations des montants des sinistres dues à l'augmentation du nombre d'assurés ou à l'inflation n'ont pas d'impact sur la distribution.

A partir des données historiques, nous calculons les rapports S/P et nous ajustons par maximum de vraisemblance quatre lois de probabilités :

- la loi Log-Normale ;

- la loi de Weibull ;
- la loi Gamma ;
- la loi Exponentielle.

Plusieurs tests statistiques, tels que le test de Cramer-von Mises et le test de Kolmogorov-Smirnov, sont effectués afin de vérifier le bon ajustement des lois. Ces tests sont présentés dans l'annexe 11.1. En fonction des résultats de ces tests nous pouvons retenir la meilleure approche.

La tarification consiste ensuite à calculer le taux pur et le taux technique à partir de ces lois. Le taux pur correspond à l'espérance du rapport S/P à la charge du stop-loss. Il peut être calculé soit à partir de la formule de l'espérance, soit par la méthode de Monte-Carlo en simulant des événements à partir de la loi théorique.

$$Taux\ pur = \mathbb{E}\left(\frac{Sinistres}{Primes}\right)$$

Le taux technique peut ensuite être calculé par application des chargements au taux pur. La prime technique correspond au produit entre le taux technique et les primes de la cédante.

$$Prime\ technique = Taux\ technique \times Primes$$

Le schéma 5.1 permet d'illustrer les différentes étapes de la tarification par expérience, à partir des données transmises par chaque cédante, jusqu'à l'obtention de la prime technique de réassurance.

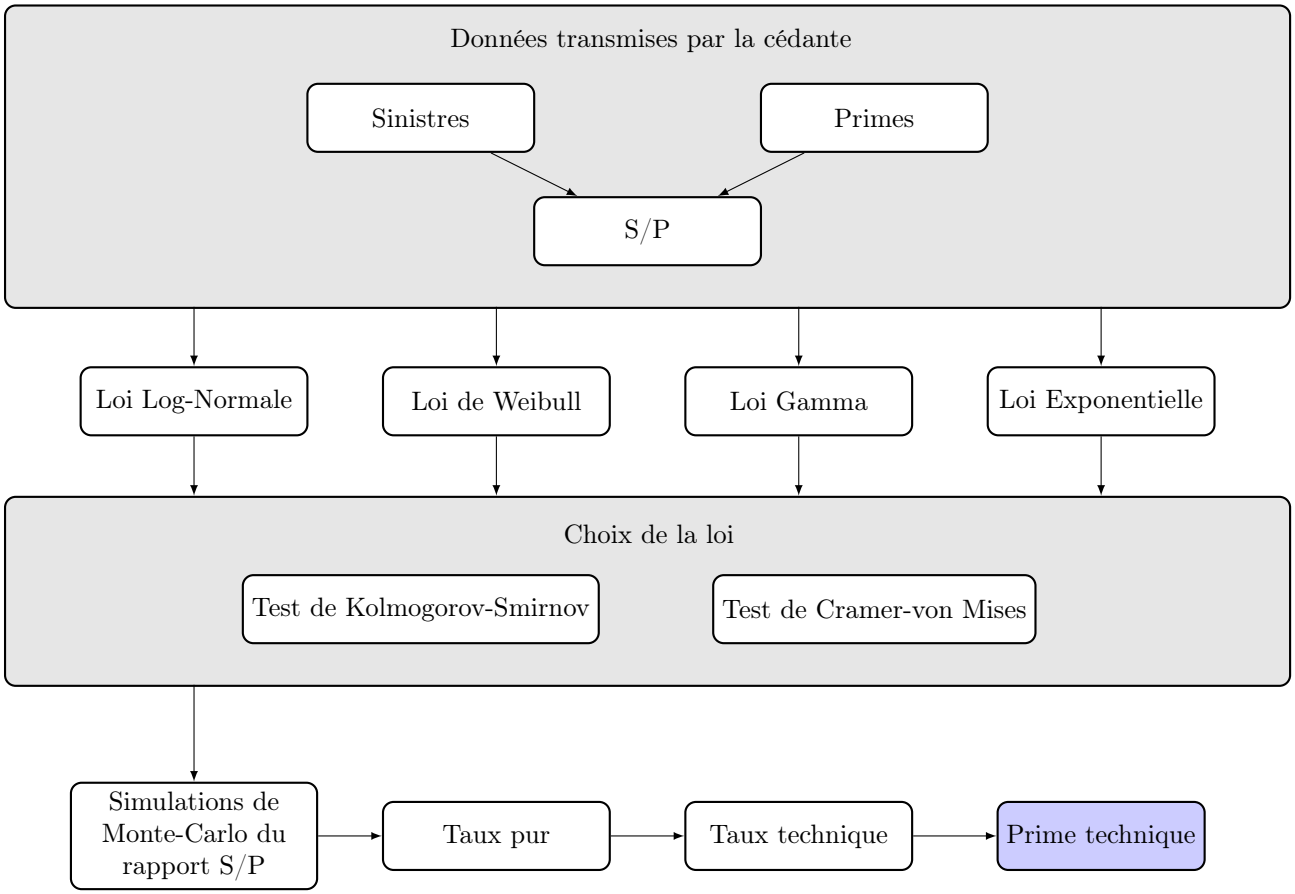

FIGURE 5.1 – Schéma de la tarification par expérience

Cependant, cette approche ne prend pas en compte l'exposition liée à la localisation des différentes polices couvertes par le traité. Nous pouvons observer sur la figure 5.2, que la sinistralité d'une cédante est fortement liée à la localisation des polices. Prenons l'exemple de deux cédantes A et B ayant souscrit 10 polices à Paris. La cédante A a souscrit des polices dans une zone dite "à risque" puisque dans le cas d'un débordement de la Seine, ces dernières risquent d'être touchées par l'événement. La cédante B, au contraire, a souscrit des polices qui sont hors danger dans le cas d'un tel débordement. En cas de survenance d'un débordement de la Seine tel que celui qui est présenté sur la figure 5.2, toutes les polices de la cédante A seraient touchées, alors que aucune des polices de la cédante B ne serait touchée. La sinistralité de la cédante A serait donc largement supérieure à la sinistralité de la cédante B.

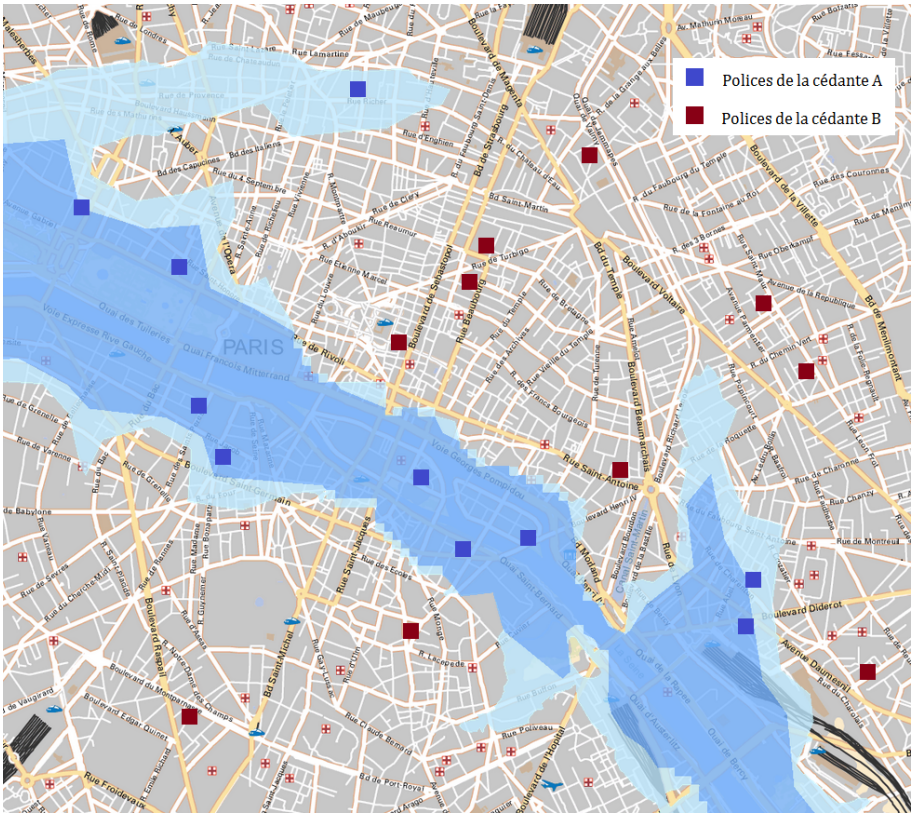

FIGURE 5.2 – Exemple d'exposition associé à la localisation des polices

Si nous utilisons la méthode de tarification sur expérience décrite ci-dessus, la prime de réassurance de ces deux cédante prendrait en compte la sinistralité historique de ces deux dernières qui reflète l'exposition puisque les cédantes ayant souscrit des polices dans des zones à risque ont vraisemblablement plus de sinistres dans leur historique. Cependant, si aucun débordement de la Seine est pris en compte dans l'historique des cédantes, l'exposition à ce risque en particulier n'est pas pris en compte dans la tarification.

En outre, cette méthode de tarification ne prend pas en compte la souscription de nouveaux risques non pris en compte dans la prime calculée en fonction de l'historique. En effet, l'historique de ces risques n'est pas inclus dans l'historique de la cédante.

Nous avons présenté ici un cas qui évidemment extrême. Toutefois, il est important pour un réassureur tel que la CCR de modéliser son exposition au travers celle de ses cédantes afin de prendre en compte les sinistres qui ne sont pas encore survenus ou qui ne sont pas pris en compte dans

l'historique des cédantes.

Pour cette raison, la CCR a développé une deuxième approche qui est utilisée afin de modéliser le risque encouru en fonction de la localisation des risques du portefeuille de la cédante : la tarification par exposition. Il est important de noter que même si cette tarification est appelée par exposition, la tarification par expérience prend également en compte l'exposition des cédantes mais avec moins de précision.

5.2 Tarification par exposition

Nous avons vu précédemment que la modélisation physique des périls couverts par les traités, permet de prendre en compte la localisation des polices et par conséquent l'exposition de chaque police. Les modèles physiques renvoient des *Year Loss Table*. A partir de ces dernières on procède à la tarification par simulation à partir d'un modèle appelé le Modèle Interne Partiel (ci-après "MIP").

Deux modèles type MIP ont été développés à la CCR :

- le *Modèle Top Down* permet de modéliser le coût par péril associé à l'ensemble des cédantes CCR et de le répartir en fonction des parts de marché ;
- le *Modèle Bottom Up* permet de modéliser le coût de chaque cédante par péril, renvoyant ainsi des résultats plus précis mais avec un temps de calcul plus important.

Dans cette étude nous allons nous intéresser au *Modèle Bottom Up* (ci-après "MBU"). Ce dernier a été développé afin d'évaluer le plus finement possible l'exposition des cédantes, de la CCR et de l'Etat. En effet, au vu de son rôle dans le régime des catastrophes naturelles, la CCR est tenue d'estimer les engagements de ses clients ainsi que de son actionnaire. Ce modèle permet d'évaluer les tarifs des cédantes, ainsi que le besoin en capital réglementaire pour les catastrophes naturelles. En outre, ce modèle permet d'apprécier facilement des éventuelles évolutions au niveau du régime et au niveau des traités délivrés par la CCR.

Le *Modèle Bottom Up* a été créé à l'aide du logiciel Remetrica. Ce logiciel permet de simuler la survenance des événements ainsi que leur coût et d'appliquer les conditions contractuelles. Le MBU permet ainsi d'effectuer des simulations de type Monte-Carlo à partir des *Year Loss Tables* de chaque cédante provenant des modèles physiques. La modélisation de la charge des cédantes, de la CCR et de l'Etat se fait donc par étapes :

- chargement des générateurs de sinistralité des cédantes à partir de la fréquence et du coût des événements provenant des YLT ;
- modélisation de la sinistralité annuelle des cédantes pour chaque péril à l'aide des générateurs décrits ci-dessus ;
- agrégation au niveau de chaque cédante pour obtenir une charge brute unique associée aux catastrophes naturelles ;
- application des conditions contractuelles afin d'obtenir la charge nette de chaque cédante, c'est-à-dire la rétention de la cédante ;
- agrégation de la cession de toutes les cédantes pour obtenir la charge brute annuelle de la CCR associée aux catastrophes naturelles ;
- application du seuil d'intervention de l'État permettant d'obtenir la charge nette annuelle de la CCR d'une part, et la charge de l'État d'autre part.

Il est important de noter que les YLT sont corrélées à 100% de sorte à modéliser les mêmes sinistres pour toutes les cédantes.

Le graphique 5.2 permet d'illustrer les différentes étapes décrites précédemment et de mettre en évidence les différences entre le *Modèle Top Down* et le *Modèle Bottom Up*.

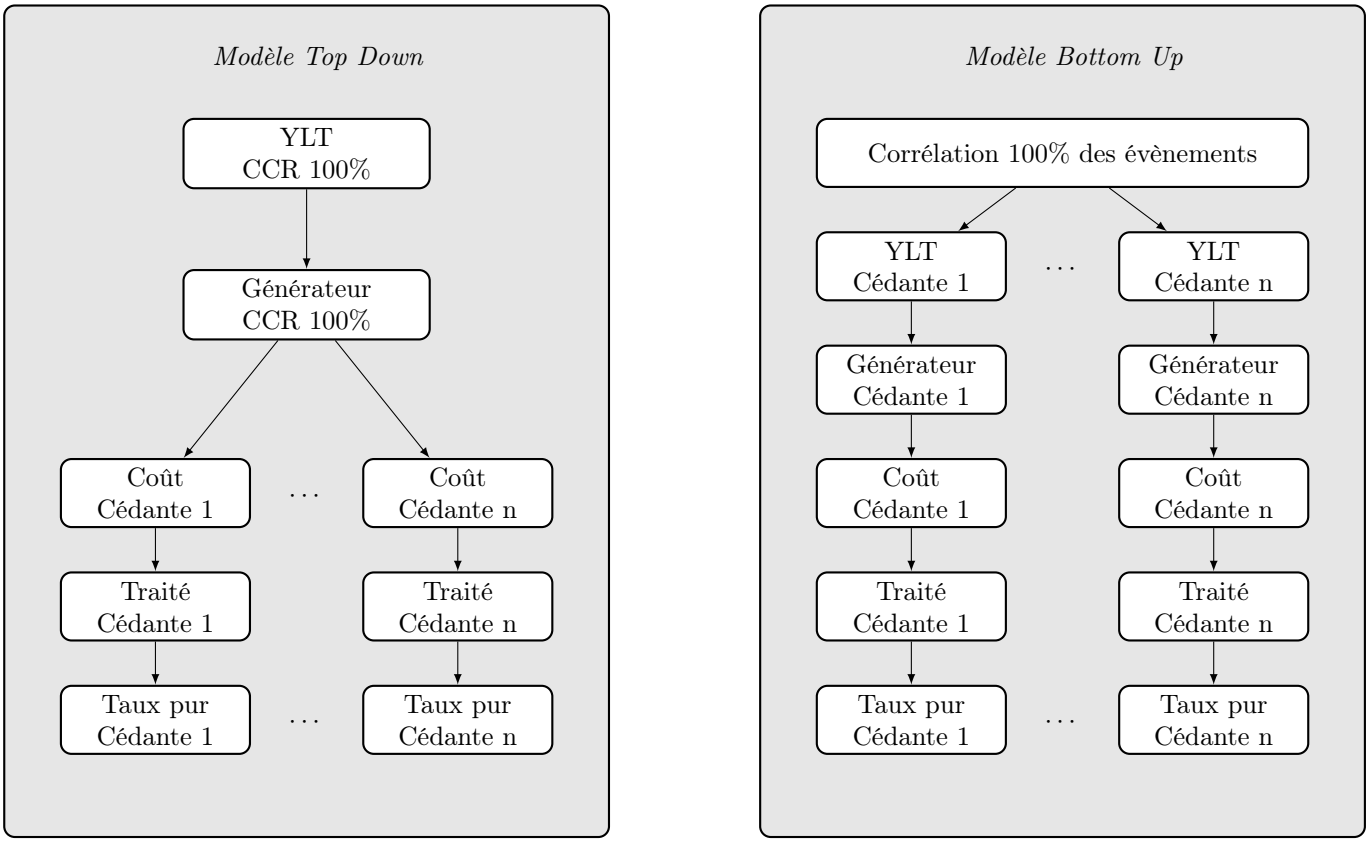

FIGURE 5.3 – Schéma de la tarification par exposition

Toutefois, les modèles physiques aboutissant aux modèles MTD et MBU sont calibrés afin de bien s'ajuster aux événements extrêmes dont les périodes de retour sont importantes et qui ne sont pas reflétés dan l'historique des données. Pour cette raison, la distribution provenant de ces modèles n'est pas considérée comme étant suffisamment fiable sur les petites périodes de retour pour pouvoir baser la tarification uniquement sur cette distribution. Ainsi, la tarification par exposition est uniquement utilisée à titre indicatif.

Chapitre 6

Objectif de l'étude

Nous avons vu qu'actuellement, les traités de réassurance des catastrophes naturelles de la CCR sont tarifés par expérience. En effet, la méthode de tarification par exposition est uniquement utilisée à titre indicatif. Par ailleurs, le besoin en capital réglementaire de solvabilité (ci-après "SCR") brut pour les catastrophes naturelles est estimé à partir du MBU. Comme le SCR brut dépasse largement le seuil d'intervention de l'État, la CCR retient un SCR net égal au seuil d'intervention de l'État.

Le but de cette étude est donc de trouver un compromis entre les deux méthodes de modélisation (historique et exposition) et de créer un modèle hybride qui permette à la CCR de :

- prendre en compte les informations provenant des modèles physiques concernant l'exposition des cédante en fonction de la localisation des polices ;
- prendre en compte l'historique de la sinistralité de la cédante ;
- tarifer des traités en prenant en compte l'historique des cédantes et leur exposition calculée à partir du MBU ;
- calculer le SCR brut à partir l'historique des catastrophes naturelles et l'exposition de la CCR.

Pour cela nous disposons de deux jeux de données : un jeu de données sur la sinistralité historique des cédantes et un jeu de données provenant du MIP qui correspond à l'exposition de la cédante modélisée par des modèles physiques. Le but est de trouver des données compatibles entre les deux jeux de données afin de pouvoir les exploiter et de créer le modèle hybride.

6.1 Données historiques

Dans le cas de la sinistralité historique nous avons accès aux données qui sont communiquées à la CCR par les cédantes dans le cadre du régime d'indemnisation des catastrophes naturelles. S'agissant d'une garantie sur la sinistralité annuelle de tous les événements reconnus comme des catastrophes naturelles, la sinistralité communiquée par la cédante est la sinistralité annuelle agrégée de tous les événements reconnus par la commission interministérielle. Nous ne disposons donc pas du détail par péril. Toutefois nous disposons du détail selon s'il s'agit de contrats d'assurance automobile ou pas.

Nous disposons également des données sur la sinistralité brute agrégée de toutes les cédantes de la CCR que nous appellerons la sinistralité "marché CCR". Il s'agit ici du coût brut pour les cédantes de la CCR avant application des traités de réassurance, à ne pas confondre avec le coût CCR qui correspond au coût net pour la CCR après application des traités de réassurance. Si nous notons $\mathbf{I}$ est l'ensemble des cédantes de la CCR, nous obtenons :

$$\text{Coût marché CCR} = \sum_{i \in \mathbf{I}} \text{Coût brut de la cédante } i$$

A ce niveau nous disposons du détail selon les catégories suivantes :

- péril inondation hors contrats d'assurance automobile ;
- péril sécheresse hors contrats d'assurance automobile ;
- autres périls hors contrats d'assurance automobile ;
- contrats d'assurance automobile.

6.2 Données sur l'exposition

Dans le cas de l'exposition, nous avons accès aux données modélisées par le MIP. Il est donc possible d'avoir la sinistralité annuelle pour chaque cédante selon les catégories suivantes :

- péril inondation hors contrats d'assurance automobile ;
- péril sécheresse hors contrats d'assurance automobile ;
- séismes hors contrats d'assurance automobile ;

- vents cycloniques hors contrats d'assurance automobile ;

- contrats d'assurance automobile.

Il est également possible d'avoir la sinistralité marché CCR avec le même détail que pour les cédantes.

6.3 Compatibilité des données

D'après les deux paragraphes précédents, il est possible d'avoir des données historiques et l'exposition de chaque cédante selon les catégories suivantes :

- sinistralité des contrats d'assurance automobile associée à des catastrophes naturelles, que nous appelerons par la suite "sinistralité automobile" ;

- sinistralité liée aux catastrophes naturelles hors contrats d'assurance automobile, que nous appelerons par la suite "sinistralité non automobile".

Il est également possible d'avoir des données la sinistralité marché CCR selon les catégories :

- péril inondation hors contrats d'assurance automobile ;

- péril sécheresse hors contrats d'assurance automobile ;

- autres périls hors contrats d'assurance automobile ;

- contrats d'assurance automobile.

6.4 Approches utilisées

Dans un premier temps, nous allons choisir un modèle qui corresponde aux critères de la CCR qui sont précisés dans la partie suivante. Il s'agira ici d'un modèle permettant de générer des distributions hybrides à partir des distributions historique et d'exposition. Pour cela nous utiliserons les données « marché CCR » afin de calibrer le modèle.

Ensuite, nous utiliserons ces modèles sur les données de chaque cédante afin de créer des distributions hybrides par cédante en distinguant les contrats d'assurance automobile des contrats hors automobile. A partir de ces distributions hybrides nous allons ensuite créer un Modèle Interne Partiel Hybride.

6.5 Critères basés sur l'expertise CCR

Nous avons vu précédemment que le but principal de la création d'une distribution hybride est la prise en compte des informations concernant l'historique de la sinistralité et l'exposition des cédantes. Toutefois, l'expertise de la CCR permet de définir des contraintes supplémentaires.

Au vu de la taille de l'historique, nous supposons que ce dernier est fiable que dans une certaine limite. Toutefois, les modèles physiques qui modélisent l'exposition sont quant-à-eux calibrés afin de modéliser au mieux les grandes périodes de retour et de prendre en compte des événements extrêmes qui ne sont pas présents dans l'historique.

De cette manière, l'historique est plus fiable que l'exposition puisqu'il reflète directement la sinistralité des cédantes pour les petites périodes de retour, mais à partir d'une certaine période de retour il faut tenir compte uniquement de l'exposition, puisque l'historique n'est plus fiable. Pour ces raisons, dans notre analyse, nous allons nous intéresser aux périodes de retour et aux niveaux de retour associés.

