

I A F

C E A

PROMOTION 1996

Mémoire présenté devant

L'Institut des Actuaires Français

pour l'obtention du

Diplôme de l'Institut des Actuaires Français

**Par : MM Daniel DESCHEEMAEKERE
Bernard PERRON**

sur le sujet

**Les événements naturels en France :
étude du risque tempête pour une compagnie d'assurances**

devant un jury composé de

Liste du jury IAF

Directeur du mémoire :

Arnaud CLEMENT-GRANDCOURT

Philippe BORNE

Jean-Luc AURE

Jean BERTHON

Jean-Luc BESSON

Jacques CHEVALIER

Arnaud COHEN

Gérard CROSET

Jean-Pierre DIAZ

Michel FROMENTEAU

Richard KOMARNICKI

Patrick LEMOINE

Charles LEVI

Pierre PETAUTON

Florence PICARD

Alain TOSETTI

Martine VAREILLES

Marc WISZNIAK

Philippe ZAOUATI

Invités :

Christine DREVETON

Philippe GAUDIBERT

Christian PARTRAT

REMERCIEMENTS

Nous tenons à remercier **Monsieur DUBY**, Directeur Scientifique de l'UAP, pour avoir financé l'achat de l'actualisation des données météorologiques et accepté de nous communiquer ses études relatives à un éventuel réchauffement de notre planète et à ses conséquences pour les Assureurs.

Nous remercions **Monsieur BORNE** qui a bien voulu nous faire partager son expérience de directeur technique d'une très importante compagnie d'assurances.

Nous remercions également **Madame DREVETON**, Responsable du bureau d'études de MÉTÉO – FRANCE, qui a mis à notre disposition toutes les données complémentaires dont nous avons besoin pour constituer une base de données météorologiques fiable sur les tempêtes.

Nous remercions bien entendu notre ami **Charles EL NOUTY**, Maître de conférences à l'Université de PARIS VI - ISUP / JUSSIEU qui a bien voulu valider l'ensemble des travaux statistiques présentés dans ce mémoire.

Enfin, nous exprimons notre profonde gratitude à **Monsieur GAUDIBERT**, Président Directeur Général, Société Parisienne de Souscription de Réassurance, qui, par ses connaissances très approfondies du sujet traité, nous a fait part, avec beaucoup de gentillesse, de ses remarques et suggestions tout au long de nos recherches.

SOMMAIRE

INTRODUCTION.....	11
PREMIERE PARTIE : LES ASPECTS THEORIQUES DU RISQUE TEMPETE	11
INTRODUCTION	15
CHAPITRE 1 : LES CONNAISSANCES SCIENTIFIQUES DU RISQUE TEMPETE	17
I - QU'EST-CE QU'UNE TEMPÊTE ?	17
I.1 - Nature et cause du phénomène.....	17
I.2 - Facteurs influençant le vent	17
II - DÉFINITION DES TERMES EMPLOYÉS.....	18
II.1 - Coup de vent, tempête, cyclone.....	18
II.2 - Mesures utilisées.....	19
II.3 - Fréquences, durées de retour.....	19
III - DIFFICULTÉS DE MESURE DU RISQUE TEMPÊTE	20
III.1 - Hétérogénéité de la vitesse du vent au sol.....	20
III.2 - Autres facteurs d'hétérogénéité caractéristiques de la tempête.....	20
IV - LES PRINCIPAUX PHÉNOMÈNES DE TEMPÊTE.....	21
IV.1 - Le cyclone tropical.....	21
IV.1.1 - Origine du cyclone tropical.....	21
IV.1.2 - Type de dommages occasionnés par le cyclone	22
IV.2 - La tornade.....	23
IV.2.1 - Origine de la tornade.....	23
IV.2.2 - Type de dommages occasionnés par la tornade.....	23
IV.3 - Le cyclone extra-tropical ou tempête d'hiver.....	25
IV.4 - Les autres phénomènes de tempête.....	26
IV.4.1 - Tempêtes orographiques (vents descendants)	26
IV.4.2 - Les Tempêtes orageuses, orages de grêle, tempêtes de sable	27
IV.4.3 - Tempête de mousson.....	28
V - EFFET DE SERRE ET CHANGEMENT CLIMATIQUE.....	28
V.1 - Liens incertains entre l'activité économique et le changement climatique	28
V.2 - Effets incertains du réchauffement climatique.....	30
VI - LA CHARGE FINANCIÈRE DES ÉVÉNEMENTS NATURELS	32
VI.1 - Le bilan financier dans le monde	32
VI.2 - Le bilan financier en France.....	35
CHAPITRE 2 : LES CONNAISSANCES METEOROLOGIQUES DU RISQUE TEMPETE.....	39
I - PRÉSENTATION DE L'ÉTUDE	39
I.1 - Méthodologie.....	39
I.2 - Caractérisation des tempêtes	40
I.3 - Le fichier d'expertise	41
I.4 - Typologie des tempêtes	43
I.4.1 - Arborescence de la typologie.....	43
I.4.2 - Caractéristiques des différents types obtenus.....	44
II - PHASE DE CLASSIFICATION	46
II.1 - La méthode de classification.....	46
II.2 - Recherche des règles décisionnelles.....	46
II.3 - Constitution du fichier élargi.....	47
II.4 - Test des règles décisionnelles	49
II.5 - Résultats définitifs de la classification.....	50
III - DISTRIBUTION DES INTENSITÉS DES TEMPÊTES	50
III.1 - Présentation et définition des paramètres descriptifs.....	50
III.2 - Sélection des paramètres descriptifs	52
III.3 - Détermination des stations caractéristiques pour chaque type de tempête	53
III.4 - Calcul des corrélations entre les stations.....	55
III.4.1 - Principe du calcul.....	55

III.4.2 - Qualité de la reconstitution	56
III.5 - <i>Distribution bidimensionnelle des paramètres descriptifs</i>	56
III.5.1 - Présentation	56
III.5.2 - Résultats	57
III.5.2.1 - Table de contingence	57
III.5.2.2 - Histogrammes.....	58
III.6 - <i>Conclusion</i>	60
IV - ETUDE DES CORRÉLATIONS	60
IV.1 - <i>Les corrélations temporelles</i>	61
IV.1.1 - Fréquence et variabilité temporelle des tempêtes	61
IV.1.1.1 - Fréquence annuelle des tempêtes	61
IV.1.1.2 - Répartition mensuelle des tempêtes	63
IV.1.1.3 - Homogénéité temporelle des tempêtes	64
IV.1.2 - Enchaînement de plusieurs événements	66
IV.1.2.1 - Succession d'événements.....	66
IV.1.2.2 - Enchaînements possibles.....	69
IV.1.3 - Composante liée aux clauses horaires	70
IV.1.3.1 - Principaux résultats de la classification sur une ou deux journées	70
IV.1.3.2 - Persistance des différents types	71
IV.1.4 - Durées de retour	72
IV.1.4.1 - Méthode de calcul.....	72
IV.1.4.2 - Résultats	73
IV.2 - <i>Les corrélations spatiales</i>	74
IV.3 - <i>Récapitulatifs</i>	74
V - CALCUL DE L'INTENSITÉ GLOBALE D'UN ÉVÉNEMENT	75
V.1 - <i>La méthode de calcul</i>	75
V.2 - <i>Stations retenues pour le calcul de l'intensité globale</i>	76
V.3 - <i>Classes de durée et de précipitations</i>	76
V.4 - <i>Cartes représentatives</i>	77
VI - CONCLUSION : PRINCIPAUX RÉSULTATS	78
VI.1 - <i>La classification</i>	78
VI.2 - <i>Paramètres descriptifs</i>	78
VI.3 - <i>Corrélation dans le temps et dans l'espace</i>	79
VI.4 - <i>Notion d'intensité globale</i>	79

CHAPITRE 3 : ETUDE DE LA VULNERABILITE DES CONSTRUCTIONS 81

I – ORIGINE ET NATURE DES DÉGÂTS OCCASIONNÉS PAR LES TEMPÊTES	81
II - ANALYSE DES DÉGÂTS	82
III - PRÉVENTION ET RÉDUCTION DES SINISTRES	83
III.1 - <i>Actions du vent sur un bâtiment</i>	83
III.2 - <i>Importance d'un bon dimensionnement du bâtiment</i>	84
IV - PRINCIPALES RÈGLES DÉFINISSANT LES EFFETS DU VENT SUR LES CONSTRUCTIONS	85
IV.1 - <i>Historique de la mise en place des principales règles</i>	85
IV.2 - <i>Généralités concernant les règles de construction</i>	88
IV.3 - <i>Valeurs</i>	91
IV.4 - <i>Modification des pressions dynamiques de base</i>	91
IV.4.1 - L'effet de hauteur par rapport au sol.....	92
IV.4.2 - Effet du site	92
IV.4.3 - Effet de masque	93
IV.5 - <i>Conclusion</i>	94

CHAPITRE 4 : L'EVOLUTION DES COUVERTURES D'ASSURANCE 97

I - L'INTERVENTION DE L'ÉTAT EN MATIÈRE D'ÉVÉNEMENTS NATURELS	97
I.1 - <i>La garantie tempêtes et grêle avant la loi du 13 juillet 1982</i>	97
I.2 - <i>La genèse de la loi du 13 juillet 1982</i>	97
I.3 - <i>Le contenu de la loi du 13 juillet 1982</i>	98
I.4 - <i>Les interférences tempêtes - catastrophes naturelles</i>	100
I.5 - <i>Vers un régime de croisière</i>	101
I.6 - <i>La genèse de la loi du 25 juin 1990</i>	102
I.7 - <i>Les derniers développements : l'articulation "tempêtes - catastrophes naturelles"</i>	106
I.8 - <i>Quel peut être le devenir des garanties tempêtes et catastrophes naturelles?</i>	107
I.9 - <i>Les régimes étrangers de catastrophes naturelles</i>	111

I.9.1 - Les pays dotés d'un système facultatif.....	111
I.9.2 - Les pays dotés d'un système obligatoire.....	112
II - L'ÉVOLUTION DES GARANTIES CONTRACTUELLES CONTRE LES EFFETS DU VENT	113
II.1 - <i>L'évolution commune à tous les types de garanties</i>	113
II.1.1 - Difficulté d'harmonisation de la définition de la tempête par les assureurs.....	113
II.1.2 - Evolution du contenu des garanties	115
II.1.2.1 - Couverture des biens contre les effets du vent avant 1983	115
II.1.2.2 - Conséquences de la pollicitation de 1983	116
II.1.2.3 - Conséquences de la loi du 25 juin 1990.....	117
II.2 - <i>Description des biens à usage d'habitation</i>	120
II.2.1 - Evaluation de l'exposition au danger des biens assurés	120
II.2.2 - Etude de la vulnérabilité des biens assurés.....	121
II.3 - <i>Description des risques d'entreprise</i>	122
II.3.1 - Evaluation de l'exposition au danger des biens assurés	122
II.3.2 - Vulnérabilité des biens assurés.....	123
II.4 - <i>Description des biens agricoles</i>	124
II.4.1 - Objectifs de la loi du 10 juillet 1964.....	124
II.4.2 - Modifications apportées par l'arrêté du 28 mars 1975	125
II.4.3 - Application de la loi du 25 juin 1990 à la garantie tempête des risques agricoles.....	125
II.4.4 - La description des risques assurés	126
II.4.4.1 - Le parc des bâtiments agricoles.....	126
II.4.4.2 - Comparaison des risques garantis selon les régimes d'indemnisation.....	127
II.4.5 - <i>Vulnérabilité des risques</i>	128
II. 5 - AUTRES RISQUES SUBISSANT LES EFFETS DU VENT	130
CHAPITRE 5 - LA REASSURANCE	131
I - LA RÉASSURANCE TRADITIONNELLE.....	131
I.1 - <i>La réassurance des catastrophes naturelles en France</i>	131
I.1.1 - La caisse Centrale de Réassurance (C.C.R.).....	131
I.1.2 - Modification du Traité	132
I.2 - <i>La réassurance des catastrophes naturelles dans le monde</i>	133
I.2.1 - Présentation du traité de réassurance CatXL.....	133
I.2.2 - Sinistres de référence : charge de sinistre des assureurs en cas de grosse catastrophe naturelle.....	134
I.2.3 - La notion de rate on line.....	134
I.2.4 - Scénarii de tempêtes	136
I.2.5 - Informations sur les cumuls et sur les sinistres	137
II - LA RÉASSURANCE FINANCIÈRE	139
II.1 - <i>Présentation</i>	139
II.1.1 - Définition.....	139
II.1.2 - Intérêt de la réassurance financière.....	140
II.2 - <i>Utilisation des contrats à terme et options sur indice d'assurances pour couvrir les catastrophes naturelles aux États-Unis</i>	141
II.2.1 - L'indice de référence	141
II.2.2 - Exemples	143
II.2.2.1 - Cas d'une adéquation parfaite des sinistralités de l'indice et de la compagnie	143
II.2.2.2 - Cas d'une adéquation imparfaite des sinistralités de l'indice et de la compagnie	145
II.2.3 - Etude des cotations effectuées sur le C.B.O.T.....	148
II.2.4 - Les risques couverts.....	148
II.2.5 - Intervenants du marché.....	150
II.3 - <i>Intérêts des contrats à terme pour les assureurs</i>	150
II.3.1 - Composition de l'indice ISO.....	151
II.3.2 - Solvabilité.....	151
II.3.3 - Réglementation	151
II.3.4 - Audit, arbitrage, jugement.....	152
II.3.5 - Liquidité et capacité.....	152
II.3.6 - Marché ouvert contre marché fermé	153
II.4 - <i>Autres perspectives</i>	153
II.4.1 - Les nouvelles générations d'options du C.B.O.T.....	153
II.4.2 - La négociation des risques catastrophiques à la bourse CATEX	154
II.4.3 - Placements privés par le biais de la titrisation directe.....	155
III - CONCLUSION	156

DEUXIÈME PARTIE : MODELISATION DU RISQUE TEMPETE.....157

CHAPITRE 1 : CONSTITUTION DE LA BASE DE DONNÉES SUR LES TEMPÊTES.....157

I – RECONSTITUTION DU PORTEFEUILLE D’ASSURANCES	157
I.1 - Périmètre de l’étude.....	157
I.1.1 - Périmètre d’observation	157
I.1.2 - Définition des risques observés	158
I.1.3 - Description des variables observées.....	159
I.2 - Reconstitution du portefeuille pour les années 1992 à 1996	160
I.2.1 - Présentation des données utilisées	160
I.2.2 - Présentation des méthodes de reconstruction du portefeuille	160
I.2.2.1 - Méthode d’agrégation des données brutes	160
I.2.2.2 - Reconstitution du chiffre d’affaires pour l’exercice 1992	164
I.2.3 - Mesure de la qualité de la méthode d’estimation	169
I.2.3.1 - Méthode de localisation des contrats non localisés	169
I.2.3.2 - Suppression du contrat “ mortalité du bétail ” du risque agricole	173
I.2.3.3 - Autres corrections apportées au chiffre d’affaires 1992	175
I.3 - Reconstitution du portefeuille pour les années 1979 à 1991	178
I.3.1 - Détermination du chiffre d’affaires 1992 par entité et zone postale	178
I.3.2 - Principe de l’utilisation d’une variable corrélée avec le lieu du risque	182
I.3.3 - Comparaison des méthodes de reconstitution des portefeuilles	183
I.3.3.1 - Méthode de reconstitution des portefeuilles utilisant l’agence	184
I.3.3.2 - Réseau agents : résultats comparés des méthodes d’estimation à évolution homogène	188
I.3.3.3 - Remarques complémentaires sur la méthode d’estimation à évolution homogène	194
I.3.4 - Mise en œuvre pratique de la reconstitution du portefeuille	196
I.3.4.1 - Reconstitution des marges par entité et risque	196
I.3.4.2 - Redistribution du portefeuille de Michelin	202
I.3.4.3 - Autres pistes d’amélioration de la précision de l’estimation du portefeuille	203
II - CONSTITUTION DE LA BASE DE DONNÉES “ SINISTRES ”.....	209
II.1 - Périmètre de la base de données “ sinistres ”	209
II.1.1 - Définitions	209
II.1.2 - Description de la base de données “ sinistres ”	210
II.2 - Identification de sinistres tempête réglés au titre de la loi sur les catastrophes naturelles	212
II.2.2.1 - Justification de l’existence de ces sinistres tempête	212
II.2.2.2 - Méthode de recherche des sinistres tempête réglés en 1982 au titre de la loi	212
II.2.2.3 - Caractérisation des sinistres tempêtes réglés au titre de la loi	219
II.3 - Enrichissement de l’information sur les sinistres tempête	227
II.3.1 - Redressement du code postal indéterminé	227
II.3.1.1 - Constat chiffré	227
II.3.1.2 - Recueil d’information complémentaire	227
II.3.1.3 - Algorithme de redressement du code postal indéterminé	230
II.3.1.4 - Résultats et validation de l’algorithme de redressement	234
II.3.2 - Autres enrichissements de la base de données "Sinistres"	236
III - BASE DE DONNÉES MÉTÉOROLOGIQUES.....	236
III.1 - Présentation de la base de données météorologiques	237
III.1.1 - Historique de la constitution de la base de données	237
III.1.2 - Evolutions de la base de données originelle	239
III.1.3 - Derniers enrichissements apportés à la base de données	239
III.2 - Travaux préliminaires à l’utilisation de la base de données.....	240
III.2.2 - Critères retenus lors du choix des stations météorologiques	242
III.2.3 - Méthode de redressement des historiques des stations météorologiques.....	245
IV - CONSTITUTION DE LA BASE DE DONNÉES “CONSTRUCTION”.....	245
IV.1 - Importance de la constitution de la base de données “ Construction ”	245
IV.2 - Constitution de la base de données relative aux normes de construction	246
IV.3 - Constitution de la base de données relative à la vulnérabilité des biens assurés	248
IV.4 - Constitution de la base de données relative au parc immobilier	251

CHAPITRE 2 : DÉTERMINATION ET ACTUALISATION DU PRÉJUDICE DÛ À LA TEMPÊTE

..... 253

I - PRINCIPES DE DÉTERMINATION ET D’ACTUALISATION DU PRÉJUDICE DÛ À LA TEMPÊTE.....	254
---	-----

I.1 - Eléments constitutifs du préjudice dû à la tempête	254
I.2 - Facteurs d'actualisation.....	255
I.3 - Modalités de calcul	256
I.3.1 - Ordre de réalisation des calculs	257
I.3.2 - Degré de finesse des calculs	258
II - ACTUALISATION LIÉE À L'ÉVOLUTION DES PRIX.....	259
III - ACTUALISATION LIÉE À L'ÉVOLUTION DE LA SOUSCRIPTION DE LA GARANTIE TEMPÊTE.....	260
III.1 - Bref historique.....	260
III.2 - Difficultés de calcul du taux de présence de la garantie tempête	261
III.3 - Estimation du taux de présence de la garantie tempête	262
III.3.1 - Nature de la présence de la garantie tempête pour chaque branche-risque	262
III.3.1.1 - Exploitation des bases informatiques	262
III.3.1.2 - Exploitation de la base documentaire	264
III.3.2 - Estimation du taux de présence de la garantie tempête pour certaines branches-risques	267
III.3.2.1 - Contraintes informatiques	267
III.3.2.2 - Réalisation d'un sondage	267
III.3.3 - Précision du calcul du taux de présence de la garantie tempête	270
III.4 - Calcul de la charge de sinistres à taux de présence de la garantie tempête égale à 100%	271
III.5 - Validation des calculs précédents	273
III.6 - Modalités pratiques de l'actualisation	274
IV - PRISE EN COMPTE DE LA FRANCHISE.....	276
IV.1 - Intérêt du calcul de l'indemnisation à niveau de franchise nul	276
IV.2 - Aspects théoriques	276
IV.2.1 - Adéquation de la distribution du montant des sinistres à une loi log-normale tronquée	276
IV.2.1 - Recherche des moments de la loi log-normale	278
IV.3 - Modalités pratiques du calcul	281
IV.3.1 - Recherche de la formule du montant des franchises	281
IV.3.1.1 - Exploitation des bases informatiques	281
IV.3.1.2 - Exploitation de la base documentaire	282
IV.3.2 - Regroupement des produits selon le montant minimal des franchises	284
IV.3.3 - Degré de finesse des calculs.....	285
IV.3.4 - Structure de redistribution de l'indemnité à niveau de franchise nul	285
V - ACTUALISATION LIÉE À L'ÉVOLUTION DU NIVEAU D'EXCLUSION	286
V.1 - Importance du calcul d'un indice d'exclusion	286
V.2 - Méthode de calcul de l'indice d'exclusion	287
V.2.1 - Echantillonnage des sinistres tempête	287
V.2.2 - Grille de lecture des dossiers sinistres	288
V.2.3 - Constitution d'un tableau des exclusions par produit	289
V.3 - Structure de redistribution du montant induit par l'indice d'exclusion	290
VI - ACTUALISATION LIÉE À L'ÉVOLUTION DU TAUX DE VÉTUSTÉ APPLIQUÉE.....	291
VI.1 - Intérêt du calcul d'un indice de vétusté	291
VI.2 - Etude de la distribution du taux de vétusté	293
VI.3 - Etude du taux de règlement	295
VI.3.1 - Cas des risques professionnel et d'habitation	296
VI.3.2 - Cas du risque agricole	296
VI.4 - Etude de la double vétusté	298
VI.3 - Hypothèses de redistribution du surcoût lié au paiement de la vétusté	299
VII - AUTRES OPÉRATIONS D'ACTUALISATION.....	299
VII.1 - Calcul de l'indice lié à la limite de garantie	299
VII.2 - Calcul de l'indice lié au taux de T.V.A.	302
VII.3 - Autres facteurs.....	303
VIII - VALIDATION DES CALCULS ET DE L'ACTUALISATION DU PRÉJUDICE.....	304
CHAPITRE 3 : LA TARIFICATION DU RISQUE TEMPÊTE.....	307
I - LA TARIFICATION CLASSIQUE	307
I.1 - La tarification avant la loi du 25 juin 1990	307
I.1.1 - Avant la création de l'UAP	307
I.1.1.1 - Pour l'Union	308
I.1.1.2 - Pour l'Urbaine	308
I.1.2 - Après la création de l'UAP	309
I.1.2.1 - Risque d'habitation	309

