

SCOR Paper

PRIX DE L'ACTUARIAT

Avril
2018
T
N°41

Modélisation du risque gel en France

Par **Laura Cohen et
Dulcy Gninghaye**
Lauréates du Prix français
de l'actuariat 2015
ENSAE ParisTech

SCOR
The Art & Science of Risk



Modélisation du risque gel en France

Laura Cohen et Dulcy Gninghayé
ENSAE ParisTech

Résumé

Ces dernières années, les répercussions économiques des catastrophes naturelles sur les populations se sont clairement aggravées, compromettant la solidité financière des sociétés d'assurance. Il est par conséquent nécessaire pour les assureurs de mesurer le risque sous-jacent ainsi que les pertes associées pour leurs clients. Toutefois, en raison de la nature imprévisible des catastrophes naturelles, des modèles catastrophes spécifiques doivent être mis en place afin de permettre aux assureurs de mieux comprendre ces risques et leurs conséquences en termes d'assurance.

La présente étude a pour objectif de construire un modèle catastrophe pour le péril gel en France. Compte tenu que le gel dépend directement de la température observée, la première partie consiste à réduire la quantité d'informations à disposition en définissant des régions homogènes en température afin de pouvoir modéliser chacune d'entre elles comme un tout. Dans la deuxième partie, le gel a été modélisé à partir de données historiques, permettant une simulation stochastique des futurs événements possibles en France. Après avoir procédé à un calibrage sur mesure des courbes de vulnérabilité, nous avons pu déduire les pertes économiques causées par des événements simulés sur le portefeuille d'AXA. Enfin, les résultats associés à chaque événement ont été regroupés afin de déterminer la distribution des pertes liées aux événements gel en France.

Un texte paraissant dans SCOR Papers n'engage que son/ses auteur(s). En publiant un article dans SCOR Papers, SCOR ne prend pas position au sujet des opinions exprimées par son auteur et dégage toute responsabilité pour les informations inexactes, erreurs de droit et opinions émises dans SCOR Papers par l'/les auteur(s).

1. Introduction

L'entrée en vigueur prochaine de Solvabilité II impose à tous les assureurs d'avoir la meilleure connaissance possible de l'ensemble des risques auxquels ils s'exposent. Dans cette directive, un traitement spécifique est notamment prévu pour les catastrophes naturelles dont les impacts peuvent s'avérer dramatiques pour les compagnies d'assurance. A titre d'exemple, l'ouragan Andrew survenu en 1992 en Floride, a causé la faillite de onze assureurs.

Lorsque l'on évoque les catastrophes naturelles, il est rare de mentionner le risque gel car sa survenance n'a généralement pas un impact aussi spectaculaire que les tremblements de terre, les ouragans ou encore les éruptions volcaniques. Or, la durée d'un événement gel ainsi que la forte accumulation des polices touchées peuvent entraîner de lourdes pertes comme ce fut le cas pour AXA en 2012, année particulièrement froide. L'objet de la présente étude est de construire un modèle catastrophe pour le péril gel qui permettra de simuler la survenance de ce risque et les pertes associées.

Cela sera effectué en quatre étapes. Premièrement, nous effectuerons une classification des régions de France en fonction de leurs températures afin de réduire la quantité d'information à disposition. Ensuite, dans les deux parties suivantes, nous détaillerons la construction de deux modules d'un modèle catastrophe, en les adaptant spécifiquement au péril gel. Le premier module permettra de générer des événements de gel en France sur une année et le deuxième module aura pour objectif de calculer l'impact de ces événements sur le portefeuille d'AXA, traduisant les événements physiques en pertes économiques. Enfin, dans la dernière partie, nous étudierons la distribution des pertes causées par le gel obtenue en sortie du modèle.

2. La modélisation catastrophe

Généralement, pour évaluer la distribution des pertes liées à un risque spécifique en assurance non vie, l'approche « fréquence-coûts » est utilisée. Cependant, cette approche n'est pas adaptée aux risques catastrophes puisque l'hypothèse sous-jacente est que la distribution des événements passés est représentative de la distribution des événements futurs. Or, les catastrophes naturelles sont par définition des événements extrêmes et rares, dépendants de phénomènes physiques, qui ne peuvent être modélisés par cette approche statistique traditionnelle.

Afin d'estimer plus précisément les risques catastrophes, il est donc nécessaire d'utiliser des approches par exposition plutôt que par historique. Les modèles catastrophes combinent la représentation mathématique de la survenance des sinistres avec les caractéristiques scientifiques du risque assuré pour générer une distribution de pertes.

Les modèles catastrophes sont devenus indispensables car ils permettent aux assureurs de :

- tarifier les contrats de réassurance pour optimiser le transfert de risque à un réassureur ;
- contrôler et diversifier les risques ;
- estimer les provisions à constituer pour faire face à une perte ;
- minimiser le capital requis par Solvabilité 2 ;
- anticiper les catastrophes naturelles et prévenir les risques pour augmenter la résilience.

Un modèle catastrophe est structuré en trois modules indépendants : le module Aléa, le module Vulnérabilité et le module Financier.

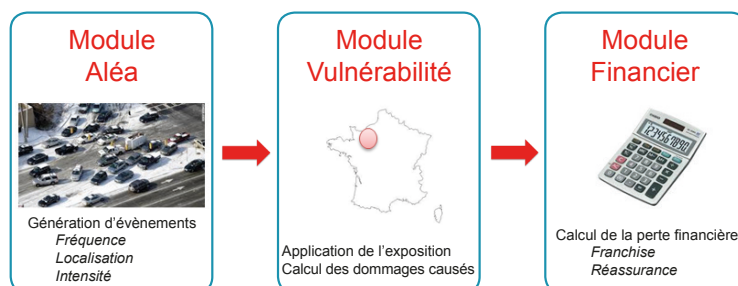


Figure 1 Structure d'un modèle catastrophe

Le module **Aléa**, élément central dans tout modèle catastrophe, a pour objectif de reproduire le phénomène physique sous-jacent au péril. Dans le cas du péril gel, nous chercherons à reproduire des phénomènes physiques liés à la température. Cette étude se focalisera principalement sur la construction de ce module qui regroupe les principales difficultés de la modélisation catastrophe.

Le module Aléa ayant constitué un dictionnaire conséquent d'évènements physiques, le module **Vulnérabilité** permet ensuite de quantifier la sinistralité de chaque évènement à travers une estimation de la perte.

Le module **Financier** a pour objectif de déterminer les pertes finales de l'assureur nettes des conditions assurantielles (caractéristiques de chaque contrat souscrit, conditions de réassurance et part de la coassurance). Pour des raisons de confidentialité, ce module n'a pas été modélisé.