Chapitre 7

Théorie bayésienne

La théorie bayésienne est une approche très répandue dans le milieu de l'assurance. Elle est notamment utilisée en crédibilité, pour prendre en compte des données ou des informations provenant de sources différentes. La crédibilité est principalement utilisée dans le cas des contrats d'assurance automobile afin de prendre en compte la sinistralité historique de la cédante par rapport à tous ses assurés mais aussi l'historique de chaque individu. La crédibilité permet ainsi de prendre en compte l'expérience de l'assureur et les particularités de chaque assuré.

Dans cette étude, nous allons utiliser la théorie bayésienne afin de créer une distribution prenant en compte les informations provenant des modèles d'exposition et l'historique de la sinistralité de la CCR. Ce modèle permettrait ainsi d'utiliser l'expertise du pôle Recherche & Développement au travers des informations provenant des modèles d'exposition de la CCR et l'expérience de la CCR en tant que réassureur au travers de l'historique de ses cédantes. Cette section permettra de présenter des notions et des résultats de la théorie bayésienne afin de pouvoir l'appliquer par la suite.

7.1 Généralités sur la théorie bayésienne

Soient $X_1, ..., X_n$ des variables aléatoires indépendantes et identiquement distribuées. Nous supposons que pour tout $i \in [1, ..., n]$, la variable X_i suit une loi $\mathcal{L}$ de paramètre θ. Le paramètre θ est une variable aléatoire qui contient des informations a priori, il s'agit ici de la première source d'information.

$$X_i \sim \mathcal{L}(\theta) \qquad \forall i \in [1, ..., n]$$

Soit $\mathbf{y} = (y_1, ..., y_k)$ le jeu de données correspondant à la deuxième source d'information. Le but de l'approche bayésienne est de prendre en compte les nouvelles informations provenant de cette source dans le paramètre θ afin de pouvoir modéliser une nouvelle variable aléatoire $Z_i = (X_i|\mathbf{y})$ prenant en compte les informations provenant de θ et de $\mathbf{y}$. De telle manière que :

$$Z_i \sim \mathcal{L}(\theta|\mathbf{y}) \qquad \forall i \in [1, ..., n]$$

7.1.1 Loi a priori

La loi a priori est la loi associée au paramètre θ que nous souhaitons crédibiliser. Cette loi comporte l'information de base provenant de la source principale. La densité de cette loi est notée $\pi(\theta)$. Les paramètres de la distribution de θ seront appelés les hyperparamètres.

7.1.2 Loi a posteriori

La loi a posteriori est la loi comportant l'information apportée par les deux sources. Il s'agit ici de la loi de θ sachant $\mathbf{y}$: $\pi_{\mathbf{y}}(\theta)$. Cette dernière dépend de la loi des observations y_i de densité $f(y)$ et de la loi a priori.

$$\pi_{\mathbf{y}}(\theta) = \frac{f(\mathbf{y}|\theta) \cdot \pi(\theta)}{\displaystyle\int_{\Theta} f(\mathbf{y}|\theta) \cdot \pi(\theta)\ d\theta}$$

7.1.3 Loi conjuguée

Afin d'appliquer la théorie bayésienne, il faut définir la loi a priori π que nous allons utiliser. Ici, nous avons utilisé la loi conjuguée dont la définition se trouve ci-après. Le but de cette approche est d'avoir la possibilité d'exploiter facilement la distribution a posteriori dès qu'une nouvelle donnée est recueillie.

Définition 1. *La loi conjuguée est définie telle que la loi a priori qui permet de garder la même distribution entre l'a priori et l'a posteriori.*

$$\pi(\theta) \in \mathcal{F} \Rightarrow \pi_{\mathbf{y}}(\theta) \in \mathcal{F}$$

Il est donc nécessaire de choisir $\mathcal{F}$ telle que :

$$\pi_{\mathbf{y}}(\theta) \in \mathcal{F}$$

Ceci équivaut à choisir $\mathcal{F}$ telle que :

$$f(\mathbf{y}|\theta) \cdot \pi(\theta) \in \mathcal{F}$$

En effet, si on raisonne de manière proportionnelle, en notant "$\propto$" le symbole proportionnel nous obtenons :

$$\pi_{\mathbf{y}}(\theta) \propto f(\mathbf{y}|\theta) \cdot \pi(\theta)$$

Pour une explication plus détaillée sur le raisonnement proportionnel voir l'annexe 11.2.

Dans le but de créer une distribution crédibilisée nous allons procéder par étapes. Un schéma sera présenté à la fin de chaque sous-section permettant ainsi d'identifier les différentes étapes. D'abord nous allons déterminer la loi que nous souhaitons étudier et sa loi conjuguée, pour cela nous étudierons la forme de la fonction de vraisemblance. Ensuite, nous estimerons les paramètres de la loi a priori, que nous appellerons les hyperparamètres a priori. Par la suite nous estimerons les hyperparamètres a posteriori, c'est-à-dire les paramètres de la loi a posteriori. Enfin, nous effectuerons des simulations, pour cela nous allons simuler les paramètres de la distribution crédibilisée et ensuite nous simulerons la sinistralité crédibilisée.

Dans cette étude, nous allons nous intéresser à la crédibilisation des lois Log-Normale et Gamma. En effet, il s'agit de deux lois qui sont actuellement utilisées pour la tarification par expérience.

7.2 Loi Log-Normale

Pour utiliser l'approche bayésienne, il est nécessaire de choisir le paramètre que nous souhaitons crédibiliser. Dans le cas de la loi Log-Normale il est possible de crédibiliser l'espérance de la loi et l'inverse de la variance que nous appellerons par la suite le paramètre de précision. Dans cette section, nous allons présenter les deux approches.

7.2.1 Espérance

Cette première approche consiste à crédibiliser l'espérance d'une loi Normale. Dans le cas d'une loi Log-Normale, il suffit d'utiliser la même approche sur le logarithme des données (étape 0 du schéma 7.2.1). Soient $X_1, ..., X_n$ des variables indépendantes et identiquement distribuées telles que :

$$X_i \sim \text{Normale}(\mu_{\text{ a priori }}, \sigma^2) \qquad \forall i \in [1, ..., n]$$

Nous allons déterminer la variance à partir de ces données et nous procéderons ensuite par étapes pour crédibiliser ces variables, les différentes étapes sont présentées dans le schéma 7.2.1. Première-

ment, nous allons déterminer la forme de la famille de la loi a priori $\pi(\mu)$. Dans un deuxième temps, nous allons estimer les hyperparamètres a priori de la loi en question. Enfin, nous estimerons les hyperparamètres a posteriori afin de modéliser la nouvelle variable :

$$Z \sim \text{Log-Normale}(\mu_{\text{ a posteriori }}, \sigma^2)$$

Loi a priori

Pour déterminer la forme de la loi a priori $\pi(\mu)$, il est nécessaire d'établir la forme de la fonction de vraisemblance. Nous rappelons que k est la taille de l'échantillon $\mathbf{y}$ que nous souhaitons utiliser pour crédibiliser. La fonction de vraisemblance est la suivante :

$$f(\mathbf{y}|\mu) = \frac{1}{\sqrt{2\pi\sigma^2}^k} \cdot \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{k}(y_i - \mu)^2\right)$$

Si on note : $\begin{cases} \bar{y} = \dfrac{1}{k}\displaystyle\sum_{i=1}^{k} y_i \\ \\ s^2 = \dfrac{1}{k}\displaystyle\sum_{i=1}^{k}(y_i - \bar{y})^2 \end{cases}$

En remplaçant, nous obtenons la fonction de vraisemblance suivante :

$$f(\mathbf{y}|\mu) = \frac{1}{\sqrt{2\pi\sigma^2}^k} \cdot \exp\left(-\frac{ks^2 + k(\mu - \bar{y})^2}{2\sigma^2}\right)$$

Il est important de noter que σ^2 est le paramètre de la loi qui est supposé connu, il ne doit pas être confondu avec la variance s^2 de l'échantillon $\mathbf{y}$. Si on réfléchit de manière proportionnelle, comme s^2 et σ^2 sont constants, nous avons :

$$f(\mathbf{y}|\mu) \propto \exp\left(-\frac{(\mu - \bar{y})}{2\sigma^2/k}\right)$$

On remarque dans l'expression précédente, le noyau d'une loi Normale. La fonction de vraisemblance appartient donc à la famille des lois Normales. Le détail des calculs se trouve dans l'annexe 11.2.2.

Nous pouvons démontrer maintenant que la loi conjuguée est une loi Normale (pour plus d'infor-

mations se référer aux sources [1], [4] et [7]). Prenons la loi a priori suivante :

$$\mu \sim \text{Normale}(m, v)$$

Alors nous avons :

$$\begin{aligned} \pi_{\mathbf{y}}(\mu) &\propto f(\mathbf{y}|\mu) \cdot \pi(\mu) \\ &\propto \exp\left(-\frac{(\mu-\bar{y})^2}{2\sigma^2/k}\right) \times \exp\left(-\frac{(\mu-m)^2}{2\cdot v}\right) \end{aligned}$$

En développant, nous obtenons :

$$\pi_{\mathbf{y}}(\mu) \propto \exp\left(-\frac{\mu^2}{2}\left(\frac{k}{\sigma^2}+\frac{1}{v}\right)+\mu\left(\frac{k\bar{y}}{\sigma^2}+\frac{m}{v}\right)-\frac{1}{2}\left(\frac{k\bar{y}^2}{\sigma^2}+\frac{m^2}{v}\right)\right)$$

Si on prend : $\begin{cases} \dfrac{1}{A} = \dfrac{\sigma^2+vk}{\sigma^2+v} \\ B = \dfrac{k\bar{y}}{\sigma^2}+\dfrac{m}{v} \\ C = \dfrac{k\bar{y}^2}{\sigma^2}+\dfrac{m^2}{v} \end{cases}$

Alors nous obtenons une loi a posteriori telle que (le détail des calculs se trouve dans l'annexe 11.2.2) :

$$\pi_{\mathbf{y}}(\mu) \propto \exp\left[-\frac{1}{2A}\left(\mu - BA\right)^2\right]$$

Nous remarquons encore une fois le noyau d'une loi Normale. La loi conjuguée est donc une loi Normale car elle permet de conserver la même famille de loi. Nous avons donc choisi d'utiliser une loi a priori de la forme :

$$\mu_{\text{ a priori}} \sim \text{Normale}(m_{\text{ a priori}}, v_{\text{ a priori}}) \tag{7.1}$$

On rappelle que les paramètres de la loi de μ sont appelés les hyperparamètres, ici $m_{\text{ a priori}}$ et $v_{\text{ a priori}}$ sont les hyperparamètres a priori.

Estimation des hyperparamètres a priori

Nous connaissons maintenant la forme de la loi a priori. Cependant, il nous faut déterminer les hyperparamètres a priori $m_{\text{ a priori}}$ et $v_{\text{ a priori}}$ afin de pouvoir les crédibiliser. De manière générale, les hyperparamètres sont souvent pré-déterminés et sont ensuite crédibilisés à partir d'un échantillon. Dans cette étude, les hyperparamètres doivent être déterminés à partir d'un échantillon $\mathbf{x} = (x_1, \cdots, x_n)$ des variables $X_1, ..., X_n$. Il s'agit ici de l'étape n°1 du schéma 7.2.1.

Pour cela, nous allons utiliser la méthode séquentielle, présentée dans la thèse [11], qui consiste à supposer que les données de l'échantillon sont des données a posteriori par rapport à $\mu_{\text{ a priori}}$. Nous supposons alors que $\mu_{\text{ a priori}}$ suit la loi définie telle que :

$$f(\mathbf{x}|\mu_{\text{ a priori}}) \propto \exp\left(-\frac{(\mu_{\text{ a priori}} - \bar{x})^2}{2 \cdot \sigma^2/n}\right)$$

Nous pouvons reconnaître là le noyau d'une loi Normale. Comme nous savons que le paramètre $\mu_{\text{ a priori}}$ suit une loi Normale de paramètres $m_{\text{ a priori}}$ et $v_{\text{ a priori}}$, alors nous pouvons en déduire les estimateurs suivants :

$$\begin{cases} \hat{m}_{\text{ a priori}} = \bar{x} = \mathbb{E}(X) \\ \\ \hat{v}_{\text{ a priori}} = \dfrac{\sigma^2}{n} \end{cases} \tag{7.2}$$

Modélisation a posteriori

Les hyperparamètres a posteriori sont ensuite calculés à partir des hyperparamètres a priori et du jeu de données $\mathbf{y}$ (étape n°2 du schéma 7.2.1). Pour cela nous allons considérer la loi a posteriori de l'espérance sachant l'échantillon $\mathbf{y}$.

$$\pi_{\mathbf{y}}(\mu) \propto f(\mathbf{y}|\mu) \cdot \pi(\mu)$$

Nous avons vu lors du calcul de la loi a priori que :

$$\pi_{\mathbf{y}}(\mu) \propto \exp\left(-\frac{\mu^2}{2}\left(\frac{1}{\hat{v}_{\text{ a priori}}} + \frac{k}{\sigma^2}\right) + \mu\left(\frac{\hat{m}_{\text{ a priori}}}{\hat{v}_{\text{ a priori}}} + \frac{\sum y_i}{\sigma^2}\right) - \frac{1}{2}\left(\frac{\hat{m}^2_{\text{ a priori}}}{\hat{v}_{\text{ a priori}}} + \frac{\sum y_i^2}{\sigma^2}\right)\right)$$

Par ailleurs, nous savons que par définition :

$$\pi_{\mathbf{y}}(\mu) \propto \exp\left(-\frac{(\mu - \hat{m}_{\text{ a posteriori}})^2}{2 \cdot \hat{v}_{\text{ a posteriori}}}\right)$$

Si nous développons, nous obtenons :

$$\pi_{\mathbf{y}}(\mu) \propto \exp\left(-\frac{\mu^2 - 2\mu \cdot \hat{m}_{\text{ a posteriori}} + \hat{m}^2_{\text{ a posteriori}}}{2 \cdot \hat{v}_{\text{ a posteriori}}}\right)$$

Nous pouvons égaliser les termes en μ^2, nous obtenons une première équation :

$$-\frac{\mu^2}{2}\left(\frac{1}{\hat{v}_{\text{ a priori}}} + \frac{k}{\sigma^2}\right) = -\frac{\mu^2}{2 \cdot \hat{v}_{\text{ a posteriori}}}$$

De la même manière, nous pouvons égaliser les termes en μ, pour avoir l'équation suivante :

$$\mu\left(\frac{\hat{m}_{\text{ a priori}}}{\hat{v}_{\text{ a priori}}} + \frac{\sum y_i}{\sigma^2}\right) = -\frac{\mu \cdot \hat{m}_{\text{ a posteriori}}}{\hat{v}_{\text{ a posteriori}}}$$

A partir des deux équations précédentes, nous avons l'expression des hyperparamètres a posteriori :

$$\begin{cases} \hat{v}_{\text{ a posteriori}} = \dfrac{\sigma^2 \cdot \hat{v}_{\text{ a priori}}}{\sigma^2 + k \cdot \hat{v}_{\text{ a priori}}} \\ \\ \hat{m}_{\text{ a posteriori}} = \hat{v}_{\text{ a posteriori}} \cdot \left(\dfrac{\hat{m}_{\text{ a priori}}}{\hat{v}_{\text{ a priori}}} + \dfrac{k \cdot \bar{y}}{\sigma^2}\right) \end{cases} \tag{7.3}$$

A partir des estimations, nous allons simuler la sinistralité crédibilisée par étapes. Nous rappelons ici que k est la taille de l'échantillon $\mathbf{y}$, et σ^2 correspond à la variance de l'exposition.

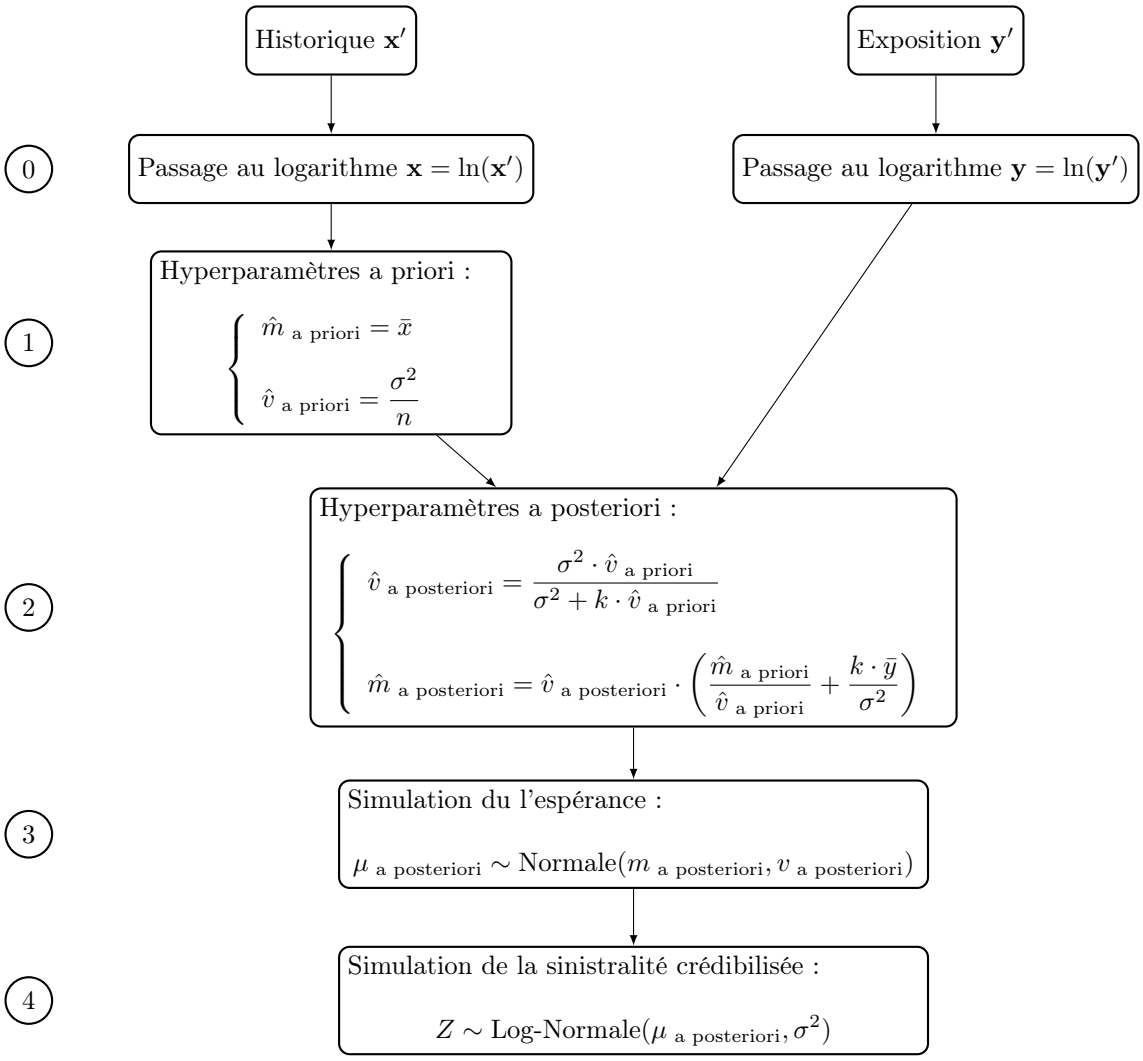

FIGURE 7.1 – Théorie bayésienne à partir d'une loi Normale-Normale

7.2.2 Paramètre de précision

Il est également possible de crédibiliser l'inverse de la variance que nous appellerons le paramètre de précision η :

$$\eta = \frac{1}{\sigma^2}$$

Soient $X_1, ..., X_n$ des variables indépendantes et identiquement distribuées telles que :

$$\ln(X_i) \sim \text{Normale}\left(\mu, \frac{1}{\eta_{\text{ a priori}}}\right) \qquad \forall i \in [1, ..., n]$$

Nous allons premièrement déterminer l'espérance μ et ensuite nous allons crédibiliser le paramètre de précision η. Pour cela nous allons procéder de la même manière que précédemment, nous allons déterminer la forme de la loi a priori, ensuite nous allons estimer les hyperparamètres a priori et, finalement, nous allons estimer les hyperparamètres a posteriori afin de modéliser la nouvelle variable :

$$Z \sim \text{Log-Normale}\left(\mu, \frac{1}{\eta_{\text{ a posteriori}}}\right)$$

Loi a priori

Pour déterminer la forme de la loi a priori $\pi(\eta)$, il est nécessaire d'établir la forme de la fonction de vraisemblance. Nous rappelons que k est la taille de l'échantillon $\mathbf{y}$ que nous souhaitons utiliser pour crédibiliser. La fonction de vraisemblance est la suivante :

$$f(\mathbf{y}|\eta) = \frac{1}{\sqrt{2\pi}^k} \cdot \eta^{k/2} \cdot \exp\left(-\frac{\eta}{2}\sum_{i=1}^{k}(y_i - \mu)^2\right)$$

Si nous raisonnons de manière proportionnelle nous avons :

$$f(\mathbf{y}|\eta) \propto \eta^{k/2} \cdot \exp\left(-\frac{\eta}{2}\sum_{i=1}^{k}(y_i - \mu)^2\right)$$

Dans ce cas, nous observons le noyau d'une loi Gamma, la fonction de vraisemblance a donc la forme d'une loi Gamma.

Afin de montrer que la loi conjuguée est une loi Gamma, nous allons montrer que lorsqu'on la loi a priori est une loi Gamma, la loi a posteriori l'est aussi. Prenons la loi a priori suivante :

$$\eta \sim \text{Gamma}(a, b)$$

Nous obtenons alors la loi a posteriori suivante :

$$\pi_{\mathbf{y}}(\eta) \propto f(\mathbf{y}|\eta) \cdot \pi(\eta)$$

En remplaçant les densités par leurs expressions, nous avons :

$$\begin{aligned}
\pi_{\mathbf{y}}(\eta) &\propto f(\mathbf{y}|\eta) \cdot \pi(\eta) \\
&\propto \eta^{k/2} \cdot \exp\left(-\frac{\eta}{2}\sum_{i=1}^{k}(y_i-\mu)^2\right) \cdot \eta^{a-1} \cdot \exp\left(-\eta \cdot b\right) \\
&\propto \eta^{a+k/2-1} \cdot \exp\left[-\eta\left(b+\frac{1}{2}\sum_{i=1}^{k}(y_i-\mu)^2\right)\right]
\end{aligned}$$

Nous pouvons reconnaître le noyau d'une loi Gamma, loi conjuguée est donc une loi Gamma. Nous choisissons alors une loi a priori de la forme suivante :

$$\eta_{\text{ a priori}} \sim \text{Gamma}(a_{\text{ a priori}}, b_{\text{ a priori}}) \tag{7.4}$$

Les hyperparamètres a priori sont ici $a_{\text{ a priori}}$ et $b_{\text{ a priori}}$.