I.1.2.2 - Risque professionnel	310
I.1.2.3 - Risque agricole	311
I.1.3 - La pollicitation de 1983	312
I.2 - La tarification après la loi du 25 juin 1990	314
I.2.1 - La conséquence de cette nouvelle loi à l'UAP	314
I.2.2 - Amélioration de la garantie Tempête	315
I.2.2.1 - Risque d'habitation	315
I.2.2.2 - Risque professionnel	316
I.2.2.3 - Cas particulier des contrats agricoles	316
I.3 - La tarification actuelle.....	318
I.3.1 - La tarification du risque Tempête à l'UAP avant la fusion avec AXA	318
I.3.1.1 - Risque d'habitation	318
I.3.1.2 - Risque professionnel	319
I.3.1.3 - Risque agricole	320
I.3.2 - La tarification du risque Tempête après la fusion de l'UAP avec AXA	322
II - PROPOSITION DE TARIFICATION À PARTIR DES DONNÉES D'ASSURANCES.....	323
II.1 - Analyse statistique des données d'assurances	323
II.1.1 - Etude de la série chronologique du nombre de sinistres	323
II.1.1.1 - Nombre et fréquence annuels des sinistres	323
II.1.1.2 - Répartition mensuelle des sinistres	326
II.1.2 - Répartition géographique de la fréquence	326
II.1.2.1 - Risque agricole	327
II.1.2.2 - Risque d'habitation	328
II.1.2.3 - Risque professionnel	329
II.1.3 - Répartition géographique de la prime pure	330
II.1.3.1 - Risque agricole	330
II.1.3.2 - Risque d'habitation	331
II.1.3.3 - Risque professionnel	332
II.1.4 - Analyse de la distribution du montant des sinistres	333
II.1.4.1 - Risque agricole	333
II.1.4.2 - Risque d'habitation	334
II.1.4.3 - Risque professionnel	335
II.1.5 - Réduction de la prime pure par l'application d'une franchise	336
II.2 - Etude des événements tempête	337
II.2.1 - Définition d'un événement tempête	337
II.2.2 - Etude de la série chronologique des tempêtes	338
II.2.2.1 - Nombre et fréquence.....	338
II.2.2.2 - Répartition mensuelle des tempêtes	340
II.2.3 - Analyse de la distribution du nombre de sinistres des tempêtes	340
II.3 - Proposition d'une tarification du risque tempête	343
CHAPITRE 4 : MODELISATION DU RISQUE TEMPÊTE	345
I - PRÉSENTATION DU MODÈLE.....	345
I.1 - Aspects théoriques	345
I.2 - Contraintes liées au modèle	347
I.2.1 - Incohérence des dates et des lieux de tempête.....	347
I.2.2 - Hétérogénéité des zones météorologiques.....	349
I.2.3 - Exposition au risque distincte au sein du portefeuille habitation	349
I.2.4 - Suppression de l'effet de taille pour les phénomènes localisés	350
I.2.5 - Détermination et actualisation du capital sous risque	351
II - VALIDATION DE LA TYPOLOGIE DES TEMPÊTES DÉFINIE PAR MÉTÉO-FRANCE	352
II.1 - Appariement des événements de Météo-France et des tempêtes du fichier " sinistres ".....	353
II.2 - Analyse descriptive de la typologie des tempêtes	354
II.3 - Recherche d'une quantification de la typologie des tempêtes	357
III - UTILISATION DU MODÈLE POUR UNE NOUVELLE APPROCHE TARIFAIRE.....	359
III.1 - Etude de l'impact local de la tempête	359
III.1.1 - Présentation du modèle d'analyse par station	359
III.1.1.1 - Détermination des événements	359
III.1.1.2 - Choix final des stations météorologiques	360
III.1.1.3 - Construction des modèles de régression linéaire	362
III.1.1.4 - Présentation des résultats.....	365
III.1.1.5 - Détermination de la vitesse d'apparition des dommages	366

III.1.2 - Utilisation des distributions bidimensionnelles par station	370
III.1.2.1 - Hypothèses simplificatrices	370
III.1.2.2 - Recherche d'une adéquation à une loi exponentielle	371
III.1.3 - Détermination d'une prime pure tempête	372
III.1.3.1 - Formulation mathématique d'une prime pure tempête	372
III.1.3.2 - Résultats du calcul d'une prime pure tempête	373
<i>III.2 - Etude de l'impact de différentes variables sur le montant de la prime pure</i>	<i>377</i>
III.2.1 - Présentation du modèle de régression	377
III.2.2 - Validation du modèle de régression	377
III.2.3 - Principaux résultats	378
III.2.3.1 - Résultats généraux	378
III.2.3.2 - Mesure de l'impact des variables sur le montant de la prime pure par événement	379
IV - CONCLUSION.....	385
CONCLUSION.....	387

CHAPITRE 1 : CONSTITUTION DE LA BASE DE DONNÉES SUR LES TEMPÊTES

Dans ce chapitre préliminaire, il convient de réunir une très importante base de données concourant à une meilleure connaissance du risque tempête. Ce travail n'a, à notre connaissance, jamais encore été mené à terme en France pour un niveau de granularité géographique égal à la zone postale.

Cette base de données se compose de quatre bases distinctes :

- les données *sur le portefeuille*,
- les données *relatives aux sinistres*,
- les données *d'origine météorologique*,
- les données *liées à l'habitat ou aux normes de construction*.

La valeur ajoutée de chacune de ces bases est de contenir des indicateurs numériques dont les variations ont un impact direct sur le risque de tempête, en terme de coût moyen ou de fréquence.

I – Reconstitution du portefeuille d'assurances

La première étape consiste à reconstituer une base de données sur le portefeuille d'assurances de l'UAP. Cette tâche se révèle particulièrement délicate comme les paragraphes suivants vont le montrer.

I.1 - Périmètre de l'étude

I.1.1 - Périmètre d'observation

La période d'observation du portefeuille d'assurances est 1980 - 1997. En effet, les informations d'origine " Sinistres " antérieures à 1980 n'étaient plus disponibles lors de la création d'un fichier " historique sinistres " décidée en 1986. L'observation des informations " portefeuille " s'est donc limitée à cette même période.

L'UAP, compagnie internationale, dispose d'un portefeuille réparti entre de nombreuses filiales ou représentations à travers le monde entier. Néanmoins, les données fournies par Météo-France se limitant à la seule France métropolitaine, il nous a paru logique de circonscrire l'étude à cette zone géographique.

Les contrats retenus dans l'étude sont ceux susceptibles d'accueillir la garantie contre les effets du vent : l'incendie et les multirisques constituent son support naturel. Ils couvrent les dommages, de toutes sortes, occasionnés aux bâtiments ou au mobilier.

Pour cette raison, les contrats spécifiques aux dommages aux biens (vol, bris de machine, dégâts des eaux, etc...), qui ne contiennent qu'une garantie, ne font pas partie du périmètre retenu. C'est aussi le cas des contrats :

- perte d'exploitation, qui couvrent les conséquences indirectes des effets du vent,
- automobile, qui sont le support à des garanties plus larges couvrant l'ensemble des dommages causés à un véhicule terrestre à moteur par les événements naturels,
- et grêle, qui ne couvrent que les conséquences de ce phénomène climatique.

Enfin, les contrats spécifiques assurant les conséquences économiques des dommages causés par la tempête aux animaux, aux récoltes non engrangées et aux bois sur pied n'entrent pas dans notre périmètre d'étude.

I.1.2 - Définition des risques observés

Il nous a semblé nécessaire de distinguer, à l'intérieur du périmètre de l'étude, trois familles de contrats :

- les contrats à *usage d'habitation*,
- les contrats couvrant les *risques d'entreprises*,
- les contrats couvrant les *risques agricoles*.

En effet, les paragraphes du chapitre 4 de la 1^{ère} partie, décrivant les risques assurés, montrent, selon les familles de contrats, une grande variabilité des dommages occasionnés. L'environnement, urbain ou rural, de ces risques joue souvent un rôle déterminant.

Cette segmentation des contrats se base sur l'analyse d'un critère déterminant : le risque principal que l'assuré, par son activité, souhaite garantir. La codification des contrats, présente dans les fichiers d'assurances, permet d'établir cette division. On distingue ainsi les contrats couvrant les risques :

- d'habitation : les contrats incendie et multirisque habitation, la multirisque immeuble,
- professionnel : les contrats incendie et multirisque professionnels, l'incendie et la multirisque industriels, l'incendie et la multirisque hôteliers, les contrats multirisque des collectivités locales et des associations,
- agricole : les contrats incendie et multirisque agricoles, à l'exclusion de la mortalité du bétail, des dommages sur récolte ou sur le bois sur pied.

I.1.3 - Description des variables observées

La reconstruction du portefeuille d'assurances doit permettre au final de disposer des variables, dites quantitatives, suivantes :

- **nombre de contrats** en cours à l'inventaire de l'année n, n appartenant à [1979 ;1996],
- montant en F des **primes émises nettes d'annulations** au cours du même exercice n,

ventilées selon la structure croisée :

- **zone postale** du lieu du risque,
- **type de risque** (habitation, professionnel, agricole).

Cet objectif se révèle très ambitieux, car une grande partie du portefeuille de l'UAP est à créer. Pour surmonter ces difficultés, plusieurs étapes bien distinctes les unes des autres ont été suivies. A ce titre, on verra ci-dessous qu'il a été nécessaire de procéder successivement :

- au retraitement et à l'agrégation des informations extraites des bases les plus récentes, donc les plus riches,
- à la recherche, au résultat aléatoire, des tableaux de bord détaillant les variables de l'étude, pour les années les plus anciennes,
- à la saisie de ces informations non conservées par l'informatique,
- à l'éclatement de ces données très agrégées selon les deux axes de la structure croisée.

I.2 - Reconstitution du portefeuille pour les années 1992 à 1996

I.2.1 - Présentation des données utilisées

L'exploitation du fichier "de synthèse" de l'UAP, qui rassemble les informations relatives au portefeuille, aux primes et aux sinistres d'un contrat, est à la base des travaux de constitution du portefeuille incendie : l'étude des fichiers informatiques a montré en effet que seul ce fichier offrait une ventilation des variables quantitatives selon la structure croisée code postal / risque.

Le fichier de synthèse contient l'ensemble des contrats en cours ou dont la date de résiliation est postérieure à la date de création de ce fichier moins 5 ans. Son contenu est réactualisé à la fin de chaque semestre, lorsque les contrats résiliés depuis plus de 5 ans par rapport à la date de constitution du fichier sortent de son périmètre et tombent définitivement dans l'oubli.

Ainsi, lors de l'exploitation du fichier de synthèse créé à la fin juin 1997, il a été possible d'extraire les données relatives aux années 1992 à 1996 : ce fichier ne permet donc pas de retrouver le profil de portefeuille pour les années antérieures à 1992.

Il est à noter une difficulté particulière, nécessitant un traitement spécifique, apparue lors de l'extraction des primes émises nettes d'annulations des contrats incendie pour l'exercice 1992 : les données relatives au risque agricole contiennent en effet le risque de mortalité du bétail, non retenu dans le périmètre d'étude.

I.2.2 - Présentation des méthodes de reconstruction du portefeuille

I.2.2.1 - Méthode d'agrégation des données brutes

La méthode suivie pour passer de données brutes à une information utilisable se décompose en cinq étapes :

- *l'extraction des données* relatives à chaque contrat appartenant au périmètre de l'étude et présent dans le fichier des contrats,
- *le passage du code INSEE de la commune du risque au code postal correspondant,*
- *l'agrégation de ces informations* contractuelles par exercice, risque, zone postale et entité,
- *le redressement opéré sur le code postal* pour remédier aux non-réponses concernant le lieu du risque,
- *et, enfin, l'agrégation de ces données* par exercice, risque et zone postale.

La **1ère étape** consiste en une extraction des contrats de type incendie présents sur le fichier de “ synthèse ” à la fin juin 1997. On y trouve, pour chaque contrat :

- une variable qualitative, correspondant à la présence, l’absence pour résiliation, ou la prochaine apparition du contrat au 31 décembre des exercices 1992 à 1996,
- et les primes émises nettes d’annulations (P.E.N.A.), pour les exercices 1993 à 1996.

On doit considérer comme absolument nécessaire la présence dans la sélection de la “ branche-risque ”, codification propre à l’UAP, permettant de répartir les contrats entre les trois catégories de risques “ habitation ”, “ agricole ” et “ entreprise ”.

De même, on ne peut envisager de bâtir une tarification de la garantie tempête sans disposer de la dimension géographique : dans le fichier de “ synthèse ”, celle-ci repose sur la codification INSEE des communes, retenue pour localiser le bien assuré.

Enfin, l’entité gestionnaire est l’autre composante à retenir : l’entité s’identifie au réseau auprès duquel l’affaire a été souscrite. On a distingué ici cinq types d’entités :

- le **réseau d’agents généraux** (réseau A), fort de 1.500 agences, qui est mandaté pour vendre les produits de la compagnie auprès de futurs assurés,
- le **réseau des courtiers** (réseau C), dont les membres sont des commerçants représentant leurs clients auprès de la société d’assurances,
- le **réseau des producteurs salariés** de la compagnie (réseau S), touchant un fixe et une commission d’apport fonction de leur production personnelle,
- le **réseau des guichets de la Banque Nationale de Paris** (B.N.P.), utilisé depuis 1991 pour la vente de produits d’assurance,
- enfin, la **délégation siège**, qui regroupe tous les risques gérés au siège administratif (ceux-ci font exception).

Le réseau d’agents généraux dispose de relais, appelés directions régionales, permettant un dialogue plus fréquent avec la compagnie. Celle-ci accorde d’ailleurs une certaine autonomie en matière tarifaire aux douze directions régionales mais, en revanche, la politique de règlement de sinistres reste très centralisée.

Nous verrons dans le paragraphe I.3 le rôle joué par l’entité gestionnaire, d’autant plus important que l’entité est corrélée avec le lieu du risque. A ce titre, une réflexion sera menée pour connaître la taille optimale pour laquelle ce lien est maximum.

La *deuxième étape* concerne le passage du code INSEE du risque au code postal correspondant. Cette opération est nécessaire car, nous le verrons plus loin, les bases “sinistres” et “météo” disposent seulement de cette division postale. Or, s’il est possible de passer d’un code INSEE à un code postal, moyennant certaines précautions, en revanche l’inverse est impossible dans la majorité des cas : en effet, la France métropolitaine compte 36 500 communes et environ 6 000 zones postales, ce qui explique qu’il faille généralement rattacher plusieurs codes INSEE à un seul code postal.

Le passage du code INSEE au code postal présente cependant une difficulté lorsque un même code INSEE correspond à plusieurs codes postaux : c’est le cas de nombreuses grandes villes, à l’exception de Paris, Lyon et Marseille, qui disposent d’un code INSEE par arrondissement.

La Poste fournit chaque mois à L’UAP une table de correspondances CPOSINS entre les codes INSEE et les codes postaux (les codes CEDEX en sont exclus). Dans cette table figurent les 38 959 couples codes INSEE - codes postaux existant à fin novembre 1997 : à titre d’illustration, Tours (code INSEE 37261) apparaît 3 fois sous les codes postaux 37000, 37100 et 37200. Un programme écrit sous APL, présent dans l’annexe B sous le nom de DEPTTEMP, permet de ne conserver dans la table CPOSINS qu’une seule occurrence de chaque code INSEE, soit un nombre total de 38 653, correspondant au code postal le plus petit (dans le cas présent : 37000).

De même, une attention toute particulière a dû être portée aux informations contenues dans la zone “PSIT”, situation du risque, du fichier des contrats en raison de réponses absentes ou erronées. Si la 4^{ème} étape de la méthode d’agrégation reviendra sur le problème des non-réponses, la 2^{ème} étape a tenté d’apporter une correction aux réponses erronées ou caduques : il n’est pas rare de trouver, dans la zone PSIT, le code postal à la place du code INSEE, ou un code INSEE, valide à l’époque de la souscription de l’affaire, mais plus en cours à la date de la constitution de la table CPOSINS, du fait de la fusion ou de la disparition de la commune du risque. On peut trouver aussi des contrats sans code INSEE.

C’est dans la procédure PORTEMP, programme écrit dans le langage HARRY, permettant l’interrogation de la base des contrats et présent dans l’annexe B, que les solutions sont apportées aux questions soulevées précédemment :

- si le contenu de la zone “ lieu du risque ” est un code INSEE, absent de la table CPOSINS, on lui affecte le code postal du chef-lieu du département auquel il appartient;
- si PSIT contient effectivement un code postal - ce que l’on peut vérifier par l’appartenance de PSIT à la 2^{ème} colonne de la table CPOSINS -, on le laisse en l’état;

- si PSIT = 0, le code postal affecté est mis à zéro; ce problème sera résolu au cours de la 4^{ème} étape.

La situation la plus fréquente est bien évidemment la présence dans la zone PSIT d'un code INSEE certifié, auquel il correspond un code postal unique dans la table univoque CPOSINS. Ce cas est la règle dans la mesure où le lieu du risque constitue une variable de tarification pour l'incendie; les seules traces d'erreurs concernent des produits non mécanisés, pour lesquels le code INSEE n'était pas déduit, comme aujourd'hui, du nom de la commune et du code postal du risque.

La *troisième étape* est l'agrégation des variables numériques, booléennes ou montant des P.E.N.A., extraites de chaque contrat lors de la 1^{ère} étape, selon la structure croisée exercice / risque / zone postale / entité. Les données se présentent sous la forme d'une matrice par entité, d'environ 10.000 lignes pour 12 colonnes pour les quatre entités à couverture nationale. Pour le réseau agents, une subdivision très précise à l'intérieur du fichier de " synthèse " offre la possibilité de constituer des matrices à l'échelle de directions régionales, de départements d'agences ou d'agences : le choix des départements sera justifié dans le paragraphe I.3 de ce chapitre.

La *quatrième étape* opère un redressement des informations présentes dans les matrices obtenues lors de la troisième étape; cette phase se justifie par la présence de contrats dont on ignore le lieu du risque. Les raisons de cette absence peuvent être les suivantes :

- l'adresse du souscripteur est très souvent identique au lieu du risque, qui de ce fait, par commodité, n'a pas été recopié par le gestionnaire du contrat (cette carence ne concerne heureusement que les contrats les plus anciens, dont la souscription se faisait par des chaînes non " mécanisées "); or l'adresse du souscripteur n'est pas contenue dans le fichier de " synthèse ",
- en risque d'entreprise, un contrat pouvant couvrir différents immeubles d'habitation ou établissements industriels, une tarification globale est alors appliquée sans référence à la tarification de chacun des risques de ce contrat.

Le redressement consiste, pour un risque et une entité donnés, à répartir le montant des variables agrégées " nombre de contrats au 31 décembre de l'année n " et " chiffre d'affaires de l'année n " en présence de la modalité " code postal nul ", sur le montant de ces variables pour les modalités au code postal non nul. Cette répartition se fait par péréquation du montant des variables agrégées de chaque modalité.

L'exemple du réseau S, celui des salariés, est utile pour illustrer ces propos : le tableau 2.1.1 présente la répartition par risque du nombre de contrats à la fin de l'année n et du chiffre d'affaires de cette même année, pour la modalité " code postal nul " et pour la modalité " code postal=1000 " avant et après redressement.

L'hypothèse sous-jacente sur laquelle repose le redressement est une répartition géographique identique, pour un risque et une entité donnés, entre les contrats avec code postal nul, et ceux avec un code postal différent de zéro. Nous reviendrons sur cette hypothèse au paragraphe suivant.

Tableau 2.1.1 : Illustration de la méthode de redressement des matrices avec code postal absent

Etape	Modalité du code postal	RISQUE					
		AGRICOLE		HABITATION		ENTREPRISE	
		nombre	C.A.(en kF)	nombre	C.A.(en kF)	nombre	C.A.(en kF)
Avant redressement	nul	60	122	1233	1972	562	39106
	1000	1	4	193	232	92	2562
Après redressement	nul	0	0	0	0	0	0
	1000	1	4	193	233	92	2603

On répartit en outre les contrats affectés à Paris, Lyon et Marseille sans autre précision, entre les différents arrondissements de ces 3 villes.

La *cinquième étape* est l'agrégation de ces données par exercice, risque et zone postale : l'intervention de l'entité ne visait que l'objectif d'une plus grande précision lors du redressement de la 4ème étape. L'entité ne joue qu'un rôle intermédiaire et ne doit donc pas être conservée.

I.2.2.2 - Reconstitution du chiffre d'affaires pour l'exercice 1992

Dans les variables portefeuille reconstituées, ne figuraient pas au paragraphe précédent les P.E.N.A. pour l'exercice 1992, le fichier de synthèse à fin 1997 ne les contenant plus. Fort heureusement, ce fichier a fait l'objet d'une extraction antérieure à la précédente, par laquelle le chiffre d'affaires de l'exercice 1992 a été ventilé par code commune. Cependant, la définition du périmètre sélectionné diffère légèrement de celle que nous avons retenue :

- le chiffre d'affaires n'est généré que par les contrats en cours à l'inventaire 1992, et non par ceux résiliés au cours de cet exercice; cet écart représente près de 5 % du total;
- le chiffre d'affaires est réparti selon une typologie comptable des contrats appelée " groupe " insatisfaisante : en effet, cette codification ne correspond pas au niveau de détail souhaité puisqu'elle englobe, dans le risque agricole, le contrat " mortalité du bétail ". Le périmètre de l'étude, défini par la suite, n'a pas retenu ce contrat.

Le chiffre d'affaires de l'exercice 1992 étant une donnée nécessaire à l'étude, il faut procéder à la suppression du contrat " mortalité du bétail " du périmètre défini pour le risque agricole, soit environ 2,7 % du nombre de contrats agricoles.

Pour extraire le chiffre d'affaires des contrats " mortalité du bétail " de la distribution du risque agricole de 1992, une idée consiste à recréer la ventilation par code INSEE de ces contrats indésirables pour l'exercice le plus ancien à notre disposition, 1993. De la comparaison de ces deux distributions (contrats agricoles 1992 d'une part, contrats " mortalité du bétail " d'autre part), il sera possible de procéder à une identification au moins partielle des contrats " mortalité du bétail ".