3. La classification des régions

Le gel dépendant directement de la température observée, différents scénarios de température doivent être simulés afin de pouvoir en déduire la survenance d'évènements gel. Pour cela, nous disposons des données historiques (sur 23 193 jours) de températures journalières minimales relevées en 250 points localisés dans toute la France. Or, modéliser la température en chacun de ces 250 points serait fastidieux. Il est donc nécessaire de réduire cette quantité d'informations dans l'espace et dans le temps pour obtenir des régions de France homogènes en température.

A cet effet, les méthodes statistiques d'analyse factorielle de données et/ou de classification sont appropriées. Dans certains cas, quelques auteurs recommandent d'appliquer une méthode d'analyse factorielle puis une classification sur les résultats de la première méthode. Cela a l'avantage d'affiner l'information et les résultats obtenus.

Dans un premier temps, nous appliquerons la méthode Empirical Orthogonal Function (EOF) qui se rapproche d'une Analyse en Composantes Principales (ACP), mais plus indiquée pour des données spatio-temporelles. Dans un second temps, nous effectuerons une méthode de classification sur les premières coordonnées principales obtenues avec la méthode EOF.

3.1. L'Empirical Orthogonal Function (EOF)

La méthode EOF vise à déterminer un nouvel échantillon de variables qui permettent de capturer une partie importante de l'information observée sur les données à travers une combinaison linéaire des variables initiales. Elle est généralement utilisée par les météorologues.

Après application de la méthode EOF, on obtient une matrice de données contenant les 250 points initiaux et 250 axes principaux en colonnes. Les axes principaux ont été obtenus par combinaison linéaire des variables temporelles initiales. Le premier axe contient plus d'information que le second, et le second plus que le troisième et ainsi de suite. Compte tenu de la décroissance de l'information contenue dans chaque axe, il convient de focaliser l'analyse sur les premiers axes principaux.

La figure suivante présente l'éboulis des valeurs propres ainsi que les pourcentages d'inertie associés aux dix premières composantes principales.

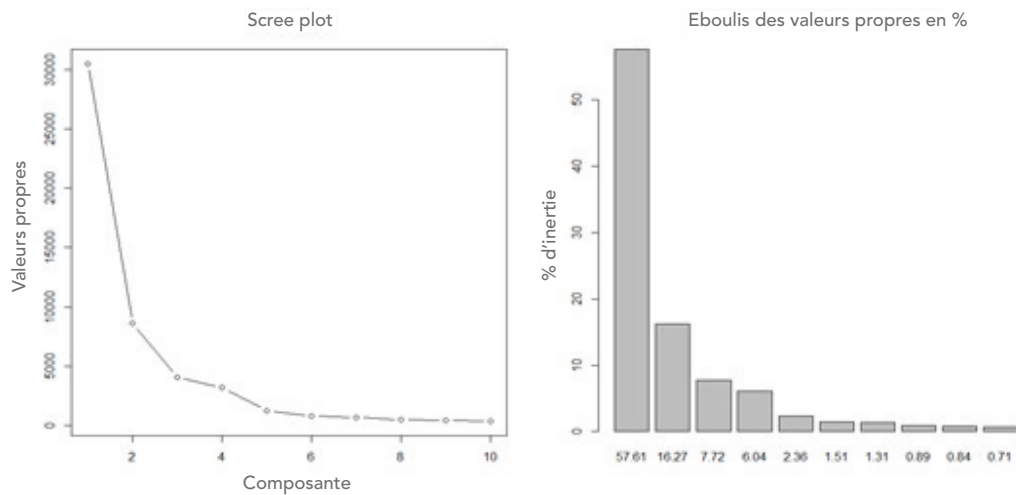


Figure 2 Eboulis des valeurs propres

A la vue du premier graphique, la règle du coude revient à choisir les 4 premières valeurs propres. A l'aide du second graphique, on constate que l'inertie expliquée par les 4 premiers axes principaux est de 87,63 %.

Afin de décrire rapidement les grandes tendances associées à ces quatre axes principaux, nous représentons sur la France les points ayant une contribution significative à la formation de chaque axe principal considéré. La Figure 3 permet de mettre en évidence différentes régions qui sont mises en opposition sur les différents axes.

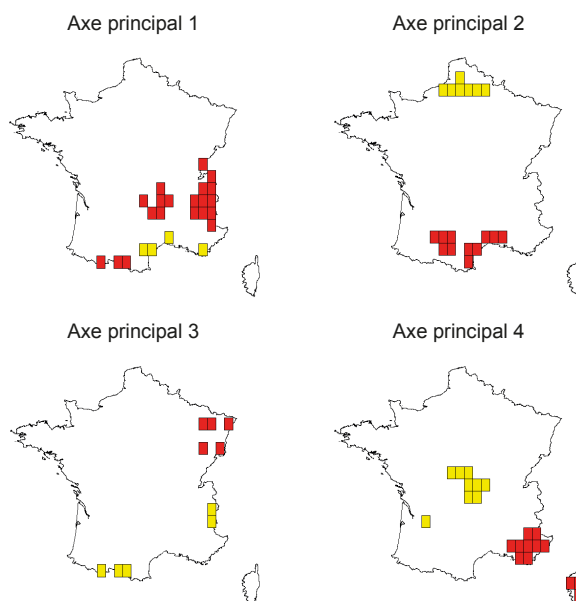


Figure 3 Points géographiques les mieux représentés sur les 4 premiers axes principaux

Sur le premier axe factoriel, nous notons une opposition entre les massifs montagneux (points en rouge) et les régions ayant un climat méditerranéen et des hivers plutôt doux (points en jaune). Le second axe factoriel met clairement en évidence une opposition entre les points du Nord avec un climat océanique et les points du Sud ayant un climat plus méditerranéen. Sur le troisième axe factoriel, on a une opposition entre les régions avec un climat sous influence montagnarde (points en jaune) et les zones ayant plutôt un climat sous influence semi-continentale (points en rouge). Enfin, le quatrième axe factoriel fait apparaître une opposition entre les zones ayant un climat méditerranéen, notamment sur la Corse et la côte d'azur (points en rouge) et le massif central (points en jaune).

3.2. La classification

A la suite de la méthode EOF, nous obtenons une matrice représentant nos 250 localisations géographiques sur les 4 premiers axes principaux retenus. Une Classification Ascendante Hiérarchique (CAH) est alors appliquée à ces points en utilisant le critère de Ward.