Estimation des hyperparamètres a priori

Nous avons vu que la loi conjuguée à la loi Normale pour le paramètre de précision est une loi Gamma. Nous souhaitons maintenant estimer les hyperparamètres correspondants a priori $a_{\text{ a priori}}$ et $b_{\text{ a priori}}$: étape n°1 du schéma 7.2.2. Dans le but d'estimer les hyperparamètres à partir de l'échantillon $\mathbf{x}$, nous allons utiliser la méthode séquentielle. Nous supposons que l'échantillon $\mathbf{x}$ est un échantillon des variables $(X|\eta_{\text{ a priori}})$. On rappelle que n est la longueur de l'échantillon $\mathbf{x}$, alors nous avons :

$$f(\mathbf{x}|\eta_{\text{ a priori}}) \propto \eta_{\text{ a priori}}^{n/2} \cdot \exp\left(-\frac{\eta_{\text{ a priori}}}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

Nous pouvons reconnaître le noyau d'une loi Gamma. Il est ainsi possible de déterminer les estimateurs des hyperparamètres a priori :

$$\begin{cases}
\hat{a}_{\text{ a priori}} = \dfrac{n}{2} - 1 \\
\hat{b}_{\text{ a priori}} = \dfrac{1}{2}\displaystyle\sum_{i=1}^{n}(x_i-\mu)^2 = \mathbb{V}(X)
\end{cases} \tag{7.5}$$

Modélisation a posteriori

Nous allons maintenant nous intéresser à l'estimation des hyperparamètres a posteriori : étape n°2 du schéma 7.2.2. Pour cela nous nous intéressons à la loi a posteriori du paramètre de précision sachant $\mathbf{y}$. Nous avons vu lors du calcul de la loi conjuguée que la loi a posteriori peut s'écrire de la manière suivante :

$$\pi_{\mathbf{y}}(\eta) \propto \eta^{\hat{a}_{\text{ a priori}}+k/2-1} \cdot \exp\left[-\eta\left(\hat{b}_{\text{ a priori}} + \frac{1}{2}\sum_{i=1}^{k}(y_i-\mu)^2\right)\right]$$

Par ailleurs, nous avons choisi une loi a priori telle que la loi a posteriori soit une loi Gamma de la forme suivante :

$$\pi_{\mathbf{y}}(\eta) \propto \eta^{\hat{a}_{\text{ a posteriori}}-1} \cdot \exp\left(-\eta \cdot \hat{b}_{\text{ a posteriori}}\right)$$

A partir des deux expressions précédentes nous obtenons les estimateurs des hyperparamètres a posteriori :

$$\begin{cases} \hat{a}_{\text{ a posteriori}} = \hat{a}_{\text{ a priori}} + \dfrac{k}{2} \\ \\ \hat{b}_{\text{ a posteriori}} = \hat{b}_{\text{ a priori}} + \dfrac{1}{2}\displaystyle\sum_{i=1}^{k}(y_i-\mu)^2 \end{cases} \tag{7.6}$$

A partir des estimations, nous allons simuler la sinistralité crédibilisée par étapes. Nous rappelons que k est la taille de l'échantillon $\mathbf{y}$, et σ^2 correspond à la variance de l'exposition.

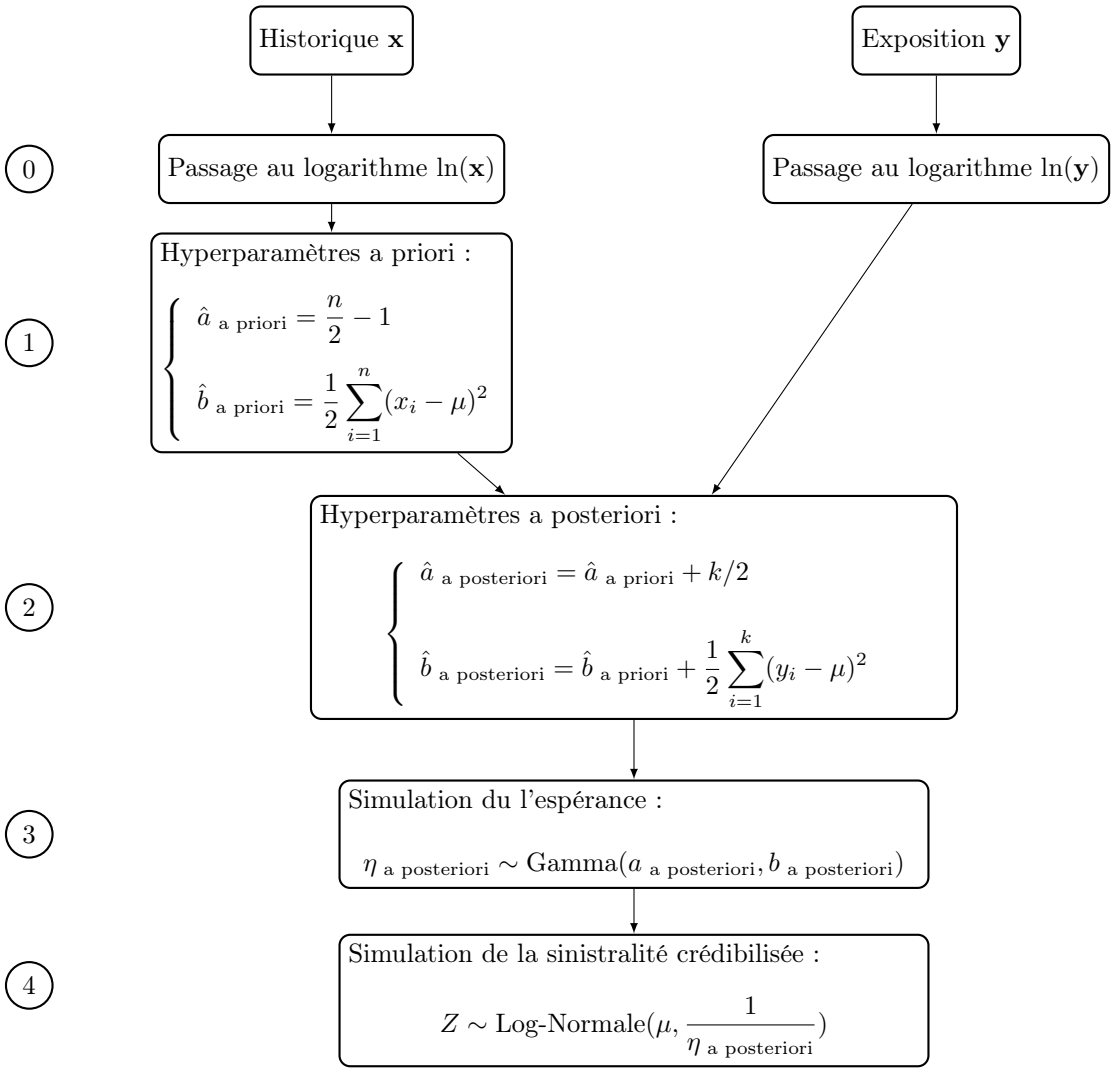

FIGURE 7.2 – Théorie bayésienne à partir d'une loi Normale-Gamma

7.3 Loi Gamma

Il est également possible de crédibiliser les paramètres de la loi Gamma. Nous allons présenter ici la crédibilisation du paramètre d'intensité de la loi Gamma. Soient $X_1, ..., X_n$ des variables aléatoires indépendantes et identiquement distribuées selon :

$$X_i \sim \text{Gamma}(\alpha, \beta_{\text{ a priori}}) \qquad \forall i \in [1, ..., n]$$

Nous allons procéder de la même manière que pour la loi Log-Normale. Nous déterminerons la forme de la famille de la loi a priori $\pi(\beta)$ et ensuite nous estimerons les hyperparamètres a priori de cette loi. Enfin nous estimerons les hyperparamètres a posteriori pour modéliser la variable crédibilisée :

$$Z \sim \text{Gamma}(\alpha, \beta_{\text{ a posteriori}})$$

7.3.1 Loi a priori

Nous allons déterminer la forme de la fonction de vraisemblance qui nous permettra ensuite de déterminer la forme de la loi a priori $\pi(\beta)$. On rappelle que k est la taille de l'échantillon $\mathbf{y}$ qui nous permettra de crédibiliser la variable. Alors la fonction de vraisemblance est telle que :

$$\begin{aligned} f(\mathbf{y}|\beta) &= \prod_{i=1}^{k} \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot y_i^{\alpha-1} \cdot \exp(-\beta y_i) \\ &= \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^k \left(\prod_{i=1}^{k} y_i\right)^{\alpha-1} \exp\left(-\beta \sum_{i=1}^{k} y_i\right) \end{aligned}$$

Si nous raisonnons de manière proportionnelle, nous obtenons :

$$f(\mathbf{y}|\beta) \propto \beta^{\alpha k} \exp\left(-\beta \sum_{i=1}^{k} y_i\right)$$

Il s'agit du noyau d'une loi Gamma, la fonction de vraisemblance est donc une loi Gamma.

Nous allons maintenant démontrer que la loi conjuguée est également une loi Gamma. Pour cela nous allons montrer que lorsque la loi a priori est une loi Gamma alors la loi a posteriori est aussi

une loi Gamma. Prenons la loi a priori suivante :

$$\beta \sim \text{Gamma}(a, b)$$

Alors nous obtenons la loi a posteriori suivante :

$$\begin{aligned}
\pi_{\mathbf{y}}(\beta) &\propto f(\mathbf{y}|\beta) \cdot \pi(\beta) \\
&\propto \beta^{\alpha k} \exp\left(-\beta \sum_{i=1}^{k} y_i\right) \cdot \beta^{a-1} \exp\left(-\beta \cdot b\right) \\
&\propto \beta^{a+\alpha k-1} \exp\left[-\left(b + \sum_{i=1}^{k} y_i\right) \beta\right]
\end{aligned}$$

Nous remarquons encore une fois qu'il s'agit du noyau d'une loi Gamma, la loi conjuguée est donc une loi Gamma. Nous choisissons donc d'utiliser la loi a priori définie par :

$$\beta_{\text{ a priori}} \sim \text{Gamma}(a_{\text{ a priori}}, b_{\text{ a posteriori}}) \tag{7.7}$$

Les hyperparamètres sont ici les paramètres $a_{\text{ a priori}}$ et $b_{\text{ a priori}}$.

7.3.2 Estimation des hyperparamètres a priori

Comme nous l'avons démontré auparavant, la loi conjuguée à la loi Gamma par rapport au paramètre d'intensité est aussi une loi Gamma. Nous obtenons donc une loi que nous appelons Gamma-Gamma. Dans cette partie nous allons estimer les hyperparamètres a priori ($a_{\text{ a priori}}$ et $b_{\text{ a priori}}$). Ces derniers seront estimés avec la méthode séquentielle et à partir de l'échantillon $\mathbf{x}$. Nous rappelons que n est la taille de l'échantillon. En supposant que $\mathbf{x}$ est un échantillon de la variable $(X|\beta_{\text{ a priori}})$, nous avons :

$$f(\mathbf{x}|\beta_{\text{ a priori}}) \propto \beta_{\text{ a priori}}^{\ \alpha n} \exp\left(-\beta_{\text{ a priori}} \sum_{i=1}^{n} x_i\right)$$

Nous savons qu'il s'agit ici d'une loi Gamma, il est donc possible d'en déduire les hyperparamètres

a priori (première étape du schéma 7.3.3) :

$$\begin{cases} \hat{a}_{\text{ a priori}} = \alpha \cdot n + 1 \\ \hat{b}_{\text{ a priori}} = \displaystyle\sum_{i=1}^{n} x_i = n \cdot \mathbb{E}(X) \end{cases} \tag{7.8}$$

7.3.3 Modélisation a posteriori

A partir des hyperparamètres a priori et de l'échantillon $\mathbf{y}$, nous pouvons estimer les hyperparamètres a posteriori (deuxième étape du schéma 7.3.3). Nous avons vu lors du calcul de la loi conjuguée que la loi a posteriori peut s'écrire de la manière suivante :

$$\pi_{\mathbf{y}}(\beta) \propto \beta^{\hat{a}_{\text{ a priori}} + \alpha k - 1} \exp\left[- \left(\hat{b}_{\text{ a priori}} + \sum_{i=1}^{k} y_i \right) \beta \right]$$

Comme nous savons que la loi a posteriori est une loi Gamma, alors nous pouvons en déduire les estimateurs des hyperparamètres a posteriori :

$$\begin{cases} \hat{a}_{\text{ a posteriori}} = \hat{a}_{\text{ a priori}} + \alpha k \\ \hat{b}_{\text{ a posteriori}} = \hat{b}_{\text{ a priori}} + \displaystyle\sum_{i=1}^{k} y_i = \hat{b}_{\text{ a priori}} + k \cdot \mathbb{E}(Y) \end{cases} \tag{7.9}$$

Le schéma suivant montre les différentes étapes effectuées pour modéliser la sinistralité a posteriori. Nous rappelons ici que k est la taille de l'échantillon $\mathbf{y}$, et σ^2 correspond à la variance de l'exposition.

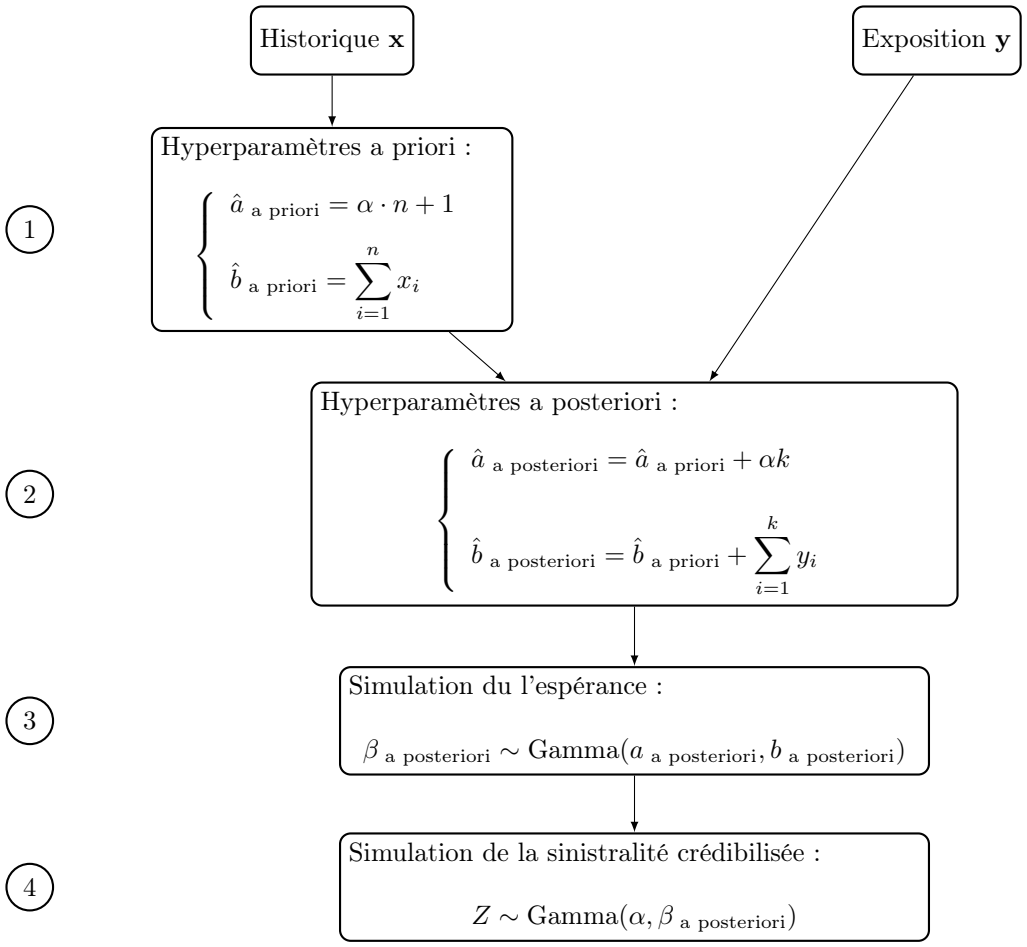

FIGURE 7.3 – Théorie bayésienne à partir d'une loi Gamma-Gamma

Chapitre 8

Crédibilisation au niveau de la charge du marché de la CCR

Cette section présente les résultats de la crédibilisation du coût brut agrégé de toutes les cédantes de la CCR pour différents périls. Nous rappelons ici que le montant agrégé est appelé le coût marché CCR.

$$\text{Coût marché CCR} = \sum_{i \in \{\text{Cédantes CCR}\}} \text{Coût brut cédante i}$$

Nous allons nous intéresser séparément à la sinistralité associée aux inondations, aux sécheresses et aux contrats d'assurance automobile.

Nous allons premièrement présenter quelques résultats concernant l'ajustement des lois utilisées aux jeux de données présentés dans la section 6. Dans un deuxième temps, nous allons présenter une approche bayésienne classique à partir des résultats de la section 7. Ensuite, nous allons présenter une deuxième approche développée dans le but d'obtenir un modèle hybride plus adapté aux besoins de la CCR. Enfin nous comparerons les résultats obtenus via ces deux méthodes.

8.1 Approche bayésienne

La théorie bayésienne a été présentée dans la partie 7. Dans cette partie, nous allons crédibiliser le coût marché CCR à partir de deux lois : la loi Log-Normale et la loi Gamma. En effet, ces lois présentent des avantages divers :

- nous avons vu précédemment, que ces lois sont actuellement utilisées pour tarifer des contrats par expérience à la CCR. Il est donc intéressant de garder les mêmes lois d'un point de vue analytique et pratique ;

- comme il s'agit de la perte agrégée annuelle et pas du coût d'un événement, ces lois semblent plus adaptées que les lois des extrêmes ;

- ces lois sont pratiques du point de vue des calculs au niveau de la théorie de la crédibilité.

Le tableau 8.1 montre les différences entre l'exposition et l'historique pour les différents périls étudiés. On remarque un écart entre les moyennes mais également un écart important au niveau des écart-types. Cet écart entre les moyennes et les écart-types des modèles est dû à la prise en compte d'événements extrêmes par le modèle d'exposition. La période de retour de tels événements étant très élevée, ces derniers sont rarement pris en compte dans l'historique. Le coût important de tels événements a un impact considérable sur la moyenne et l'écart-type.

	Historique		Exposition	
Données	Espérance (M€)	Écart-type (M€)	Espérance (M€)	Écart-type (M€)
Inondation	534	400	574	802
Sécheresse	360	379	441	684
Automobile	49	123	67	166

TABLE 8.1 – Statistiques des données

La crédiblisation va permettre de prendre en compte les deux jeux de données dans le but de trouver un équilibre entre les statistiques. L'objectif est de modéliser le coût marché CCR a posteriori par une loi Log-Normale ou par une loi Gamma dont les paramètres contiennent l'information provenant des deux sources. Tout d'abord, nous allons vérifier le bon ajustement de la loi Log-Normale et de la loi Gamma aux données. Nous présenterons, ensuite, les résultats associés à la crédibilisation par rapport aux différents paramètres des lois.

8.1.1 Ajustement des lois de probabilités

Afin de vérifier l'ajustement des lois de probabilité sur nos données nous allons effectuer des tests statistiques et des tests graphiques. Premièrement, au niveau des tests graphiques nous allons tracer les *Quantile-Quantile Plot* à l'aide de la fonction `qqplot` du package `stats` sur le logiciel `R`. Ensuite nous allons tester les hypothèses :

(H_0) : L'échantillon suit une loi Log-Normale

(H_1) : L'échantillon ne suit pas une loi Log-Normale

Nous testerons également les hypothèses :

(H_0) : L'échantillon suit une loi Gamma

(H_1) : L'échantillon ne suit pas une loi Gamma

Pour tester ces hypothèses, nous allons faire appel à la fonction `ks.test` du package `stats` pour le test de Kolmogorov-Smirnov et à la fonction `cvm.test` du package `goftest` pour le test de Cramér-von Mises. Une brève explication de ces tests est présentée dans l'annexe 11.1.

Inondation

Dans le cas de l'inondation, nous remarquons sur les graphiques de la figure 8.1, que l'historique s'ajuste bien aux lois Log-Normale et Gamma d'après les QQplots.

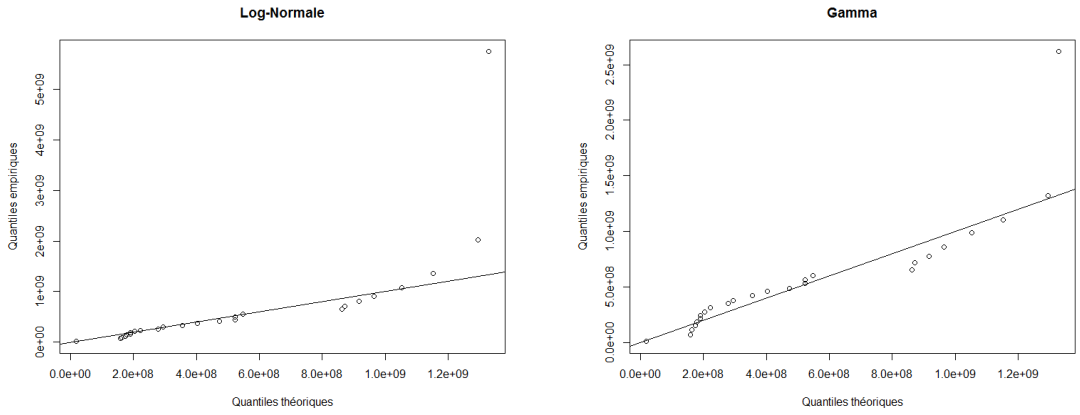

FIGURE 8.1 – Ajustement des lois à l'historique des inondations

Si nous effectuons les tests statistiques précédents sur l'historique des inondations nous obtenons des résultats favorables. En effet, pour les deux lois nous obtenons des écarts faibles par rapport à la distribution théorique et la probabilité d'obtenir de tels écarts est élevée. Ainsi, d'après les résultats présentés sur le tableau 8.2, les tests ne permettent pas de rejeter les hypothèses (H_0).

	Kolmogorov-Smirnov		Cramér-von Mises	
Loi	Statistique	p-value	Statistique	p-value
Log-Normale	0,14	0,69	0,07	0,76
Gamma	0,15	0,61	0,10	0,59

TABLE 8.2 – Résultats des tests sur l'historique des inondations

Nous nous intéressons ensuite au modèle par l'exposition associée au péril inondation. De la même manière, l'exposition semble bien s'ajuster aux lois d'après la figure 8.2, cependant les tests statistiques ne permettent pas de confirmer ses hypothèses. En absence d'un meilleur ajustement, nous avons décidé de crédibiliser avec ces deux lois afin de comparer les résultats avec la distribution historique utilisée pour la tarification. En effet, nous avons vu précédemment que ces lois sont actuellement utilisées pour tarifer les traités des catastrophes naturelles de la CCR.