Une extraction issue du fichier de " synthèse " permet la création d'un fichier contenant, pour chaque contrat " mortalité du bétail " :

- le code INSEE de la commune du lieu du risque,
- la prime émise nette d'annulation du contrat, pour l'exercice 1993,
- la date d'entrée en risque du contrat : si celle-ci est antérieure à l'inventaire 1992, le contrat était par conséquent en cours à cette date et il convient de le retirer de la distribution de 1992,
- et, enfin, le numéro du contrat " mortalité du bétail ", permettant d'en déduire le code entité d'appartenance en vigueur à l'inventaire 1993 (il faut savoir en effet que la répartition du chiffre d'affaires 1992 par entité est connue, ce qui permet d'établir un lien entre la répartition par entité et celle par code INSEE, et de valider a posteriori, au niveau régional, la reconstitution du chiffre d'affaires par code INSEE).

Ce fichier recense tous les contrats indésirables et en cours aux inventaires 1992 et 1993, soit près des trois quarts. Seuls les contrats en cours à l'inventaire 1992 et résiliés avant celui de 1993 restent inaccessibles.

Un outil, mis à notre disposition par le service Etudes de la Direction de l'informatique, permet d'accéder avec plus de facilité à l'énorme fichier contenant l'ensemble des mouvements, de production, de résiliation, de remplacement, etc...., ayant affecté les contrats dommages produits depuis 1987 ou résiliés depuis 1983. Y figurent bien évidemment les mouvements de résiliation, enregistrés en 1993, des contrats " mortalité du bétail ", et les informations utiles à leur étude, à l'exception de la prime nette d'annulation 1993 : nous aurions pu la calculer à partir de la technique du *prorata temporis*, mais cela exigeait un temps précieux.

Une fois la ventilation par code INSEE des contrats " mortalité du bétail " présents à l'inventaire 1992 recréée, l'étape suivante consiste à la comparer à la distribution du chiffre d'affaires agricole 1992 pour en extraire la part afférente aux contrats

“ mortalité du bétail ”. La technique d’identification utilisée ici repose sur l’idée sous-jacente que, pour un grand nombre de codes INSEE, il se trouve un même nombre de contrats agricoles (n_A^n) et “ mortalité du bétail ” (n_{MB}^n) entre les deux distributions, ce qui permet d’extraire sans estimation de la ventilation agricole les contrats indésirables.

Il est aussi possible d’utiliser cette technique grâce à la connaissance des P.E.N.A. “ mortalité du bétail ” 1992 par entité : comme l’entité d’appartenance est connue pour chacun des contrats “ mortalité du bétail ”, un lien peut être fait entre le code INSEE et l’entité et il devient possible d’identifier les indésirables dès que le chiffre d’affaires d’une entité est exclusivement réalisé sur un seul code INSEE (on a alors : $n_{MB}^n = n_{entité}^n$).

Dans tous les autres cas, soit près d’une fois sur deux, il faut procéder à l’estimation de la prime 1992 par code commune, tenant compte de la part du risque “ mortalité du bétail ” en 1992 relativement au risque agricole et à ce même risque en 1993. La formule mathématique est la suivante :

$$\hat{P}_{MB}^n = p_A P_A^n + p_{MB} P_{MB}^{n+1}, \text{ où :}$$

- \hat{P}_{MB}^n est l’estimation de la prime nette d’annulation du contrat “ mortalité du bétail ” (MB) pour l’année 1992 (n),
- P_A^n représente la prime nette d’annulation du contrat agricole pour l’année 1992,
- P_{MB}^{n+1} représente la prime nette d’annulation du contrat “ mortalité du bétail ” pour l’année 1993 ($n+1$),
- p_A et p_{MB} représentent la pondération respective, dont la somme égale l’unité, des primes agricole 1992 et “ mortalité du bétail ” 1993; elle fait intervenir le rapport du nombre de contrats “ mortalité du bétail ” pour l’année 1993 (n_{MB}^{n+1}), et du nombre de contrats agricoles pour l’année 1992 (n_A^n).

Concrètement, p_A et p_{MB} ont la formulation suivante :

$$P_A = \frac{\binom{n}{n_A}^{n+1}}{\binom{n}{n_A}^n} \left(\frac{\binom{n}{n_A}^{n+1}}{\binom{n}{n_A}^n} \right)^r, \quad P_{MB} = \frac{1}{\binom{n}{n_A}^{n+1} \left(\frac{\binom{n}{n_A}^{n+1}}{\binom{n}{n_A}^n} \right)^r + 1}, \quad \text{avec } r \in \mathbb{N}^*.$$

La prime agricole aura une importance d'autant plus grande que le nombre de contrats agricoles sera petit relativement au nombre de contrats " mortalité du bétail " 1993 : on comprend aisément que, plus n_A^n est faible, plus n_{MB}^n est grand en proportion et influe sur le niveau finalement atteint par la prime P_A^n . L'utilisation du coefficient r permet de renforcer le poids de la prime pour laquelle l'influence du contrat " mortalité du bétail " est la plus grande (i.e. celle dont le rapport $\frac{n_{MB}^n}{n_{MB}^{n+1}}$, pour les contrats " mortalité du bétail " 1993, ou le rapport $\frac{n_{MB}^n}{n_A^n}$, pour les contrats agricoles 1992, est le plus proche de 1).

Pour fixer les esprits, nous pouvons prendre un cas concret du calcul effectué, qui montre l'impact variable des deux primes selon le coefficient r .

Dans l'exemple présenté dans le tableau 2.1.2, l'égalité du nombre de contrats " mortalité du bétail " entre 1992 et 1993 (4) renforce la probabilité que les contrats présents en 1992 le soient aussi en 1993, avec une prime nette légèrement réévaluée selon l'évolution de l'indice FNB.

Tableau 2.1.2 : Illustration de la méthode de calcul de la prime 1992 du contrat " mortalité du bétail "

n_A^n	n_{MB}^n	n_{MB}^{n+1}	P_A^n	P_{MB}^{n+1}	r	P_A	\hat{P}_{MB}^n
10	4	4	26766	2169	1	0,40	4545
					2	0,16	3270
					3	0,064	2599
					4	$2,56 \cdot 10^{-2}$	2296
					5	$1,024 \cdot 10^{-2}$	2168
					6	$4,96 \cdot 10^{-3}$	2116
					∞	0	2080

D'autre part, la connaissance de la valeur du rapport $\frac{n_A^n}{n_{MB}^{n+1}}$ relativement à 1 permet d'en déduire quelle prime est la plus influente et l'emporte sur l'autre lorsque r tend vers ∞ : ainsi, si $\frac{n_A^n}{n_{MB}^{n+1}}$ est supérieur à 1 (i.e. si le poids représenté par le nombre des contrats MB en 1992 est plus fort relativement au nombre de contrats MB en 1993 qu'au nombre de contrats agricoles en 1992), le numérateur de p_A tend donc vers 0 lorsque r croît. Sans idée a priori, r est fixée à 1.

(Par suite, r a été estimée a posteriori à 3, comme étant la valeur minimisant l'écart entre la somme des \hat{P}_{MB}^n sur l'ensemble des codes commune et le montant du chiffre d'affaires généré par les contrats " mortalité du bétail " en 1992.)

On réajuste alors les \hat{P}_{MB}^n de telle manière que le total de ces estimations par entité soit égal au volume des P.E.N.A. 1992 par entité dont on dispose par ailleurs. Ce rééquilibrage est possible en raison de l'existence d'une territorialité, du moins au sein du réseau agents : en effet, dans la majorité des cas, une entité regroupe plusieurs codes INSEE qu'elle est la seule à exploiter.

On retire alors ce volume de contrats et de P.E.N.A. à la série des contrats agricoles.

On verra dans le paragraphe suivant la qualité de ces estimations : la répartition réelle par entité du chiffre d'affaires pour l'exercice 1992 du produit " mortalité du bétail ", mise à notre disposition, autorise en effet sa comparaison avec la distribution de cette variable pour les contrats identifiés précédemment.

On trouvera dans l'annexe C le détail du programme exécuté sous SAS et appelé SUP36134, en référence à la codification comptable du produit " mortalité du bétail ",

Un ajustement ayant été apporté aux données du risque agricole, il est alors possible d'apporter les modifications suivantes à l'ensemble de la matrice 1992 du chiffre d'affaires 1992 ventilé par risque et code INSEE :

- utilisation de la table de correspondances univoques entre le code INSEE et le code postal pour substituer au code INSEE le code postal, selon les dispositions définies au paragraphe précédent,
- prise en compte dans le chiffre d'affaires 1992 des primes nettes d'annulations générées par les contrats résiliés au cours de cet exercice, au prorata du chiffre d'affaires connu pour chaque zone postale,
- en raison de la présence du code postal " zéro ", application de la technique de redressement, exposée plus haut, à la série 1992, et ceci pour chacun des 3

risques incendie, au prorata du nombre de contrats et du montant des P.E.N.A. pour 1992 de chaque code postal.

Une autre technique d'estimation du chiffre d'affaires 1992 n'a pas finalement été retenue : elle consiste à utiliser les nombres de contrats, pour 1992 et 1993, et le chiffre d'affaires, pour 1993, et à considérer que ces deux types d'indicateurs ont une évolution convergente. Mais l'évolution souvent erratique du chiffre d'affaires par rapport au nombre de contrats en portefeuille ne conduit pas à de bons résultats.

I.2.3 - Mesure de la qualité de la méthode d'estimation

Au paragraphe précédent, le résultat que nous avons obtenu découle des hypothèses faites et des techniques utilisées. Il convient de justifier ces hypothèses et techniques et de voir s'il est possible de mesurer le niveau de précision de ces méthodes.

I.2.3.1 - Méthode de localisation des contrats non localisés

L'agrégation des données brutes pour les années 1992 à 1996 nous a orientés vers l'emploi d'une méthode visant à localiser les contrats pour lesquels le lieu du risque est indéterminé, c'est-à-dire à leur attribuer le code postal le plus *vraisemblable*. La vraisemblance du code postal repose sur le bon sens qui a prévalu lors du choix de l'hypothèse retenue pour mener les calculs à terme : la distribution des contrats " non localisés " (i.e. avec code INSEE indéterminé) est supposée identique à celle des contrats " localisés ".

Il convient, à ce stade de notre réflexion, de valider autant que faire se peut l'hypothèse d'*équidistribution*. Concrètement, la tâche revient à calculer la distance du khi-2 entre les distributions de contrats localisés et de contrats non localisés. Or, si la première de ces distributions est parfaitement connue, par définition la seconde, objet de nos recherches, ne l'est pas.

Pour passer outre cet écueil, une échappatoire consiste à limiter la validation au réseau agents (qui représente 65 % du chiffre d'affaires global), et à comparer ces 2 distributions, non pas selon la zone postale du risque, mais selon l'entité départementale : en effet, comme nous le verrons plus loin en détail, il existe une très forte concentration du portefeuille d'une agence autour du lieu d'implantation de celle-ci. Il n'est donc pas vain de tenir pour probable l'appartenance du contrat non localisé au département de l'agence à laquelle il est rattaché. C'est d'ailleurs cette proximité entre le risque et l'agence gestionnaire qui a prévalu dans le choix d'une répartition des contrats non localisés au niveau départemental et non national.

Dans une première phase, un tableau de contingence est constitué (cf. les notations dans le tableau 2.1.3), renfermant le nombre de contrats par département de l'agence et département du risque, pour *les contrats localisés*. Le taux de contrats non-localisés tend à décroître au cours des années : le choix de l'exercice retenu, 1993, n'est donc

pas neutre. La distance du χ^2 , notée d^2 et calculée sur ce tableau croisé, mesure l'écart à l'indépendance entre les deux variables à partir de la formule classique :

$$d^2 = \sum_i \sum_j \frac{(n_{ij} - \frac{n_{i.} \cdot n_{.j}}{n})^2}{\frac{n_{i.} \cdot n_{.j}}{n}}$$

Dans le cas présent, d^2 prend les valeurs suivantes :

- $d^2 = 4.267.913$ pour le risque agricole,
- $d^2 = 11.769.902$ pour le risque professionnel,
- $d^2 = 95.538.308$ pour le risque d'habitation.

Tableau 2.1.3 : Tableau de contingence de deux variables X et Y qualitatives à r et s modalités

X : département de l'agence

Y : département du risque

r = s = 95

X	Y	y ₁	y ₂	y _j	y _s	
x ₁		n ₁₁	n ₁₂				n _{1s}	n _{1.}
x ₂		n ₂₁	n ₂₂				n _{2s}	n _{2.}
⋮								
x _i					n _{ij}			n _{i.}
⋮								
x _r		n _{r1}	n _{r2}				n _{rs}	n _{r.}
		n _{.1}	n _{.2}		n _{.j}		n _{.s}	n

On fixe le risque d'erreur à $\alpha = 1\%$, c'est-à-dire à une faible probabilité de trouver, en cas d'indépendance des 2 variables, une valeur supérieure au c_n^2 correspondant, v étant très supérieur à 100, il est d'usage de recourir à l'approximation de Fisher :

$$\sqrt{2c_n^2} - \sqrt{2n-1} \approx U_\alpha,$$

où U_α est le fractile de la loi normale centrée réduite tel que :

$$\alpha = P(U > U_\alpha) = 1\%.$$

La lecture de la table de la loi normale donne : $U_{\alpha} = 2,3$; l'on en déduit la valeur critique c_n^2 à partir de laquelle l'indépendance sera rejetée :
 $c_n^2 = 9147,9$.

Ce résultat montre la quasi identité des 2 variables, car d^2 est très supérieur à la valeur critique.

Par la suite, ne disposant que de la répartition, par département de l'agence, du nombre de *contrats non localisés (distribution a)*, on tire, du tableau de contingence 2.1.3 croisant le nombre de contrats localisés 1993 par département de l'agence et département du risque, la distribution marginale par département de l'agence (*distribution b, égale à la série (n_j)* du tableau 2.1.3). Une distance entre ces deux distributions *a* et *b* peut être construite sur la base de la distance du χ^2 dont la définition est la suivante :

$$d_{ab}^2 = \sum_{i=1}^k \sum_{j=1}^r \frac{(n'_{ij} - n'_{i.} p'_{.j})^2}{n'_{i.} p'_{.j}}$$

Il est en effet fait l'hypothèse que les deux distributions constituent des échantillons issus de deux variables aléatoires ventilées par département de l'agence $j, j=1, \dots, 95$:

- nombre de contrats localisés ($n'_{1.}$), égale à la série ($n_{i.}$) du tableau 2.1.3,
- nombre de contrats délocalisés ($n'_{2.}$).

n'_{ij} est le nombre d'individus de l'échantillon appartenant au département de l'agence j de la variable $n'_{i.}$, effectif de l'échantillon i ; $n'_{.j} = \sum_{i=1}^k n'_{ij}$ est le nombre total des individus appartenant au département de l'agence j .

Il s'agit de montrer que ces deux échantillons proviennent d'une même population : c'est l'hypothèse H_0 que l'on fait lors de l'écriture du test du χ^2 , contre H_1 : " les échantillons sont significativement différents ".

Dans l'hypothèse H_0 on peut parler de probabilités $p'_{.1}, p'_{.2}, \dots, p'_{.r}$ de posséder les modalités 1, 2, ..., r . Il s'agit alors de comparer les effectifs constatés n'_{ij} aux effectifs espérés $n'_{i.} p'_{.j}$, qui ne doivent pas en être très éloignés.

Dans l'hypothèse H_0 , d_{ab}^2 est une réalisation d'une variable d_0^2 suivant un χ^2 dont le nombre de degrés de liberté est égal au nombre de termes de d^2 (2.r), auquel on doit

retirer k , le nombre de relations indiquant que les sommes de lignes sont constantes et égales :

$$\sum_j n'_{ij} = \sum_j n'_i \cdot p'_{ij} = n'_i.$$

Donc d_0^2 est un c_{kr-k}^2 .

Cependant, comme les $p'_{1j}, p'_{2j}, \dots, p'_{rj}$ ne sont pas connus, il faut les estimer par $\hat{p}'_{ij} = \frac{n'_{ij}}{n'_i}$, où n' est le nombre d'individus d'un échantillon. Ceci représente $r - 1$ relations indépendantes : en effet, pour estimer les r probabilités, on n'a besoin que de $r - 1$ relations car $\sum_{j=1}^r p'_{ij} = 1$. d_0^2 est alors un $c_{(kr-k)-(r-1)}^2 = c_{(k-1)(r-1)}^2$ si H_0 est vraie.

Le test consistera à rejeter H_0 si d_{ab}^2 constaté est trop grand, comme pour un test habituel du χ^2 .

Par la suite, une table peut nous fournir la précision correspondant à d_{ab}^2 , c'est-à-dire la probabilité que les échantillons soient considérés à tort comme ayant une origine identique.

Le calcul de d_{ab}^2 sur l'année 1993 donne, tous risques confondus, 201,1, n' étant exprimé en dizaines d'unités et valant 3076. Pour un χ^2 à 94 degrés de liberté, cette valeur correspond à une probabilité inférieure à 1 % de considérer à tort les deux échantillons comme émanant d'une même source : ce risque est donc très faible. En revanche, à un seuil de 5 %, comme l'on a $P(c_{94}^2 > 118) = 5 \%$, nous ne pouvons accepter l'hypothèse d'une équidistribution géographique des contrats non localisés et de ceux qui ne le sont pas.

Néanmoins, le rapport entre la distance du khi-deux et le nombre d'observations, qui est seulement de 6,5 %, montre le bien-fondé de la disposition consistant à « localiser » les contrats délocalisés dans le département de l'agence à laquelle ils sont rattachés.

I.2.3.2 - Suppression du contrat “ mortalité du bétail ” du risque agricole

La précision de notre estimation du chiffre d'affaires par zone postale débarrassé du contrat “ mortalité du bétail ” est d'autant meilleure qu'est faible la part de ces contrats indésirables.

Contrairement au paragraphe précédent, la précision liée à l'opération d'extraction des contrats indésirables n'est pas évaluée à partir de l'hypothèse de répartition des P.E.N.A. proportionnellement au poids de chaque modalité. En effet, pour près de la moitié des modalités, il est possible de connaître la valeur exacte des P.E.N.A.. Pour l'autre moitié, l'estimation des primes 1992 du contrat “ mortalité du bétail ”, faite au niveau de chaque code INSEE, est une moyenne des primes nettes d'annulations, pondérée par le poids respectif des primes du risque agricole en 1992 et de celles du contrat indésirable en 1993 (cf. la formulation mathématique présentée au paragraphe I.2.2.2).

Pour cette opération, la qualité de l'estimation se fait naturellement par la comparaison des ventilations réelle et estimée du chiffre d'affaires par entité régionale. En effet, par définition cette ventilation n'est pas connue par zone postale. Néanmoins, l'entité régionale étant corrélée avec la région du risque, il peut être astucieux d'utiliser cette variable pour évaluer la qualité de l'estimation.

Le tableau 2.1.4 suivant permet de mieux étayer nos propos.

Ainsi, puisqu'il n'est pas possible de calculer la distance entre \tilde{N}_Y et N_Y , ce dernier étant inconnu, il faut s'en remettre au calcul de la distance entre \tilde{N}_X et N_X , qui, lui, est connu. Ce calcul se justifie par l'existence d'un écart à l'indépendance très important entre X et Y , comme le montre le tableau 2.1.5 fournissant la répartition du chiffre d'affaires du contrat “ mortalité du bétail ” selon l'entité régionale et la région du risque. On se limite pour ce tableau de contingence au réseau Agents qui est le seul réseau dont la structure de distribution soit régionalisée.

La distance du liée à cette matrice est de : 33.422. Ce chiffre est à comparer à la borne supérieure de l'écart à l'indépendance des deux variables, égale au produit de la somme de la matrice (3.863 kF) par le nombre de colonnes amputé de 1 (11), soit : 44.493. La relation entre X et Y est donc presque fonctionnelle.

A ce niveau de dépendance, il est donc légitime de mesurer la qualité de l'estimation à travers l'écart entre les distributions réelle et estimée du chiffre d'affaires par entité régionale. Cet écart est égal à : 216,4. Or, dans l'hypothèse où les distributions sont issues d'une même source, le χ^2 à 14 degrés de libertés est égal à 23,7 pour un risque d'erreur de 5% : on ne peut donc retenir l'hypothèse que les distributions sont équivalentes. Néanmoins, le rapport entre la distance du k hi-deux et la somme des distributions, qui est seulement de 3,5 %, montre la légitimité de la méthode d'identification des contrats « mortalité du bétail ».

Tableau 2.1.4 : Principe d'utilisation de l'entité régionale pour l'évaluation de la qualité de l'estimation du chiffre d'affaires du contrat " mortalité du bétail " réparti selon la région du risque

X : entité régionale
 Y : région du risque
 r = s = 12

\tilde{N}_{XY}	y_1	y_2	y_j	y_s	\tilde{N}_X (estimé)
x_1	\tilde{N}_{11}	\tilde{N}_{12}				\tilde{N}_{1s}	\tilde{N}_1
x_2	\tilde{N}_{21}	\tilde{N}_{22}				\tilde{N}_{2s}	\tilde{N}_2
\vdots							
x_i				\tilde{N}_{ij}			\tilde{N}_i
\vdots							
x_r	\tilde{N}_{r1}	\tilde{N}_{r2}				\tilde{N}_{rs}	\tilde{N}_r
\tilde{N}_Y (estimé)	$\tilde{N}_{.1}$	$\tilde{N}_{.2}$		$\tilde{N}_{.j}$		$\tilde{N}_{.s}$	n

N_X (réel connu)
n_1
n_2
n_i
n_r
n

N_Y (réel inconnu)	n_1	n_2		n_j		n_s	n

Tableau 2.1.5 : Répartition du chiffre d'affaires du contrat " mortalité du bétail " selon l'entité régionale et la région du risque (en kF)

Entité / région	Ouest	Est	Rhône-Alpes	Sud-ouest Mon.	Sud-Est	Alsace-Moselle	Lorraine	Sud-ouest Toul.	Centre	Nord	Bassin parisien	Paris
Ouest	184		1					2	3			
Est		360	1			1			9			2
Rhône-Alpes			360		1				8			
Sud-ouest Montauban				208	2			1	3			
Sud-Est			5	3	219			1	3			3
Alsace-Moselle						23	5					
Lorraine					1		77					
Sud-ouest Toulouse								1196				8
Centre	2	1	8						491			
Nord		7								117		
Bassin parisien	2	5							33		79	1
Paris	9	2	10	3	15	9	7			9	128	478

I.2.3.3 - Autres corrections apportées au chiffre d'affaires 1992

En revanche, la nécessité du redressement de la ventilation du chiffre d'affaires 1992 due :

- à la modalité “ zone postale = 0 ”,
- à un chiffre d'affaires généré par les seuls contrats résiliés durant l'exercice 1992,

ne peut se faire qu'en se fondant sur l'hypothèse, déjà retenue pour les exercices 1993 à 1996, d'une répartition :

- du chiffre d'affaires 1992 de la modalité “ zone postale = 0 ”,
- ou du chiffre d'affaires 1992 généré par les contrats résiliés en 1992,

selon le poids des autres modalités.