Le dendrogramme issu de la CAH est ci-après présenté :

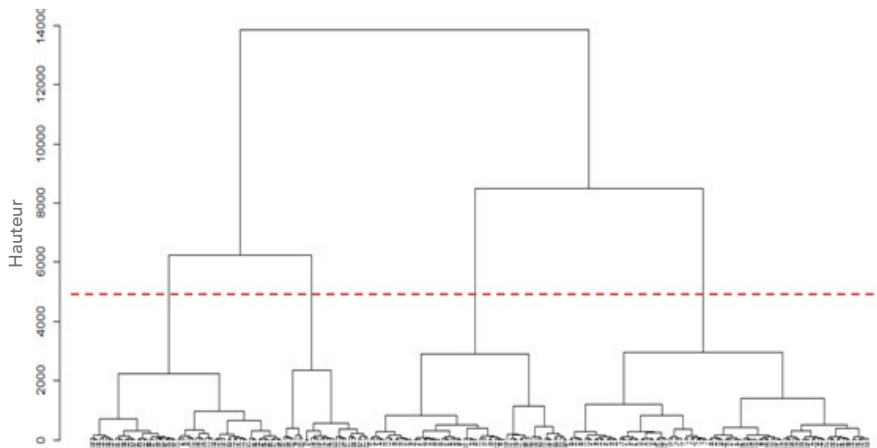


Figure 4 Dendrogramme associé à la Classification Ascendante Hiérarchique (CAH) sur les coordonnées des 4 premiers axes principaux de l'EOF

Suivant le niveau où le dendrogramme est « coupé », différentes classes sont obtenues. En considérant le critère de parcimonie, nous souhaiterions avoir le plus petit nombre de classes possible. Un critère supplémentaire afin de s'assurer de l'homogénéité des températures pour chaque région consiste à opter pour une variance intraclasse faible, soit un écart type intrarégional inférieur ou proche de 2° Celsius.

Ces différents critères nous conduisent à sélectionner quatre classes ci-après représentées.



Figure 5 Segmentation de la France en 4 régions (résultats de la CAH à la suite de l'EOF)

La première classe, en vert, regroupe une partie du Sud de la France. Elle est caractérisée par des villes ayant un climat tempéré avec des hivers plutôt doux. La zone au Sud-Ouest de la France est caractérisée par un climat océanique aquitain. On y retrouve des villes comme Bordeaux ou Limoges. Le reste des localisations géographiques dans cette première région est caractérisé par un climat méditerranéen. On y retrouve des villes telles que Nice, Marseille, Perpignan, Toulon et toute la Corse.

La deuxième classe, en rouge, englobe le Nord-Ouest de la France et prend en compte les zones sous influence du climat océanique. On y retrouve des villes comme Paris, Lille, Caen, Rennes.

La troisième classe, en violet, regroupe le Nord-Est de la France et une partie du Centre. Cette zone est sous influence d'un climat semi-continental et contient des régions d'altitude moyenne. On y retrouve des villes telles que Strasbourg, Nancy, Dijon. Elle contient également un point au Sud-Est de la France, qui correspond à une région d'altitude moyenne proche des Pyrénées.

La quatrième classe, en bleu cyan, caractérise des zones sous influence montagnarde. On y retrouve des villes telles que Saint-Etienne et Grenoble.

4. Construction du module Aléa

Le module Aléa vise à simuler plusieurs scénarios de gel en France. La classification effectuée dans la partie précédente a permis de réduire la dimension du problème pour pouvoir simuler des scénarios de gel dans chaque région, et donc uniquement en quatre points.

La construction du module aléa se fera en trois étapes :

- la caractérisation d'un évènement gel, fortement dépendant des températures ;
- la modélisation statistique de la température régionale en tenant compte des corrélations inter-régions ;
- la simulation de plusieurs scénarios de température afin de déduire les évènements de gel.

4.1. Caractérisation de l'évènement gel

Physiquement, le gel est dû à des températures négatives. Or, la survenance de températures négatives ne génère pas forcément de sinistres pour l'assureur. Ce qui nous intéresse est un évènement gel au sens assurantiel, c'est-à-dire un évènement gel qui engendre des pertes pour un assureur. Cela complexifie donc la caractérisation du phénomène.

Il est intuitif de supposer que la survenance d'un évènement gel au sens assurantiel est aussi lié à la température mais il faut se questionner sur la façon dont la température influe sur ce phénomène. Par exemple, le gel serait-il lié à la température du jour d'occurrence du sinistre ou plutôt aux écarts de température observée sur une période ? Et pour quel nombre de sinistres ?

La caractérisation du gel se fera donc en deux phases notamment :

- la recherche des variables expliquant la survenance du gel ;
- la modélisation des sinistres historiques en lien avec les variables explicatives précédemment trouvées.

Dans la première phase, nous prenons comme variable à expliquer la survenance du gel ou non. La survenance d'un évènement gel est matérialisée par une déclaration de sinistres. Comme variables explicatives, nous utilisons les températures minimales journalières en chaque localisation géographique ainsi que les écarts de température sur différentes périodes. A l'aide de régressions logistiques, nous testons plusieurs combinaisons de ces variables explicatives. Nous retenons finalement les écarts de température sur 20 jours, la température minimale journalière et la région.

Dans la seconde phase, les déterminants précédemment mis en exergue sont utilisés afin de modéliser la fréquence journalière des sinistres donnée par :

$$\text{fréquence} = \frac{\text{Nombre de sinistres journaliers}}{\text{Exposition}}$$

L'exposition étant connue, cela revient à modéliser le nombre de sinistres journaliers. Dans l'approche classique de modélisation de variables de comptage, les régressions de poisson ou binomiale négative sont utilisées.

La régression de poisson adaptée pour les phénomènes rares ne permet pas de prendre en compte la surdispersion du nombre de sinistres dans notre étude. La régression binomiale négative, adaptée en cas de surdispersion, sous-estime les petits nombres de sinistres (notamment les sinistres nuls correspondant aux jours où il n'y a pas eu de déclaration de sinistres). En effet, la binomiale négative n'est pas très performante compte tenu des variations importantes d'amplitude sur le nombre de sinistres observés (minimum : 0, maximum : 700, écart type : 50 sinistres). En outre, les données présentent un nombre important de zéros. En effet, sur les 1646 observations de la table de données, 716 ont un nombre de sinistres nul.

Compte tenu de ce nombre important de zéros, nous nous tournons vers les modèles de comptage avec excès de zéros : zero inflated et hurdle. Les analyses effectuées permettent de retenir un modèle hurdle. Après un backtesting du modèle simulant le nombre de sinistres au travers d'une régression hurdle, il ressort que ce modèle sous-estime le nombre de sinistres sur l'ensemble des années. En effet, le total estimé de sinistres sur l'ensemble des années est égal à 81% du total historique de sinistres sur les années 2008 à 2012. En faisant le détail de cette estimation par année, on se rend compte que l'année 2012, qui à elle seule compte près de 76 % du nombre total de sinistres, est très largement sous-estimée dans le modèle, l'écart de prédiction étant de -72 %. Le modèle hurdle a du mal à prendre en compte ces extrêmes. Il a en fait lissé les observations : il surestime très fortement les années où il y a eu peu de sinistres et sous-estime celles où il y a eu de nombreux sinistres. Or, nous souhaitons justement capter les évènements extrêmes dans un modèle catastrophe. Cette modèle ne convient donc pas pour l'étude.