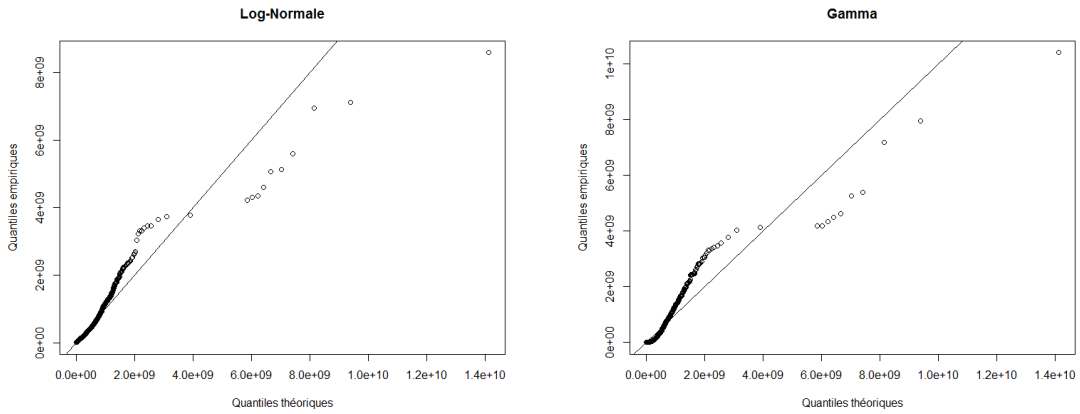

FIGURE 8.2 – Ajustement des lois à l'exposition des inondations

Sécheresse

En ce qui concerne la sécheresse, la figure 8.3 permet de déduire un bon ajustement des deux lois sur l'historique des données.

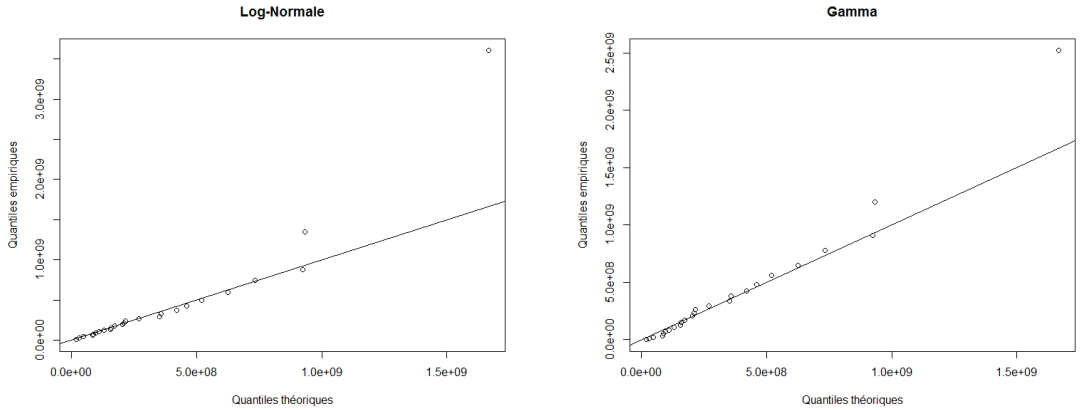

FIGURE 8.3 – Ajustement des lois à l'historique des sécheresses

Les résultats des tests statistiques que nous avons effectués sont présentés sur le tableau 8.3. Nous obtenons un meilleur ajustement que pour les inondations, les écarts sont plus faibles et les probabilités sont plus élevées.

	Kolmogorov-Smirnov		Cramér-von Mises	
Loi	Statistique	p-value	Statistique	p-value
Log-Normale	0,07	1	0,02	1
Gamma	0,11	0,86	0,04	0,95

TABLE 8.3 – Résultats des tests sur l'historique des sécheresses

Cependant au niveau de l'ajustement des lois sur l'exposition associée aux sécheresses, nous pouvons remarquer que la loi Gamma semble s'ajuster aux données d'après le QQplot, or la loi Log-Normale ne semble pas s'ajuster aux données. De la même manière que pour l'exposition associée aux inondations, les tests statistiques ont rejeté les deux lois. Nous avons donc décidé de baser notre choix sur les tests graphiques et de conserver la loi gamma.

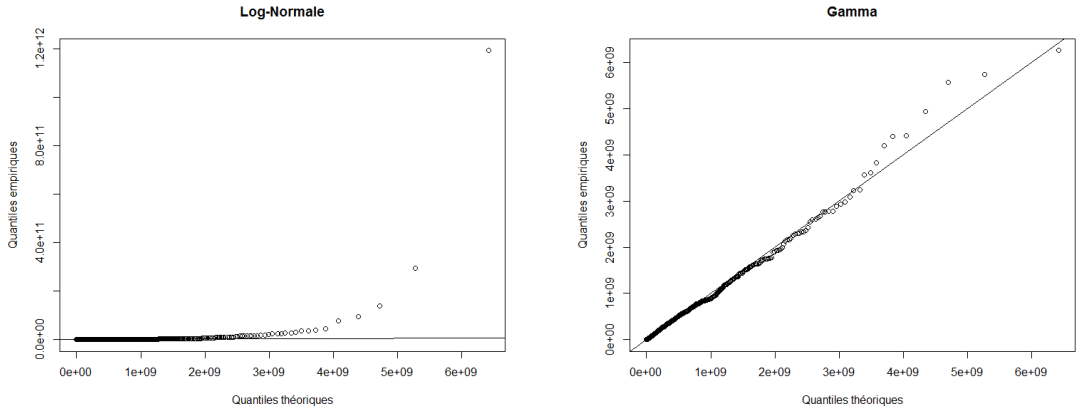

FIGURE 8.4 – Ajustement des lois à l'exposition des sécheresses

Automobile

Dans le cadre des contrats d'assurance automobile, la figure 8.5 permet de déduire un bon ajustement de la loi Log-Normale d'après les QQplots. Néanmoins, la loi Gamma ne semble pas s'ajuster correctement aux données. Les tests statistiques permettront encore une fois de trancher.

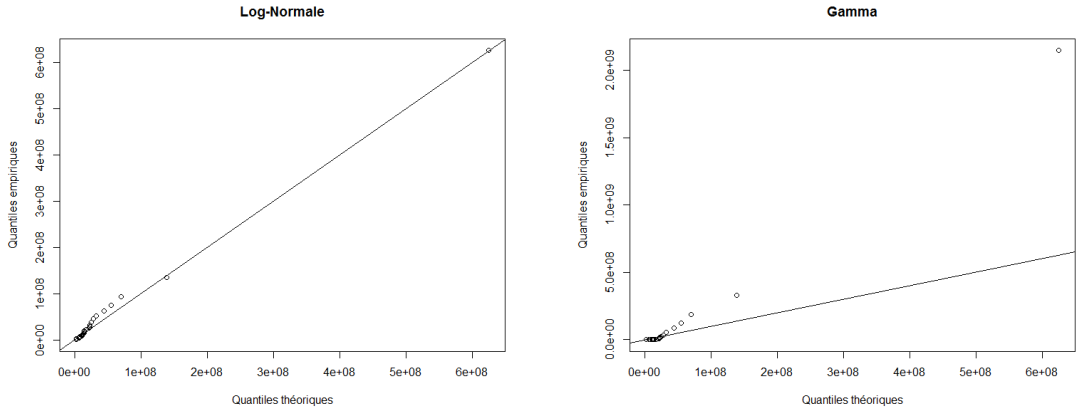

FIGURE 8.5 – Ajustement des lois à l'historique des contrats automobile

Comme nous aurions pu constater à partir des QQplots, les tests statistiques présentés sur le tableau 8.4, permettent de confirmer le bon ajustement de la loi Log-Normale et de rejeter l'hypothèse

d'une loi Gamma.

	Kolmogorov-Smirnov		Cramér-von Mises	
Loi	Statistique	p-value	Statistique	p-value
Log-Normale	0,13	0,76	0,09	0,66
Gamma	0,51	1,7e-06	1,78	2,2e-05

TABLE 8.4 – Résultats des tests sur l'historique des contrats automobile

En ce qui concerne l'exposition, d'après les QQplots présentés sur la figure 8.6, les deux lois semblent s'ajuster aux données et d'après tests statistiques les deux lois sont rejetées. Le choix a été fait de modéliser la sinistralité automobile à partir d'une loi log-normale qui n'est pas rejetée par les tests effectués sur l'historique.

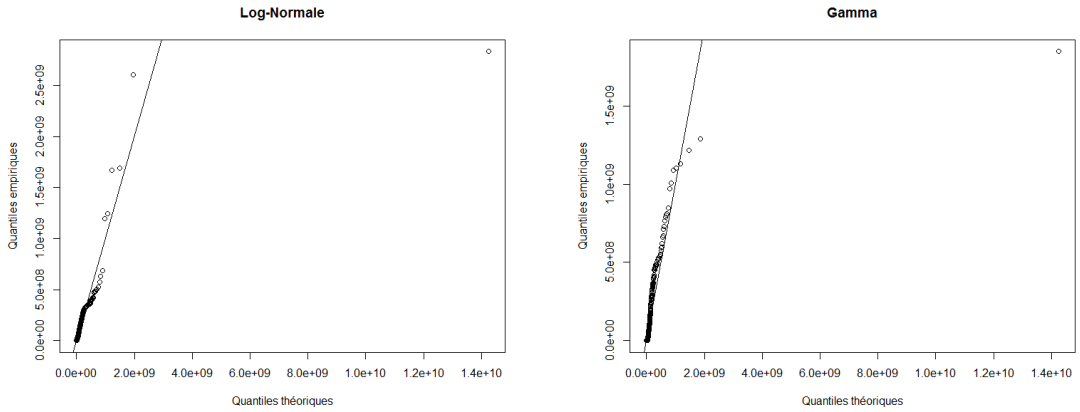

FIGURE 8.6 – Ajustement des lois à l'exposition des contrats automobile

8.1.2 Application de la théorie bayésienne

Dans cette partie nous rappellerons brièvement les différentes étapes de la crédibilisation ainsi que les formules que nous allons utiliser par la suite. Comme nous l'avons vu précédemment, nous allons crédibiliser les paramètres à partir des trois méthodes suivantes :

- crédibilisation de l'espérance de la loi Log-Normale en supposant que les données suivent une loi Log-Normale et que la variance est connue ;

- crédibilisation du paramètre de précision de la loi Log-Normale en supposant que les données suivent une loi Log-Normale et que l'espérance est connue ;
- crédibilisation du paramètre d'intensité de la loi Gamma en supposant que les données suivent une loi Gamma et que le paramètre de forme est connu.

Nous rappelons par ailleurs que nous partons de la loi a priori qui correspond à l'historique pour ensuite incorporer l'information provenant de l'exposition.

Loi a priori

Il est nécessaire de connaître la loi a priori de chaque paramètre crédibilisé. Le tableau 8.5 rappelle les résultats dans la partie 7 concernant la loi a priori de chaque paramètre crédibilisé.

Loi des X_i	Paramètre	Loi du paramètre
Log-Normale(μ, σ^2)	μ a priori	Normale$(m$ a priori$, v$ a priori$)$
Log-Normale $\left(\mu, \frac{1}{\eta}\right)$	η a priori	Gamma$(a$ a priori$, b$ a priori$)$
Gamma(α, β)	β a priori	Gamma$(a$ a priori$, b$ a posteriori$)$

TABLE 8.5 – Loi a priori de chaque paramètre crédibilisé

Estimation des hyperparamètres a priori

Dans un premier temps, nous allons estimer les hyperparamètres a priori à partir de l'historique, à l'aide des estimateurs définis dans le chapitre 7 qui sont rappelés dans le tableau 8.6.

Loi des X_i	Paramètre	Estimateurs des hyper-paramètres	
Log-Normale(μ, σ^2)	μ a priori	$\hat{m}$ a priori $= \mathbb{E}(X)$	$\hat{v}$ a priori $= \frac{\sigma^2}{n}$
Log-Normale $\left(\mu, \frac{1}{\eta}\right)$	η a priori	$\hat{a}$ a priori $= \frac{n}{2} - 1$	$\hat{b}$ a priori $= \mathbb{V}(X)$
Gamma(α, β)	β a priori	$\hat{a}$ a priori $= \alpha \cdot n + 1$	$\hat{b}$ a priori $= n \cdot \mathbb{E}(X)$

TABLE 8.6 – Estimateurs des hyper-paramètres a priori

Modélisation a posteriori

Nous allons ensuite estimer les hyperparamètres a posteriori à partir des formules suivantes :

Loi des X_i	Paramètre	Estimateurs des hyper-paramètres a posteriori
Log-Normale(μ, σ^2)	μ a priori	$\hat{m}_{\text{ a posteriori}} = \hat{v}_{\text{ a posteriori}} \cdot \left(\frac{\hat{m}_{\text{ a priori}}}{\hat{v}_{\text{ a priori}}} + \frac{k \cdot \bar{y}}{\sigma^2} \right)$ $\hat{v}_{\text{ a posteriori}} = \frac{\sigma^2 \cdot \hat{v}_{\text{ a priori}}}{\sigma^2 + k \cdot \hat{v}_{\text{ a priori}}}$
Log-Normale $\left(\mu, \frac{1}{\eta}\right)$	η a priori	$\hat{a}_{\text{ a posteriori}} = \hat{a}_{\text{ a priori}} + k/2$ $\hat{b}_{\text{ a posteriori}} = \hat{b}_{\text{ a priori}} + \frac{1}{2} \sum_{i=1}^{k} (y_i - \mu)^2$
Gamma(α, β)	β a priori	$\hat{a}_{\text{ a posteriori}} = \hat{a}_{\text{ a priori}} + \alpha k$ $\hat{b}_{\text{ a posteriori}} = \hat{b}_{\text{ a priori}} + k \cdot \mathbb{E}(Y)$

TABLE 8.7 – Estimateurs des hyper-paramètres a posteriori

Il sera ensuite possible de simuler le paramètre crédibilisé a posteriori ainsi que la sinistralité crédibilisée à partir des lois qui figurent sur le tableau 8.8. Pour cela nous simulerons 1 000 fois le paramètre a posteriori et pour chaque paramètre crédibilisé nous simulerons 1 000 fois la sinistralité crédibilisée.

Loi des Z_i	Paramètre	Loi du paramètre
Log-Normale(μ, σ^2)	μ a posteriori	Normale$(m_{\text{ a posteriori}}, v_{\text{ a posteriori}})$
Log-Normale $\left(\mu, \frac{1}{\eta}\right)$	η a posteriori	Gamma$(a_{\text{ a posteriori}}, b_{\text{ a posteriori}})$
Gamma(α, β)	β a posteriori	Gamma$(a_{\text{ a posteriori}}, b_{\text{ a posteriori}})$

TABLE 8.8 – Loi a posteriori de chaque paramètre crédibilisé

8.1.3 Crédibilisation du péril inondation

Dans cette partie, nous allons présenter la crédibilisation de la sinistralité liée aux inondations sur le territoire français. Comme il a été dit précédemment, nous allons procéder par étapes. Tout d'abord nous estimer les hyperparamètres a priori. Ensuite nous allons estimer les hyperparamètres a posteriori. Enfin, nous simulerons le paramètres crédibilisé a posteriori afin de simuler ensuite la sinistralité liée au péril inondation.

Estimation des hyperparamètres a priori

Les résultats des estimations de ces hyperparamètres figurent dans le tableau 8.9.

	LogNormale-Normale		LogNormale-Gamma		Gamma-Gamma	
Hyperparamètre	m	v	a	b	a	b
Estimation a priori	19,8	0,037	13,5	23,2	45,5	1,3 e+10

TABLE 8.9 – Hyperparamètres a priori du péril inondation

Modélisation a posteriori

Après crédibilisation des hyperparamètres a posteriori, nous obtenons les résultats figurant dans le tableau 8.10.

	LogNormale-Normale		LogNormale-Gamma		Gamma-Gamma	
Hyperparamètre	m	v	a	b	a	b
Estimation a posteriori	19,7	1,8 e-05	25006	30730	89103	2,9 e+13

TABLE 8.10 – Hyperparamètres a posteriori du péril inondation

Nous remarquons des différences importantes, or ces écarts peuvent difficilement être analysés puisque ce sont de hyperparamètres. Nous avons donc simulé les paramètres à partir des lois a priori et a posteriori afin de pouvoir comparer les écarts. Nous avons les représentations graphiques des simulations des paramètres a priori et a posteriori sur les graphiques de la figure 8.7. Ces trois graphiques permettent de comparer le paramètre sans aléa, c'est-à-dire la moyenne et le paramètre de précision pour la loi Normale et le paramètre d'intensité pour la loi Gamma, aux paramètres simulés à partir des lois a priori et a posteriori. Nous remarquons que le paramètre simulé a priori varie autour du paramètre sans aléa. Ceci permet de confirmer que les estimateurs et donc les hypothèses associées à ces estimateurs sont cohérents.

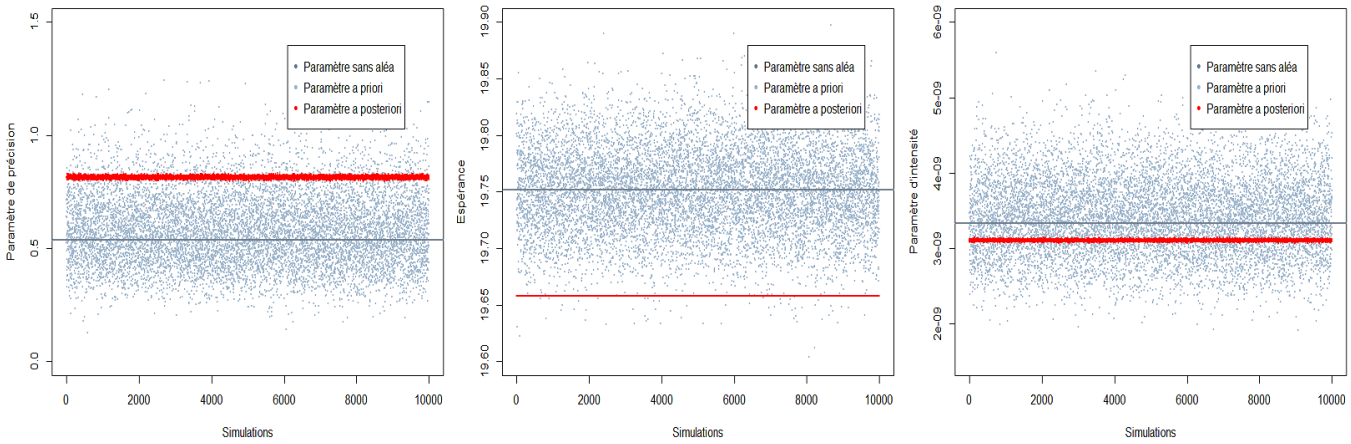

FIGURE 8.7 – Paramètres a priori et a posteriori du péril inondation

Nous remarquons ici que le paramètre de précision a posteriori est plus élevé et donc par conséquent la variance a posteriori devrait être moins élevée que celle a priori. De la même manière, l'espérance et l'intensité a posteriori sont moins élevées que celles a priori. Pour analyser plus finement ces résultats nous allons nous intéresser à la sinistralité modélisée.

Sinistralité crédibilisée

Pour chaque paramètre ayant été simulé à partir des lois a priori et a posteriori nous avons simulé la sinistralité. Nous obtenons ainsi la sinistralité a priori et a posteriori. Quelques statistiques figurent sur le tableau 8.11. Nous remarquons que l'espérance de la sinistralité dépend de la méthode choisie, ainsi une loi LogNormale-Normale donne un résultat proche de l'historique alors qu'une loi Gamma-Gamma donne un résultat très proche de l'exposition. La loi LogNormale-Gamma quant-à-elle, donne une espérance plus faible que celle de l'historique.

	Historique	Exposition	LogNormale-Normale	LogNormale-Gamma	Gamma-Gamma
Espérance (M€)	534	574	531	457	573
Écart-type (M€)	400	802	634	311	430

TABLE 8.11 – Statistiques de la sinistralité crédibilisée pour le péril inondation

En ce qui concerne l'écart-type, nous remarquons que la loi LogNormale-Normale permet d'obtenir

un écart-type qui correspond à un équilibre entre l'historique et l'exposition. Dans le cas de la loi Gamma-Gamma, l'écart-type de la loi a posteriori est très proche de l'historique. Enfin, avec la loi LogNormale-Gamma nous obtenons un écart-type moins élevé que l'historique et que l'exposition.

Il est à noter que, ces résultats ne correspondent pas aux résultats attendus. En effet, nous avons vu dans le tableau 6.1 que la moyenne et l'écart-type des données provenant du modèle d'exposition sont plus élevés. L'impact attendu de la prise en compte de ces données lors de la crédibilisation était une espérance et une variance plus élevées a posteriori.

Les statistiques précédentes ne permettent pas de conclure quant-à l'utilisation des différentes méthodes. Le but de cette étude était de trouver une distribution proche à l'historique pour les petites périodes de retour et au contraire proche de l'exposition pour les grandes périodes de retour. Pour cette raison, nous allons nous intéresser aux périodes de retour qui figurent sur le tableau 8.12.

Période de retour	Historique	Exposition	LogNormale-Normale	LogNormale-Gamma	Gamma-Gamma
2 ans	441	397	345	378	469
5 ans	817	803	754	635	870
10 ans	1 077	1 113	1 133	833	1 147
20 ans	1 328	1 448	1 588	1 041	1 411
50 ans	1 656	2 027	2 320	1 340	1 747
100 ans	1 883	3 927	2 985	1 583	2 004
200 ans	2 112	6 654	3 779	1 848	2 250
500 ans	2 395	8 145	4 989	2 237	2 582
1000 ans	2 607	9 394	6 146	2 555	2 816

TABLE 8.12 – Niveaux de retour de la sinistralité du péril inondation (M€)

Ici, les niveaux de retour de l'historique sont obtenus par la méthode de Monte-Carlo à partir d'une loi ajustée. Dans le cas du péril inondation, nous avons utilisé la loi Gamma.

D'une part, il est à noter que pour les périodes de retour de l'ordre d'une dizaine d'années, les lois LogNormale-Normale et Gamma-Gamma renvoient des niveaux de retour plus élevés que l'exposition. Dans le cas de la loi LogNormale-Gamma les niveaux de retour sont plus petits que ceux obtenus à partir de l'historique. Cependant, d'après l'expertise de la CCR il faudrait obtenir des niveaux proches des niveaux historiques. Nous pouvons remarquer que les résultats obtenus sont très sensibles au choix de la loi.