Cette estimation \hat{e} est obtenue en additionnant un vecteur r_1 , qui correspond à ce qui est *déjà réparti* (i.e. le chiffre d'affaires ventilé par zone postale « non nulle » ou le chiffre d'affaires 1992 généré par les contrats non résiliés en 1992), et \hat{r}_2 , une estimation par zone postale de *ce qui est à répartir*. En notation matricielle, \hat{e} prend la forme suivante :

$$\hat{e} = r_1 + \hat{r}_2 = \begin{bmatrix} n_i \\ \vdots \end{bmatrix} + \begin{bmatrix} n'_i \cdot \frac{n - n'}{n'} \\ \vdots \end{bmatrix} = \begin{bmatrix} n'_i \cdot \frac{n}{n'} \\ \vdots \end{bmatrix}, \text{ dans laquelle :}$$

- n est la somme des n_i , ($i \in [1; m]$) où m est le nombre de zones postales pour lesquelles $r_1(i)$ n'est pas nul), c'est-à-dire le chiffre d'affaires 1992 total (réparti ou à répartir),
- n' est la somme des n'_i , c'est-à-dire le total du chiffre d'affaires 1992 déjà réparti.

Or, contrairement au paragraphe précédent, il n'est pas possible d'apprécier l'écart entre les distributions réelle r et estimée \hat{e} par *zone postale* par la distance entre les distributions réelle et estimée par *entité*. Il faut donc procéder autrement pour juger de la qualité de cette distribution par zone postale.

En l'absence d'une variable de contrôle comme l'entité, il nous reste néanmoins la possibilité de trouver un majorant à la distance entre \hat{e} et la distribution réelle r : en effet, la distance maximale sera atteinte pour $r_{2 \max}$ ainsi défini :

- $\forall i \neq i_0, r_{2 \max}(i) = 0, i \in [1; m],$
- $r_{2 \max}(i_0) = \sum_{i=1}^n (r(i) - r_1(i)) = n - n', \text{ et } r_1(i_0) = n'_{i_0} = \underset{i}{\text{Min}} r_1(i) \geq 1.$

Dans ce cas d'espèce, alors $r_{\max} = r_1 + r_{2 \max}$: la distribution réelle présente la particularité d'être le plus possible éloignée de la distribution estimée. L'expression de r_{\max} est :

$$r_{\max} = \begin{bmatrix} n'_1 \\ \cdot \\ n'_i \\ \cdot \\ n'_{i_0-1} \\ n'_{i_0} + n - n' \\ n'_{i_0+1} \\ \cdot \\ n'_m \end{bmatrix}$$

Le calcul de leur écart se fait selon la définition précédemment donnée du k hi-deux :

$$d^2_{\max} = \sum_{i=1}^m \sum_{j=1}^2 \left(\frac{n_{ij} - \frac{n_i \cdot n_j}{n}}{\frac{n_i \cdot n_j}{n}} \right)^2 = \sum_{i \neq i_0} \left\{ \frac{\left(n'_i \cdot \frac{n}{n'} - \frac{n'_i \cdot (1 + \frac{n}{n'})}{2} \right)^2}{\frac{n'_i \cdot (1 + \frac{n}{n'})}{2}} + \frac{\left(n'_i - \frac{n'_i \cdot (1 + \frac{n}{n'})}{2} \right)^2}{\frac{n'_i \cdot (1 + \frac{n}{n'})}{2}} \right\}$$

$$+ \frac{\left(n'_{i_0} \cdot \frac{n}{n'} - \frac{n'_{i_0} \cdot (\frac{n}{n'} + 1) + n - n'}{2} \right)^2}{\frac{n'_{i_0} \cdot (\frac{n}{n'} + 1) + n - n'}{2}} + \frac{\left(n'_{i_0} + n - n' - \frac{n'_{i_0} \cdot (\frac{n}{n'} + 1) + n - n'}{2} \right)^2}{\frac{n'_{i_0} \cdot (\frac{n}{n'} + 1) + n - n'}{2}}$$

La distance maximale vaut, après calcul :

$$d^2_{\max} = \frac{(n'-n)^2 \cdot (n'-n'_{i_0})}{n' \cdot (n'+n)} + \frac{(2n'_{i_0} (1 - \frac{n}{n'}) + n - n')^2}{n'_{i_0} (1 + \frac{n}{n'}) + n - n'}$$

Dans notre cas, on peut trouver des zones postales où n'_{i0} est égal à 0, alors que le nombre de contrats n'_y est pas nul. L'expression de la distance maximale prend alors la tournure suivante :

$$d^2_{\max} = \frac{(n'-n)^2 \cdot n'}{n' \cdot (n'+n)} + \frac{(n-n')^2}{n-n'} = (n-n') \left(\frac{n-n'}{n+n'} + 1 \right) = \frac{n-n'}{n+n'} \cdot 2n$$

Il est logique que la distance maximale soit proportionnelle à la somme à répartir ($n-n'$) : plus cette différence sera faible, plus la précision de l'estimation sera grande.

Le calcul de d^2_{\max} associé au redressement du chiffre d'affaires 1992 suite à :

- la présence de la modalité “ zone postale = 0 ”,
- un chiffre d'affaires généré par les seuls contrats non résiliés durant l'exercice 1992,

conduit aux valeurs suivantes, le chiffre d'affaires étant exprimé en k ilo-franc :

- “ zone postale = 0 ” : cet écart est égal à 266,5. Or, pour un risque d'erreur de 1 % de considérer à tort que les distributions sont équivalentes, le χ^2 à 6.000 degrés de libertés, correspondant au nombre de zones postales, peut être apprécié, par approximation de Fisher, à 6.254. On peut donc faire nôtre l'hypothèse l'équivalence des distributions. D'ailleurs, le rapport entre la distance du khi-deux et la somme des distributions, qui est seulement de 3,2 %, montre la légitimité de la méthode de redressement utilisée,
- chiffre d'affaires généré par les seuls contrats non résiliés : l'écart, égal à 293,2, est très en deçà du χ^2 à 6.000 degrés de libertés au taux de confiance de 99 %. Le rapport entre la distance du khi-deux et la somme des distributions, qui est seulement de 3,6 %, montre que, en définitive, la distribution r_1 n'est pas tellement « contrariée » par l'aléa r_2 .

Comme on le voit, le calcul de l'écart maximum ne peut fournir qu'une idée minimaliste de la véritable qualité de l'estimation. L'hypothèse d'équiprobabilité des partitions du vecteur r_2 conduit à des calculs impossibles à mener à terme en raison de l'incommensurabilité du nombre de partitions. De plus, cette hypothèse est, elle aussi, très peu ancrée dans la réalité et aurait conduit à une précision très en deçà de la réalité. Il faut donc se contenter de ces estimations sans pouvoir juger vraiment de leur qualité, en se souvenant tout de même qu'elles reposent sur des hypothèses de bon sens.

I.3 - Reconstitution du portefeuille pour les années 1979 à 1991

I.3.1 - Détermination du chiffre d'affaires 1992 par entité et zone postale

Après la reconstruction du portefeuille pour 1992, exprimé en nombre de contrats et en chiffre d'affaires, dont on a la distribution par zone postale et celle par entité, il est ensuite nécessaire de procéder à celle du portefeuille de l'exercice 1991 par zone postale. Pour ce dernier, seule est connue la répartition par entité.

Pour ce faire, les conventions de notation suivantes sont adoptées :

- $P_n(ZP, entité)$ est la matrice croisant le portefeuille P (C pour le chiffre d'affaires ou N pour le nombre de contrats recensés à l'inventaire) pour l'exercice n selon la zone postale (ZP) et l'entité,

- $\hat{P}_n(ZP, entité)$ est l'estimation de cette matrice,

- $P_n(ZP, .)$ est la distribution, inconnue, du portefeuille selon la zone postale, ou encore la marge verticale de la matrice $P_n(ZP, entité)$; elle sera estimée par $\hat{P}_n(ZP, .)$,

- $P_n(., entité)$ est la distribution, connue, du portefeuille selon l'entité, ou encore la marge horizontale de la matrice $P_n(ZP, entité)$.

L'édification de $\hat{P}_n(ZP, .)$ passe par le recours à la variable entité, qui est une rupture connue du portefeuille. Le principe de cette édification est le suivant :

$$P_n(., entité) + \hat{P}_{n+1}(ZP, entité) \rightarrow \hat{P}_n(ZP, entité) \rightarrow \hat{P}_n(ZP, .)$$

La justification du choix de cette variable et le détail de cette méthode d'estimation seront apportés au paragraphe suivant. Ce qu'il importe ici de voir, c'est la nécessité de connaître la matrice $\hat{P}_{n+1}(ZP, entité)$ servant d'intermédiaire de calcul. Or, nous ne disposons pas de la matrice $C_{92}(ZP, entité)$. Il faut donc la rebâtir.

Le principe schématique de son édification peut être décrit de la manière suivante :

$$C_{92}(., entité) + \hat{C}_{92}(ZP, .) + C_{93}(ZP, entité) \rightarrow \hat{C}_{92}(ZP, entité)$$

La matrice inconnue, dont on connaît néanmoins les marges, peut être estimée grâce à l'apport décisif de cette même matrice pour 1993, qui offre une structure croisée très similaire à la matrice de 1992.

Ce type de problème trouve sa résolution par l'emploi de méthodes de redressements sur critères multiples. Ces techniques d'ajustement sont relativement récentes en raison du coût informatique de leur mise en place. En dehors de celle reposant sur l'analyse factorielle des correspondances, on en recense principalement deux : la méthode ASAM (Ajustement Statistique et Algébrique d'une Matrice) et la méthode du " Ratio Algorithm system " (RAS).

La méthode ASAM, imaginée à l'INSEE par Durieux et Payen, a pour objet de recréer une structure croisée (x_{ij}) respectant au mieux le morphisme de celle de (a_{ij}) . Elle consiste en la résolution du programme suivant :

$$\left| \begin{array}{l} \text{Min} \sum_i \sum_j \frac{1}{r_{ij}} (x_{ij} - a_{ij})^2 \\ \text{avec} \sum_j x_{ij} = r_i \\ \sum_i x_{ij} = s_j \\ \sum_i \sum_j x_{ij} = \sum_i \sum_j a_{ij} = T \end{array} \right.$$

L'intérêt de la présence des r_{ij} est de pouvoir moduler l'importance de chaque case. On peut montrer que le choix optimal des r_{ij} est de les prendre proportionnels aux variances des a_{ij} , considérés comme des variables aléatoires.

$$r_{ij} = c s^2_{ij} \text{ avec } s^2_{ij} = V(a_{ij})$$

Cette méthode, satisfaisante sur le point de vue du respect des structures, est en revanche plus coûteuse que la méthode RAS, dont le principe mécanique consiste à ajuster les lignes et colonnes de (a_{ij}) sur les marges r_i et s_j . Concrètement, la n ème itération de l'algorithme s'écrit :

$$\left| \begin{array}{l} x_{ij}^{(n)} = x_{ij}^{(n-1)} \cdot \frac{1}{1 + \frac{J_{i.}^{(n)} + J_{.j}^{(n)}}{2}} \\ \text{avec:} \\ J_{i.}^{(n)} = \frac{x_{i.}^{(n-1)}}{r_i} - 1 \\ J_{.j}^{(n)} = \frac{x_{.j}^{(n-1)}}{s_j} - 1 \end{array} \right.$$

L'évolution de chaque case constatée au cours de cette itération dépend pour une même part de l'écart :

- entre le total de la ligne à laquelle elle appartient ($x_i^{(n)}$) et l'élément de la marge verticale qui lui correspond (r_i),
- et entre le total de la colonne à laquelle elle appartient ($x_{.j}^{(n)}$) et l'élément de la marge horizontale qui lui correspond (s_j).

Nous laissons au lecteur le soin de montrer la convergence de la suite ($x_{ij}^{(n)}$) vers une et une seule matrice. Ceci peut être fait par récurrence en supposant que, si :

$$\forall n > n_0, \forall (i,j) \text{ on a : } |x_{ij}^{(n)} - x_{ij}^{(n-1)}| > |x_{ij}^{(n+1)} - x_{ij}^{(n)}|$$

alors, au rang $n+1$, l'inégalité est vérifiée :

$$|x_{ij}^{(n+1)} - x_{ij}^{(n)}| > |x_{ij}^{(n+2)} - x_{ij}^{(n+1)}|.$$

Cette méthode cale le portefeuille de 1992 sur celui de 1993. Par elle, le morphisme des matrices est respecté.

L'exemple suivant permet de voir la rapidité de la convergence :

$$A = X^{(0)} = \begin{bmatrix} 5 & 10 & 3 \\ 4 & 8 & 12 \\ 2 & 0 & 7 \\ 20 & 5 & 12 \end{bmatrix} \text{ est la matrice dont la structure est à respecter alors que}$$

$$r = \begin{bmatrix} 15 \\ 22 \\ 8 \\ 32 \end{bmatrix} \text{ et } S = [25 \quad 20 \quad 32] \text{ sont les marges sur lesquelles il faut caler } A.$$

La 1ère itération de la méthode RAS donne : $X^{(1)} = \begin{bmatrix} 4,10 & 8,51 & 2,65 \\ 3,43 & 7,14 & 11,15 \\ 1,69 & 0 & 6,40 \\ 16,69 & 4,34 & 10,82 \end{bmatrix}$ et la 2ième :

$$X^{(2)} = \begin{bmatrix} 3,99 & 8,44 & 2,67 \\ 3,39 & 7,19 & 11,39 \\ 1,65 & 0 & 6,46 \\ 16,43 & 4,35 & 11,01 \end{bmatrix}. \text{ Après la 3ième opération, } X^{(3)} = \begin{bmatrix} 3,94 & 8,42 & 2,68 \\ 3,36 & 7,20 & 11,48 \\ 1,62 & 0 & 6,46 \\ 16,33 & 4,36 & 11,13 \end{bmatrix}$$

évolue moins sensiblement et c'est au 13ième passage que la matrice a des marges identiques à celles de r et s à la précision de 10^{-2} :

$$X^{(13)} = \begin{bmatrix} 3,88 & 8,42 & 2,70 \\ 3,30 & 7,17 & 11,53 \\ 1,58 & 0 & 6,42 \\ 16,25 & 4,41 & 11,34 \end{bmatrix}$$

Même si cela peut paraître moins essentiel, la méthode d'arrondissement, imaginée par J-M Grosbras, permet d'obtenir une matrice arrondie et calée.

Elle consiste d'abord à prendre la partie entière de la matrice brute estimée. Sur l'exemple précédent, cela donne :

$$E = \begin{bmatrix} 3 & 8 & 2 \\ 3 & 7 & 11 \\ 1 & 0 & 6 \\ 16 & 4 & 11 \end{bmatrix} R_E = \begin{bmatrix} 13 \\ 21 \\ 7 \\ 31 \end{bmatrix}$$

$$S_E = [23 \quad 19 \quad 30] \quad |72|$$

Il faut donc rajouter 5 à la matrice : on choisit la ligne i_0 où le déficit est le plus grand : ici, $i_0 = 1$.

Pour $i = i_0$, on détermine le j_0 tel que $X_{ij} - E_{ij}$ soit maximum. Ici, $j_0 = 1$, correspondant à $3,88 - 3$, et le 3 est remplacé par un 4. Cela se traduit par la notation suivante :

$$X_{i_0 j_0} \leftarrow E_{i_0 j_0} + 1.$$

Le déficit est donc réduit d'une unité pour la ligne i_0 et la colonne j_0 : $L_{i_0} \leftarrow L_{i_0} - 1$ et $C_{j_0} \leftarrow C_{j_0} - 1$ et on continue ainsi jusqu'à ce que tous les L_i et les C_j soient nuls. (Quand L_{i_0} est nulle, cette ligne n'évolue plus).

Les cas où : $L_{i_1} = L_{i_2}$ ne sont pas gênants, ils conduisent à la même matrice estimée

arrondie : $X = \begin{bmatrix} 4 & 8 & 3 \\ 3 & 7 & 12 \\ 2 & 0 & 6 \\ 16 & 5 & 11 \end{bmatrix}$, et la différence entre la matrice brute estimée et la

matrice entière estimée est : $\begin{bmatrix} -0,12 & 0,42 & -0,30 \\ 0,30 & 0,17 & -0,47 \\ -0,42 & 0 & 0,42 \\ 0,25 & -0,59 & 0,34 \end{bmatrix}$.

Sa mise en œuvre se révèle relativement ardue puisqu'elle nécessite la manipulation de matrices de grandes dimensions : de l'ordre de 6000 sur 100. Dans la pratique, il faut environ 75 itérations pour que la matrice finale soit atteinte.

I.3.2 - Principe de l'utilisation d'une variable corrélée avec le lieu du risque

L'estimation de la matrice $P_{n+1}(ZP, entité)$ étant réalisée, il est alors possible de passer à l'étape suivante, consistant à évaluer la distribution du portefeuille par zone postale, $\hat{P}_n(ZP, .)$, selon le principe suivant :

$$P_n(., entité) + P_{n+1}(ZP, entité) \rightarrow \hat{P}_n(ZP, entité) \rightarrow \hat{P}_n(ZP, .)$$

L'édification de $\hat{P}_n(ZP, .)$ nécessite l'utilisation de la variable entité, qui est une rupture connue du portefeuille. Ainsi, grâce à la connaissance du **portefeuille par entité** pour l'année n et du portefeuille par entité et par zone postale pour l'année n+1, nous en déterminons le portefeuille par entité et par zone postale pour l'année n, dont la marge verticale représente la **ventilation par zone postale**.

L'invariance, quelle que soit l'année, des proportions de chaque zone postale pour les distributions de chacune des entités est l'hypothèse sous-jacente sur laquelle repose la prédiction; en langage mathématique, cela se traduit par :

$$\forall n, \forall entité, \forall ZP, \frac{\hat{P}_n(ZP, entité)}{P_n(., entité)} = \frac{P_{n+1}(ZP, entité)}{P_{n+1}(., entité)}$$

Cette méthode est appelée **méthode d'estimation à évolution homogène**.

Ce principe de conservation des proportions d'une année sur l'autre au niveau de chaque entité est à la base de tous les calculs qui suivent; il faut donc un motif suffisamment fort pour le retenir.

Cet argument est la forte corrélation qui existe entre l'entité et la zone postale : cette variable se trouve en effet très liée au lieu du risque, en particulier pour le réseau agents qui est régionalisé. L'unité de production de ce réseau est l'agence; celle-ci reste confinée à la territorialité qui lui a été attribuée, et qui est composée d'un certain nombre de cantons qu'elle est la seule à exploiter.

Ainsi, l'évolution du portefeuille d'une entité est souvent intimement liée à celle de la zone géographique qui lui est rattachée; il n'est pas exclu, d'autre part, que le portefeuille d'une entité soit dominé par quelques zones postales : dans ce cas, leur variation est très proche de celle de l'entité. L'hypothèse de conservation des proportions est alors plus facile à observer lorsqu'un nombre limité de zones postales constitue le portefeuille d'une entité.

Cette réalité n'est pas caractéristique des réseaux Salarié et Courtage, dont la couverture est nationale, même si l'on peut rencontrer ci et là des places fortes, pour le courtage notamment. La vente de contrats habitation dans le réseau des guichets de la BNP étant quasi inexistante avant 1991, l'incertitude de la méthode pour la BNP n'a que peu d'influence sur le résultat final.

I.3.3 - Comparaison des méthodes de reconstitution des portefeuilles

Quel niveau de précision offre cet estimateur ? Pour juger de sa qualité, il peut être utile de comparer le résultat de cette estimation avec le portefeuille réel pour un exercice où cette information est disponible, comme 1995. Ainsi, muni du portefeuille 1996 ventilé par entité et zone postale, et du portefeuille 1995 réparti par entité, on est en mesure d'évaluer le portefeuille 1995 par zone postale. Cette distribution est alors comparée à la véritable, selon une distance à choisir.

L'écart obtenu entre le réel et l'estimé de l'exercice 1995 donnera une idée assez précise de l'écart qui prévaut entre ces deux quantités pour des années antérieures.

A l'heure de la mise en place de la méthode, un choix s'offre à nous concernant la taille retenue pour les entités du réseau agence : si l'on désire des portefeuilles d'environ 1000 contrats et de 5 millions de Frs d'encaissement, représentant deux ou trois cantons d'un département, alors notre choix se portera sur les agences. En revanche, si l'on préfère que les portefeuilles soient de taille plus importante, on optera pour l'unité départementale, constituée des agences situées dans un même département.

N'ayant pas d'avis sur la question, le test sera effectué avec les deux types d'entités, et l'entité finalement retenue pour l'estimation des portefeuilles 1979 à 1991 sera celle pour laquelle l'expérience aura fourni l'écart le plus faible.

I.3.3.1 - Méthode de reconstitution des portefeuilles utilisant l'agence

Contrairement à la méthode utilisant l'entité départementale, la méthode d'estimation par agence présente une réelle difficulté de réalisation en raison des créations, disparitions et fusions possibles d'agences. Il peut en outre arriver qu'un flux apparaisse entre deux agences, ne conduisant ni à la disparition ni à la création de l'une d'entre elles. Or, ces flux sont bien à rattacher à une agence existante.