Les températures extrêmement froides et le nombre de sinistres importants observés en 2012 nous incitent à penser que les sinistres se déclenchent massivement dès qu'un seuil de température donné est atteint.

Afin de déterminer ce seuil de température, nous utilisons la mean excess function. En théorie, le seuil de température extrême est celui à partir duquel les températures peuvent être modélisées par une loi de Pareto généralisé. En pratique, le seuil choisi doit correspondre à un point de cassure dans l'allure de la mean excess function.

La Figure 6 illustre le tracé de la mean excess function pour la température moyenne régionale (tminMoy). Pour faciliter la lecture graphique de la mean excess function, nous avons tracé- tminMoy.

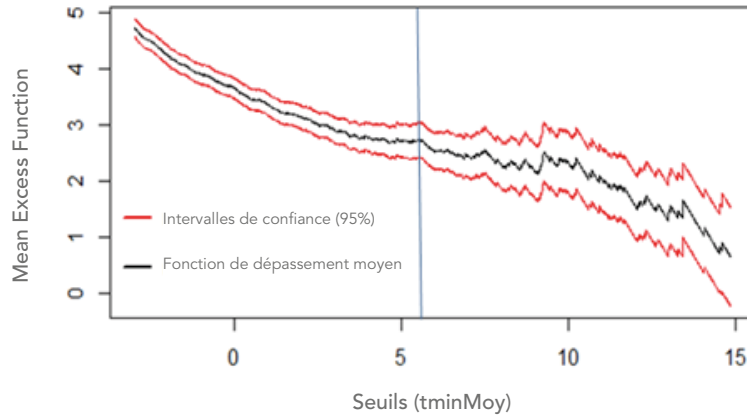


Figure 6 Mean Excess function de tminMoy

Nous fixons donc notre seuil à $-5,5^{\circ}\text{C}$, ce qui permet de partager nos données en deux sous-populations :

- ⊙ les observations pour lesquelles $-5,5 < \text{tminMoy} < 3^{\circ}$, soit 1 352 observations ;
- ⊙ les observations pour lesquelles $\text{tminMoy} \leq -5,5^{\circ}$, soit 294 observations.

Pour chaque sous-population, on cherche le modèle qui ajuste le mieux le nombre de sinistres.

Pour le premier sous-échantillon, du fait de la surdispersion du nombre de sinistres et un excès de zéros dans les données, le modèle hurdle est le mieux adapté.

Pour le second sous échantillon, les modèles de comptage classiques sont les mieux adaptés car le problème d'excès de zéros dans la base de données a disparu. Par contre, les données sont toujours surdispersées. La régression binomiale négative est la plus appropriée.

4.2. Modélisation de la température

Le gel dépendant directement des températures observées et plus particulièrement de l'amplitude des décalages de températures, il est nécessaire d'effectuer une modélisation de la température en France pour pouvoir en déduire l'apparition d'évènements gel. Cette modélisation est effectuée sur les températures moyennes journalières (depuis 1950) des quatre régions issues de la classification.

L'approche Box & Jenkins (1976) est utilisée. La série initiale de température (X_t) est décomposée de la manière suivante :

$$X_t = m_t + s_t + Y_t$$

Avec :

- ⊙ m_t : la tendance
- ⊙ s_t : la saisonnalité
- ⊙ Y_t : la série résiduelle

Dans un premier temps, nous allons tenter d'ajuster un modèle ARMA à Y_t . Un modèle ARMA(p,q) se décompose en deux processus : un processus AR (Auto Regressive) d'ordre p et un processus MA (Moving Average) d'ordre q. Soit (X_t) un processus ARMA(p,q), il vérifie l'équation suivante :

$$X_t = \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t$$

Le terme ε_t est communément appelé résidu. Pour valider ce modèle, il faut que ε_t soit un bruit blanc gaussien :

- ⦿ Les résidus doivent être stationnaires, c'est-à-dire de variance constante au cours du temps.
- ⦿ Les résidus doivent être indépendants.
- ⦿ Les résidus doivent être gaussiens.

Afin de s'assurer de la stationnarité des résidus (ε_t), nous utilisons un critère graphique. Nous traçons donc sur la Figure 7 un box plot résumant quelques caractéristiques de la série étudiée (médiane, quartiles, minimum, maximum et déciles). Pour cela, les valeurs prises par ε_t ont été regroupées par mois. On observe une dispersion des valeurs de ce processus par mois : les valeurs de ε_t sont plus dispersées pendant les mois d'hiver que pendant les mois d'été. Cela prouve donc que ε_t n'est pas un bruit blanc. Le modèle ARMA est rejeté.

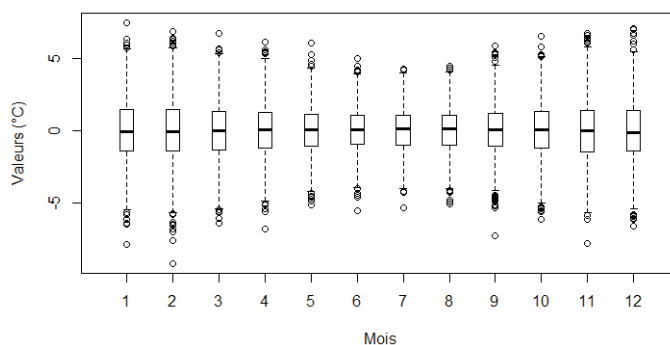


Figure 7 Dispersion des résidus par mois

Cette dispersion doit être prise en compte dans la suite de l'étude afin d'obtenir un modèle convenable. De plus, la dispersion semble saisonnière car elle est visiblement plus importante pendant les mois d'hiver. Nous testons alors un **modèle linéaire à variance périodique** pour tenir compte de la saisonnalité des résidus.

La série initiale est donc décomposée ainsi :

$$X_t = m_t + s_t + \rho_t + Y_t$$

Avec :

- ⦿ s_t : la saisonnalité
- ⦿ m_t : la tendance
- ⦿ Y_t : un processus ARMA à estimer avec $\text{Var}(Y_t) = 1$. Dans la section précédente, $\text{Var}(Y_t)$ n'était pas constant.
- ⦿ ρ_t : l'écart type de la série, supposé périodique de périodicité annuelle.

Pour ce modèle à variance périodique, l'ensemble des hypothèses (résidus stationnaires, indépendants et gaussiens) sont vérifiées en chaque région. Ce modèle a donc été retenu pour la suite de l'étude.

4.3. Modélisation de la température

Cette partie a pour but de simuler différents scénarios de température à partir des quatre séries modélisées précédemment. La température étant un phénomène homogène, la simulation dans les quatre régions ne peut s'effectuer de manière indépendante. De fait, si un pic de température est observé dans une région de France, il est très probable que les autres régions soient aussi affectées par ce pic.