D'autre part, pour des périodes de retour supérieures à 100 ans, les distributions a posteriori sont toujours très proches de l'historique et ne permettent pas de donner plus d'importance aux niveaux obtenus à partir du modèle d'exposition.

8.1.4 Crédibilisation du péril sécheresse

De la même manière, nous allons crédibiliser la sinistralité liée aux sécheresses, mais cette fois-ci nous allons uniquement crédibiliser le paramètre d'intensité de la loi Gamma. Pour cela, nous allons procéder de la même manière que précédemment.

Estimation des hyperparamètres a priori

Les hyperparamètres a priori ont été estimés à partir des mêmes estimateurs que pour le péril inondations. Les résultats de ces estimations figurent dans le tableau 8.13.

	Gamma-Gamma	
Hyperparamètre	a	b
A priori	24	9 e+09

TABLE 8.13 – Hyperparamètres a priori du péril sécheresse

Modélisation a posteriori

Nous avons ensuite estimé les hyperparamètres a posteriori dont les résultats se trouvent dans le tableau 8.14.

	Gamma-Gamma	
Hyperparamètre	a	b
A posteriori	45181	2,2 e+13

TABLE 8.14 – Hyperparamètres a posteriori du péril sécheresse

Les simulations du paramètre d'intensité a priori et a posteriori sont représentées sur la figure 8.8. Nous remarquons ici que le paramètre d'intensité obtenu a posteriori est moins élevé que le paramètre a priori. Afin de mieux interpréter ces résultats, nous allons estimer quelques statistiques.

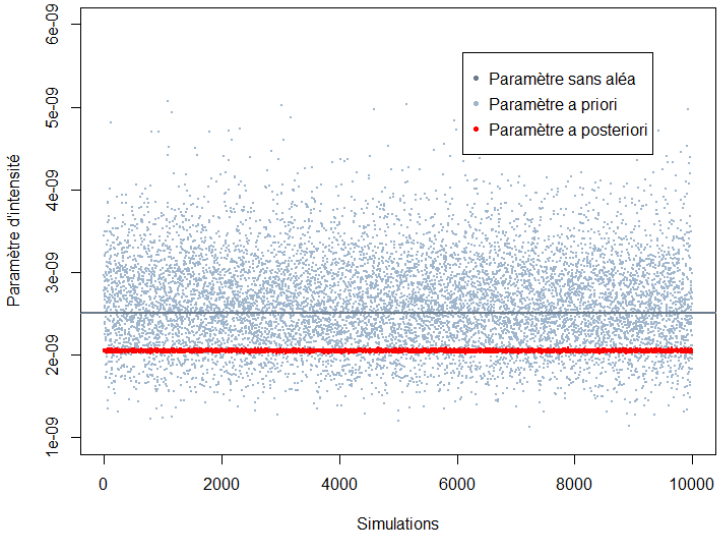

FIGURE 8.8 – Paramètres a priori et a posteriori du péril sécheresse

Sinistralité crédibilisée

Comme précédemment, nous avons simulé la sinistralité à partir de chaque paramètre simulé à partir de la loi a posteriori. Les résultats figurent sur le tableau 8.15. Nous remarquons que l'espérance de la distribution crédibilisée est plus proche de l'exposition que de l'historique. Toutefois, l'écart-type reste plus proche de l'historique.

	Historique	Exposition	Gamma-Gamma
Espérance (M€)	360	441	441
Écart-type (M€)	379	684	465

TABLE 8.15 – Statistiques de la sinistralité crédibilisée pour le péril sécheresse

Pour mieux analyser la distribution obtenue nous allons nous intéresser aux périodes de retour présentées sur le tableau 8.16.

Période de retour	Historique	Exposition	Gamma-Gamma
2 ans	237	171	293
5 ans	580	690	715
10 ans	847	1 211	1 043
20 ans	1 107	1 842	1 372
50 ans	1 449	2 705	1 803
100 ans	1 718	3 323	2 133
200 ans	2 001	3 838	2 468
500 ans	2 279	4 698	2 907
1000 ans	2 506	5 260	3 259

TABLE 8.16 – Niveaux de retour de la sinistralité pour le péril sécheresse (M€)

Dans le cas de la sécheresse, nous obtenons des résultats plus satisfaisants que pour l'inondation. En effet, pour des périodes de retour supérieures à 5 ans, les niveaux de retour correspondent à un équilibre entre l'historique et l'exposition. Néanmoins, pour les périodes de retour de l'ordre d'une centaine d'années, le poids accordé à l'exposition n'est pas suffisamment important puisque nous obtenons des niveaux de retour proches de l'historique. En outre, pour des périodes de retour de l'ordre de 2 ans, les niveaux obtenus sont plus sévères que l'historique et que l'exposition.

8.1.5 Crédibilisation de la sinistralité automobile

Dans le cas des contrats d'assurance automobile, nous avons crédibilisé la sinistralité à partir de deux méthodes : la crédibilisation de l'espérance et la crédibilisation du paramètre de précision de la loi Log-Normale.

Estimation des hyperparamètres a priori

Les hyperparamètres estimés a priori à partir des estimateurs définis précédemment figurent dans le tableau 8.17.

	LogNormale-Normale		LogNormale-Gamma	
Hyperparamètre	m	v	a	b
A priori	16,7	0,056	13	35

TABLE 8.17 – Hyperparamètres a priori de l'automobile

Modélisation a posteriori

De la même manière, les hyperparamètres estimés a posteriori figurent dans le tableau 8.18.

	LogNormale-Normale		LogNormale-Gamma	
Hyperparamètre	m	v	a	b
A posteriori	17,3	3,1 e-05	22642	47691

TABLE 8.18 – Hyperparamètres a posteriori de l'automobile

Nous avons simulé les paramètres crédibilisés comme précédemment. Nous obtenons les simulations qui figurent sur la figure 8.9.

Nous remarquons que lors de la crédibilisation de l'espérance à partir d'une loi LogNormale-Normale, nous obtenons une espérance a posteriori plus élevée que l'espérance a priori. Ce résultat est cohérent avec les données de base puisque la moyenne historique est plus faible que la moyenne d'exposition. L'introduction de données provenant de l'exposition lors de la crédibilisation a donc pour impact une augmentation de l'espérance.

Cependant, dans le cas de la loi LogNormale-Gamma, nous obtenons des résultats moins cohérents puisque le paramètre de précision a posteriori est plus élevé que le paramètre de précision a priori. Or, comme la variance est inversement proportionnelle au paramètre de précision, ceci veut dire que nous obtenons une variance moins élevée, alors que l'injection d'évènements extrêmes devrait avoir l'impact contraire.

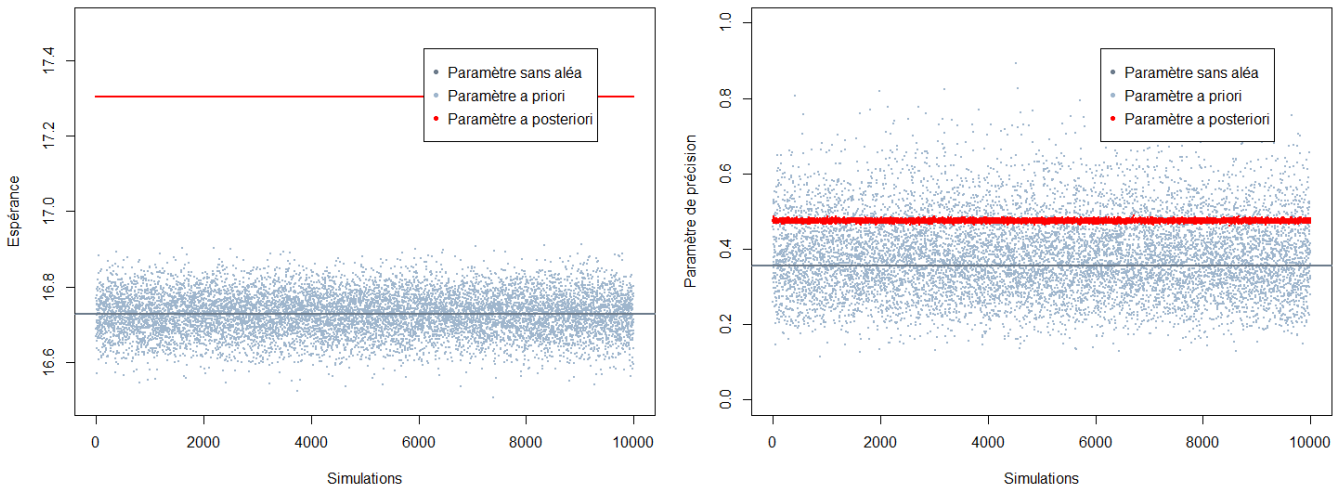

FIGURE 8.9 – Paramètres a priori et a posteriori de la sinistralité automobile

Sinistralité crédibilisée

De la même manière que pour les deux périls présentés précédemment, nous avons simulé la sinistralité à partir de chaque paramètre. Les résultats figurent sur le tableau 8.19. Ces résultats permettent de faire une analyse plus fine. A partir de la loi LogNormale-Normale, nous obtenons une espérance et une variance plus élevées que les statistiques de l'exposition et à partir de la loi LogNormale-Gamma nous obtenons des statistiques plus faibles que l'historique.

	Historique	Exposition	LogNormale-Normale	LogNormale-Gamma
Espérance (M€)	49	67	89	32
Écart-type (M€)	123	166	295	48

TABLE 8.19 – Statistiques de la sinistralité crédibilisée pour l'assurance automobile

Afin de mieux analyser ces écarts au niveau de l'espérance et de la variance, nous allons analyser les périodes de retour présentées sur le tableau 8.16.

Période de retour	Historique	Exposition	LogNormale-Normale	LogNormale-Gamma
2 ans	18	32	33	18
5 ans	60	81	107	45
10 ans	110	134	198	71
20 ans	185	216	329	104
50 ans	325	501	586	160
100 ans	480	684	856	214
200 ans	700	923	1 221	278
500 ans	1 076	1 459	1 862	382
1000 ans	1 523	1 874	2 544	480

TABLE 8.20 – Niveaux de retour de la sinistralité pour l'assurance automobile (M€)

Les niveaux de retour permettent de conclure quant à l'adéquation de cette distribution du point de vue des critères définis à partir de l'expertise de la CCR. Dans le cas de la distribution LogNormale-Normale, les niveaux de retour plus élevés que l'historique et l'exposition pour des petites périodes de retour confirment la mauvaise adéquation de cette distribution. Dans le cas de la distribution LogNormale-Gamma, des niveaux de retour inférieurs à l'historique et à l'exposition quelque soit la période de retour mettent également en évidence une mauvaise adéquation de cette méthode.

8.1.6 Conclusion sur l'approche bayésienne

L'application de la théorie bayésienne dans le but de créer une loi mélange prenant en compte l'historique et l'exposition des cédantes ne semble pas pertinente dans le cadre des catastrophes naturelles. En effet, cette section nous a permis de cerner les faiblesses de cette méthode :

- cette approche semble très sensible au choix de la loi : les résultats sont très variés selon si on choisit d'ajuster une loi Log-Normale ou une loi Gamma ;
- la méthode bayésienne est également très sensible au choix du paramètre que l'on souhaite crédibiliser : pour une même loi les résultats divergent en fonction du paramètre crédibilisé ;

- l'hypothèse selon laquelle l'historique et l'exposition suivent la même loi ne semble pas réaliste : dans aucun des cas présentés précédemment les tests n'ont permis de conclure de manière positive sur l'ajustement de la loi aux deux jeux de données ;
- les tests sont peu fiables sur un historique aussi faible, la queue de distribution de la loi ajustée n'est donc pas suffisamment fiable pour être prise en compte dans la loi mélange.

Pour cette raison, nous avons décidé de créer une distribution hybride plus pertinente dans le cadre de notre étude. Cette distribution sera présenté par la suite.

8.2 Distribution hybride

Le but du modèle hybride est de répondre aux critères fixés par l'expertise de la CCR. A cette fin, nous avons défini une distribution hybride à partir d'un mélange convexe des distributions historique et d'exposition dont les poids seront définis en fonction de la période de retour. Pour cela, nous avons défini des contraintes spécifiques à la CCR.

8.2.1 Présentation de l'approche

Le but de cette approche est de déterminer le quantile hybride à partir du quantile de la distribution de l'exposition et du quantile de la distribution de l'historique du coût marché CCR. Soient x_p^{exp} et x_p^{hist} les quantiles liés à l'exposition et à l'historique, respectivement. Ces derniers sont définis tels que :

$$\begin{cases} \mathbb{P}(X_{exp} \leqslant x_p^{exp}) = p \\ \\ \mathbb{P}(X_{hist} \leqslant x_p^{hist}) = p \end{cases}$$

Alors, on définit un quantile hybride x_p^{hyb} à partir d'un facteur de crédibilité α et des quantiles provenant de l'historique et de l'exposition par :

$$x_p^{hyb} = \alpha \cdot x_p^{exp} + (1 - \alpha) \cdot x_p^{hist} \tag{8.1}$$

Cependant, notre but est de choisir un facteur de crédibilité différent en fonction de la période de retour. De cette manière, il sera possible d'accorder plus de poids à l'historique pour les petites périodes de retour et à l'exposition pour les grandes périodes de retour. Nous allons donc définir une fonction α, que nous appellerons par la suite la fonction facteur hybride, telle que :

$$\alpha = \alpha(p) \qquad \text{où : } p = \mathbb{P}(X_{hyb} \leqslant x_p^{hyb}) \tag{8.2}$$

Ainsi, il suffit de choisir une fonction f respectant un certain nombre de contraintes pour obtenir une distribution en accord avec l'expertise de la CCR.

8.2.2 Choix de la fonction quantile hybride

Dans cette partie nous allons définir des fonctions quantile hybride permettant de respecter un certain nombre de contraintes correspondant aux besoins de la CCR. Pour cela, il faut établir ces contraintes :

- Ensemble de départ et ensemble d'arrivée
  La fonction quantile hybride doit être définie telle que :

$$\alpha \ : [0,1] \longrightarrow [0,1]$$

- Périodes de retour inférieures à 50 ans
  Dans le cas des petites périodes de retour, l'historique des données et plus fiable est permet une meilleure estimation du coût marché CCR. L'historique doit être dominant pour les périodes de retour en question, le facteur de crédibilité doit être tel que :

$$\forall\ p \leqslant 0,98\ ,\ \alpha(p) \leqslant \frac{1}{2}$$

- Périodes de retour supérieures à 200 ans
  Dans le cas des périodes de retour très élevées, l'historique n'apporte aucune information complémentaire par rapport à l'exposition. Pour cette raison, à partir d'une période de retour de 200 ans, le facteur de crédibilité doit être égal à 1.

$$\forall\ p \geqslant 0,995\ ,\ \alpha(p) = 1$$

Nous allons créer deux distributions hybrides qui respectent les contraintes vues précédemment. Pour cela, nous avons choisi deux fonctions quantile hybride différentes :

- La première, que nous appelons distribution hybride polynomiale est définie par le facteur de crédibilité suivant :
$$\alpha(p) = \min\{1, 37 \cdot p^{50}, 1\}$$
- La deuxième, est une distribution hybride exponentielle définie par la fonction quantile hybride suivante :
$$\alpha(p) = \min\{10^{-30} \cdot \exp(69,75 \cdot p), 1\}$$

La représentation graphique de ces deux fonctions (figure 8.10) montre que pour une période de retour inférieure à 20 ans, c'est-à-dire pour $p \leqslant 0,9667$, le modèle d'exposition n'est presque pas

pris en compte. En effet, plus α est petit, plus on accorde de l'importance à l'historique. De la même manière, pour une période de retour supérieure à 100 ans, c'est-à-dire pour $p \geqslant 0,99$, l'historique n'est presque pas pris en compte. On remarque également que le modèle hybride exponentiel donne plus d'importance à l'historique pour les petites périodes de retour et à l'exposition pour les grandes périodes de retour que le modèle polynomial.

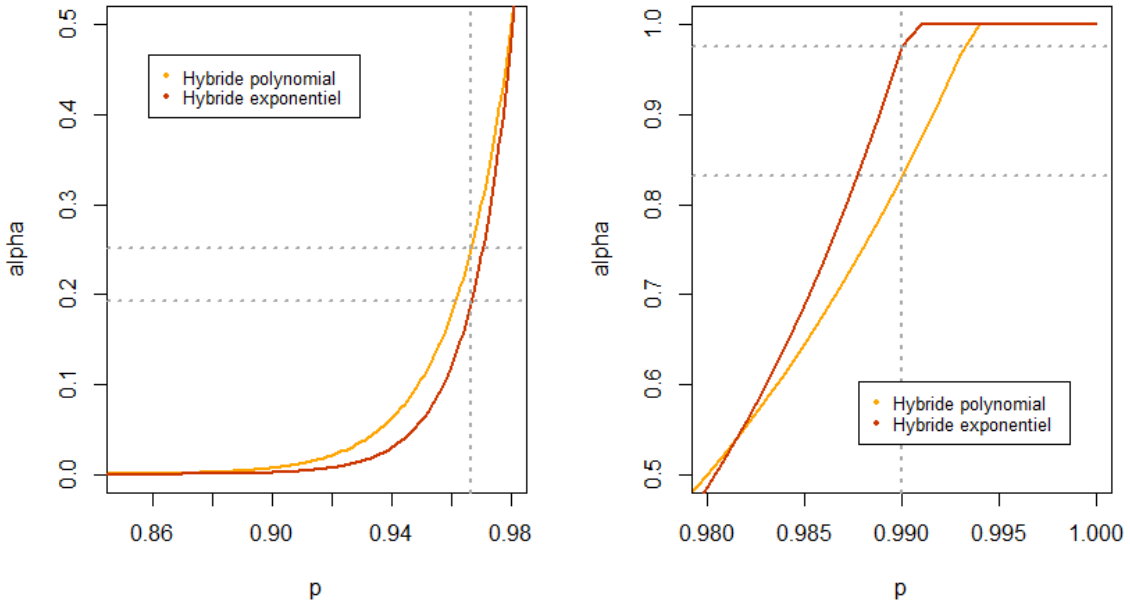

FIGURE 8.10 – Facteurs de crédibilité

8.2.3 Calcul des quantiles hybrides

Afin de calculer les quantiles hybrides, il faut avoir a disposition des quantiles provenant de l'exposition et des quantiles provenant de l'historique. Dans le cas de l'exposition, nous allons utiliser les données brutes provenant du modèle d'exposition, c'est-à-dire les 50 000 années simulées par le Modèle Interne Partiel.

Toutefois, pour obtenir des quantiles historiques il faut ajuster une loi à l'historique de la sinistralité. Ici nous allons utiliser les mêmes lois que nous avons ajustées dans la partie 8.1.1.

A partir des quantiles de l'exposition et de l'historique nous appliquons la formule 8.1 afin d'obtenir les quantiles hybrides, qui permettent de définir des distributions hybrides à partir desquelles nous

allons modéliser le coût des différents périls.

8.2.4 Crédibilisation hybride du péril inondation

Dans le cas du péril inondation, nous avons utilisé la distribution d'une loi Gamma afin d'estimer les quantiles historiques. En effet, la loi Gamma permet de conserver la moyenne de l'historique, contrairement à la loi Log-Normale qui renvoie une moyenne plus élevée. L'espérance et l'écart-type des différentes distributions sont présentés dans le tableau 8.21.

	Historique	Exposition	Hybride polynomial	Hybride exponentiel
Espérance (M€)	534	574	593	594
Écart-type (M€)	400	802	771	779

TABLE 8.21 – Statistiques de la sinistralité crédibilisée pour le péril inondation

Sur le tableau précédent nous pouvons lire que l'écart-type de la distribution hybride se trouve bien entre l'écart-type historique et l'écart-type de l'exposition. Cependant, nous remarquons que, quelque soit la méthode hybride utilisée, la sinistralité hybride moyenne est plus élevée que les sinistralités moyennes historique et d'exposition. Pour analyser ses résultats, nous allons nous intéresser aux niveaux de retour du tableau 8.22.

Période de retour	Historique	Exposition	Hybride polynomial	Hybride exponentiel
2 ans	441	397	441	441
5 ans	817	803	817	817
10 ans	1 077	1 113	1 077	1 077
20 ans	1 328	1 448	1 341	1 336
50 ans	1 656	2 027	1 841	1 836
100 ans	1 883	3 927	3 581	3 877
200 ans	2 112	6 654	6 654	6 654
500 ans	2 395	8 145	8 145	8 145
1000 ans	2 607	9 384	9 384	9 384

TABLE 8.22 – Niveaux de retour de la sinistralité pour le péril inondation (M€)

Nous remarquons, sur le tableau 8.22, que pour les petites périodes de retour, les niveaux de retour de l'exposition sont plus faibles que les niveaux de retour historiques. Au contraire, pour les grandes périodes de retour, les niveaux de retour de l'exposition sont plus élevés que les niveaux de retour

historiques.

Comme il a été expliqué précédemment, nous avons accordé un poids plus important à l'historique pour les petites périodes de retour et à l'exposition pour les grandes périodes des retour. En procédant ainsi nous accordons plus poids au montant le plus important pour la période de retour en question. La distribution hybride obtenue est donc plus sévère que les deux distributions de base. Pour cette raison nous obtenons une moyenne plus élevée.

8.2.5 Crédibilisation hybride du péril sécheresse

Dans le cas du péril sécheresse nous avons également estimé les quantiles historiques à partir d'une loi Gamma. Le tableau 8.23 contient les moyennes et les écart-types obtenus avec les différentes distributions. Nous remarquons ici que les moyennes et les variances des distributions hybrides correspondent bien à un équilibre entre l'historique et l'exposition.

	Historique	Exposition	Hybride polynomial	Hybride exponentiel
Espérance (M€)	360	441	395	394
Écart-type (M€)	379	684	549	552

TABLE 8.23 – Statistiques de la sinistralité crédibilisée pour le péril sécheresse

Les niveaux de retour des deux approches sont présentés dans le tableau 8.24. Ce tableau permet de vérifier que les contraintes établies à partir de l'expertise de la CCR sont bien respectées. Nous remarquons que pour des périodes de retour allant jusqu'à 20 ans, les niveaux de retour hybrides sont très proches des niveaux historiques. Symétriquement, pour des périodes de retour supérieures à 100 ans, les niveaux de retour hybrides sont très proches de l'exposition.

Période de retour	Historique	Exposition	Hybride polynomial	Hybride exponentiel
2 ans	237	171	237	237
5 ans	580	690	580	580
10 ans	847	1 211	850	848
20 ans	1 107	1 842	1 185	1 151
50 ans	1 449	2 705	2 077	2 059
100 ans	1 718	3 323	3 051	3 283
200 ans	2 001	3 838	3 838	3 838
500 ans	2 279	4 698	4 698	4 698
1000 ans	2 506	5 260	5 260	5 260

TABLE 8.24 – Niveaux de retour de la sinistralité pour le péril sécheresse (M€)

8.2.6 Crédibilisation hybride de la sinistralité automobile

Dans le cas de la sinistralité liée aux contrats d'assurance automobile, nous avons utilisé une distribution Log-Normale pour estimer les quantiles de l'historique. De la même manière que pour le péril sécheresse, nous remarquons que d'après le tableau 8.25, l'espérance et l'écart-type hybrides correspondent à une pondération des espérances et écart-types historiques et d'exposition.