La figure 2.1.1 recense les différents cas de transferts de l'activité incendie, rencontrés entre 1995 et 1996, d'une agence vers une autre agence :

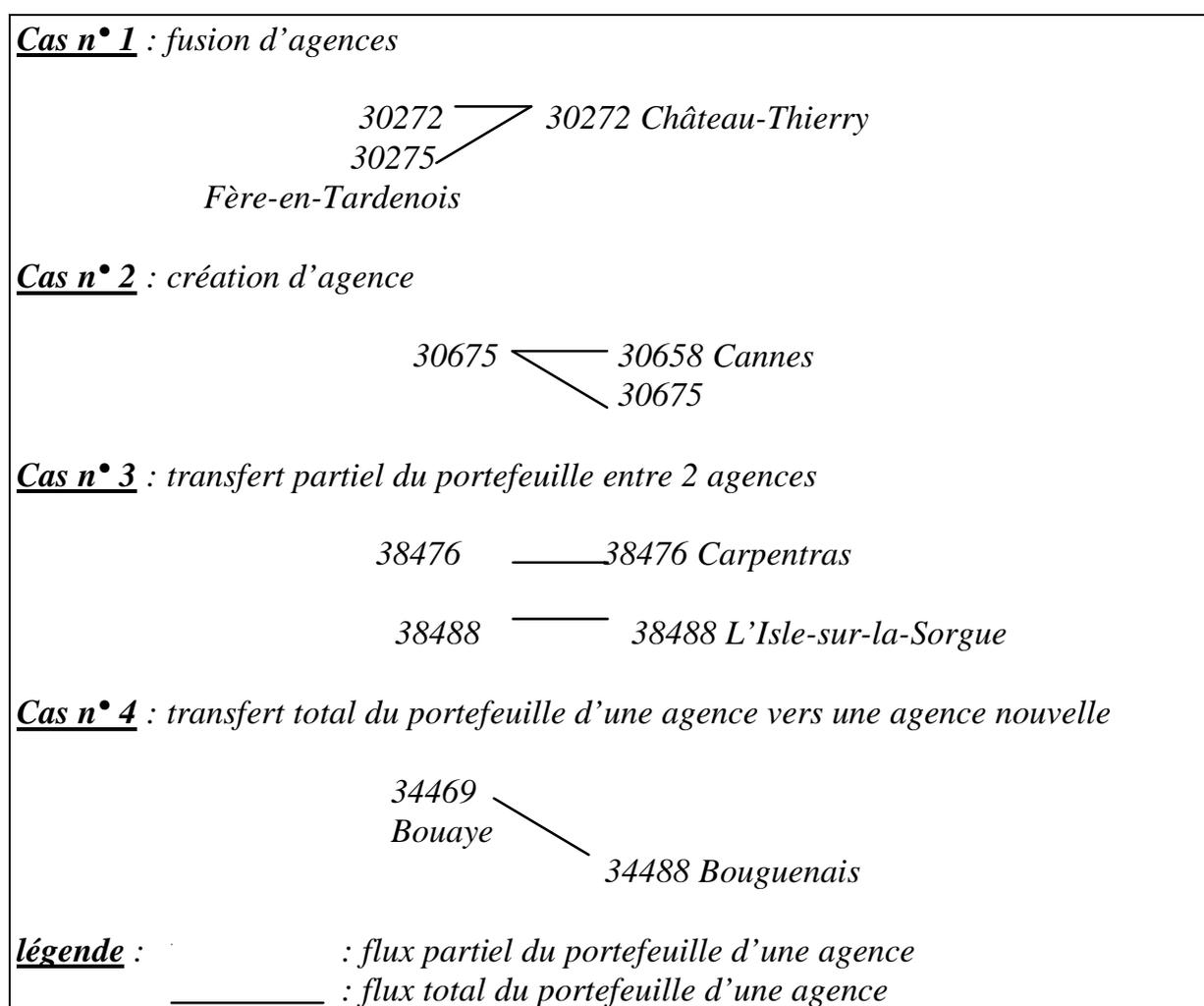


Figure 2.1.1 : Recensement des différents cas de transferts de portefeuille entre agences

- la fusion des agences de Fère-en-Tardenois (code agence 30275) et de Château-Thierry (code agence 30272), en une seule agence codifiée par le code agence 30272 (cas n°1),

- la création, à Cannes, de l'agence 30658, issue de l'agence 30675 (cas n°2),
- le transfert d'une partie du portefeuille de l'agence de Carpentras (code 38476) vers celui de l'Isle-sur-la-Sorgue (code agence 38488) (cas n°3),
- le transfert de la totalité du portefeuille de l'agence de Bouaye (code 34469) vers l'agence nouvellement créée de Bouguenais (code agence 34480) (cas n°4).

Selon le principe d'estimation à évolution homogène défini au paragraphe 1.3.2, on

$$a : \forall n, \forall entité, \forall ZP, \frac{P_n^{\wedge}(ZP, entité)}{P_n(., entité)} = \frac{P_{n+1}(ZP, entité)}{P_{n+1}(., entité)} ;$$

lorsque l'entité a zéro contrat en 1996, ce qui est le cas de l'agence 30275 (cas n°1) et l'agence 34469 (cas n°4) sur la figure précédente, il y a indétermination du rapport et il est impossible d'estimer la répartition du portefeuille de cette agence en 1995. (Cette difficulté ne se rencontre pas lorsque la rupture choisie pour l'estimation du portefeuille est le département). On se trouve alors dans l'absolue nécessité d'identifier l'agence absorbante, sous peine de ne pouvoir retrouver la distribution de l'agence absorbée. La distribution estimée sera celle issue du regroupement des deux agences, et ceci pour toutes les années antérieures au regroupement : ce dernier a un caractère irréversible.

Si, pour les deux situations précédentes, il s'avère nécessaire de procéder à la réunion des agences, une disparaissant au profit de l'autre, en revanche, le cas n° 3 nous montre la possibilité d'un transfert partiel de portefeuille entre deux agences. La question se pose alors de savoir si l'importance du flux justifie l'appariement des deux agences, ou au contraire si leur distribution géographique décalée rend cette opération vide de sens.

Pour obtenir une réponse à cette question, il nous faut définir des règles précises en matière d'appariement d'agence, qu'il convient de respecter :

- un transfert de portefeuille entre deux agences doit être mis en évidence,
- les agences doivent présenter une grande proximité quant à la distribution géographique de leur portefeuille respectif.

La *1ère condition nécessaire* au rapprochement est donc l'existence d'un flux de portefeuille entre deux agences. Ceci impose de dresser, d'une part la liste des agences disparues au cours de l'exercice 1995, d'autre part celle des agences apparues au cours de cette période. Ces listes ne sont bien évidemment pas exhaustives, puisque, comme

on l'a vu, des flux partiels peuvent exister. Pour les mettre en évidence, la technique utilisée consiste à repérer les agences, dans le fichier de synthèse « portefeuille » :

- dont la variation *relative* du portefeuille dépasse un certain seuil, positif, ou est inférieur à un certain autre, négatif,
- ou dont la variation *absolue* du portefeuille dépasse un certain seuil, positif, ou est inférieur à un certain autre, négatif.

Cette 1ère phase de recherche des agences, faisant l'objet d'un transfert de contrats entrants ou sortants, ou susceptibles d'en présenter un, est suivie d'une seconde, plus délicate, visant à apparier les agences, c'est-à-dire à associer, à l'agence à l'origine de ce flux, l'agence vers laquelle ce flux se dirige.

Cette seconde phase de recherche de flux entre agences peut être menée de manière automatique, à partir d'un outil informatique déjà utilisé, la transaction 4W4NLH, permettant la consultation de l'ensemble des mouvements administratifs ayant affecté les contrats en portefeuille. Ainsi, la recherche d'écritures relatives au mouvement 02 « transfert de portefeuille entre agence » de chacune des agences présentes sur la liste des agences dressée lors de la 1ère phase permet de juger de l'existence ou non de flux entre agences : l'absence d'écriture pour ce type de mouvement est synonyme d'absence de transfert entre agence.

Pour les contrats des agences ayant des écritures recherchées, la conséquence est un changement de leur numéro : l'interrogation de l'écran portefeuille de l'ancien numéro de contrat nous fournit le nouveau numéro, dans la structure duquel est contenu le code agence recevant le flux.

Cet appariement est subordonné à l'existence d'un lien fort entre les deux agences : c'est tout naturellement l'objet de la *deuxième condition nécessaire* au rapprochement de deux agences. Un indicateur de proximité entre deux agences *a* et *b* doit être construit pour l'occasion. Il peut être défini comme suit :

$$IP(a,b) = \frac{\sum_{i \in ZP} \text{Min}(P(i,a), P(i,b))}{\frac{1}{2} \sum_{i \in ZP} (P(i,a) + P(i,b))} \quad \text{où :}$$

- *i* est le numéro de la zone postale appartenant à l'ensemble des zones postales communes à *a* et *b*,
- *P(i,a)* (respectivement *P(i,b)*) est le nombre de contrats en portefeuille dans la zone postale *i* pour l'agence *a* (respectivement l'agence *b*).

IP est compris entre 0 et 1, 1 désignant une parfaite proximité entre les deux agences ($IP(a,a)=1$) et 0 une dissymétrie totale. Cet indicateur tient compte :

- d'une part, de la taille relative d'une agence par rapport à une autre ($\text{Min}(P(i,a),P(i,b))$) : plus elles sont disproportionnées, plus IP tend vers 0,
- d'autre part, de la répartition géographique des risques pour les deux agences : moins ces risques sont situés sur des zones communes, plus IP tend 0.

IP se compose donc d'un effet *taille*, qui tient compte du rapport entre $\sum_{i \in ZP_a} P(i,a)$ et $\sum_{i \in ZP_b} P(i,b)$, et d'un effet *similarité*, qui est influencé par le nombre de zones postales communes, et non par leur poids relatif. Comme le suggère la figure 2.1.2, IP évoque en quelque sorte la part représentée par l'aire géographique commune des deux agences, pondérée par le nombre de contrats de ces zones postales, $A(a \cap b)$, sur l'aire totale couverte par les deux portefeuilles, $A(a \cup b)$.

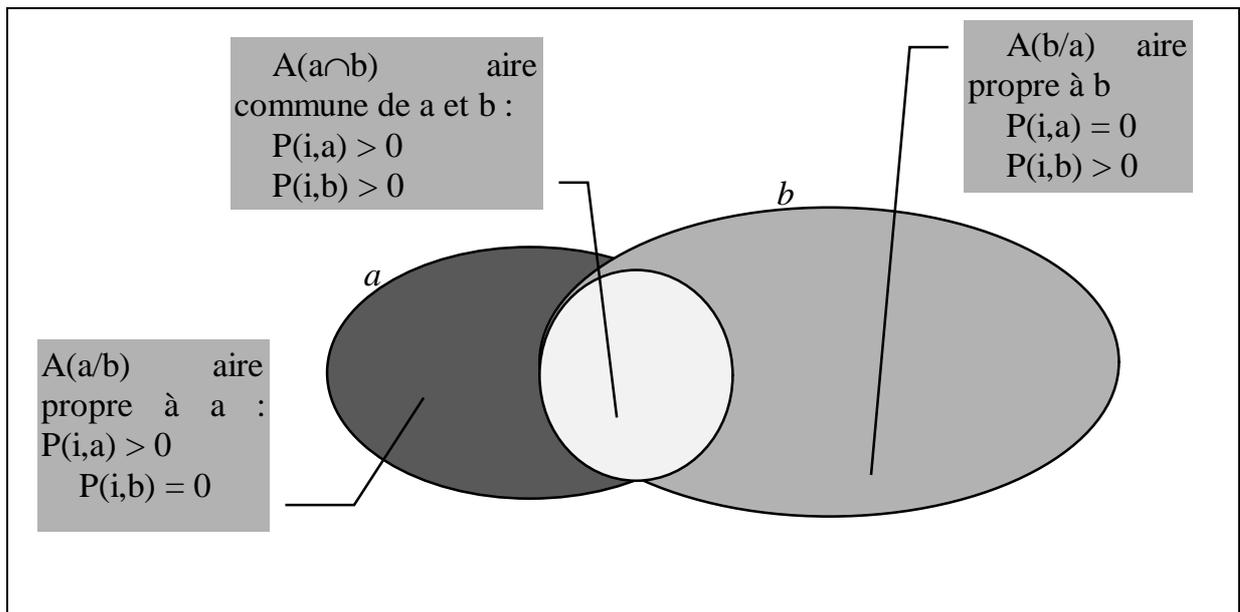


Figure 2.1.2 : Interprétation géographique de l'indicateur de proximité de deux agences

La décision du rapprochement de deux agences pour un flux partiel de portefeuille se juge donc à la valeur prise par l'indicateur de proximité associé, ce dernier devant être suffisamment élevé pour un appariement. Lorsque le flux est total, la valeur de

l'indicateur de proximité est bien sûr sans objet, même s'il est préférable qu'elle soit élevée.

Signalons que le calcul de l'indicateur de proximité se fait :

- sur l'année 1995 pour le cas n° 1 (fusion d'agences),
- sur l'année 1996 pour le cas n° 2 (création d'agence),
- indifféremment sur les années 1995 et 1996 pour le cas n° 3 (transfert partiel de portefeuille),
- et, pour le cas n° 4 (transfert total de portefeuille), sur l'année 1995 pour l'agence fermée et l'année 1996 pour l'agence nouvelle.

En situation réelle, et non dans le cadre de cette simulation, la distribution par zone postale des agences n'est pas connue pour l'année n : elle l'est seulement pour l'année $n+1$. L'indicateur ne pourrait donc être calculé pour les cas n° 1 et n° 4. Néanmoins, le recours à l'outil de consultation du portefeuille 4W4NLH offre une solution toute trouvée, d'autant que le calcul de cet indicateur n'est pas nécessaire lors d'un transfert total de portefeuille.

I.3.3.2 - Réseau agents : résultats comparés des méthodes d'estimation à évolution homogène

Après avoir présenté les développements spécifiques requis pour l'utilisation de la méthode d'estimation par agence, il faut maintenant comparer les résultats fournis par cette dernière avec la méthode ayant recours au département, pour le seul réseau des agents, les autres réseaux ne disposant pas de cette structure départementale.

C'est l'écart entre les distributions estimée e et réelle r qu'il importe de mesurer. Pour cela, la distance utilisée est celle du k hi-deux :

$$d^2 = \sum_{i=1}^2 \sum_{j=1}^s \frac{(n_{ij} - n_{i.} p_j)^2}{n_{i.} p_j}$$

Il est en effet fait l'hypothèse que les deux distributions r et e constituent des échantillons issus de deux variables aléatoires :

- nombre estimé de contrats en portefeuille en 1995 (n_1),
- nombre de contrats réellement en portefeuille en 1995 (n_2).

En notation matricielle, cela donne :

$$\begin{bmatrix} e_1 & \dots & e_j & \dots & e_s \\ r_1 & \dots & r_j & \dots & r_s \end{bmatrix} = \begin{pmatrix} e \\ r \end{pmatrix}$$

n_{ij} est le nombre d'individus de l'échantillon appartenant au département de l'agence j de la variable $n_{i.} = \sum_{j=1}^s n_{ij}$, effectif de l'échantillon i ; $n_{.j} = \sum_{i=1}^2 n_{ij}$ est le nombre total des individus appartenant au département de l'agence j .

Il s'agit de montrer que ces deux échantillons proviennent d'une même population : c'est l'hypothèse H_0 que l'on fait lors de l'écriture du test du χ^2 , contre H_1 : " les échantillons sont significativement différents ".

Dans l'hypothèse H_0 on peut parler de probabilités p_1, p_2, \dots, p_s de posséder les modalités $1, 2, \dots, s$. Il s'agit alors de comparer les effectifs constatés n_{ij} aux effectifs espérés $n_{i.} p_{.j}$, qui ne doivent pas en être très éloignés.

Dans l'hypothèse H_0 , d_0^2 est une réalisation d'une variable D_0^2 suivant un χ^2 dont le nombre de degrés de liberté est égal au nombre de termes de d^2 ($2.s$), auquel on doit retirer 2, le nombre de relations indiquant que les sommes de colonnes sont constantes et égales :

$$\sum_j n_{ij} = \sum_j n_{i.} p_{.j} = n_{i.} = n.$$

Donc D_0^2 est un $c_{2(s-1)}^2$.

Cependant, comme les p_1, p_2, \dots, p_s ne sont pas connus, il faut les estimer par $\hat{p}_j = \frac{n_{.j}}{2n}$, ce qui représente $s - 1$ relations indépendantes : en effet, pour estimer les s probabilités, on n'a besoin que de $s - 1$ relations car $\sum_{j=1}^s p_{.j} = 1$. D_0^2 est alors un

$$c_{2(s-1) - (s-1)}^2 = c_{(s-1)}^2 \text{ si } H_0 \text{ est vraie.}$$

Le test consistera à rejeter H_0 si d^2 constaté est trop grand, comme pour un test habituel du χ^2 .

Par la suite, une table peut nous fournir la précision correspondant à d_0^2 , c'est-à-dire la probabilité que les échantillons soient considérés à tort comme ayant une origine identique.

Avant d'effectuer nos calculs, il peut être intéressant de donner une expression simplifiée de d^2 dans l'hypothèse H_0 :

$$d^2 = \sum_{i=1}^2 \sum_{j=1}^s \frac{(n_{ij} - n_i \cdot p_j)^2}{n_i \cdot p_j} = \sum_{j=1}^s \frac{((e_j - \frac{e_j + r_j}{2})^2 + (r_j - \frac{e_j + r_j}{2})^2)}{\frac{e_j + r_j}{2}}, \text{ soit encore :}$$

$$d_0^2 = \sum_{j=1}^s \frac{(e_j - r_j)^2}{e_j + r_j}.$$

La seule base de comparaison commune entre ces deux méthodes étant le département, c'est à ce niveau que se calculeront les χ^2 . Il faut donc agréger les résultats de la méthode par agence. Les χ^2 à calculer sont donc à 94 degrés de liberté.

Le tableau 2.1.6 suivant compare les χ^2 des méthodes d'estimation par département de l'agence et par agence, pour les risques agricole, d'habitation et professionnels. Il indique une *quasi équivalence des résultats des deux méthodes, la méthode départementale se révélant un peu meilleure*, puisque le χ^2 départemental est un peu plus faible que le χ^2 par agence.

Tableau 2.1.6 : Comparaison des χ^2 des méthodes d'estimation par département ou agence

risque	c ² département	c ² agence
agricole	1,69 %	1,69 %
habitation	0,15 %	0,17 %
professionnel	1,25 %	1,29 %
tous risques confondus	0,32 %	0,34 %

Le calcul de d_0^2 départemental, tous risques confondus, donne 0,32 %. Pour un χ^2 à 94 degrés de liberté, cette valeur correspond à une probabilité de considérer à tort les deux échantillons comme émanant d'une même source à moins de 0,1 % : ce risque est donc très faible. *Nous pouvons donc accepter l'hypothèse d'une équivalence des distributions géographiques réelle et estimée, quelle que soit la méthode d'estimation considérée.*

Si l'on se livre à une analyse comparative des deux méthodes, on se rend compte que chacune possède des atouts et des faiblesses : c'est ce que montre le tableau 2.1.7, qui donne, par risque, la répartition du nombre de départements selon le signe de la quantité : χ^2 (département) - χ^2 (agence); celle-ci est calculée avec une décimale, ce qui explique les cas d'égalité.

Il est à noter le nombre important, pour le risque agricole, de départements pour lesquels la méthode d'éclatement par agence est plus précise que celle à rupture départementale (47 contre 41), alors qu'ils ne sont que :

- 5 contre 84 pour le risque d'habitation,
- 25 contre 63 pour le risque professionnel.

Tableau 2.1.7 : répartition du nombre de départements selon le risque et le signe de la quantité : χ^2 (département) - χ^2 (agence)

risque	nombre de départements avec		
	χ^2 (département) > χ^2 (agence)	χ^2 (département) = χ^2 (agence)	χ^2 (département) < χ^2 (agence)
agricole	47	7	41
habitation	5	6	84
professionnel	25	7	63
tous risques confondus	4	10	81

Cette singularité de la méthode d'éclatement par agence pour le risque agricole s'explique par la faiblesse des nombres de contrats agricoles dans certaines zones rurales : les évolutions locales sont en effet mieux retranscrites par cette méthode. En revanche, dans les zones urbaines, où les risques d'habitation et professionnels prédominent, les volumes plus importants de contrats assurent une moins grande variabilité des stocks, aidée en cela par la faible évolution du portefeuille : une notable partie des contrats est en portefeuille depuis de fort nombreuses années sans qu'aucune évolution des garanties n'ait été proposée, de peur que l'assuré ne résilie, suite à la baisse consentie de sa prime trop élevée...

Le phénomène des petites catégories est donc moins répandu pour la méthode d'estimation à rupture départementale. C'est d'ailleurs une chance car cette méthode, et à un degré moindre la méthode par agence, retranscrit moins bien les évolutions des petites quantités. En effet, si une zone postale avait 1 contrat en 1996, pour une entité donnée (agence ou département), il n'y a pratiquement aucune chance que l'estimation pour 1995 conduise à une valeur différente de 1.

On trouvera en annexe D les programmes de reconstitution du portefeuille du réseau agence selon le principe d'évolution homogène de chaque zone postale exploitée par l'entité : celle-ci peut être une agence (programme PORTESTAG), ou un département (programme PORTESTDEP). Ce dernier programme peut aussi être utilisé pour la reconstitution du portefeuille des autres réseaux.

On trouvera d'ailleurs la solution à un problème touchant le portefeuille lorsqu'il est exprimé en volume : l'arrondissement des nombres estimés. Les zones postales dont les valeurs absolues des différences d'arrondis sont les plus importantes (i.e. sont proches de 0,5) seront prioritairement traitées.

Cependant, une question se pose : les conclusions précédentes ne seraient-elles pas battues en brèche si l'on avait utilisé une autre distance entre l'estimateur et le réel ? En d'autres termes, la distance utilisée - celle du χ^2 - n'influence-t-elle pas l'interprétation des résultats ?

Pour tenter d'apporter une réponse à cette interrogation, nous avons recours à **une autre distance : celle dite de l'écart absolu**. Sa définition est la suivante :

$$d_{abs}^2 = \frac{\sum_{j=1}^s |e_j - r_j|}{\sum_{j=1}^s r_j},$$

où (e_j) et (r_j) sont respectivement les distributions estimée et réelle du nombre de contrats 1995 par département, et s le nombre de départements.

La particularité de cette distance est de donner un poids équivalent à chacun des termes du numérateur : elle met spécialement l'accent sur les zones postales à fort volume, et néglige par conséquent les petites catégories.