Afin de tenir compte de cette dépendance géographique, nous avons recours à la théorie des copules permettant de caractériser la dépendance entre plusieurs variables aléatoires. Nous voulons déterminer la copule paramétrique la mieux adaptée à nos données et reflétant au mieux les relations de dépendance entre résidus. Il existe deux grandes familles de copules paramétriques : les copules elliptiques et les copules archimédiennes. La recherche de la copule optimale est réalisée sur un ensemble de quatre copules paramétriques sélectionnées a priori : (i) gaussienne et Student (copules elliptiques) et (ii) Gumbel et Clayton (copules archimédiennes).

Grâce à la méthode d'estimation CML (Canonical Maximum Likelihood), nous avons trouvé pour chacune des quatre copules paramétriques sélectionnées, les paramètres correspondants maximisant la vraisemblance. L'objectif à présent est de comparer les résultats obtenus avec l'utilisation de ces quatre copules avec les données de résidus empiriques. Nous sélectionnerons ainsi la copule paramétrique qui modélise au mieux la structure de dépendance des résidus.

Nous utilisons dans un premier temps le dépendogramme qui représente la structure de dépendance sous la forme d'un nuage de points des simulations d'une copule théorique. Nous comparons le dépendogramme de la copule empirique à celui des quatre copules paramétriques sélectionnées. Le dépendogramme permet d'observer aussi le caractère plus ou moins simultané des réalisations issues de l'échantillon. Dans les queues plus précisément, il sera utile d'analyser si la simultanéité est forte et donc s'il est nécessaire de calibrer sur notre échantillon une copule avec une dépendance de queue.

Nous avons représenté sur la Figure 8 les dépendogrammes associés à la copule empirique et aux copules paramétriques pour les résidus correspondants à la région 1 (région Sud) et à la région 2 (région Nord-Ouest) pour 20 000 simulations. Les résultats sont similaires pour les cinq autres couples.

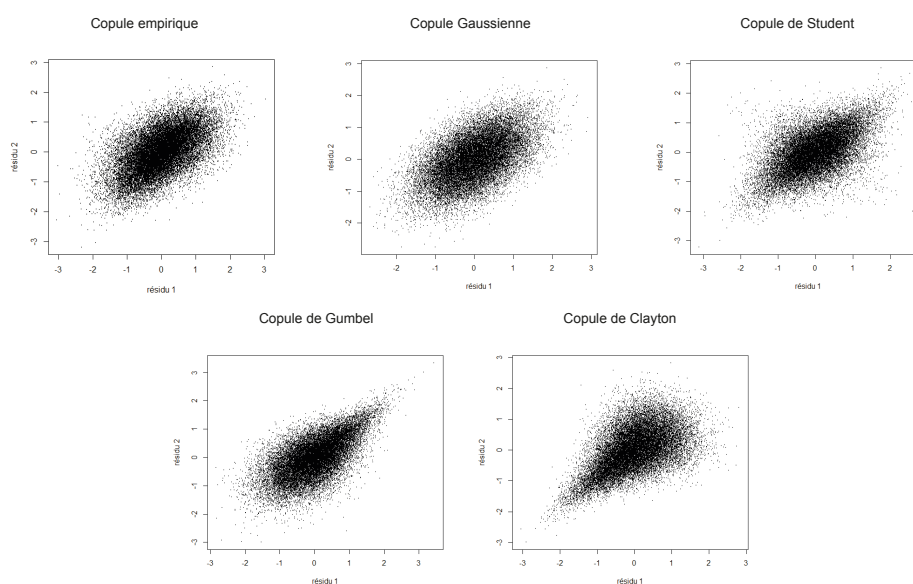


Figure 8 Dépendogrammes associés à la copule empirique et aux copules paramétriques pour les résidus 1 et 2

Les copules gaussiennes et de Student se rapprochent du nuage de points de la copule empirique et semblent convenir. Nous pouvons également noter que la copule de Student permet de mieux modéliser la dépendance des points extrêmes et semble donc mieux adaptée aux données empiriques.

Les critères AIC et BIC, prenant en compte le nombre de paramètres, confirment le choix de la copule de Student. Bien qu'elle ait le plus grand nombre de paramètres, la vraisemblance du modèle est suffisamment grande pour qu'elle ne soit pas pénalisée par son nombre élevé de paramètres.

4.4. Simulations de scénarios

L'objectif de cette section est de regrouper l'ensemble des résultats précédents pour construire des scénarios de gel sur une année en France. Pour cela, nous simulons d'abord plusieurs scénarios de température sur un an pour pouvoir en déduire ensuite des événements gel.

Scénarios de température

Nous nous servons de la modélisation des températures régionales ainsi que de la théorie des copules afin de simuler conjointement les résidus associés à chaque région. Nous simulons alors 10 000 scénarios de température pour l'année suivante. Nos données de température s'arrêtant au 1er juillet 2013, nous construisons 10 000 scénarios de température du 2 juillet 2013 au 2 juillet 2014 pour chacune des quatre régions. Les températures simulées correspondent à la moyenne des températures minimales au sein de chaque région.

La Figure 9 représente les températures prédites dans chaque région pour le premier scénario. La courbe noire représente les températures observées depuis août 2011 et la courbe rouge un scénario de températures pour l'année suivante.

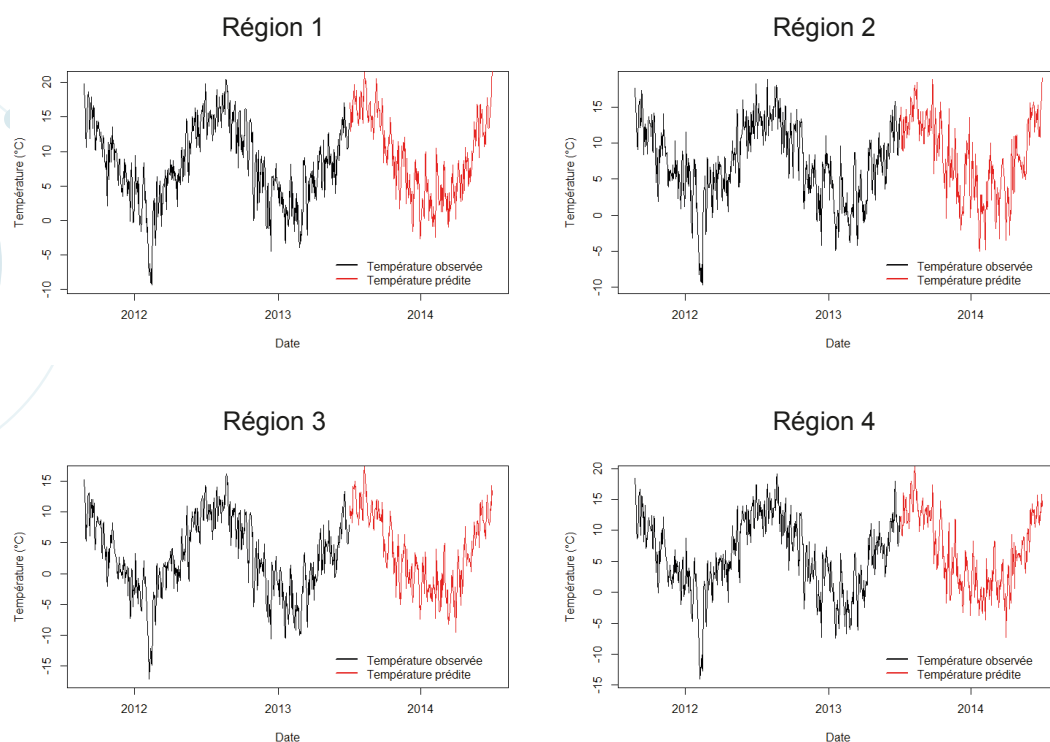


Figure 9 Exemple d'un scénario de température par région de juillet 2013 à juillet 2014