	Historique	Exposition	Hybride polynomial	Hybride exponentiel
Espérance (M€)	49	67	54	54
Écart-type (M€)	123	166	161	162

TABLE 8.25 – Statistiques de la sinistralité crédibilisée pour l'assurance automobile

De la même manière que précédemment, le tableau 8.26 permet de vérifier le respect des contraintes établies dans la partie 6. Encore une fois, le poids accordé à l'historique est bien équilibré en fonction des périodes de retour et permet d'obtenir des niveaux de retour hybrides proches des niveaux historiques pour les petites périodes de retour et proches de l'exposition pour les grandes périodes de retour.

Période de retour	Historique	Exposition	Hybride polynomial	Hybride exponentiel
2 ans	18	32	18	18
5 ans	60	81	60	60
10 ans	110	134	110	110
20 ans	185	216	188	187
50 ans	325	501	413	411
100 ans	480	684	650	679
200 ans	700	923	923	923
500 ans	1 076	1 459	1 459	1 459
1000 ans	1 523	1 874	1 874	1 874

TABLE 8.26 – Niveaux de retour de la sinistralité pour l'assurance automobile (M€)

8.3 Comparaison des méthodes de crédibilisation

Nous avons vu dans les deux sections précédentes, deux méthodes de crédibilisation différentes : la méthode bayésienne et la méthode hybride. Dans cette section, nous allons comparer ces deux méthodes afin de décider quelle distribution nous allons utiliser dans le Modèle Interne Partiel Hybride.

8.3.1 Péril inondation

Dans le cas du péril inondation, la figure 8.11 permet de comprendre plus clairement la différence entre les deux approches. Sur le graphique de gauche, nous remarquons que la méthode bayésienne permet de prendre en compte les montants modélisés à partir de l'exposition sous la forme d'une augmentation du niveau de retour historique proportionnelle à la période de retour. Toutefois, cette approche ne prend pas en compte le comportement de la courbe provenant de l'exposition.

Ceci met en évidence un des principaux inconvénients de cette approche : l'hypothèse de la loi du montant des sinistres. En effet, pour appliquer cette méthode il est nécessaire d'ajuster la même loi aux deux jeux de données, or il n'est pas toujours possible d'ajuster une loi aux données provenant du modèle d'exposition.

L'utilisation de la méthode bayésienne suppose donc que le modèle d'exposition pour les inondations suit la même loi que l'historique. La loi mélange obtenue se base donc sur une approximation de l'exposition et non pas sur les données provenant du modèle.

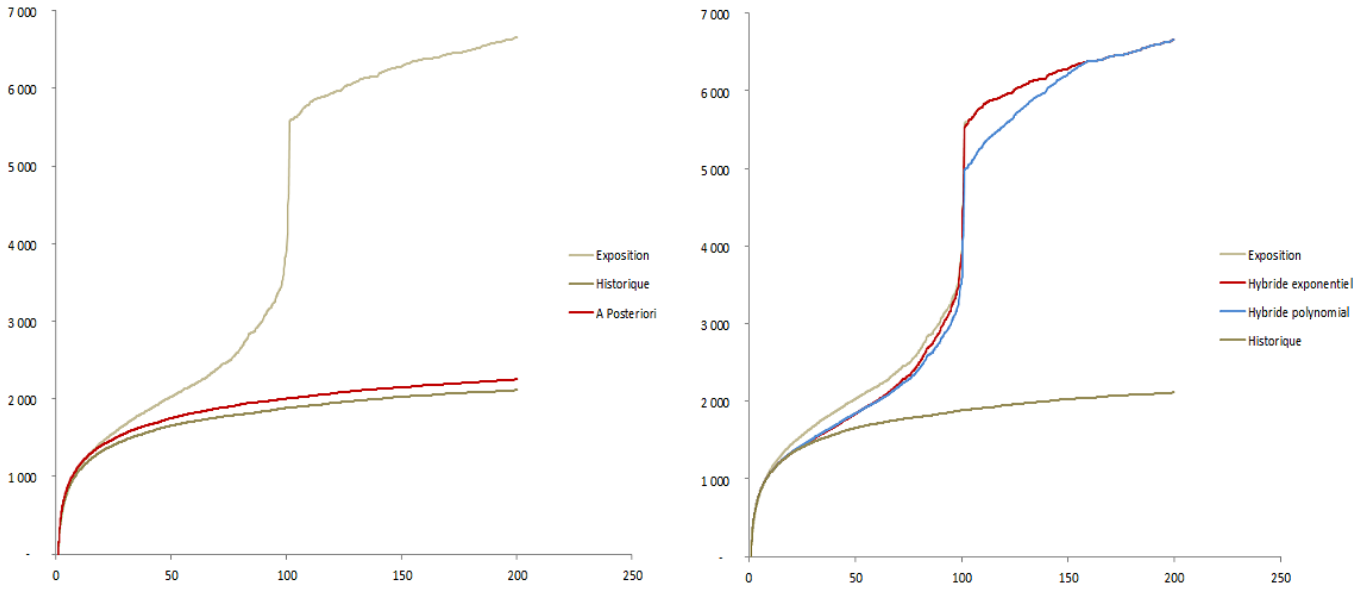

FIGURE 8.11 – Niveaux de retour des inondations

Au contraire, nous remarquons sur le graphique de droite, que l'approche hybride permet de bien prendre en compte le comportement des deux courbes. La courbe hybride prend ainsi la forme de la courbe historique pour les petites périodes de retour et s'ajuste ensuite à la courbe d'exposition.

Nous remarquons également que la courbe provenant de l'exposition aux inondations a une allure très particulière. En effet, il y a un saut au niveau des événements centennaux. Il est donc important de prendre en compte la forme de cette courbe lors de la modélisation et la tarification des traités.

8.3.2 Péril sécheresse

Les courbes des niveaux de retour pour le péril sécheresse ont des allures plus classiques. Sur la figure 8.13, nous pouvons observer les niveaux de retour obtenus à partir de la méthode bayésienne à gauche et à partir de la méthode hybride à droite.

Le graphique de gauche permet d'observer l'effet de la méthode bayésienne plus clairement. Nous remarquons que dans cas où l'exposition s'ajuste bien à une loi, nous n'observons pas l'effet de l'absence de prise en compte de la forme de la courbe. Toutefois, la courbe mélange prend en compte les deux distributions indépendamment de la période de retour. De cette manière, l'historique est toujours pris en compte. Le choix de la loi peut donc avoir un impact important alors

qu'il se base sur un historique court qui n'est pas fiable pour modéliser les périodes de retour élevées.

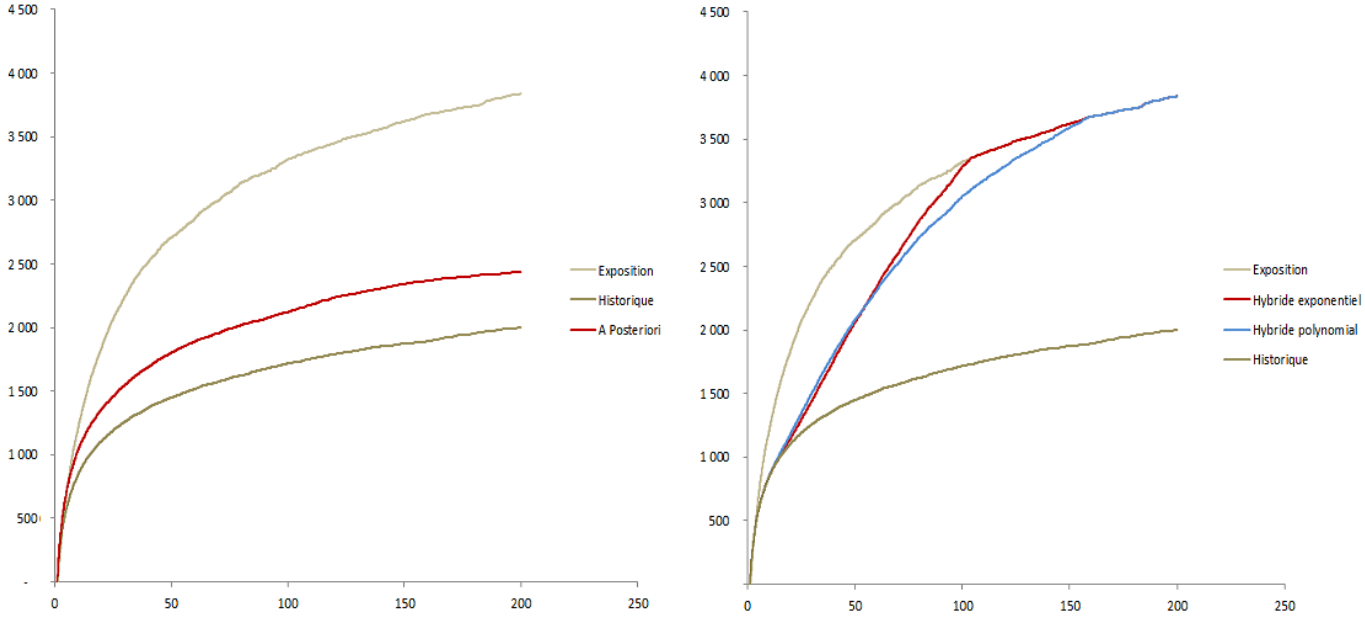

FIGURE 8.12 – Niveaux de retour des sécheresses

L'approche hybride permet de mieux prendre en compte la fiabilité des données provenant de l'historique pour la modélisation des différentes périodes de retour. Nous remarquons que le modèle hybride exponentiel permet de converger plus rapidement vers le modèle d'exposition. De cette manière, nous pouvons neutraliser l'historique à partir d'une certaine période de retour afin d'obtenir des résultats plus fiables.

8.3.3 Sinistralité automobile

Enfin, dans le cas de la sinistralité associée aux contrats d'assurance automobile, nous pouvons observer un dernier inconvénient de la méthode bayésienne. Le graphique à gauche de la figure 8.13 permet d'illustrer le cas d'une loi mélange qui ne correspond pas à un équilibre entre les deux jeux de données. En effet, quelque soit la période de retour, le niveau de retour mélange est moins sévère que les niveaux de retour historique et d'exposition.

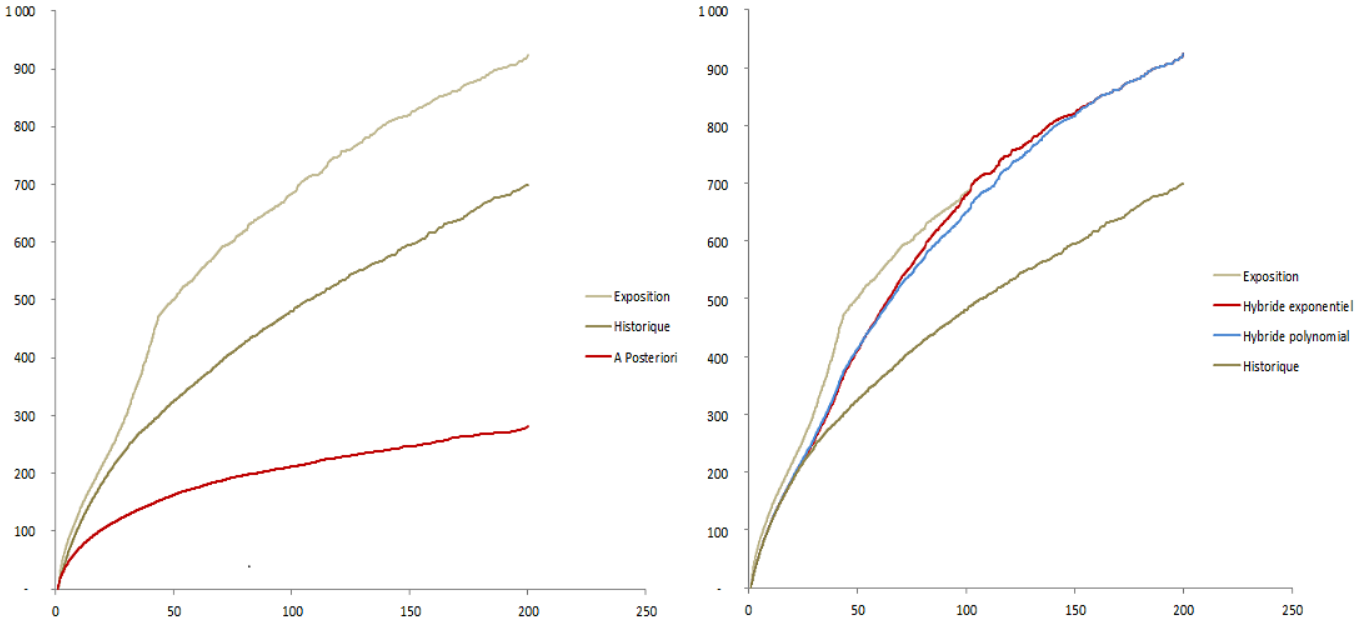

FIGURE 8.13 – Niveaux de retour de la sinistralité automobile

La méthode hybride permet d'assurer un équilibre entre les deux lois afin d'obtenir des niveaux de retour hybrides adaptés.

8.4 Conclusion sur la comparaison des méthodes de crédibilisation

Nous avons crédibilisé le coût marché CCR à partir d'une approche bayésienne classique pour différents périls. Nous avons remarqué que, selon l'expertise de la CCR, l'approche bayésienne ne s'adapte pas aux données que nous avons à notre disposition pour les différentes raisons qui sont présentées ci-dessous.

Le principal problème concernant la théorie bayésienne est l'ajustement d'une loi. La sensibilité de cette méthode au choix de la loi et au choix du paramètre que l'on souhaite crédibiliser rend les résultats très instables. Cette instabilité est ici accentuée par un historique faible qui ne permet pas de déterminer quelle est la loi la plus adaptée.

Par ailleurs, les jeux de données historique et d'exposition doivent être modélisés par une même loi.

Il s'agit ici d'une hypothèse très forte dès lors que nous sortons du cadre classique de la crédibilité. Dans notre étude, au vu de la grandeur des montants des sinistres, les données historiques et les données provenant de l'exposition peuvent difficilement être approchées par la même loi. En effet, les données historiques ne comportent pas les sinistres extrêmes qui sont modélisés par le modèle par exposition.

Enfin, les informations fournies par l'exposition sont ici résumées par des statistiques. En effet, nous avons vu dans la section 7 que les paramètres de la loi a posteriori et donc par conséquent de la loi mélange, sont estimés à partir des statistiques provenant des deux jeux de données. De cette manière, cette approche ne prend pas en compte les informations apportées par chaque sinistre mais uniquement la moyenne, la variance et le nombre de données de chaque jeu de données. L'information que ces statistiques apportent concernant les quantiles extrêmes est donc très limitée.

Les poids accordés aux quantiles historique et d'exposition sont ainsi indépendants de la période de retour associée. Cette approche s'adapte donc mal à notre problématique puisque l'historique est modélisé à partir d'une loi qui est ajustée à des données qui ne sont pas fiables pour estimer des périodes de retour élevées.

Nous rappelons que le but de cette étude est de trouver une distribution proche de la distribution historique pour les petites périodes de retour et qui converge vers la distribution de l'exposition pour les grandes périodes de retour. Nous avons donc développé une deuxième approche en fonction des particularités des périls et des données étudiés et plus particulièrement de l'expertise CCR. A partir de cette méthode nous obtenons une distribution hybride qui permet de remédier aux inconvénients de la méthode bayésienne cités ci-dessus.

Une combinaison convexe permet de créer une distribution hybride à partir des données de l'exposition sans passer par le choix d'une loi. Toutefois, il est nécessaire d'ajuster une loi aux données historiques pour obtenir suffisamment de quantiles afin de définir la distribution hybride. Ainsi, cette approche permet de réduire les inconvénients liés à l'ajustement des lois et de prendre en compte la distribution de l'exposition sans passer par des modélisations peu fiables.

Nous avons vu précédemment que les quantiles hybrides sont définis en fonction de la période de retour. Le poids accordé aux quantiles historique et d'exposition est défini en fonction de la période de retour, permettant ainsi d'éliminer les problèmes liés à la taille de l'historique et sa fiabilité pour les grandes périodes de retour.

Enfin, cette approche permet de prendre en compte toute l'information contenue de chaque donnée modélisée à partir du modèle par exposition. En effet, pour chaque donnée d'exposition nous avons défini une donnée hybride. Cependant, la prise en compte de l'historique se fait via des quantiles modélisés à partir d'une loi qui est définie en fonction des statistiques provenant des données historiques. Cette approche permet donc de réduire ces problèmes mais pas de les éliminer.

Nous pouvons donc conclure que l'approche la plus adaptée dans le cadre de cette étude est la méthode hybride. Pour cette raison, nous avons développé un modèle de tarification hybride semblable au modèle par exposition à partir de cette approche. Ce dernier permet de simuler des événements à partir de la distribution hybride qui a été décrite ci-dessus. Le modèle en question est décrit dans la section suivante.

Chapitre 9

Création d'un Modèle Interne Partiel Hybride

Dans cette section nous allons détailler le fonctionnement du Modèle Interne Partiel Hybride. Ce modèle pourrait être utilisé par la suite pour la tarification des traités, ou bien comme le modèle MBU à titre indicatif.

9.1 Mise en place des données

Dans un premier temps, il a fallu mettre en place une base de données permettant d'alimenter le modèle. Dans ce modèle nous avons séparé la sinistralité provenant des contrats automobile de la sinistralité des polices dommages aux biens professionnels, particuliers et agricoles. Pour chaque cédante nous avons donc créé deux distributions hybrides permettant de modéliser la sinistralité annuelle de la cédante pour chacune des catégories. Nous appellerons ces dernières la distribution hybride automobile et la distribution hybride non automobile.

Dans le but de créer les distributions hybrides, nous avons extrait les distributions provenant du modèle historique et du modèle d'exposition sous la forme d'un tableau de quantiles. Dans le cas de l'exposition, nous avons simulé 50 000 années à partir du modèle MBU à partir desquelles nous avons extrait les quantiles. Dans le cas de l'historique, nous avons utilisé la méthode de Monte-Carlo en simulant des années à partir d'une loi log-normale. Le tableau 9.1 montre un exemple des tableaux de données que nous utiliserons.

	Historique		Exposition	
p	Automobile	Non automobile	Automobile	Non automobile
0,001	$x_{0,001}^{H,A}$	$x_{0,001}^{H,N}$	$x_{0,001}^{E,A}$	$x_{0,001}^{E,N}$
0,002	$x_{0,002}^{H,A}$	$x_{0,002}^{H,N}$	$x_{0,002}^{E,A}$	$x_{0,002}^{E,N}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

TABLE 9.1 – Quantiles extraits des modèles d'exposition et historique

A partir du tableau de quantiles provenant des deux distributions, nous estimons les quantiles hybrides à l'aide de la fonction quantile hybride et de la formule 8.1. Nous obtenons ainsi un tableau similaire de quantiles hybrides pour chaque catégorie.

9.2 Description du modèle

Une fois que nous avons créé les distributions hybrides de chaque cédante, nous avons procédé à la création du modèle hybride sur ReMetrica. Comme le modèle MBU, le modèle hybride est composé de modules ayant différentes fonctions. Tout d'abord nous avons des générateurs permettant de modéliser la sinistralité de chaque cédante à partir d'un tableau de quantiles ou d'une loi de probabilités pour laquelle on fournit les paramètres. Ici nous avons alimenté les générateurs avec les tableaux de quantiles tels qu'ils ont été présentés précédemment. Comme les traités de réassurance ne distinguent pas les contrats automobile du reste des polices dommages, il faut ensuite faire une agrégation par cédante.

Les conditions contractuelles du traité de réassurance de la cédante sont ensuite appliquées au coût brut de cette dernière. Tout d'abord, nous appliquons le quote-part qui divise le coût entre l'assureur et le réassureur de manière proportionnelle à la répartition des primes. L'excédant de perte annuelle est ensuite appliqué à la part qui reste à la charge de la cédante. Nous obtenons ainsi la coût net pour l'assureur et le coût brut pour la CCR. Enfin, nous appliquons le seuil d'intervention de l'État pour obtenir le coût net pour la CCR et coût pour l'État. Le schéma 9.2 résume les différentes étapes du modèle.