Le tableau 2.1.8 établit une comparaison des méthodes d'estimation par rupture départementale ou par agence, en utilisant la distance de l'écart absolu.

Tableau 2.1.8 : Comparaison des distances de l'écart absolu des méthodes d'estimation par département ou agence

risque	d_{abs}^2 département	d_{abs}^2 agence
agricole	6,88 %	6,61 %
habitation	2,70 %	2,67 %
professionnel	6,66 %	6,53 %
tous risques confondus	3,27 %	3,22 %

Il en ressort que *la méthode par agence, du point de vue de l'écart absolu, est meilleure pour tous les types de risques, et pour tous les risques confondus* : d_{abs}^2 est légèrement plus faible pour l'agence (3,22 %) que pour le département (3,27 %). L'avantage est encore plus marqué pour le risque agricole (6,88 % contre 6,61 %).

Ce résultat n'est certes pas contradictoire avec celui du tableau 2.1.6 mais montre plutôt, s'il en était besoin, l'importance que revêt le choix de la distance qui agit en miroir déformant. Ce tableau confirme les conclusions logiques tirées de la définition de la distance de l'écart absolu : en effet, les zones postales au volume important contribuent de manière essentielle à la formation de la distance de l'écart absolu; une estimation éloignée pour une zone postale à fort volume a alors pour conséquence l'apparition d'un écart important. Or, pour tous les risques, le tableau précédent indique une distance plus petite pour la méthode d'estimation par agence. Cette méthode répond donc mieux que celle par département à l'évolution du nombre de contrats des zones postales au volume important.

En revanche, la distance du χ^2 , en ne privilégiant aucun type de zone postale, constitue un outil de mesure de la *qualité d'ensemble* de l'estimation. Cette caractéristique n'est pas partagée par la distance de l'écart absolu, pour laquelle un écart de 1 au numérateur pour une zone postale de 5 unités compte autant qu'un même écart de 1 pour une zone postale de 100 unités.

D'autre part, la distance de l'écart absolu, même si elle nous a permis de mettre en exergue les atouts respectifs des méthodes d'estimation, n'est pas, à notre connaissance, tabulée : elle ne nous permet donc pas de donner une idée précise de la qualité d'estimation.

Nos priorités étant :

- d'une part, une transcription aussi fidèle que possible de la qualité d'ensemble de l'estimation,
- d'autre part, une mesure de la précision de l'estimation grâce à une distance tabulée,

il nous faut préférer la distance du c^2 à la distance de l'écart absolu.

Dans cette perspective, le bon sens nous pousse à exploiter plutôt les résultats du tableau 2.1.6 que ceux du tableau 2.1.8 dont on peut tirer la réflexion suivante : la meilleure prise en compte, par la méthode d'estimation par agence par rapport à la méthode à rupture départementale, des évolutions locales ne compense pas la présence plus importante de zones postales à faible volume, dont l'estimation est plus sujette à caution.

Il en résulte que notre préférence va à la méthode d'estimation par département pour reconstituer le portefeuille du réseau agents.

I.3.3.3 - Remarques complémentaires sur la méthode d'estimation à évolution homogène

Pour les raisons exposées précédemment, le choix de la méthode d'estimation se portera donc, pour le réseau agents, sur celle utilisant le département. Sa mise en œuvre se révèle en effet moins délicate que celle par agence, dont on a pu entrevoir au paragraphe 1.3.3.1 la complexité, l'existence de transferts de portefeuille du courtage vers certaines agences rendant plus difficiles encore les choses.

Tableau 2.1.9 : Comparaison des χ^2 selon le risque et le réseau (méthode à évolution homogène)

risque	réseau A	réseau C	réseau S	BNP	siège	tous réseaux
agricole	1,69 %	19,82 %	13,25 %	-	-	1,74 %
habitation	0,15 %	1,04 %	0,77 %	4,76 %	10,24 %	0,12 %
professionnel	1,25 %	5,27 %	9,22 %	-	-	1,05 %
tous risques confondus	0,32 %	1,57 %	1,59 %	4,76 %	10,24 %	0,25 %

L'analyse des résultats de la méthode d'estimation pour les autres réseaux de l'UAP, peut se faire à partir du tableau 2.1.9, qui ventile la valeur du χ^2 selon le réseau et le risque. Ses résultats, riches en enseignements, sont les suivants :

- le réseau agents présente les meilleurs résultats, les χ^2 étant les plus faibles pour chacun des trois risques : *un apport de précision supplémentaire* - qui n'est pas démontré -, *par la méthode d'estimation par éclatement par agence, se serait donc révélé vain*, les autres réseaux présentant des χ^2 au moins 5 fois plus élevés;
- l'agrégation des résultats sur l'ensemble des réseaux offre des χ^2 encore plus faibles que ceux du réseau agents, à l'exception du risque agricole : tous risques confondus, le χ^2 vaut 1,05 % contre 1,25 % pour le réseau agents. Une explication en est un *phénomène de compensation entre réseaux* particulièrement présent pour les zones postales à faible volume. Celles-ci, par l'agrégation des résultats, gagnent en importance et se comportent de façon moins erratique et sont moins sujettes aux fluctuations. Elles ont ainsi une évolution plus conforme à l'ensemble du réseau. Il se peut aussi que le phénomène de compensation se manifeste à la fois par une surestimation du portefeuille, pour un réseau donné, et par une sous-estimation pour un autre

réseau. Il faut savoir en effet que les places fortes d'un réseau correspondent en général aux « déserts » pour les autres réseaux;

- de même, il est à signaler que les cases croisant le réseau et le risque au volume le moins important sont celles dont les χ^2 sont les plus élevés : le χ^2 du risque agricole du réseau C est ainsi à 19,82 %. Cela signifie que **la méthode d'estimation utilisée suppose des entités à faible variabilité de leur portefeuille**, ce qui n'est pas généralement la caractéristique des cases à faible volume. Mais leur poids étant par définition plus faible, les conséquences néfastes sont réduites; (le réseau de la BNP n'a véritablement pris de l'importance qu'après 1991, et la qualité de l'estimation proposée par la méthode pour ce réseau n'a que peu d'importance.)
- enfin, on remarquera que, parmi les trois risques, le risque agricole présente les résultats les plus dégradés : le χ^2 vaut 1,74 % contre 0,25 % pour l'ensemble des risques. Ceci révèle la petitesse des nombres estimés et le **handicap présenté par la méthode d'estimation à évolution homogène pour cette catégorie de contrats** : en effet, de nombreuses zones postales ont 0, 1 ou 2 contrats agricoles en 1996 et l'évolution de l'entité départementale, pour le réseau agents, ou du réseau, pour les autres, sur laquelle repose l'estimation, conduit pratiquement toujours à une estimation 1995 égale au chiffre réel 1996.

En conclusion, la solution conduisant aux résultats les plus précis aurait consisté à retenir :

- la méthode d'estimation à rupture par agence pour le risque agricole et le réseau agents, ce réseau dominant très largement ce risque,
- et la méthode d'estimation à évolution homogène pour les autres réseaux (à rupture départementale pour le réseau agents), pour les risques d'habitation et professionnel.

Cependant, en raison d'une mise en œuvre particulièrement délicate, la méthode d'estimation conservera l'entité départementale pour le réseau agents, quel que soit le risque.

La simulation faite sur l'année 1995, à partir des données 1995-1996, conduit à un χ^2 de 0,25 % ce qui, pour un χ^2 à 94 degrés de liberté, correspond à une probabilité de considérer à tort les deux échantillons comme émanant d'une même source à moins de 0,1 % : ce risque est donc très faible. **Nous pouvons donc accepter l'hypothèse d'une équivalence de la distribution géographique réelle, et de celle estimée par la méthode d'estimation à évolution homogène.**

I.3.4 - Mise en œuvre pratique de la reconstitution du portefeuille

Après avoir réfléchi sur la qualité de l'estimateur servant à la reconstruction du portefeuille, il est nécessaire de passer maintenant à sa réalisation pratique. Nous verrons tour à tour les difficultés concrètes auxquelles nous avons été confrontés.

I.3.4.1 - Reconstitution des marges par entité et risque

Au paragraphe I.3.2, nous avons exposé le principe de l'utilisation d'une variable corrélée, l'*entité*, avec le *lieu du risque (ZP)* pour reconstituer la distribution du portefeuille par zone postale, $\hat{P}_n(ZP,.)$:

$$P_n(.,entité) + P_{n+1}(ZP,entité) \rightarrow \hat{P}_n(ZP,entité) \rightarrow \hat{P}_n(ZP,.)$$

La matrice $P_{n+1}(ZP,entité)$ est supposée connue (pour 1993), ou estimée de proche en proche par la méthode à évolution homogène pour les années précédentes. Celle-ci donne $\hat{P}_n(ZP,.)$.

Intervenant aussi dans le processus d'estimation, les marges par entité $P_n(.,entité)$ sont à reconstituer. Ces données ont nécessité des recherches de longue haleine, au sein des différents secteurs de l'UAP où l'information pouvait se situer : secteur des études sur les produits, secteur des études statistiques, contrôle de gestion. Très souvent, l'information recueillie s'est révélée parcellaire ou très pauvre, car seuls les exercices les plus récents étaient disponibles. Cette collecte d'information s'est par la suite orientée vers les lieux d'archivage situés en province.

Le support de l'information recueillie était de deux types :

- soit le papier, sous la forme de suivis mensuels de production et de portefeuille, ou sous la forme d'états annuels dressant l'inventaire du portefeuille, à travers le nombre de contrats et les primes émises nettes d'annulations,
- soit le format informatique, à travers des bibliothèques de données privées utilisées pour une autre fin que la présente étude, ou à travers des fichiers centraux informatiques archivés et à lire par le langage de programmation APL.

Le résultat de cette collecte, en partie couronnée de succès, a permis de remonter à l'année 1975 pour le portefeuille et les primes ventilées par risque et réseau. De même, pour le réseau agents, il a été possible de retrouver ce même portefeuille ventilé par :

- classification comptable du risque et entité régionale, sous un format papier,
- agence, le risque d'incendie étant confondu avec la catégorie de dommages aux biens (hors la construction) à partir de 1987, sous un support informatique.

L'agrégation de la vingtaine de modalités de la classification comptable du risque, précédée d'une longue phase de saisie, permet de retrouver sans difficulté le portefeuille des risques agricole, d'habitation et d'entreprise, établi à l'inventaire des années 1979 à 1991. Nous avons ainsi 13 matrices (12,3) parfaitement connues - une par exercice -, dont la représentation suivante utilise x comme symbolique d'une information de type portefeuille incendie :

$$\begin{array}{c}
 \text{entité régionale (12)} \\
 \downarrow \\
 \begin{array}{c}
 \xrightarrow{\text{risque(3)}} \\
 \left[\begin{array}{ccc}
 x & x & x \\
 \vdots & \vdots & \vdots \\
 \vdots & \vdots & \vdots \\
 \vdots & \vdots & \vdots \\
 x & x & x
 \end{array} \right]
 \end{array}
 \end{array}$$

D'autre part, nous disposons pour le réseau agents du portefeuille total de l'incendie (et des dommages aux biens, à partir de 1987), le risque construction étant exclu, par exercice et par agence. Une agrégation de ces données d'agence au niveau de l'entité départementale conduit aux 13 matrices (95,1) suivantes, empruntant la lettre y pour indiquer une information de type portefeuille incendie + (éventuellement) les dommages aux biens :

$$\begin{array}{c}
 \text{département (95)} \\
 \downarrow \\
 \begin{array}{c}
 \xrightarrow{\text{incendie+DAB(1)}} \\
 \left[\begin{array}{c}
 y \\
 \vdots \\
 \vdots \\
 \vdots \\
 \vdots \\
 y
 \end{array} \right]
 \end{array}
 \end{array}$$

De prime abord, ces informations ne semblent pas avoir de point commun. Cependant, il faut savoir qu'une entité régionale du réseau agents représente plusieurs entités départementales, ou départements. Une table de correspondances entre les régions - aux frontières mobiles au cours de la période 1979 - 1991 - et les départements est à constituer. Elle permettra de représenter, pour le portefeuille en volume (i.e. nombre de contrats) et en valeur (i.e. primes émises nettes d'annulations), pour chaque exercice et chacune des 12 régions, 312 matrices (i.e. 2 x 13 x 12) ayant le format suivant :

$$\begin{array}{c}
 \text{risque (3 ou 4)} \rightarrow \\
 \left[\begin{array}{cccc}
 \cdot & \cdot & \cdot & \cdot \\
 \vdots & \vdots & \vdots & \vdots \\
 \vdots & \vdots & \vdots & \vdots \\
 \vdots & \vdots & \vdots & \vdots \\
 \cdot & \cdot & \cdot & \cdot
 \end{array} \right] \left[\begin{array}{c}
 y \\
 \vdots \\
 \vdots \\
 \vdots \\
 y
 \end{array} \right] \\
 \downarrow \\
 \text{département} \\
 \left[\begin{array}{cccc}
 x & x & x & y-x \\
 \vdots & \vdots & \vdots & \vdots \\
 \vdots & \vdots & \vdots & \vdots \\
 \vdots & \vdots & \vdots & \vdots \\
 y & \vdots & \vdots & \vdots
 \end{array} \right] \left[\begin{array}{c}
 y \\
 \vdots \\
 \vdots \\
 \vdots \\
 y
 \end{array} \right]
 \end{array}$$

Les accolades et les « . » de la matrice (exercice, région) à structure croisée (département x risque) symbolisent son caractère inconnu, contrairement aux marges horizontale (total d'une région par risque) et verticale (total tous risques confondus par département de la région), qui sont entre crochets.

Si le nombre de colonnes est parfaitement déterminé (il est de 3 avant 1987 et de 4, après cette date avec le risque « autres dommages aux biens »), en revanche, le nombre de lignes diffère selon les régions, sa moyenne avoisinant 8.

Signalons le problème particulier des cinq succursales de représentation du courtage : en effet, très longtemps, le réseau C s'est réduit au seul bureau de Paris, les succursales de Marseille, Lyon, Bordeaux, Strasbourg et Lille étant rattachées aux directions régionales auxquelles ces métropoles étaient rattachées. La question se pose alors de savoir s'il vaut mieux :

- isoler le portefeuille des succursales dans chacune des matrices des régions de rattachement,
- ou l'inclure dans le département où elles se trouvent.

Deux arguments se contrebalancent : d'un côté, une répartition par risque très différente entre le portefeuille de la succursale et celui du département où elle est localisée plaide pour la création d'une ligne supplémentaire dans la matrice par entité et risque. En outre, le poids de ces succursales, souvent équivalent à celui d'un département, plaide pour cette séparation.

De l'autre, le portefeuille des succursales faisant souvent l'objet de transferts vers les agences environnantes, il peut être préférable de laisser les succursales dans les départements où elles se trouvent. D'ailleurs, la définition des entités départementales englobait déjà ces unités de souscription, pour les années 1993 à 1996.

La réponse à cette interrogation consiste tout simplement à privilégier la matrice que l'argument le plus fort conseille de prendre : ainsi, si l'on a constaté des transferts

importants entre la succursale et les agences environnantes, on n'isolera pas cette entité. En revanche, si la composition du portefeuille diffère de celle du département de rattachement, il faut donc ajouter provisoirement à ces matrices une ligne relative à la succursale. Après calculs, la ligne supplémentaire sera ensuite agrégée à la ligne du département où se trouve géographiquement située la succursale.

Une question se pose alors : comment peut-on estimer les matrices croisant le département et le risque ? La résolution de ce problème se trouve tout naturellement dans l'application d'une méthode bien rodée, celle du *raking ratio*, ou méthode RAS. L'estimation de la matrice de l'exercice n s'effectuera en calant la matrice connue n+1 sur les marges de la matrice n, pour ainsi respecter au mieux la structure croisée de la matrice n+1 :

$$\begin{array}{c}
 \begin{array}{ccc}
 \xrightarrow{\text{risque}(4)} & \xrightarrow{\text{risque}(4)} & \xrightarrow{\text{risque}(4)} \\
 & n & n+1 & n \\
 \downarrow \text{département} & \left[\begin{array}{cccc} \cdot & \cdot & \cdot & \cdot \\ \vdots & \vdots & \vdots & \vdots \\ \cdot & \cdot & \cdot & \cdot \end{array} \right] + & \left[\begin{array}{cccc} x & x & x & y-x \\ \vdots & \vdots & \vdots & \vdots \\ \cdot & \cdot & \cdot & \cdot \\ x & x & x & y-x \end{array} \right] & \rightarrow & \left[\begin{array}{cccc} x & x & x & y-x \\ \vdots & \vdots & \vdots & \vdots \\ \cdot & \cdot & \cdot & \cdot \\ x & x & x & y-x \end{array} \right]
 \end{array}
 \end{array}$$

Quelle précision doit-on accorder à cet estimateur ? L'exactitude de nos évaluations dépend de différents facteurs :

- le respect, d'une année sur l'autre, de la structure croisant le département et le risque,
- la taille de la matrice : plus le nombre de cases est important, plus nos calculs seront sujets à caution,
- l'importance du risque « autres dommages aux biens » : celui-ci vient contrarier nos estimations sur l'incendie pur à partir de 1987; avant cette date, le facteur est bien sûr sans objet.

Si la portée des deux derniers facteurs peut facilement être mesurée (il y a en moyenne 32 cases dans les matrices mais seules 24 nous intéressent, et le risque indésirable représente entre 5 et 10 % du chiffre d'affaires total d'un département), en revanche, la justesse du choix qui nous a conduits à retenir le premier facteur ne pourra être vérifiée qu'après avoir procédé à quelques simulations.

Celles-ci seront menées en comparant les structures croisées estimée et réelle, pour l'année n , dans l'hypothèse où les informations de deux années successives n et $n+1$ sont disponibles.

Compte tenu du fait que chacun des termes de la matrice est utilisé par la suite comme élément de la marge par entité et risque, on appréciera la distance entre les matrices estimée et réelle grâce à la distance du χ^2 suivante :

$$d^2 = \sum_{i=1}^2 \sum_{j=1}^{3.n_r} \frac{(n_{ij} - n_i.p_j)^2}{n_i.p_j}$$

Les matrices ont fait auparavant l'objet d'une linéarisation, et se présentent sous la forme d'un vecteur $(1, 3.n_r)$, n_r représentant le nombre de départements de la région considérée.

L'hypothèse est en effet faite que les deux distributions r et e constituent des échantillons issus de deux variables aléatoires :

- nombre estimé de contrats en portefeuille par entité et département pour l'année n et la région r (n_1),
- nombre de contrats réellement en portefeuille pour l'année n et la région r (n_2).

En notation matricielle, cela donne :

$$\begin{bmatrix} e_1 & \dots & e_j & \dots & e_{3.n_r} \\ r_1 & \dots & r_j & \dots & r_{3.n_r} \end{bmatrix} = \begin{pmatrix} e \\ r \end{pmatrix}$$

Il s'agit de montrer que ces deux échantillons proviennent d'une même population : c'est l'hypothèse H_0 que l'on fait lors de l'écriture du test du χ^2 , contre H_1 : " les échantillons sont significativement différents ".

Dans l'hypothèse H_0 on peut parler de probabilités $p_1, p_2, \dots, p_{3.n_r}$ de posséder les modalités $1, 2, \dots, 3.n_r$. Il s'agit alors de comparer les effectifs constatés n_{ij} aux effectifs espérés $n_i.p_j$, qui ne doivent pas en être très éloignés.

Dans l'hypothèse H_0 , d_0^2 est une réalisation d'une variable D_0^2 suivant un χ^2 dont le nombre de degrés de liberté est égal au nombre de termes de d^2 ($2.3.n_r$), auquel on doit retirer $3 + n_r$, le nombre de relations liant les termes de la distribution estimée, c'est-à-dire celles qui lient :

- la somme de chaque colonne : 3,

- la somme de chaque ligne : n_r ,

des matrices croisant les départements et les risques.

Donc D_0^2 est un $c_{6n_r - n_r - 3}^2 = c_{5n_r - 3}^2$.

Cependant, comme les $p_1, p_2, \dots, p_{3.nr}$ ne sont pas connus, il faut les estimer par $\hat{p}_j = \frac{n_j}{2n}$, ce qui représente $3n_r - 1$ relations indépendantes. D_0^2 est alors un $c_{5n_r - 3 - (3n_r - 1)}^2 = c_{2(n_r - 1)}^2$ si H_0 est vraie. Si $n_r = 8$, il y a alors 14 degrés de liberté.

Le test consistera à rejeter H_0 si d^2 constaté est trop grand, comme pour un test habituel du χ^2 .

Le tableau 2.1.10 donne les résultats des différents χ^2 des 12 régions, calculés pour le chiffre d'affaires de l'année 1992 à partir de 1993, et la valeur du χ^2 correspondant à accepter à tort l'hypothèse H_0 pour un degré de confiance de 95 %. Les montants sont exprimés en million de F.

Tableau 2.1.10 : Comparaison selon la région des χ^2 relatifs au calcul du chiffre d'affaires par département et risque (1992)

région	Nombre de départements	χ^2 calculé	χ^2 théorique	χ^2 calculé / chiffre d'affaires
Ouest	11	0,29	18,3	0,09 %
Est	8	0,17	14,1	0,09 %
Rhône-Alpes	8	0,43	14,1	0,13 %
Sud-Ouest Montauban	7	0,64	12,6	0,05 %
Sud-Est	12	0,16	19,7	0,04 %
Alsace-Moselle	4	0,05	7,8	0,04 %
Lorraine	6	0,02	11,1	0,02 %
Sud-Ouest Toulouse	8	0,18	14,1	0,07 %
Centre	12	0,09	19,7	0,04 %
Nord	4	0,07	7,8	0,03 %
Bassin parisien	8	0,55	14,1	0,23 %
Paris	7	0,11	12,6	0,04 %

Il ressort de l'analyse de ce tableau une grande proximité des distributions réelle et estimée pour l'ensemble des régions, puisque le χ^2 calculé est très inférieur au χ^2 théorique : *ceci justifie a priori l'utilisation de la méthode R.A.S. pour le calcul des marges par entité et département pour les années 1979 à 1991.*

La précision liée à ce calcul peut en outre être approchée par le calcul du rapport de la distance du χ^2 calculé sur la moyenne des chiffres d'affaires des distributions estimée et réelle : cet indicateur est toujours inférieur à 0,25 %.