La Figure 10 représente les températures simulées dans les quatre régions. Ce graphique fait bien apparaître une corrélation entre les régions qui a été obtenue grâce à l'utilisation de copules.

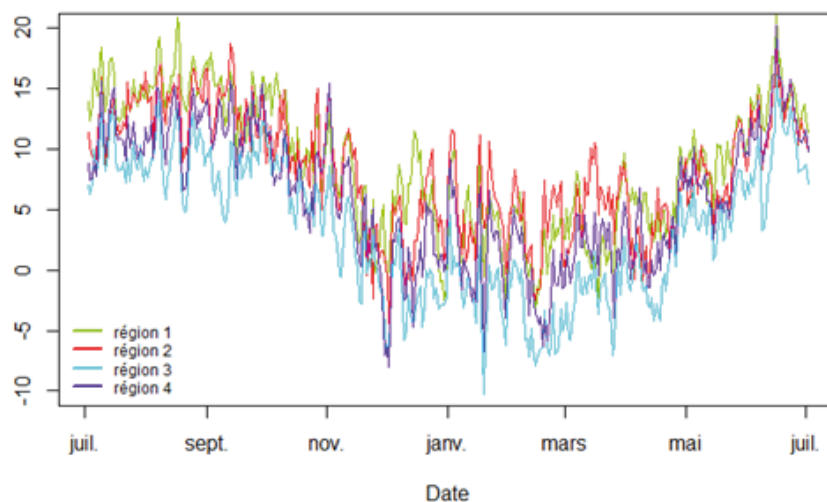


Figure 10 Exemple d'un scénario de température dans les quatre régions

Les températures minimale et maximale sur l'ensemble des scénarios sont respectivement de -30°C et de 34°C . Ces deux valeurs sont plus extrêmes que les valeurs historiques (respectivement à -19°C et à 23°C). Cela semble néanmoins cohérent puisque 10 000 scénarios/années ont été simulés et que nous disposons d'un historique de 63 années. Il convient de noter que la température minimale observée en France depuis 1947 était de -32°C .

Scénarios d'évènements

Les températures ayant été simulées, nous disposons de la température minimale moyenne dans chaque région. A partir de cette information, nous déduisons les écarts de température sur 20 jours. Ensuite, il s'agit de prédire les nombres de sinistres dans chaque région en tenant compte du seuil de $-5,5^{\circ}\text{C}$ le jour du sinistre et de la saison. En effet, la prédiction doit être effectuée pendant l'hiver uniquement et lorsque la température est inférieure à 3°C . Lorsque cette température est strictement supérieure à $-5,5^{\circ}\text{C}$, les prédictions de sinistres utilisent la régression hurdle. Sinon, la prédiction utilise la régression binomiale négative. Dans le reste des cas, en été ou lorsque la température est supérieure à 3°C , on prédit un nombre nul de sinistres.

Nous faisons également l'hypothèse que l'exposition par région est constante d'une année à l'autre. Finalement, nous obtenons 10 000 scénarios donnant la probabilité de sinistralité journalière pour une police donnée au sein de chaque région pendant une période annuelle.

5. Construction du module Vulnérabilité

Cette partie vise à connaître la perte associée aux différents biens assurés lorsqu'un évènement gel survient. Ce module fournit des courbes de destruction (ou courbes de vulnérabilité) qui modélise la fonction de répartition du taux de destruction¹ observé suite à un évènement gel.

¹Montant du sinistre divisé par la valeur du bien assuré

Dans un modèle catastrophe usuel, les courbes associent un taux de destruction différent selon l'intensité de l'évènement considéré. Pour le risque gel, nous considérons que les montants de sinistres causés par le gel ne dépendent pas de l'intensité de l'évènement. En effet, quand il survient, un évènement gel cause généralement des sinistres très similaires quel que soit le seuil de déclenchement. Nous faisons donc l'approximation ici que le montant de perte lors de la survenance d'un sinistre gel dépend uniquement du type et de la valeur du bien assurée et non de l'intensité de l'évènement gel qui est prise en compte dans la probabilité de sinistralité dans le module Aléa.

Pour modéliser la vulnérabilité des biens au gel, nous avons calibré des courbes de la famille paramétrique MBBEFD (Maxwell-Boltzman, Bose-Einstein, Fermi-Dirac). Cette méthode a été popularisée par un article de S. BERNEGGER² et est devenue un standard dans le secteur de l'assurance pour plusieurs raisons :

- ⊙ Les courbes MBBEFD permettent de modéliser la probabilité de destruction d'un bien en prenant en compte des discontinuités. Par exemple, certains biens peuvent subir des destructions soit mineures soit totales et la fonction de répartition du taux de destruction présente donc des discontinuités.
- ⊙ Elles peuvent prendre des formes très diverses : convexes, concaves, points d'inflexion, ...

Deux méthodes permettent de définir une courbe MBBEFD. Tout d'abord, il y a la méthode générale qui nécessite deux paramètres. Il existe aussi un cas particulier de la méthode générale, appelée Hyperbolic MBBEFD, qui nécessite un seul paramètre. La méthode MBBEFD à un paramètre permet de calibrer plus facilement les courbes de destruction d'une part et d'obtenir des résultats plus robustes d'autre part.

Nous avons donc choisi d'adopter cette méthode dans l'étude. Par ailleurs, c'est la méthode la plus utilisée en pratique.

La première méthode permet ainsi de définir la fonction de densité associée au taux de destruction d'un bien par :

$$f(x) = \frac{1}{m \cdot \left(1 + \frac{x}{m}\right)^2} \cdot \mathbf{1}_{0 \leq x < 1} + p \cdot \mathbf{1}_{x=1}$$

Ainsi, cette fonction ne dépend que du paramètre m qui est le taux de destruction³ médian associé au bien. Pour chaque type de bien, nous avons pu calculer le paramètre m associé à l'aide de régressions linéaires sur l'historique des pertes d'AXA entre 2008 et 2012.