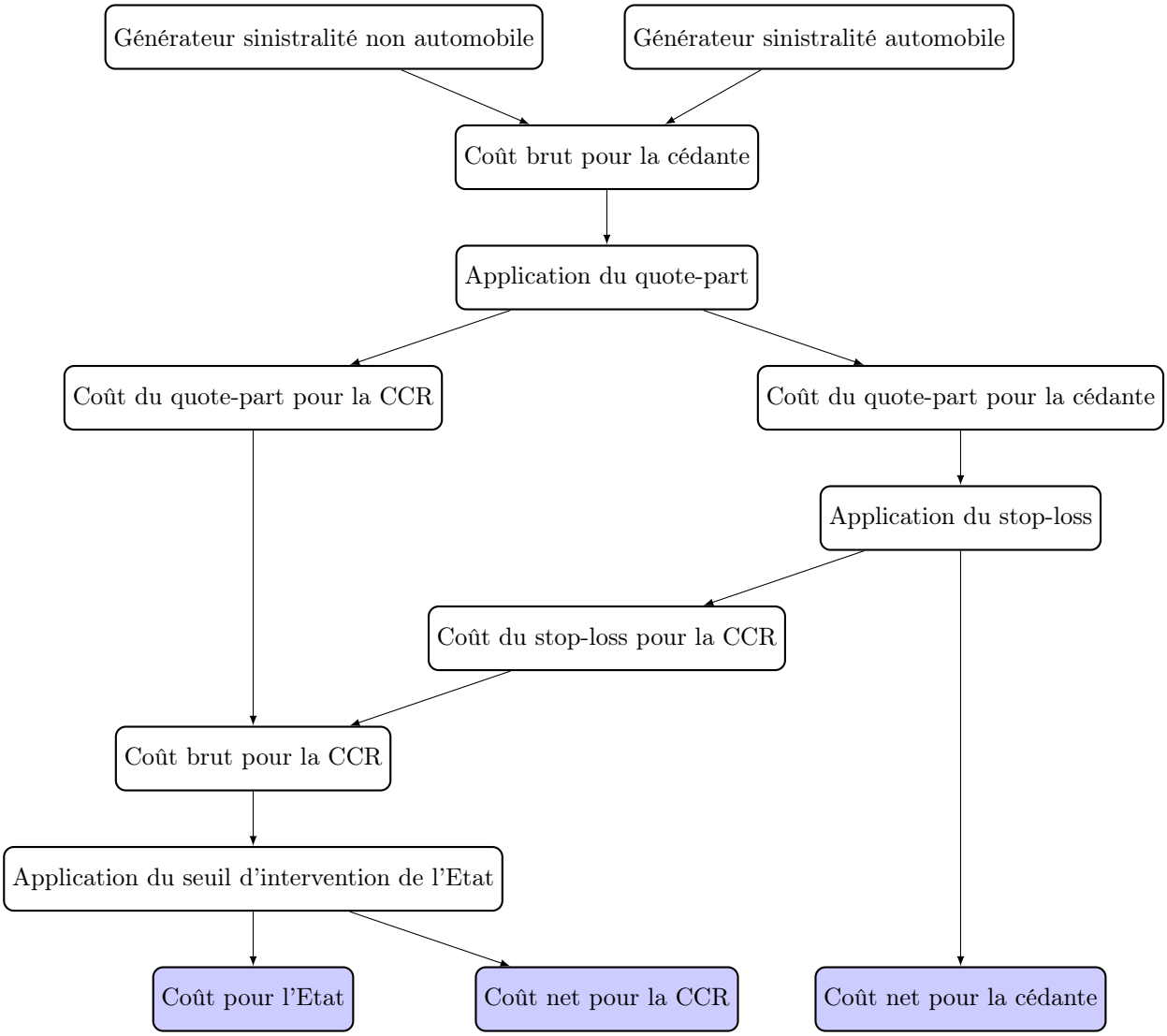

FIGURE 9.1 – Schéma du modèle hybride pour une cédante

9.3 Résultats au niveau de la CCR

Dans cette section, nous allons présenter des résultats qui concernent la sinistralité à la charge de la CCR dans le cadre du régime des catastrophes naturelles pour les différents modèles. Dans un premier temps, nous allons étudier les moyennes et les écart-types des différents modèles (hybride polynomial et hybride exponentiel). Ensuite, nous allons procéder à une analyse plus fine à travers des niveaux de retour.

9.3.1 Moyenne et écart-type

Les principales statistiques à étudier sont la moyenne et l'écart-type. En effet, nous cherchons un modèle qui nous permette de prendre en compte l'exposition des cédantes tout en réduisant l'écart avec les tarifs actuels. Le but est de prendre en compte les événements extrêmes modélisés par le modèle d'exposition, il est donc important d'avoir un écart-type qui reflète la sinistralité extrême et par conséquent proche de l'écart-type du modèle d'exposition.

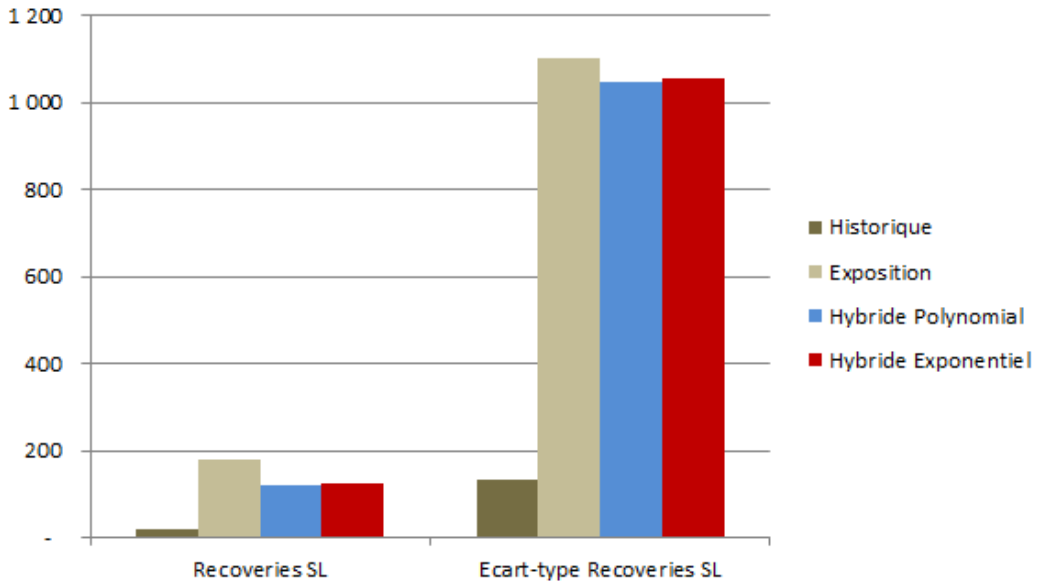

FIGURE 9.2 – Sinistralité à la charge de la CCR (M€)

Nous pouvons observer les résultats obtenus sur la figure 9.2. Nous remarquons que l'espérance des deux modèles hybrides se trouve bien entre les modèles historique et d'exposition. De la même

manière, nous pouvons observer que l'écart-type des modèles hybrides se rapproche de celui l'exposition.

Le tableau 9.2 présente les statistiques chiffrées que nous avons vues précédemment sur le graphique. Nous remarquons que la charge moyenne de la CCR des modèles hybrides représente bien un équilibre entre les modèles historique et d'exposition. Ces modèles permettent ainsi de trouver une espérance raisonnable tout en prenant en compte l'exposition.

	Historique	Exposition	Hybride polynomial	Hybride exponentiel
Charge moyenne CCR	504	902	721	723
Écart-type CCR	487	2 363	2 213	2 236

TABLE 9.2 – Sinistralité à la charge de la CCR (M€)

Par ailleurs, nous remarquons que l'écart-type des modèles hybrides est beaucoup plus proche du modèle d'exposition. Ce résultat est cohérent avec les hypothèses des modèles hybrides. En effet, le but était d'introduire des événements extrêmes modélisés à partir du modèle d'exposition, tels que des événement centennaux ou millénaires. La prise en compte de telles catastrophes augmente systématiquement l'écart-type.

En outre, qu'il s'agisse de la moyenne ou de l'écart-type, la différence entre les deux modèles hybrides est considérée comme étant négligeable. Les deux modèles sont donc équivalents à ce niveau.

9.3.2 Niveaux de retour

Afin d'analyser plus en détail ces résultats, nous allons nous intéresser aux différents niveaux de retour. Ces derniers nous permettront de vérifier la prise en compte des contraintes définies lors de la création des distributions hybrides. Une représentation graphique pour les périodes de retour inférieures à 200 ans est présentée sur la figure 9.3. En théorie les deux modèles hybrides sont identiques au modèle d'exposition pour les périodes de retour supérieures à 200 ans.

Sur la figure en question nous pouvons principalement observer les phénomènes suivants :

- pour les petites périodes de retour, les deux modèles hybrides ont des niveaux de retour qui se rapprochent du modèle historique ;

- au contraire, pour les grandes périodes de retour, les deux modèles hybrides ont des niveaux de retour semblables à ceux du modèle d'exposition ;

- le modèle hybride exponentiel se rapproche plus rapidement du modèle d'exposition que le modèle hybride polynomial.

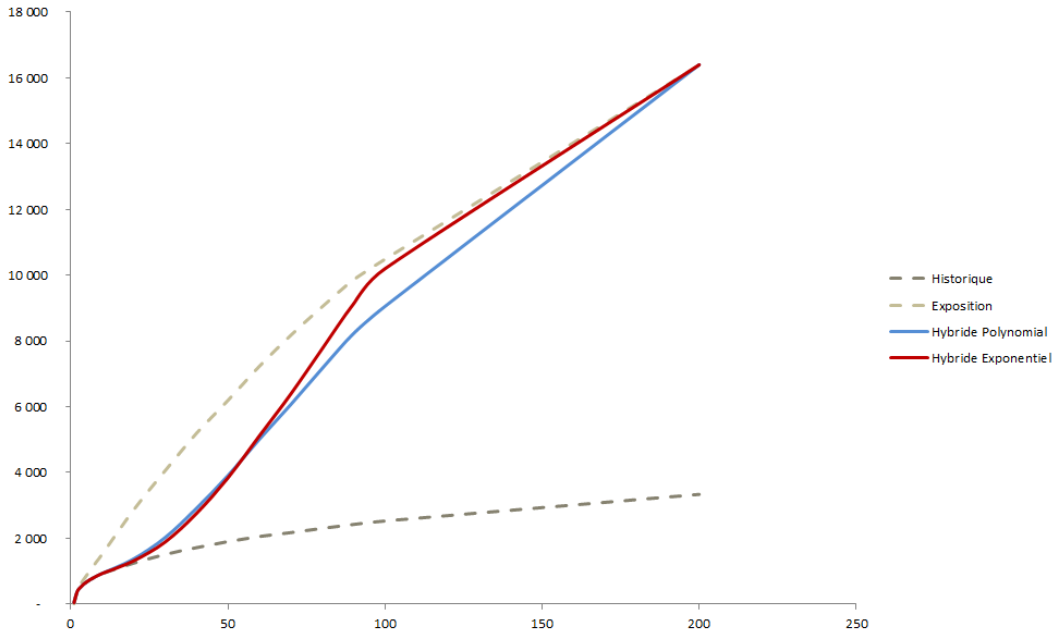

FIGURE 9.3 – Niveaux de retour de la sinistralité à la charge de la CCR (M€)

Nous allons vérifier ces résultats à partir du tableau 9.3. Ce dernier présente les niveaux de retour de la sinistralité à la charge de la CCR pour des périodes de retour allant de 1 an à 1000 ans.

Période de retour	Historique	Exposition	Hybride polynomial	Hybride exponentiel
1 ans	49	10	49	49
2 ans	375	405	377	377
5 ans	674	877	676	676
10 ans	931	1524	937	934
20 ans	1 247	2 869	1 376	1 320
50 ans	1 902	6 195	3 910	3 840
100 ans	2 528	10 485	9 053	10 198
200 ans	3 338	16 388	16 397	16 397
500 ans	4 532	26 250	26 239	26 239
1000 ans	5 529	35 518	35 516	35 516

TABLE 9.3 – Niveaux de retour de la sinistralité de la CCR (M€)

Le tableau en question permet de confirmer les observations que nous avons pu faire sur la figure précédente. En effet, pour les petites périodes de retour jusqu'à 20 ans, les niveaux de retour hybrides sont très proches de l'historique. Ceci est cohérent avec les hypothèses que nous avons prises, puisque l'historique est prépondérant par rapport à l'exposition.

Pour une période de retour de 50 ans, nous avons un équilibre entre l'historique et l'exposition. Encore une fois ceci semble cohérent parce que nous avons pris un facteur de crédibilité au alentours de 50% pour les événements quinquennaux, c'est-à-dire pour les catastrophes ayant lieu en moyenne une fois tous les 50 ans.

Symétriquement, dans le cas des périodes de retour élevées, les niveaux de retour hybrides sont proches du modèle d'exposition. En effet, l'exposition est prépondérante à partir d'une période de retour de 50 ans. Pour les événements centennaux, nous remarquons que les niveaux de retour hybrides sont presque égaux à l'exposition. En effet, comme l'historique ne contient que 30 années il a été décidé de ne plus prendre en compte l'historique pour les périodes de retour supérieures à 100 ans, puisque ce dernier n'est pas pertinent.

9.3.3 Conclusion par rapport aux résultats CCR

Nous avons vu dans cette section, que les principales statistiques sont satisfaisantes. En effet, l'espérance des modèles hybrides correspond à un équilibre entre l'historique et l'exposition, et l'écart-type se rapproche du modèle d'exposition ce qui correspond à la prise en compte des événe-

ments extrêmes.

Nous avons également vu que les niveaux de retour pour les différentes périodes de retour sont cohérents avec les hypothèses que nous avons prises pour définir les distributions hybrides : la prépondérance de l'historique pour les petites périodes de retour, et au contraire, la prépondérance du modèle d'exposition pour les grandes périodes de retour.

En outre, en ce qui concerne les différences entre les deux modèles hybrides nous avons remarqué que l'écart au niveau de la moyenne et de l'écart-type était négligeable. Toutefois, nous avons vu que le modèle hybride exponentiel suit le modèle d'exposition à partir des événements centennaux alors que le modèle polynomial le suit uniquement à partir des événements bicentennaux. Comme cet écart n'affecte pas la moyenne et l'écart-type pour le calcul de la prime, il semble plus cohérent de garder le modèle hybride exponentiel. En effet, l'information apportée sur les catastrophes centennales par un historique de 30 ans est très réduite, et le modèle d'exposition est plus fiable pour de telles périodes de retour.

9.4 Résultats au niveau d'une cédante

Au niveau des cédantes nous pouvons obtenir des résultats très variés. Nous savons qu'au niveau du marché CCR le modèle d'exposition est plus sévère que le modèle historique, mais l'écart entre les résultats du modèle d'exposition et l'historique de la cédante dépend de chaque cédante et de la précision des données transmises par la cédante.

Nous allons illustrer les résultats à partir d'une cédante que nous appellerons la cédante A pour des raisons de confidentialité.

De manière générale nous obtenons des graphiques semblables aux graphiques que nous avons vus au niveau de la CCR. Nous obtenons une moyenne qui correspond à un équilibre entre l'exposition et l'historique de la cédante ainsi qu'un écart-type du même ordre de grandeur que l'écart-type du modèle d'exposition. Ceci est le cas de la cédante A dont les résultats sont présentés sur la figure 9.4.

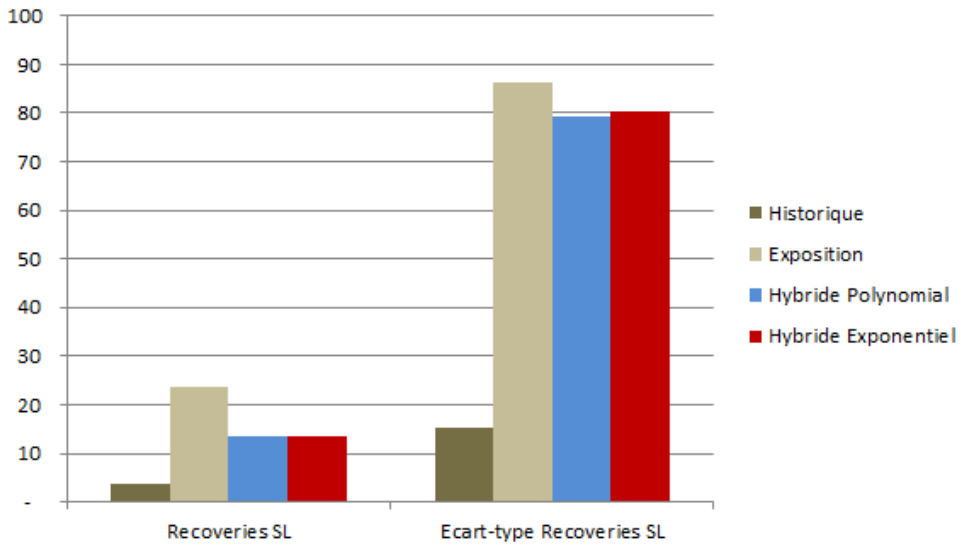

FIGURE 9.4 – Sinistralité à la charge du Stop-Loss de la cédante A (M€)

Le tableau 9.4 permet de confirmer ses résultats. Nous remarquons également que nous obtenons la même moyenne avec les deux modèles hybrides et la différence au niveau de l'écart-type est négligeable.

	Historique	Exposition	Hybride polynomial	Hybride exponentiel
Charge moyenne	4	24	13	13
Écart-type	15	86	79	80

TABLE 9.4 – Sinistralité à la charge du Stop-Loss de la cédante A (M€)

En ce qui concerne les niveaux de retour, pour les petites périodes de retour nous avons la même allure que pour le modèle historique. Le niveaux de retour s'élèvent ensuite pour atteindre le niveau de retour d'exposition pour une période de retour de 100 ans pour le modèle exponentiel et 200 ans pour le modèle polynomial.

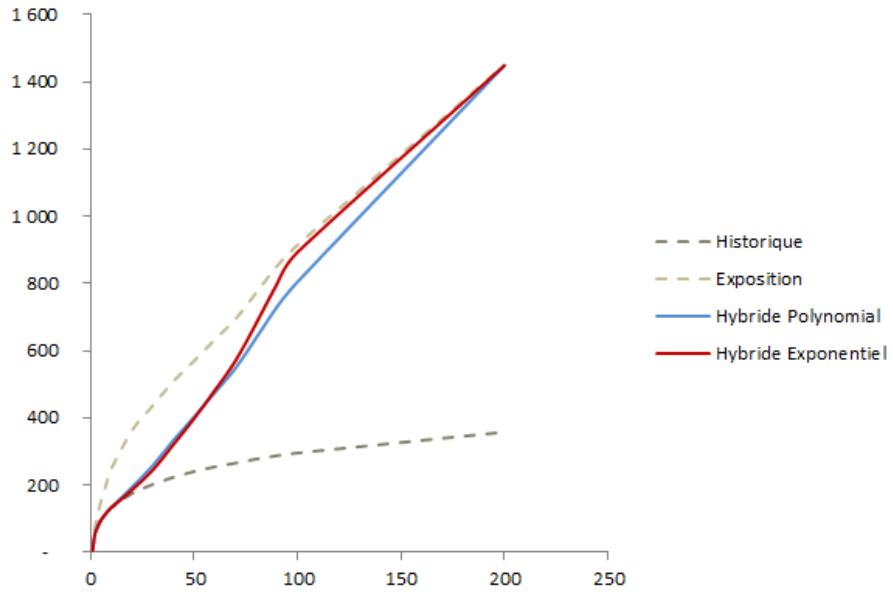

FIGURE 9.5 – Niveaux de retour de la charge du Stop-Loss de la cédante A(M€)

A partir du tableau 9.5, nous pouvons tirer les mêmes conclusions que pour les résultats au niveau de la CCR et vérifier la cohérence de ces résultats avec les hypothèses des modèles. Nous remarquons la prépondérance du modèle historique pour des périodes inférieures à 50 ans et ensuite la prépondérance du modèle d'exposition.

Période de retour	Historique	Exposition	Hybride polynomial	Hybride exponentiel
1 ans	5	1	5	5
2 ans	51	66	51	51
5 ans	95	150	95	95
10 ans	133	248	134	133
20 ans	175	362	194	185
50 ans	241	569	403	398
100 ans	296	915	804	893
200 ans	358	1 450	1 448	1 448
500 ans	442	2 214	2 215	2 215
1000 ans	508	2 434	2 429	2 429

TABLE 9.5 – Niveaux de retour de la sinistralité brute de la cédante (M€)

Chapitre 10

Conclusion générale et perspectives

Le Régime des Catastrophes Naturelles a un impact important dans l'activité de la Caisse Centrale de Réassurance. Ce dernier représente 90% du chiffre d'affaires de la réassurance publique. Il est donc crucial pour la CCR de modéliser avec précision le risque sous-jacent. La difficulté de cette activité réside dans la modélisation à la fois des sinistres fréquents tels que des inondations dans le Var mais également des évènements extrêmes qui peuvent aller des crues centennales aux sécheresses, séismes ou même vents cycloniques bicentenaires.

Pour cette raison la CCR a développé, à l'aide d'équipes d'experts dans les différents domaines concernés, des modèles permettant d'étudier les sinistres extrêmes. Les évènements extrêmes sont ainsi modélisés à partir de leurs caractéristiques physiques pour lesquels nous disposons d'un historique suffisamment important. Ces modèles sont calibrés sur les périodes de retour importantes et sont donc peu fiables sur les petites périodes de retour. Dans le cadre des sinistres fréquents, nous disposons également de l'historique des cédantes. Cet historique permet d'ajuster des lois de probabilités afin de modéliser de tels évènements. Le but de cette étude était d'apporter une base commune afin d'obtenir une unique distribution contenant les informations apportées par l'historique des cédantes et par l'expertise des équipes CCR au travers de la modélisation physique des événements.

La théorie bayésienne est généralement utilisée dans le cadre des contrats d'assurance automobile. Cette méthode permet de prendre en compte l'expérience de l'assureur contenue dans l'historique de la sinistralité de l'assureur, mais également les spécificités de l'assuré qui sont comprises dans sa sinistralité personnelle. De cette manière, la théorie bayésienne permet de prendre en compte deux sources d'informations afin de créer une distribution plus riche. Toutefois, les limites de cette

méthode sont accentuées par les particularités des périls étudiés ici.

Les particularités des deux jeux de données rendent le choix d'une loi commune difficile. Une étude approfondie serait nécessaire, mais cela ne semble pas idéal puisqu'il serait nécessaire de faire une étude spécifique pour chaque cédante. Par ailleurs, les informations provenant des deux jeux de données sont résumées par des statistiques. Les informations concernant les quantiles extrêmes sont donc restreintes. En outre, la sensibilité des résultats au choix de la loi les rend peu fiables. La méthode bayésienne ne semble donc pas adaptée dans le cadre de cette étude puisqu'elle ne permet pas d'exploiter pleinement les informations provenant des deux sources d'information.

Une combinaison convexe des distributions historique et d'exposition semblait être la solution idéale. Nous avons donc défini des poids en fonction de la période de retour. Pour cela nous avons défini une fonction que nous avons appelée la fonction quantile hybride permettant de calculer un poids pour chaque quantile de manière à obtenir un quantile mélange que nous appelons le quantile hybride. Cette méthode nous a ainsi permis de prendre en compte la sinistralité extrême simulée par les modèles physiques à travers la prise en compte des quantiles extrêmes. Par ailleurs, cette méthode permet de réduire les erreurs dues à une modélisation peu fiable de la loi historique sur les queues de distribution. Pour cela il suffit d'accorder un poids suffisamment important à l'exposition pour les grandes périodes de retour. Pareillement, cette méthode permet de réduire les erreurs provenant d'un mauvais calibrage des modèles physiques sur les petites périodes de retour en accordant un poids suffisamment important à l'historique pour les petites périodes de retour. Enfin, la combinaison convexe permet d'écarter les risques liés au choix de la loi et du paramètre crédibilisé, et les mauvais ajustements des lois.

La distribution hybride prend en compte des critères spécifiques aux modèles étudiés tels que la fiabilité de la distribution historique et d'exposition en fonction de la période de retour étudiée. Du point de vue pratique, elle peut être mise en place avec moins de difficultés et ne nécessite pas d'une étude approfondie sur chaque cédante. Par ailleurs, cette méthode intuitive ne nécessite pas une connaissance approfondie de la théorie bayésienne pour la compréhension, ce qui est important pour une plus grande transparence des résultats des modèles.