I.3.4.2 - Redistribution du portefeuille de Michelin

Un autre aspect de la reconstitution du portefeuille est la prise en compte de la redistribution du portefeuille de Michelin : en effet, le contrat liant l'UAP à cette entreprise prévoit, entre autres, la couverture du risque d'habitation des employés de la firme. Or, si le montant de la prime pour ce contrat est bien connu, le nombre de risques géographiquement distincts qu'il couvre n'a fait l'objet d'une comptabilisation correcte qu'à partir de 1991, le nombre de contrats étant auparavant égal à 1.

Comme il se trouve que le portefeuille de Michelin est géographiquement réparti sur l'ensemble du portefeuille du Puy-de-Dôme, la solution de bon sens consiste à :

- estimer le nombre de contrats de la « branche-risque » 1402 (celle réservée à Michelin), à partir de la connaissance des primes annuelles et d'une estimation de la prime moyenne afférente à chaque risque,
- réintégrer le nombre de contrats estimés à la distribution du portefeuille exprimé en volume, pour le risque d'habitation et le département du Puy-de-Dôme.

La répartition des risques du portefeuille de Michelin se fera alors naturellement, dans une troisième étape, et sera identique à la distribution par zone postale des contrats du Puy-de-Dôme.

Il reste cependant à préciser les règles d'évaluation du nombre de contrats Michelin. La technique la plus simple consiste à dresser l'historique de la prime annuelle moyenne par risque, sur la période réduite 1991-1996. On constate une évolution très voisine de l'indice des prix ou de l'indice FNB : c'est d'ailleurs cet indice qui est utilisé pour l'actualisation des années 1979 à 1990.

A cette évaluation de la prime moyenne correspond une estimation du nombre de contrats, dont la vraisemblance dépend directement des valeurs prises : des valeurs trop éloignées de l'actuel nombre auraient décrédibilisé l'estimateur.

Dans l'impossibilité de mesurer la qualité de l'approximation du nombre de contrats, il faut néanmoins se rappeler que le risque d'erreur maximum, égal au rapport

de ce nombre sur le total du nombre de contrats habitation en France sur la période, est inférieur à 1 %.

I.3.4.3 - Autres pistes d'amélioration de la précision de l'estimation du portefeuille

Tous les développements précédents ont eu pour principal objectif l'approximation, de la manière la plus rigoureuse possible compte tenu des moyens à notre disposition, des distributions par zone postale du nombre de contrats et des P.E.N.A., pour les trois risques incendie.

Un indicateur de précision, mesurant le phénomène étudié mais appliqué sur une période différente, a permis de valider la méthode retenue.

Cependant, d'autres pistes d'amélioration de la précision de l'estimateur existent, mais leur mise en œuvre coûteuse en temps n'a pu être finalisée.

Il faut signaler cependant l'existence de descriptifs départementaux, trouvés lors de la collecte d'informations, présentant l'activité et l'implantation des réseaux Agents, S et tous réseaux pour la branche incendie. Ils donnent entre autres un indice d'implantation, rapportant le nombre de contrats en portefeuille à l'inventaire 1984 au nombre d'habitants de ce département.

L'exploitation de ces descriptifs est malaisée car :

- seul le risque habitation et la totalisation des risques sont fournis pour le réseau S et le total des réseaux; les risques agricole et d'entreprise ne peuvent donc être isolés;
- les départements de la région parisienne ne figurent pas au palmarès, de même qu'une totalisation de la France;
- en outre, sont aussi exclus de cette statistique les contrats des réseaux S et C dont on ne peut situer le lieu du risque.

Néanmoins, les totalisations par réseau ou par risque sont disponibles par l'intermédiaire d'autres états déjà cités.

Le descriptif du réseau agents ne retient pas, comme variable de rupture, le département du lieu du risque mais celui de l'agence. Cet état est donc issu de la même source que celle utilisée pour la détermination des distributions par département et risque (cf. le paragraphe I.3.4.1) à partir de la méthode du raking ratio. Le réseau agents nous a déjà offert une bonne partie de ses « secrets » et l'apport de précision supplémentaire ne justifie donc pas un changement de méthode : cet état ne servira qu'à en ratifier les résultats.

En revanche, il semble digne d'intérêt d'exploiter ces sources pour les réseaux S et C, malgré leur imperfection : la détermination de la distribution par département du lieu du risque du nombre de contrats par risque pour le réseau S sera la première étape. Puis, lors d'une deuxième étape, se fera l'estimation, par différence entre le total France et les réseaux A et S, de cette même distribution pour le courtage.

Le tableau 2.1.11 présente l'architecture des données du réseau S à notre disposition, à l'issue de l'incontournable phase de saisie. Il donne, sur un exemple créé pour la circonstance, le résultat de la méthode retenue pour l'estimation des cases du tableau croisant les risques (agricole, d'entreprise) avec les départements d'une région donnée, la marge verticale, en gras, étant connue.

Tableau 2.1.11 : Exemple de reconstitution par risque et département du portefeuille régional du réseau S

<i>département</i>	<i>risque agricole</i>	<i>risque d'entreprise</i>	<i>total</i>
département 1	2	4	6
département 2	4	8	12
département 3	6	12	18
département 4	2	3	5
région	14	27	41

Cette méthode, qui pose l'hypothèse d'une indépendance parfaite entre la variable *risque* et la variable *département*, retient la formule suivante pour l'estimation de chaque case n_{ij} :

$$n_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$$

où $n_{i.}$ et $n_{.j}$ sont respectivement la i-ème valeur de la marge verticale (i.e. le total par département) et la j-ème valeur de la marge horizontale (i.e. le total par risque).

Cette hypothèse s'avère réaliste quand la région est homogène, mais peu crédible lorsqu'il existe en son sein une opposition entre des départements urbains, où le risque agricole est peu représenté à l'inverse du risque d'entreprise, et d'autres ruraux. Dans ce cas, le recours à une variable d'intérêt, telle que le nombre d'entreprises par département ou le nombre d'exploitations agricoles, est un facteur d'amélioration sensible de la crédibilité des résultats.

Auparavant, la marge horizontale du tableau 2.1.11, inconnue, aura été appréciée grâce à la méthode d'indépendance des variables précédemment utilisée et appliquée au tableau 2.1.12 croisant le nombre de contrats selon deux axes :

- le risque (agricole, professionnel),
- l'une des régions, ainsi que l'I le-de-France et les contrats « délocalisés ».

Tableau 2.1.12 : Architecture du portefeuille du réseau S selon la nature et le risque

<i>nature</i>	<i>risque agricole</i> (1)	<i>risque d'entreprise</i> (2)	<i>sous-total</i> (1)+(2)	<i>risque d'habitation</i> (3)	<i>total</i>
province (c)	3 514	15 560	19 074	251 545	270 619
Ile-de-france (a) contrats « délocalisés » (b)					
sous-total à affecter (h=a +b)	583	2 579	3 162	66 505	69 667
total réseau S	4 097	18 139	22 236	318 051	340 286

Ce tableau synoptique présente :

- dans les cases vides, les données inconnues du réseau S : les contrats « délocalisés » et ceux de l'Ile-de-France,
- en caractère normal, les données estimées par la méthode d'indépendance des variables « nature » et « risque »,
- en caractère gras, les données réelles.

Si la méthode précédente conduit à des résultats satisfaisants, il faut néanmoins en reconnaître les limites :

- elle concerne les 88 départements appartenant aux 10 régions de province,
- elle ne règle pas le sort des contrats « délocalisés »,
- elle exclut les 7 départements de la région Ile-de-France, dont on ne connaît pas la part dans l'ensemble formé avec les risques « délocalisés ».

On se heurte en effet à une double indétermination, due au nombre inconnu des contrats « délocalisés » (*b*) et à la méconnaissance de leur distribution géographique. La première de ces indéterminations est levée en considérant que le poids de ces contrats non localisés est peu différent de celui de 1993, connu. De même, on pourra s'appuyer sur la distribution des proportions de contrats « délocalisés » par département en 1993 pour régler la deuxième indétermination. Bien entendu, chaque risque est à traiter séparément, les risques agricole et d'habitation étant moins affectés par la méconnaissance du lieu du risque.

Ainsi, le modelage du tableau 2.1.12 sera achevé lorsque la ligne « province » sera complétée de la part $(1-p).b$ des contrats « délocalisés » situés en province, *p* étant la

part du nombre de contrats dont le lieu du risque est inconnu et en fait situés en Ile-de-France, sur le nombre de contrats au lieu de risque inconnu; et lorsque la ligne « Ile-de-France » sera remplie des contrats de l'Ile-de-France (a) et du reste des contrats non affectés et situés dans cette région ($p.b$).

Le portefeuille estimé de l'Ile-de-France \hat{a}' vaut alors :

$$\hat{a}' = \hat{a} + \hat{p}.\hat{b} = h + (\hat{p} - 1).\hat{b},$$

où h est le portefeuille non intégré dans le descriptif. L'estimation globale du portefeuille de province \hat{c}' est égale à :

$$\hat{c}' = c + (1 - \hat{p}).\hat{b}$$

Il est possible de se ramener à un seul niveau d'indétermination en supposant que la part p de contrats *non localisés* et en réalité situés en Ile-de-France est proche de la part p' de contrats *localisés* situés dans cette région. Or, cette proportion vaut exactement :

$$p' = \frac{a}{a + c}.$$

L'expression du portefeuille estimé de l'Ile-de-France \hat{a}' n'est alors plus fonction que de l'estimateur de b :

$$\hat{a}' = h + (\hat{p}' - 1).\hat{b} = h - \frac{c.\hat{b}}{h + c - \hat{b}},$$

de même que celle du portefeuille de province :

$$\hat{c}' = c + (1 - \hat{p}').\hat{b} = c + \frac{c.\hat{b}}{h - \hat{b} + c}$$

Puis vient la phase d'arrondissement de la matrice avec respect des marges, qui a été exposée au paragraphe I.3.1 lors de la présentation de la méthode R.A.S..

La deuxième étape vient alors, où se fait l'estimation, par différence entre le total France et les réseaux A et S, de cette même distribution pour le courtage. Les mêmes opérations :

- de reconstitution par risque (agricole et professionnel) et département du portefeuille régional du réseau C (cf. tableau 2.1.11),
- d'estimation de la part représentée par l'Ile-de-France sur le total du courtage,
- de répartition des contrats « non localisés » entre la province et Paris,

sont de nouveau mises en œuvre.

Il reste alors à comparer, tour à tour pour les réseaux S et C, les distributions départementales du risque obtenues :

- par la méthode d'estimation à évolution homogène,
- par la méthode utilisant les descriptifs départementaux.

Le tableau 2.1.13 suivant relève les distances du χ^2 pour les réseaux S et C et pour chacun des risques, calculées sur la base du nombre de contrats en portefeuille à l'inventaire de 1984.

Tableau 2.1.13 : Comparaison des distributions départementales des réseaux S et C fournies par deux méthodes grâce à la distance du χ^2 , selon le risque (portefeuille 1984)

<i>réseau</i>	<i>risque agricole</i>	<i>risque d'habitation</i>	<i>risque d'entreprise</i>
Salariés	245,8	14.312	998
Courtage	30,2	10.458	1.828

Compte tenu des résultats et du nombre de degrés de liberté - 94 -, on peut dire que les deux distributions sont dans tous les cas significativement différentes : la probabilité, pour ces valeurs du χ^2 , de considérer à tort ces deux échantillons comme issus d'une même loi est très faible.

L'apport de la méthode issue du descriptif, très significatif du point de vue de la précision, est donc primordial.

Mais quel crédit peut-on réellement accorder à cette méthode ? Bien qu'elle contienne en elle une part d'incertitude, on peut néanmoins maximiser le risque d'erreur en faisant correspondre au nombre de contrats délocalisés b un intervalle de valeurs auquel il appartient de manière certaine :

$$b \in \left[\hat{b} - \frac{1}{2}\hat{b}; \hat{b} + \frac{1}{2}\hat{b} \right].$$

On en déduit alors les distributions extrêmes résultantes, dont on calcule la distance du χ^2 avec la distribution estimée lorsque $b = \hat{b}$. Ces χ^2 correspondent à une probabilité p_1 (pour le nombre le plus faible) et p_2 (pour le nombre le plus fort) de considérer à tort la distribution estimée avec l'une ou l'autre des distributions extrêmes comme issues d'une même source.

La méthode issue du descriptif départemental permet donc une amélioration sensible de la distribution départementale des portefeuilles des réseaux S et C. Elle est donc à privilégier pour la province.

Deux conséquences sont à tirer de ce perfectionnement : il faut d'une part « caler » la distribution 1984 par zone postale des départements de province, estimée par la

méthode à évolution homogène, sur le total départemental fourni par la méthode issue du descriptif. Cela revient en fait à opérer, pour chaque zone postale d'un département, une simple règle de trois proportionnelle au rapport du portefeuille évalué par la méthode issue du descriptif à celui estimé par la méthode à évolution homogène $(\frac{P_{84}^d}{P_{84}^{eh}})$.

D'autre part, la réévaluation de la distribution du portefeuille de 1984 doit être accompagnée de celle des années 1985 à 1992, selon une formule modulant l'impact du rapport des portefeuilles $(\frac{P_{84}^d}{P_{84}^{eh}})$ selon le rapport du nombre d'années séparant 1993 et l'année courante n et 9 :

$$P_n^d = P_n^{eh} \left(\frac{P_{84}^d}{P_{84}^{eh}} \right)^{\frac{93-n}{9}}, n = 84, \dots, 93$$

Cette formule permet un *lissage exponentiel de la différence existant sur les années 1984 et 1993 entre les deux estimateurs du portefeuille*, représentés par deux courbes se rejoignant en 1993, comme la figure 2.1.3 en témoigne.

Quant à la période antérieure à 1984, il convient, lors de la nouvelle estimation de la valeur du portefeuille de chaque zone postale, de *conserver la croissance annuelle enregistrée pour chacun des réseaux*, et de prendre pour référentiel le portefeuille 1984 estimé par l'utilisation du descriptif départemental, de sorte que l'on déduit P_{n-1}^d par la formule suivante :

$$P_{n-1}^d = P_{n-1}^{eh} \frac{P_n^d}{P_n^{eh}}, n = 80, \dots, 84$$

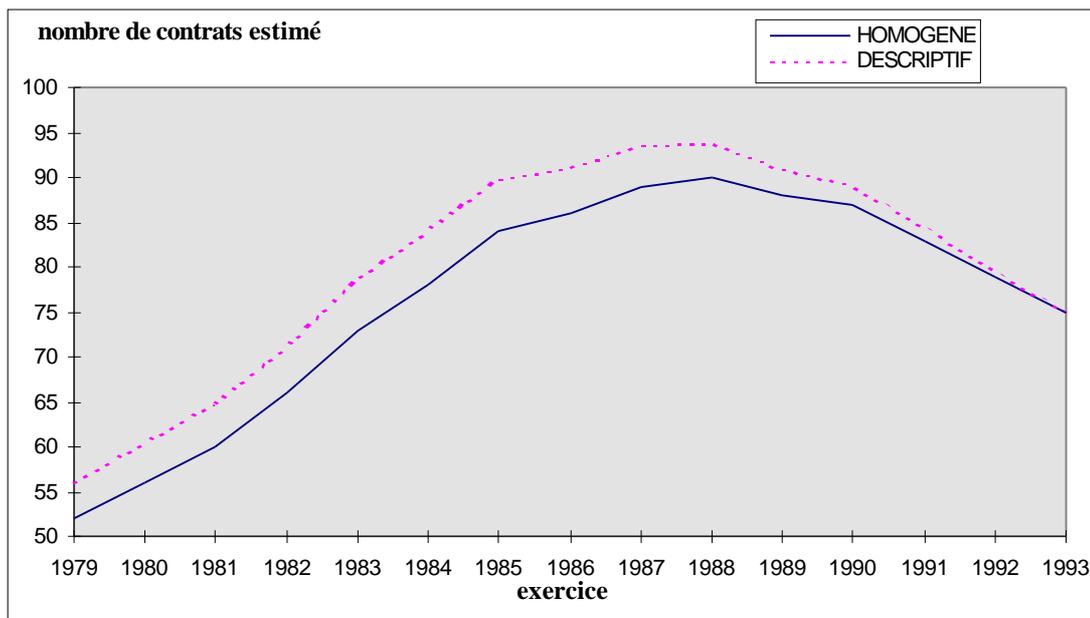


Figure 2.1.3 : Exemple du passage de l’estimateur à évolution homogène à celui issu du descriptif départemental pour le portefeuille des réseaux S et C

II - Constitution de la base de données “ Sinistres ”

Parallèlement à la reconstitution du portefeuille d’assurances, il est nécessaire de réunir, pour le calcul d’une fréquence ou d’un taux de destruction due à la tempête, un ensemble d’informations relatives aux sinistres dont la cause est la tempête, c’est-à-dire une base de données “ sinistres ”.

Nous verrons dans un premier temps le périmètre de l’étude, puis les modalités de collecte de ces dossiers tempêtes ayant affecté le portefeuille de l’UAP de 1980 à 1996. Enfin, nous envisagerons les possibles enrichissements et les nécessaires redressements à apporter à cette base.

II.1 - Périmètre de la base de données “ sinistres ”

II.1.1 - Définitions

La base de données “ **sinistres** ” constitue un élément incontournable à l’étude des tempêtes. On appelle ici *sinistre* toute atteinte fortuite portée à l’intégrité d’un *bâtiment* ou d’un *bien*, en un endroit et en un lieu précis, et dont les conséquences sont éventuellement garanties par un *contrat d’assurance* souscrit auprès de l’UAP. On veillera donc à ne pas confondre cette notion avec le concept d’*événement*, qui englobe plusieurs sinistres et dont les conséquences peuvent être garanties par un *traité de réassurance*. De même, il ne sera question ici que des conséquences *matérielles*

directes du sinistre, les conséquences corporelles ou indirectes (tels que la perte d'exploitation, le préjudice financier,...) sortant de notre champ d'investigation.

En outre, comme cela a déjà été dit, la période d'observation des sinistres provoqués par les effets du vent est limitée à l'intervalle 1980 - 1997. Cela correspond aux années pour lesquelles l'information relative aux sinistres a pu être collectée, non sans certaines difficultés, comme nous le verrons par la suite. Pour les années antérieures à 1980, il n'existe plus de données informatiques disponibles puisque l'UAP n'a décidé la création du fichier " historique sinistres " qu'en 1986.

II.1.2 - Description de la base de données " sinistres "

Offrant une information synthétique mais néanmoins relativement riche, la base de données " sinistres " sera au final constituée d'environ 300.000 articles précisant les caractéristiques administratives et les dommages occasionnés par la tempête *au niveau de chaque sinistre*.

Chaque article doit contenir au final:

- le numéro d'identifiant du *sinistre*,
- le numéro d'identifiant du *contrat*,
- le *jour de survenance* du sinistre,
- la *zone postale* du lieu de survenance du sinistre,
- le code " *branche-risque* ", codification propre à l'UAP suffisamment fine pour retrouver les trois risques,
- le code numérique des " *ventilations sinistrées* ", c'est-à-dire des garanties affectées par le sinistre,
- le montant, en franc courant, du *coût réel du sinistre pour la société d'assurance*,
- le montant, en franc courant, des *honoraires versés à l'expert*.

Même si, par certains aspects, cette tâche paraît fastidieuse, la collecte d'informations est un passage obligé avant une étape déterminante, l'actualisation de la base de données, objet du chapitre 2 de cette partie. Le contenu de cette base ainsi modifiée se révélera très utile lors de la mise en place d'une tarification spécifique à la tempête (chapitre 3).

II.2- Extraction des sinistres dus à la tempête

II.2.1 - Sélection des sinistres dont la garantie tempête est affectée

Pour extraire les informations d'origine sinistres dus à la tempête, des travaux d'exploitation informatique se sont révélés nécessaires sur la période 1980 – 1997. Le département de l'informatique a ainsi mis à la disposition des rédacteurs de ce mémoire deux transactions contenant la liste des sinistres dus à la tempête. Ces listes, portant les noms barbares de 2STEMP et 2STEMQ, résultent de la sélection, des fichiers “ historique sinistres ” et “ Fichier général des sinistres ”, des seuls sinistres dont la garantie tempête est affectée (i.e. dont la valeur de la “ ventilation sinistrée ” vaut 12).

Grâce à l'aide du département de l'informatique, un temps précieux a pu être épargné et plus de 290.000 sinistres dus à la tempête extraits.

La figure 2.1.4 reprend l'image de la transaction 2STEMP, dont les variables sont exactement celles présentées au paragraphe II.1.2, et qui plus est dans un ordre identique.

R. SINISTRE	R. CONTRAT	SURV	LIEU	BRII	VS	VS2- VS3	TOTALPT.	PTS. EXP
8200310023	200000047545	820108	20999	5080	12		1600	1600
8200310115	200000061015	820129	38999	5201	12		2960	
8200310155	200000060454	820305	30999	5271	12		2097	315
8200310258	200000048090	820310	95999	5261	12		2995	467
8200310311	200003962597	820310	92999	1581	12			
8200310317	200000061274	820210	92999	1575	12		6552	660
8200310339	200009032595	820101	63999	1580	12		3410	
8200310346	200000049443	820310	78999	5211	12			
8200310354	200006508311	820310	77999	1575	12		27430	2000
8200310396	200006506824	820310	92999	1583	12		400	400
8200310404	200006508311	820310	77999	1575	12			
8200310422	200000052573	820301	94999	1534	12			
8200310423	200000046615	820302	75999	1575	12			
8200310427	200000030902	820301	80999	5101	12			
8200310440	200000035098	820305	93999	5201	12		580	580
8200310441	200000035097	820310	94999	5201	12		880	600
8200310477	200000061274	820310	92999	1575	12		2948	616
8200310527	200003644947	820301	76999	5201	12		6325	600
8200310557	200006507746	820310	95140	1550	12		580	580
8200310559	200000061274	820310	92999	1575	12			

Figure 2.1.4 : Image de la transaction 2STEMP

II.2.2 - Identification de sinistres tempête réglés au titre de la loi sur les catastrophes naturelles

II.2.2.1 - Justification de l'existence de ces sinistres tempête

Pourtant, le dénombrement des sinistres dus à la tempête n'est pas achevé : en effet, le chapitre 4 de la 1ère partie a présenté en détail la loi du 13 juillet 1982 relative aux catastrophes naturelles; en particulier, son champ d'intervention n'excluait pas, jusqu'en 1990, la tempête. C'est ainsi que des arrêtés interministériels ont été pris, permettant l'indemnisation de nombreux dommages dus au vent en 1982, 1983, et de 1986 à 1989.