A titre d'exemple, nous traçons les courbes de destruction associées à cinq valeurs de somme assurée pour la branche d'activité Habitation.

²"Swiss RE exposure curves and the MBBEFD distribution class", 1997

³Pour plus de détails sur la méthode : BERNEGGER S., 1997, The Swiss RE exposure curves and the MBBEFD distribution class

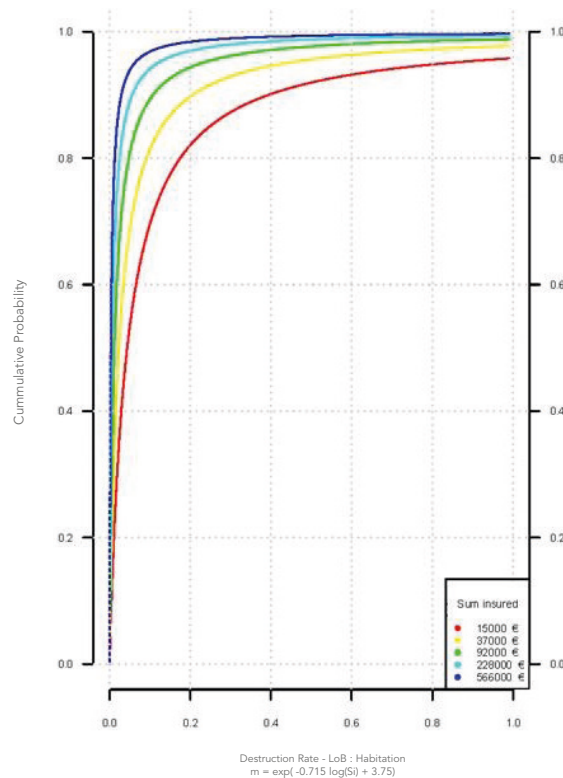


Figure 11 Courbes de destruction pour la LoB Habitation

6. Résultats du modèle

Les principaux résultats d'un modèle catastrophe sont synthétisés par les courbes AEP et OEP :

- **La courbe OEP** (*Occurrence Exceedance Probability*) : associe une période de retour⁴ (en années) à la perte maximale engendrée par un événement sur une année. L'OEP caractérise donc la probabilité annuelle qu'un événement unique entraîne des pertes supérieures à un certain montant.
- **La courbe AEP** (*Aggregate Exceedance Probability*) : associe une période de retour (en années) à la perte totale engendrée par l'ensemble des événements sur une année. L'AEP caractérise donc la probabilité annuelle que l'ensemble des événements sur une année entraîne des pertes supérieures à un certain montant. Ainsi, une courbe AEP est toujours au-dessus d'une courbe OEP.

Ces courbes permettent de définir respectivement le capital réglementaire requis dans le cadre de Solvabilité II et la couverture de réassurance optimale à souscrire.

⁴Une période de retour correspond à l'inverse de la fréquence statistique d'un événement parmi un ensemble d'événements (annuels) équiprobables. Elle doit donc être interprétée comme une probabilité statistique. Elle est exprimée en années et correspond à la probabilité d'occurrence d'un scénario. Par exemple, un événement de période de retour de 200 ans a une probabilité de 0,5 % de se produire sur une année.

Les courbes AEP et OEP de notre modèle sont obtenues en couplant les modules Aléa, Vulnérabilité et Financier. Pour une question de confidentialité, le module Financier permettant de prendre en compte les caractéristiques des contrats souscrits et les conditions de réassurance n'a pas été modélisé. Nous considérons donc ici la perte brute avant l'application des conditions financières.

Les évènements gel, générés de façon quotidienne pour chaque région avec le module Aléa, doivent être appliqués au portefeuille de polices pour traduire la probabilité d'occurrence du risque en une perte financière. A cette fin, les polices du portefeuille d'AXA France ont été regroupées selon leur appartenance géographique à l'une des 4 régions de notre segmentation. Pour chaque jour de chaque scénario régional annuel, nous commençons par évaluer la perte causée par le gel sur les polices du portefeuille régional, puis nous sommes les pertes associées à chaque police pour obtenir la perte sur le portefeuille d'AXA pour chaque scénario.

La Figure 12 représente les courbes AEP et OEP prises sur des périodes de retour de 1 à 250 ans. Pour des raisons de confidentialité, les montants de pertes ne sont pas communiqués. Cependant, la valeur x associée à l'axe des ordonnées représente la perte brute totale d'AXA causée par le gel pendant l'année 2012. Nous pouvons donc déterminer que la période de retour associée à cet évènement sur la courbe AEP est de huit ans.

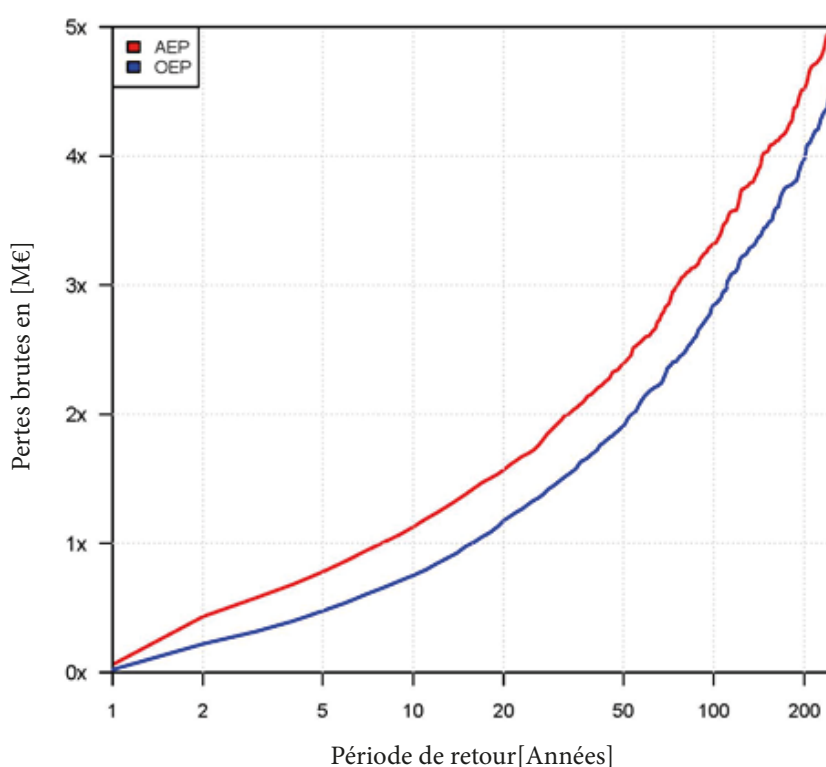


Figure 12 Courbes AEP et OEP

Conclusion

L'objectif de ce mémoire était de construire un modèle catastrophe modélisant le risque gel. Un modèle catastrophe combine trois modules, chaque module apportant une information précise. Ce mémoire est focalisé sur la construction des deux premiers : le module Aléa et le module Vulnérabilité. Le module financier est réalisé par l'équipe interne d'AXA pour une question de confidentialité des clauses. Au terme de l'étude, nous obtenons une distribution des pertes brutes issue de l'application de la modélisation du risque gel sur le portefeuille exposé.