Un Modèle Interne Partiel Hybride a ensuite été créé à partir de la distribution hybride. Ce dernier permet de modéliser la sinistralité de chaque cédante et d'appliquer ensuite les traités de réassurance. Toutefois, en absence d'informations concernant le détail par péril sur la sinistralité historique de chaque cédante, il n'a pas été possible de créer un générateur par péril. Ainsi, la sinistralité modélisée pour chaque cédante correspond à la sinistralité prise en charge par le Régime

des Catastrophes Naturelles tous périls confondus. Cette agrégation peut potentiellement introduire une corrélation entre les différents périls. A ce jour, nous ne disposons pas des données nécessaires pour remédier à ce problème. Toutefois, il serait intéressant de demander aux cédantes le détail par péril mais ce processus pourrait s'avérer très long.

Le modèle décrit ci-dessus permettra d'évaluer avec plus de précision les éventuelles modifications du Régime des Catastrophes Naturelles, mais également au niveau des couvertures proposées dans le cadre de la garantie d'État. Une étude au niveau du chargement de la prime permettrait d'utiliser ce modèle comme base pour la tarification des traités de réassurance proposés par la CCR.

Actuellement, les équipes du Département Études Techniques Garantie d'État travaillent sur la création d'un outil de tarification hybride en prenant comme base cette étude.

Bibliographie

[1] J. Depuis. Statistique bayésienne et algorithmes mcmc. Mémoire, IMAT, 2007.

[2] G. E. Dicka. Tarification de traités de réassurance non proportionnels dans le cadre du régime des catastrophes naturelles français. Mémoire, ISUP, 2013.

[3] V. Goulet. Mathématiques actuarielles IARD II (théorie de la crédibilité). Mémoire, Université de Laval, 2010.

[4] A. G. H. Buhlmann. A Course in Credibility Theory and its Applications. Springer, 2005.

[5] F. Lévêque. Le risque d'accident nucléaire majeur : calcul et perception des probabilités, 2013.

[6] G. Mao. Modélisation du risque automobile pour le péril inondation dans le cadre du régime français d'indemnisation des catastrophes naturelles. application à la tarification. Mémoire, ISFA, 2014.

[7] K. Murphy. Conjugate bayesian analysis of the gaussian distribution, 2007.

[8] A. Quantin. Modélisation du péril inondation : passage d'une approche déterministe à une approche probabiliste en vue d'une tarification à l'exposition des couvertures de réassurance dans le cadre du régime des catastrophes naturelles. Mémoire, CEA, 2011.

[9] P. Therond. Théorie de la crédibilité, 2015. http ://www.therond.fr/wp-content/uploads/cours/Credibilite.pdf.

[10] Z. Wang. Tarification des couvertures de réassurance des catastrophes naturelles en excédent de sinistre par risque, comparaison des approches à l'exposition et à l'expérience. Mémoire, EURIA, 2012.

[11] Q. Y. Zhu. Modèles bayésiens et application à l'estimation des caractéristiques de produits finis et au contrôle de la qualité. Thèse, Ecole Nationale des Ponts et Chaussées, 1191.

Chapitre 11

Annexes

11.1 Annexe A : Tests statistiques

Les explications ci-dessous sont tirées de "http ://www.stat.ualberta.ca/ hooper/teaching/misc/P-value.pdf"

Les tests statistiques utilisés dans cette étude permettent de tester une hypothèse nulle contre une hypothèse alternative. L'hypothèse alternative correspond à ce que nous pensons qui serait vrai dans le cas où l'hypothèse nulle serait fausse. Ces hypothèses concernent un modèle statistique pour les données en question.

Il est important de tenir en compte que les tests ne permettent pas de prouver qu'une hypothèse est vraie mais uniquement de démontrer qu'une hypothèse est plus plausible que l'autre en supposant qu'au moins une des hypothèses est vraie. Il est donc important de bien choisir l'hypothèse alternative. En effet, si nous testons une hypothèse nulle contre une hypothèse alternative qui est fausse, les résultats du tests vont forcément indiquer que l'hypothèse nulle est plus plausible mais cela voudra uniquement dire que l'hypothèse alternative est très peu probable et le test n'apportera aucune information concernant l'hypothèse nulle.

Dans le cadre de notre étude, nous avons testé les hypothèses suivantes :

(H_0) : L'échantillon suit une loi Log-Normale

(H_1) : L'échantillon ne suit pas une loi Log-Normale

Nous avons également testé les hypothèses :

(H_0) : L'échantillon suit une loi Gamma

(H_1) : L'échantillon ne suit pas une loi Gamma

Dans le cadre de notre étude, nous voulions démontrer que l'hypothèse que nos données suivent une loi était plus plausible que l'hypothèse que nos données ne suivent pas cette loi. Pour cela, nous avons utilisé deux tests : le test de Kolmogorov-Smirnov et le test de Cramér-von Mises.

De manière générale, les tests renvoient une probabilité qui permet de quantifier la preuve contre l'hypothèse nulle, appelée *p-value*. Pour cela, les tests se basent sur des statistiques (chaque test calcule une statistique différente). Les tests permettent ainsi d'évaluer la statistique calculée sur les données par rapport à la probabilité de trouver une telle statistique si l'hypothèse nulle était vraie.

La *p-value* peut être interprétée de la façon suivante : supposons que l'hypothèse nulle est vraie, et que nous simulions de nouvelles données de la même manière que nous avons simulé notre jeu de données. Si nous calculions la statistique sur ce nouveau jeu de données, la *p-value* serait la probabilité que cette statistique soit plus éloignée dans la queue de distribution que la statistique de notre jeu de données.

Le tableau suivant montre un example d'interprétation des *p-value* :

$p > 0{,}10$	Les données semblent être plus cohérentes avec l'hypothèse nulle. Il n'y a pas de preuve contre l'hypothèse nulle.
$0{,}05 < p < 0{,}10$	La preuve contre l'hypothèse nulle est très faible.
$0{,}01 < p < 0{,}05$	Il y a des preuves modérées contre l'hypothèse nulle et en faveur de l'hypothèse alternative.
$0{,}001 < p < 0{,}01$	La preuve contre l'hypothèse nulle est forte.
$p < 0{,}001$	Les données semblent être plus cohérentes avec l'hypothèse alternative. La preuve contre l'hypothèse nulle est très forte.

TABLE 11.1 – Interprétation des p-values

11.1.1 Test de Kolmogorov-Smirnov

Le test de Kolmogorov-Smirnov permet de tester les hypothèses à partir des lois de répartition théorique et observée. Pour cela, les paramètres de la loi doivent être spécifiés. Ces derniers ont été

estimés lors de l'étude à partir de la méthode du maximum de vraisemblance et de la méthode des moments.

Nous avons également étudié la fonction de répartition observée donnée par la formule suivante :

$$F(x) = \frac{1}{n} \sum_{k=1}^{n} 1_{X_k \leqslant x}$$

Soit F_0 la fonction de répartition théorique, la statistique qui définit ce test est la suivante :

$$\Delta = \sup_{X \in \mathbb{R}} \mid F(x) - F_0(x) \mid$$

Sous l'hypothèse nulle, $\sqrt{n}\Delta$ converge vers la distribution de Kolmogorov : la région critique ou zone de rejet est définie pour $\sqrt{n}\Delta > K_\alpha$ où K_α est le quantile de la distribution de Kolmogorov au niveau $1 - \alpha$.

La région critique ou zone de rejet du test se constitue donc des grandes valeurs de Δ. Ce test permet ainsi d'évaluer la probabilité d'obtenir un tel écart entre les fonctions de répartition empirique et théorique lorsqu'on tire au sort des valeurs à partir de la loi statistique testée.

11.1.2 Test de Cramér-von Mises

Le test de Cramér-von Mises est équivalent au test de Kolmogorov-Smirnov pour les petits échantillons. Il se base également sur la fonction de répartition, mais sur une statistique différente.

Si nous supposons que les paramètres de la loi sont connus, la statistique est la suivante :

$$I = \int_{\mathbb{R}} [F(x) - F_0(x)]^2 dF(x)$$

Nous remarquons que ce test permet de prendre en compte tous les points de la distribution alors que le test de Kolmogorov-Smirnov prend uniquement en compte l'écart maximal. Ainsi, le test de Cramér-von Mises est moins sensible à la présence de points aberrants.

11.1.3 Exemples

L'exemple suivant permet d'illustrer les résultats. Nous avons d'abord simulé deux échantillons le premier à partir d'une loi Normale centrée réduite et le deuxième à partir d'une loi Gamma de paramètres *shape* = 1 et *scale* = 2.

Sur chacun de ses paramètres nous avons testé, à l'aide des deux tests, les hypothèses suivantes :

(H_0) : L'échantillon suit une loi Normale centrée réduite

(H_1) : L'échantillon ne suit pas une loi Normale centrée réduite

Le code `R` est le suivant :

```
param1 = 0
param2 = 1
echantillon1 = rnorm(1000, param1, param2)
echantillon2 = rgamma(1000, 1, 2)
ks.test(echantillon1, "pnorm", param1, param2)
ks.test(echantillon2, "pnorm", param1, param2)
goftest::cvm.test(echantillon1, "pnorm", param1, param2)
goftest::cvm.test(echantillon2, "pnorm", param1, param2)
```

Dans le cas du premier échantillon nous obtenons des statistiques proches de 0 avec les deux tests, les écarts entre les fonctions de répartition empirique et théorique sont donc faibles. Par ailleurs, les *p-values* sont élevées donc la probabilité d'obtenir de telles statistiques sous l'hypothèse nulle est proche de 1.

```
        One-sample Kolmogorov-Smirnov test
data:  echantillon1
D = 0.0047, p-value = 0.982
alternative hypothesis: two-sided

        Cramer-von Mises test of goodness-of-fit
        Null hypothesis: Normal distribution
data:  echantillon1
omega2 = 0.0327, p-value = 0.9666
```

Au contraire dans le cas du deuxième échantillon, le test de Kolmogorov-Smirnov renvoie une *p-value* faible, l'hypothèse nulle devrait donc être rejetée. De manière similaire, le test de Cramér-von Mises renvoie une statistique assez élevée, donc même si la *p-value* n'est pas proche de 0, l'hypothèse nulle devrait être rejetée parce que l'écart entre les deux fonctions de répartition est importat.

```
        One-sample Kolmogorov-Smirnov test
data:  echantillon2
D = 0.5002, p-value < 2.2e-16
alternative hypothesis: two-sided

        Cramer-von Mises test of goodness-of-fit
        Null hypothesis: Normal distribution
data:  echantillon2
omega2 = 555.1386, p-value = 0.1332
```

Cet exemple montre l'importance d'analyser le couple renvoyé par le test et non pas la *p-value* indépendamment de la statistique.

11.2 Annexe B : Généralités sur la théorie bayésienne

Cette annexe permet de présenter les principes de la théorie bayésienne afin de mieux comprendre les différentes étapes. Pour plus d'information, il est possible de consulter les sources [1], [3], [4], [7] et [11] dans la bibliographie.

11.2.1 Raisonnement proportionnel

Nous allons considérer la fonction de vraisemblance $L(\theta, \mathbf{x}) = f(X|\theta)$, où $\mathbf{x}$ est un échantillon de la variable aléatoire X. Cette fonction conserve toute l'information issue de l'échantillon $\mathbf{x}$ par rapport au paramètre θ. Dans cette étude le but est de conserver l'information des deux échantillons que nous souhaitons crédibiliser afin de déterminer un paramètre à partir des informations issues des deux échantillons.

Dans le but de simplifier les calculs, nous allons introduire la notion de noyau de la fonction de vraisemblance. Les principal but est de conserver la totalité de l'information issue de chaque échantillon, une statistique permettant de conserver la totalité de cette information est qualifiée d'exhaustive.

Soit X un échantillon aléatoire de paramètre θ. Une statistique T est exhaustive si la distribution : $\mathbb{P}_t = \mathbb{P}_\theta[\cdot|T(X) = t]$ est indépendante du paramètre θ.

S'il existe une statistique exhaustive T telle que la fonction de vraisemblance puisse s'écrire :

$$L(\theta, \mathbf{x}) = h(\mathbf{x}) \cdot g(T(\mathbf{x})|\theta)$$

Alors $g(T(\mathbf{x})|\theta)$ est le noyau de la fonction de vraisemblance. On peut alors écrire à l'aide du symbole proportionnel $\propto$:

$$L(\theta, \mathbf{x}) \propto g(T(\mathbf{x})|\theta)$$

Le noyau contient alors l'intégralité de l'information concernant le paramètre θ issue de l'échantillon $\mathbf{x}$. Cette notion permet de simplifier les calculs au niveau du choix de la loi a priori.

11.2.2 Loi Normale-Normale

Nous rappelons ici que le paramètre crédibilisé est l'espérance, ainsi :

$$\begin{cases} X_i \sim \text{Normale}(\mu, \sigma^2) & \forall i \in [1, ..., n] \\ \\ \mu \sim \text{Normale}(m, v) \end{cases}$$

Fonction de vraisemblance

Nous souhaitons crédibiliser l'espérance à partir de l'échantillon $\mathbf{y}$, pour cela il est nécessaire de calculer la fonction de vraisemblance $f(\mathbf{y}|\mu)$. Si nous reprenons l'expression de la fonction de vraisemblance nous obtenons :

$$f(\mathbf{y}|\mu) = \frac{1}{\sqrt{2\pi\sigma^2}^k} \cdot \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{k}(y_i - \mu)^2\right)$$

Or nous avons :

$$\begin{aligned} \sum_{i=1}^{k}(y_i - \mu)^2 &= \sum_{i=1}^{k}((y_i - \bar{y}) - (\mu - \bar{y}))^2 \\ &= \sum_{i=1}^{k}\left[(y_i - \bar{y})^2 + (\mu - \bar{y})^2 - 2 \cdot (y_i - \bar{y}) \cdot (\mu - \bar{y})\right] \\ &= \sum_{i=1}^{k}(y_i - \bar{y})^2 + \sum_{i=1}^{k}(\mu - \bar{y})^2 - 2 \cdot (\mu - \bar{y}) \cdot \sum_{i=1}^{k}(y_i - \bar{y}) \\ &= \sum_{i=1}^{k}(y_i - \bar{y})^2 + \sum_{i=1}^{k}(\mu - \bar{y})^2 - 2 \cdot (\mu - \bar{y}) \cdot (k \cdot \bar{y} - k \cdot \bar{y}) \\ &= \sum_{i=1}^{k}(y_i - \bar{y})^2 + \sum_{i=1}^{k}(\mu - \bar{y})^2 \end{aligned}$$

Prenons les notations suivantes : $\begin{cases} \bar{y} = \frac{1}{k}\sum_{i=1}^{k} y_i \\ \\ s^2 = \frac{1}{k}\sum_{i=1}^{k}(y_i - \bar{y})^2 \end{cases}$

Alors nous obtenons :

$$\sum_{i=1}^{k}(y_i-\mu)^2 = k\cdot s^2 + k\cdot(\mu-\bar{y})^2$$

En remplaçant nous obtenons la fonction de vraisemblance suivante :

$$\begin{aligned} f(\mathbf{y}|\mu) &= \frac{1}{\sqrt{2\pi\sigma^2}^k}\cdot\exp\left(-\frac{ks^2+k(\mu-\bar{y})^2}{2\sigma^2}\right) \\ &\propto \exp\left(-\frac{(\mu-\bar{y})^2}{2\sigma^2/k}\right) \end{aligned}$$

Loi conjuguée

Nous rappelons que le but est de montrer que la loi conjuguée à la loi Normale est aussi une loi Normale lorsque l'on crédibilise l'espérance, pour cela il est nécessaire de déterminer la forme de $f(\mathbf{y}|\mu)\cdot\pi(\mu)$. Prenons la loi a priori suivante :

$$\mu \sim \text{Normale}(m, v)$$

Alors nous avons :

$$\pi_{\mathbf{y}}(\mu) \propto f(\mathbf{y}|\mu)\cdot\pi(\mu) \propto \exp\left(-\frac{(\mu-\bar{y})^2}{2\sigma^2/k}\right)\times\exp\left(-\frac{(\mu-m)^2}{2\cdot v}\right)$$

Si nous développons, nous obtenons :

$$\begin{aligned} \pi_{\mathbf{y}}(\mu) &\propto \exp\left(-\frac{k}{2\sigma^2}\cdot\mu^2+\frac{k}{2\sigma^2}\cdot 2\mu\bar{y}-\frac{k}{2\sigma^2}\cdot\bar{y}^2-\frac{1}{2v}\cdot\mu^2+\frac{1}{2v}\cdot 2\mu m-\frac{1}{2v}\cdot m^2\right) \\ &\propto \exp\left(-\frac{\mu^2}{2}\left(\frac{k}{\sigma^2}+\frac{1}{v}\right)+\mu\left(\frac{k\bar{y}}{\sigma^2}+\frac{m}{v}\right)-\frac{1}{2}\left(\frac{k\bar{y}^2}{\sigma^2}+\frac{m^2}{v}\right)\right) \\ &\propto \exp\left[-\frac{1}{2}\left(\mu^2\left(\frac{\sigma^2+vk}{\sigma^2+v}\right)-2\mu\left(\frac{k\bar{y}}{\sigma^2}+\frac{m}{v}\right)+\left(\frac{k\bar{y}^2}{\sigma^2}+\frac{m^2}{v}\right)\right)\right] \end{aligned}$$

Prenons les notations suivantes : $\begin{cases} \frac{1}{A} = \frac{\sigma^2 + vk}{\sigma^2 + v} \\ B = \frac{k\bar{y}}{\sigma^2} + \frac{m}{v} \\ C = \frac{k\bar{y}^2}{\sigma^2} + \frac{m^2}{v} \end{cases}$

Alors nous obtenons une loi a posteriori telle que :

$$\begin{aligned} \pi_{\mathbf{y}}(\mu) &\propto \exp\left[-\frac{1}{2}\left(\frac{\mu^2}{A} - 2\mu B + C\right)\right] \\ &\propto \exp\left[-\frac{1}{2A}\left(\mu^2 - 2\mu BA + (BA)^2\right) - \frac{1}{2A}\left(CA - (BA)^2\right)\right] \\ &\propto \exp\left[-\frac{1}{2A}\left(\mu - BA\right)^2\right] \end{aligned}$$

Il s'agit ici du noyau de la loi Normale. La conjuguée est donc une loi Normale puisqu'elle permet de conserver la même forme entre la loi a priori et la loi a posteriori.

Table des figures

2.1 Nombre de reconnaissances de catastrophes naturelles 10
2.2 Historique des périls pris en charge par le régime des catastrophes naturelles 12
2.3 Répartition de la charge du régime catastrophes naturelles par péril 13

3.1 Répartition du chiffre d'affaires par activité 15
3.2 Application des traités 17
3.3 Exemple d'application des traités 18

4.1 Modélisation physique des catastrophes naturelles 20
4.2 Exemple d'aléa d'inondation modélisé sur Paris 23
4.3 Coût moyen des inondations pour les différentes communes 25
4.4 Coût moyen des sécheresses pour les différentes communes 28
4.5 Aléa des séismes 30

5.1 Schéma de la tarification par expérience 36
5.2 Exemple d'exposition associé à la localisation des polices 37
5.3 Schéma de la tarification par exposition 40

7.1 Théorie bayésienne à partir d'une loi Normale-Normale 52
7.2 Théorie bayésienne à partir d'une loi Normale-Gamma 56
7.3 Théorie bayésienne à partir d'une loi Gamma-Gamma 60

8.1 Ajustement des lois à l'historique des inondations 63
8.2 Ajustement des lois à l'exposition des inondations 64
8.3 Ajustement des lois à l'historique des sécheresses 65
8.4 Ajustement des lois à l'exposition des sécheresses 66
8.5 Ajustement des lois à l'historique des contrats automobile 66
8.6 Ajustement des lois à l'exposition des contrats automobile 67

8.7 Paramètres a priori et a posteriori du péril inondation 71
8.8 Paramètres a priori et a posteriori du péril sécheresse 74
8.9 Paramètres a priori et a posteriori de la sinistralité automobile 77
8.10 Facteurs de crédibilité 81
8.11 Niveaux de retour des inondations 86
8.12 Niveaux de retour des sécheresses 87
8.13 Niveaux de retour de la sinistralité automobile 88

9.1 Schéma du modèle hybride pour une cédante 93
9.2 Sinistralité à la charge de la CCR (M€) 94
9.3 Niveaux de retour de la sinistralité à la charge de la CCR (M€) 96
9.4 Sinistralité à la charge du Stop-Loss de la cédante A (M€) 99
9.5 Niveaux de retour de la charge du Stop-Loss de la cédante A(M€) 100

Liste des tableaux

4.1 *Event Loss Table* 21
4.2 *Year Loss Table* 21

8.1 Statistiques des données 62
8.2 Résultats des tests sur l'historique des inondations 64
8.3 Résultats des tests sur l'historique des sécheresses 65
8.4 Résultats des tests sur l'historique des contrats automobile 67
8.5 Loi a priori de chaque paramètre crédibilisé 68
8.6 Estimateurs des hyper-paramètres a priori 68
8.7 Estimateurs des hyper-paramètres a posteriori 69
8.8 Loi a posteriori de chaque paramètre crédibilisé 69
8.9 Hyperparamètres a priori du péril inondation 70
8.10 Hyperparamètres a posteriori du péril inondation 70
8.11 Statistiques de la sinistralité crédibilisée pour le péril inondation 71
8.12 Niveaux de retour de la sinistralité du péril inondation (M€) 72
8.13 Hyperparamètres a priori du péril sécheresse 73
8.14 Hyperparamètres a posteriori du péril sécheresse 73
8.15 Statistiques de la sinistralité crédibilisée pour le péril sécheresse 74
8.16 Niveaux de retour de la sinistralité pour le péril sécheresse (M€) 75
8.17 Hyperparamètres a priori de l'automobile 75
8.18 Hyperparamètres a posteriori de l'automobile 76
8.19 Statistiques de la sinistralité crédibilisée pour l'assurance automobile 77
8.20 Niveaux de retour de la sinistralité pour l'assurance automobile (M€) 78
8.21 Statistiques de la sinistralité crédibilisée pour le péril inondation 82
8.22 Niveaux de retour de la sinistralité pour le péril inondation (M€) 82
8.23 Statistiques de la sinistralité crédibilisée pour le péril sécheresse 83
8.24 Niveaux de retour de la sinistralité pour le péril sécheresse (M€) 84

8.25 Statistiques de la sinistralité crédibilisée pour l'assurance automobile 84
8.26 Niveaux de retour de la sinistralité pour l'assurance automobile (M€) 85

9.1 Quantiles extraits des modèles d'exposition et historique 92
9.2 Sinistralité à la charge de la CCR (M€) 95
9.3 Niveaux de retour de la sinistralité de la CCR (M€) 97
9.4 Sinistralité à la charge du Stop-Loss de la cédante A (M€) 99
9.5 Niveaux de retour de la sinistralité brute de la cédante (M€) 100

11.1 Interprétation des p-values 106