A la différence des modèles catastrophes classiques, où les évènements sont simulés ponctuellement et donc de façon non consécutive, l'approche adoptée ici permet d'apporter une information journalière sur la survenance de sinistres liés au gel. Cette information présente un atout majeur pour un assureur dans le cadre de l'optimisation de ses traités de réassurance et de sa gestion du risque de gel.

La démarche adoptée a également le mérite d'être souple et adaptable. Le modèle a été pensé de façon à pouvoir intégrer de nouveaux pays tout en tenant compte de leur corrélation. Les résultats sur la France étant crédibles et cohérents avec les pertes historiques d'AXA, ce modèle a été élargi aux pays du bassin Europe.

Enfin, le modèle construit pourrait être adapté à un autre risque climatique important : le risque de sécheresse. En s'inspirant de la méthodologie de cette étude, il faudrait cependant considérer les températures maximales et tenir compte de la typologie des sols.

Bibliographie

- BERNEGGER S., 1997, The Swiss RE exposure curves and the MBBEFD distribution class
- BJÖNRSSON H. & VENEGAS S.A., 1997, A manual for EOF and SVD – Analyses of Climatic Data, McGill University
- CAMERON, A.C. & TRIVEDI, P.K., 1998, Regression analysis of count data, Cambridge, Cambridge University Press
- CAMERON, A.C. & TRIVEDI, P.K., 2005, Microeconometrics: Methods and Applications, Cambridge, Cambridge University Press.
- CASIEZ G., Dynamic Time Warping : déformation temporelle dynamique, Université de Lille
- CHARPENTIER A., DUTANG C., L'Actuariat avec R
- CHAVEZ-DEMOULIN V. & DAVISON A.C., 2012, Modelling time series extremes, Volume 10, Number 1, P 109–133
- CREPON B. & JACQUEMET N., 2010, Econométrie : méthodes et applications, De Boeck
- DENUIT M. & CHARPENTIER A., 2005, Mathématiques de l'assurance non-vie, Tome II, Economica
- EUROPEAN CLIMATE ASSESSMENT & DATASET, <http://eca.knmi.nl/>
- FERMANIAN, J.D., 2014, Théorie des copules, support de cours ENSAE
- FFSA, 2012, Rapport annuel
- FLYNN M. & FRANCIS L. A., 2009, More flexible GLMs zero-inflated models and hybrid models, Casualty Actuarial Society E-forum
- GORGE G., 2013, Insurance Risk Management and Reinsurance
- HANNACHI A., 2004, A Primer for EOF Analysis of Climate Data, Department of Meteorology, University of Reading U.K
- HUGUES G., RAO S. & RAO T., 2006, Statistical analysis and time-series models for minimum/maximum temperatures in the Antarctic Peninsula
- KHAROUBI-RAKOTOMALALA C. 2008, "Les fonction copules en finance", Université de la Sorbonne
- LEDOLTER J., 2013, Datamining and business analytics with R, Wiley
- LOPEZ O., 2014, Econométrie de l'assurance, support de cours ENSAE
- ROBERT Y. C., Théorie des Valeurs Extrêmes, support de cours ENSAE
- RONCALLI T., 2002, Gestion des risques multiples ou copules et aspects multidimensionnels du risque, support de cours ENSAI
- RONCALLI T., FRACHOT A., 2009, La Gestion des Risques Financiers, Economica
- ROUSTANT O., 2003, Produits dérivés climatiques : aspects économétriques et financiers
- SAKOE H., CHIBA S., 1978, Dynamic programming algorithm optimization for spoken word recognition
- SAPORTA G., 2011, Probabilités, Analyse de données et Statistique, Technip
- SKLAR A., 1959, Fonctions de répartition à n dimensions et leurs marges, Publications de l'Institut de Statistique de l'Université de Paris
- STRAUSS D., 2009, Empirical Orthogonal Function Analysis (Principal Component Analysis), ICTP
- SWISS RE, 2003, Catastrophes naturelles et réassurance
- SWISS RE, 2003, Introduction à la réassurance
- ZEILEIS A., KLEIBER C. & JACKMAN S., 2008, Regression models for count data in R, Journal of statistical software, Volume 27, Issue 8

- | | | |
|----------------------------------|---|---|
| SCOR Paper N°27 - October 2013 | ☉ | Are great earthquakes clustered? |
| SCOR Paper N°28 - February 2014 | ☉ | Solar storms and their impacts on power grids - Recommendations for (re)insurers |
| SCOR Paper N°29 - September 2014 | ☉ | A game-theoretic approach to non-life insurance market cycles |
| SCOR Paper N°30 - September 2014 | ☉ | Explicit Föllmer-Schweizer decomposition of life insurance liabilities through Malliavin calculus |
| SCOR Paper N°31 - January 2015 | ☉ | The use of economic scenarios generators in unstable economic periods |
| SCOR Paper N°32 - March 2015 | ☉ | An Integrated Notional Defined Contribution (NDC) Pension Scheme with Retirement and Permanent Disability |
| SCOR Paper N°33 - April 2015 | ☉ | Exploring the Dependence between Mortality and Market Risks |
| SCOR Paper N°34 - November 2015 | ☉ | A Change of Paradigm for the Insurance Industry |
| SCOR Paper N°35 - January 2016 | ☉ | The Globalization of Infectious Diseases |
| SCOR Paper N°36 - February 2016 | ☉ | Spatial Risk Measures and Applications to Max-Stable Processes |
| SCOR Paper N°37 - March 2016 | ☉ | ALM Model: The contribution of Fuzzy Logic to behaviour modelling |
| SCOR Paper N°38 - April 2016 | ☉ | Explicit diversification benefit for dependent risks |
| SCOR Paper N°39 - December 2016 | ☉ | A General Framework for Modeling Mortality to Better Estimate its Relationship to Interest Rate and Risks |
| SCOR Paper N°40 - Juin 2017 | ☉ | On the Diversification Benefit of Reinsurance Portfolios |



Chaque année, SCOR récompense les meilleurs travaux de recherche dans le domaine de l'actuariat en attribuant des prix dans plusieurs pays à travers le monde. Ces prix, dont un certain nombre sont financés par la Fondation SCOR pour la Science, ont pour but de promouvoir la science actuarielle, de développer et d'encourager la recherche dans ce domaine et de contribuer à faire progresser la connaissance et la gestion des risques. SCOR publie certains des travaux primés dans sa collection de SCOR Papers.