Ces règlements ont pu bénéficier à certains assurés de l'UAP :

- dont le contrat " incendie " ne comprenait pas une garantie protégeant contre les effets du vent,
- dont les dommages touchaient certains éléments appartenant à la liste des exclusions de la garantie tempête de leur contrat.

En revanche, si un contrat bénéficiait d'une couverture tempête, l'arrêté stipulait généralement que la garantie légale intervenait dans un deuxième temps, dans la mesure où la couverture contractuelle n'était pas suffisante.

L'existence de sinistres dus à la tempête, mais réglés grâce au dispositif légal, ne fait pas de doute. Il reste toutefois à en mesurer l'importance pour l'UAP, en volume et en valeur.

II.2.2.2 - Méthode de recherche des sinistres tempête réglés en 1982 au titre de la loi

L'UAP a mis en place en 1982 un dispositif de comptabilisation des sinistres réglés au titre de la loi du 13 juillet 1982; précisément, une nouvelle " ventilation sinistrée " (V.S.), dont le code est 30, autorise une connaissance précise du nombre et du montant des sinistres dus à une catastrophe naturelle.

Il a été aussi créé pour l'occasion une codification caractérisant la nature des événements naturels à l'origine du sinistre : on la retrouve systématiquement dans la fiche informative de chaque dossier. A titre d'illustration, on notera que la lettre P correspond à " tempête ", alors que le N est synonyme d' " inondation " .

Par conséquent, pour mener à bien notre recherche des sinistres dus à la tempête mais réglés au titre de la loi du 13 juillet, il est nécessaire de disposer d'un fichier informatique contenant à la fois :

- la variable “ ventilation sinistrée ”, nécessaire à la sélection des seuls sinistres contenant la V.S. 30,
- et la variable “ nature de l’événement ”, autorisant le choix des seuls sinistres occasionnés par la tempête.

Une enquête menée au sein de l’UAP nous conduit au secteur réassurances, qui a demandé la création d’un fichier “ catastrophes naturelles ” disposant de tels éléments d’information. Le fichier mis à notre disposition a été constitué en décembre 1992 et permet d’établir la liste des sinistres “ cat. nat. ” pour les exercices de survenance postérieurs ou égaux à 1983.

Cependant, ce n’est qu’en 1988 qu’il a été décidé de prolonger de 5 années supplémentaires sur le fichier “ général des sinistres ” la durée de présence des sinistres portant une V.S. 30. L’exercice 1982 n’a malheureusement pas fait l’objet de telles mesures conservatoires, et l’on ne dispose plus que d’une infime partie de la liste des numéros de sinistres “ cat. nat. ” pour cette année.

La liste de ces sinistres “ cat. nat. ” est pourtant nécessaire pour mesurer l’impact de la loi du 13 juillet dans le règlement des événements tempête.

Quelle stratégie alternative nous permettrait d’arriver à nos fins ?

Avant de le savoir, il est intéressant de s’interroger sur le potentiel maximum représentés par ces sinistres. C’est ainsi que, suite à l’exploitation d’un fichier informatique “ synthèse par ventilation sinistrée ” nommé VSI, on dénombre, au total pour 1982 :

- 38 157 sinistres dont la garantie “ catastrophe naturelle ” est touchée (i.e. contenant la V.S. 30),
- dont 28 156 ont en outre leur garantie “ tempête ” affectée (i.e. contenant la V.S. 12).

Il y a donc 10 001 sinistres “ cat. nat. ” susceptibles d’avoir été déclarés en raison des dommages provoqués par les tempêtes en 1982. Celles-ci ont très durement frappé la France cette année-là, et en particulier la tempête survenue dans le sud-ouest entre le 6 et le 10 novembre 1982, et ayant fait l’objet d’un arrêté interministériel.

Même s’il n’est pas, à ce stade de notre recherche, possible d’estimer la part de sinistres “ cat. nat. ” véritablement occasionnés par la tempête, il ne saurait être question d’abandonner les investigations. Il reste à en définir les modalités.

Cette recherche se fera en deux étapes :

- en premier lieu, circonscrire les gisements potentiels de sinistres “ cat. nat. ” et en extraire les numéros,
- puis, dans un deuxième temps, vérifier a posteriori que ces numéros correspondent bien à des sinistres “ cat. nat. ”, par l’interrogation d’une transaction relative aux sinistres (cf. la figure 2.1.5 à ce sujet).

```

4NMENU, S- A- 8200310023                                     *. F?*
POL 2 0000 0047545 C B. R 5080 S. SERV 4713 MDN 001 B A DELEGATION : 51
ASS U N I C O M I L COURT 1 0027433 T CLIENT                B * SITUATION EN HISTORIQUE
SIN 82 00310023 K DU 08 01 LIEU 20999 AGNT 000058 C        D 13 310882 (310882)
TR  CIRC 62 RESP 7 RSRV 10 BLOC 1 FR D RECUP. FI 99 D S. M F 19 310882 (310882)
PRES          S    VS NS CHARGE. PAL  CHARGE. REC  S C T E PTS. PAL.  PTS. REC. F
SUBI          12          15999          0 0 0 F          1600
PROT          S
GEST
INSP          REG1
CH. MS
CAT. NAT 1
LESE 000 RGL          REC. DOS          ORIG 30 R. PRO 1 M IDA P          S
F1 VS  NS  TX  MN          MX          SF  N AIDE+ S . . . . . I
F2 VS  NS  TX  MN          MX          SF  P VS NS
F3 VS  NS  TX  MN          MX          SF  Q FRA          SUSP4
CLA
ADVERSAIRE          DOM          RC. ADV 5 S ST          SJ
TYP. VEH  IMMAT          POL          SIN          CIE  T
COAS 08 TXAXA 080 0000 TXCOA          APERIT          SIN. APE          U
IMMAT. ASS          COND          SEX          CELIB  NAIS          PERM          V
MARQ          DOM AP  EXP 1          GRAVE          OPPOSIT          LOCAL          W
T. GEST GAD EVAG          PE  AP. RD  PI  TARIF          GAR          A  Y

```

Figure 2.1.5: Fiche signalétique d’un sinistre de l’UAP

La première étape doit faire l’objet d’un soin tout particulier : en effet, il faut savoir qu’environ 1 million de sinistres sont déclarés chaque année à l’UAP. La deuxième étape s’avérerait particulièrement fastidieuse si la technique de localisation n’était pas efficace : en effet, on ne peut vérifier la présence ou l’absence de la V.S. 30 sur la fiche signalétique que pour 5.000 sinistres au maximum par jour. Cependant, ce n’est qu’à la fin de nos investigations que cette technique pourra être qualifiée.

Pour rendre plus facile la localisation de ces sinistres “ cat. nat. ” sans V.S. 12 dont on doit retrouver le numéro, il convient de connaître les filons éventuels par réseau : l’utilisation du fichier VSI précédent permet de cerner le potentiel des réseaux S et C et de chaque entité régionale du réseau agents. (Pour la BNP, la vente de contrats par son réseau des guichets n’était pas à l’époque encore apparue.) Il sera alors possible de savoir à tout moment à quel *taux de découverte* se situent nos recherches.

Tout naturellement, *la recherche des numéros de sinistres doit s’orienter vers les plages des entités ou des agences particulièrement affectées* : la territorialité de ces dernières limite en effet l’espace commercial de leur représentant à certains cantons bien déterminés, leur portefeuille étant géographiquement situé dans les immédiats alentours de leur pas de porte, selon des cercles concentriques de densité décroissante en clients. (C’est ce qu’une étude, en marge de ce mémoire, a montré). D’autre part, le

changement de région d'un assuré vers la zone d'activité d'une autre agence doit normalement conduire à un transfert du contrat de l'assuré vers le portefeuille de l'agence dans la zone de chalandise de laquelle se trouve sa nouvelle adresse. Ces deux raisons contribuent à ce que le portefeuille sinistré d'une zone appartienne dans l'écrasante majorité des cas à l'agence la plus proche. Pour les autres réseaux, il existe des plages spécifiques au risque incendie qui limitent ainsi les recherches.

La liste de ces agences peut être dressée à partir de l'exploitation du fichier des sinistres dont la garantie tempête est touchée (i.e. dont la valeur de la V.S. vaut 12), établi au paragraphe II.2.2.1.

De même, *on privilégiera les plages de numéros de sinistres succédant à celles principalement occupées par ces sinistres avec une V.S. 12*. Quelques explications sont nécessaires à la bonne compréhension de cette règle : il faut savoir en effet que le numéro d'un sinistre n'est pas attribué aléatoirement, mais qu'il est déterminé en fonction de la date de *déclaration* de l'événement, c'est-à-dire de la date à laquelle l'assureur inscrit dans ses fichiers les informations de ce sinistre. *Le délai de déclaration*, c'est-à-dire la période séparant la date à laquelle l'assuré est informé de la survenance du sinistre et la date de déclaration, étant normalement limité à 5 jours, cette date est souvent très corrélée avec la date de *survenance*, date à laquelle l'événement prend naissance.

C'est en effet sur cette idée de bon sens que repose notre technique de recherche : *les déclarations de sinistres qui suivent la déclaration d'un sinistre ayant une V.S. 12 à une date et un lieu conformes à ceux d'une catastrophe naturelle de type tempête ont de grandes chances de concerner aussi une tempête*.

Qui plus est, les dates et lieux de survenance où le nombre des sinistres avec une V.S. 12, extraits du fichier établi au paragraphe II.2.2.1, est le plus grand coïncident exactement avec ceux des catastrophes naturelles, prouvant ainsi que des sinistres "cat. nat." ont eu pour origine la tempête. Si çà n'avait pas été le cas, toute recherche sur ces plages de sinistres aurait été vaine.

Sur un plan pratique, les agences (AGENCE sur la figure 2.1.6) appartenant aux départements les plus touchés (DEPT) sont sélectionnées et, pour chacune d'elles, les numéros du premier (PREMSIN) et du dernier (DERNSIN) sinistres avec une V.S. 12. Bien entendu, les sinistres ayant une V.S. 12 et dont les dates de survenance (DATSI) coïncident avec celles des catastrophes naturelles de type tempête ont des numéros (NUSIN) figurant dans cet intervalle. On génère alors automatiquement la liste des numéros de sinistre appartenant à cet intervalle et suivant le numéro des sinistres avec une V.S. 12 : ils constituent la liste des sinistres "cat. nat." potentiels (SINCAT1 à SINCAT24). La figure 2.1.6 indique le processus de génération automatique de numéros de sinistres "cat. nat." potentiels réalisé sous Excel.

NUSIN	AGENCE	DATSI	COPOS	DEPT	PREMSIN	DERNSIN	SINCAT1	SINCAT2	SINCAT3
-------	--------	-------	-------	------	---------	---------	---------	---------	---------

8219730388	1973	821107	19999	19	8219730031	8219730527	0	0	0
8219730389	1973	821107	19510	19	8219730031	8219730527	8219730390	0	0
8219730391	1973	821107	19999	19	8219730031	8219730527	0	0	0
8219730392	1973	821107	19999	19	8219730031	8219730527	0	0	0
8219730393	1973	821107	19999	19	8219730031	8219730527	0	0	0
8219730394	1973	821107	19999	19	8219730031	8219730527	0	0	0
8219730395	1973	821107	19999	19	8219730031	8219730527	8219730396	8219730397	8219730398
8219730399	1973	821107	19999	19	8219730031	8219730527	0	0	0
8219730400	1973	821107	19999	19	8219730031	8219730527	0	0	0
8219730401	1973	821107	19999	19	8219730031	8219730527	0	0	0
8219730402	1973	821107	19999	19	8219730031	8219730527	8219730403	8219730404	8219730405
8219730406	1973	821107	19999	19	8219730031	8219730527	0	0	0
8219730407	1973	821107	19999	19	8219730031	8219730527	8219730408	8219730409	8219730410
8219730412	1973	821107	19999	19	8219730031	8219730527	8219730413	8219730414	8219730415
8219730421	1973	821107	19999	19	8219730031	8219730527	0	0	0
8219730422	1973	821107	19999	19	8219730031	8219730527	8219730423	8219730424	8219730425
8219730436	1973	821107	19999	19	8219730031	8219730527	0	0	0
8219730437	1973	821107	19999	19	8219730031	8219730527	8219730438	8219730439	8219730440

Figure 2.1.6 : Processus de génération automatique de numéros de sinistres “ cat. nat.” potentiels

La génération de cette liste donne lieu, par la suite, à la photographie, selon une procédure automatisée, de la fiche informative de chaque sinistre (cf. la figure 2.1.5). Cette opération, bien que limitant au maximum l'intervention humaine, est “chronophage” : on ne peut photographier plus de 500 fiches à l'heure. Puis, il convient de visionner chaque fiche pour voir si, oui ou non, y figure la V.S. 30. Cette opération, elle aussi, est automatiquement réalisée, mais elle est pratiquement immédiate.

Il est alors temps de tirer des enseignements de cette phase de recherche. Le tableau 2.1.14 permet de juger de l'efficacité de notre méthode à travers un indicateur très simple, le *taux de V.S. 30*. Celui-ci rapporte le nombre de fiches informatives appartenant à des sinistres “ cat. nat. ” sur le nombre total de ces fiches collectées. Il est évalué selon *l'ordre de génération* du numéro de sinistre, c'est-à-dire le rang que ce dernier occupe après le numéro du sinistre tempête à l'origine de la génération.

On constate que le taux de V.S. 30 décroît à mesure que l'ordre de génération croît, atteignant plus de 25 % pour les sinistres dont le numéro suit immédiatement celui d'un sinistre “ tempête ”. Ce résultat s'explique par la plus grande probabilité qu'un sinistre tempête suive un sinistre de même nature : les déclarations des assurés arrivent, tel un processus de Poisson, en agence et sont souvent traitées de manière groupée par le personnel d'agence. A l'inverse, pour les numéros de sinistre situés entre les 4^{ème} et 11^{ème} rangs, le taux de V.S. 30 n'est que de peu supérieur à 10 % tout en étant peu inférieur au rang précédent, montrant que le phénomène s'estompe très progressivement.

Tableau 2.1.14 : Taux de sinistres avec V.S. 30 selon l'ordre de génération de numéros suivant les sinistres tempête de 1982

ordre de génération	nombre de fiches interrogées	nombre de fiches avec V.S. 30	taux de V.S. 30
n + 1	8 453	2 255	26,7 %
n + 2	5 546	946	17,1 %
n + 3	4 313	616	14,3 %
n + 4	3 491	413	11,8 %
n + 5	2 924	319	10,9 %
n + 6	2 503	279	11,1 %
n + 7	2 175	211	9,7 %
n + 8	1 895	200	10,6 %
n + 9	1 695	164	9,7 %
n + 10	1 175	115	9,8 %
n + 11	1 002	114	11,4 %
total	35 172	5 632	16,0 %

Ces résultats encourageants démontrent a posteriori la justesse de nos analyses et l'efficacité de la technique mise en œuvre : pas moins de 5.632 numéros interrogés, sur 35.172, appartiennent à des sinistres “ cat. nat. ”, ce qui correspond à un taux de V.S. 30 de 16,0 %. *Le taux de découverte dépasse donc les 56 %.*

On peut donc continuer ces investigations plus avant, dans la mesure où l'atteinte de ces résultats n'est pas trop coûteuse.

Ces recherches prennent deux directions complémentaires :

- d'une part, la poursuite de l'examen des numéros de sinistres de la 12ème à la 24ème position derrière le numéro d'un sinistre “ tempête ” dont les coordonnées coïncident avec celles d'une catastrophe naturelle,
- d'autre part, l'exploration des numéros de sinistres succédant à ceux de sinistres “ tempête ” dont les coordonnées ne coïncident pas avec celles d'une catastrophe naturelle.

Si le premier axe de recherche n'est pas nouveau, en revanche il suppose de concentrer ses efforts sur les départements dont le taux de V.S. 30 reste à un niveau acceptable, et pour lesquels le filon n'est pas tari. Les listes de numéros interrogeables se réduisent au fur et à mesure que le dernier numéro de sinistre “ tempête ” de chaque agence est atteint.

Le deuxième axe de recherche offre une perspective nouvelle, l'analyse, par acquis de conscience, du taux de V.S. 30, de sinistres dont les numéros suivent ceux de sinistres " tempête " vérifiant l'une des trois situations suivantes :

- la date de survenance correspond à celle d'une catastrophe naturelle, mais le département de l'agence ne figure pas parmi ceux de la zone sinistrée,
- le département de l'agence correspond à l'un de ceux de la zone sinistrée, mais la date de survenance ne figure pas parmi celles d'une catastrophe naturelle,
- le département de l'agence ne figure pas parmi ceux de la zone sinistrée, de même que la date de survenance, absente de la liste des dates des catastrophes naturelles.

Les départements absents de la zone sinistrée en sont néanmoins adjacents, de manière à limiter la zone à des dimensions réalistes. C'est cette même préoccupation qui explique la raison d'une période de recherche étendue aux quelques jours précédant ou suivant la date de survenance des événements naturels.

Les fruits de ces recherches s'avèrent bien modestes. Même si ces résultats sont décevants, ils ne font que confirmer les présomptions de moindre présence de sinistres catastrophiques dans des zones et à des dates non concernées par ces cataclysmes.

L'ensemble des résultats précédents est présenté dans le tableau 2.1.15 suivant, donnant, pour chaque réseau ou entité régionale, le poids des différents types de sinistres tempête à l'origine de la découverte des sinistres avec V.S. 30, et le taux de découverte finalement atteint.

Le taux de découverte finalement atteint, plus de 70 %, est très satisfaisant. Il en est même très supérieur dans certaines entités régionales très marquées par les tempêtes en 1982. Cependant, ce taux étant plus faible pour d'autres régions, convient-il d'y poursuivre des investigations pour le faire augmenter?

Il semble falloir répondre par la négative, dans la mesure où les régions dont le potentiel de sinistres non découverts ayant une V.S. 30 reste important coïncident avec les zones très marquées par les inondations centennales de décembre 1982. D'autre part, il convient, avant d'envisager une éventuelle reprise de cette enquête, de caractériser les sinistres " cat. nat. " obtenus de manière à séparer le bon grain de l'ivraie, en d'autres termes, à distinguer les sinistres dus à la tempête de ceux ayant une autre origine.

Tableau 2.1.15 : taux de découverte des sinistres avec V.S. 30 selon les caractéristiques des sinistres tempête à l'origine de la génération de leur numéro et selon l'entité

entité	avec les caractéristiques des cat. nat.	potentiel	taux de découverte
Ouest	59	273	17,8 %
Est	121	363	25,0 %
Rhône-Alpes	1 901	408	82,3 %
Sud-Ouest Montauban	1 165	125	90,3 %
Sud-Est	1 682	221	88,4 %
Alsace-Moselle	1	150	0,7 %
Lorraine	1	102	1,0 %
Sud-Ouest Toulouse	111	90	55,2 %
Centre	1 061	720	59,6 %
Nord	2	14	12,4 %
Bassin Parisien	1	9	10,5 %
Paris	0	12	0,0 %
réseau S	670	206	76,5 %
réseau C	356	177	66,8 %
Total UAP	7 131	2 870	71,3 %

II.2.2.3 - Caractérisation des sinistres tempêtes réglés au titre de la loi

Les renseignements recueillis lors de l'étape précédente font alors l'objet d'une analyse détaillée destinée à préciser sans ambages la nature du sinistre et l'origine des dommages : la cause en est-elle la tempête ?

La réponse n'est pas toujours simple car l'on se heurte bien souvent à de nombreuses difficultés lors de l'étude des fiches signalétiques des sinistres. Ainsi n'est-il pas rare de rencontrer des fiches dont *la nature de l'événement générateur est indéterminée ou incertaine* : c'est en particulier le cas lorsque est utilisée la modalité W de la variable " nature de l'événement ", qui définit toute catastrophe naturelle autre que celles nommément désignées par une lettre :

- N : inondations,
- P : tempêtes (ouragans, cyclones, grêle sur les toitures),
- Q : séismes (tremblements de terre, éruptions, raz-de-marée),
- R : mouvements de terrain (glissements, affaissements),
- S : avalanches,
- T : poids de la neige.

Cette situation conduit à regrouper sous ce vocable des événements de nature indéterminée, dont on ignore l'origine atmosphérique ou terrestre. Pour lever le doute,

une personne du Secteur des Méthodes Administratives et Comptables de l'UAP se reporte alors au texte relatif à la catastrophe naturelle paru au Journal Officiel et précise alors les lettres à employer dans la fiche signalétique. On s'aperçoit alors que cette lettre W désigne le plus fréquemment les orages, accompagnés ou non de vent.

De plus, cette codification n'offre pas un grand confort d'utilisation car souvent on observe un *événement "mixte", c'est-à-dire de natures différentes* : la survenance d'une tempête accompagnée de fortes pluies entraînant des inondations est fréquente. Il faut alors utiliser l'une des 2 lettres N et P à l'exclusion de l'autre sur la fiche signalétique pour qualifier l'événement à l'origine du sinistre, même si l'on peut imaginer un bâtiment endommagé à la fois par le vent et par les inondations.

Enfin, il n'est pas rare de constater des *anomalies dans l'utilisation des codifications* par le rédacteur du dossier sinistre, du fait de leur complexité ou de l'absence ou de l'oubli des consignes en la matière. Ces informations contradictoires conduisent alors à sélectionner à tort un sinistre pensant qu'une tempête se trouvait à l'origine de ce dernier. Le rejet d'un sinistre tempête est aussi un cas envisageable.

Ces faits militent donc pour la mise en place d'un *score*, permettant de mieux apprécier la probabilité qu'un sinistre ayant une V.S. 30 sans V.S. 12 ait été engendré par une tempête. Il permettra de régler les cas litigieux où des renseignements d'un même sinistre sont contradictoires.

Les indicateurs intégrant ce score sont choisis pour leur lien supposé avec le phénomène :

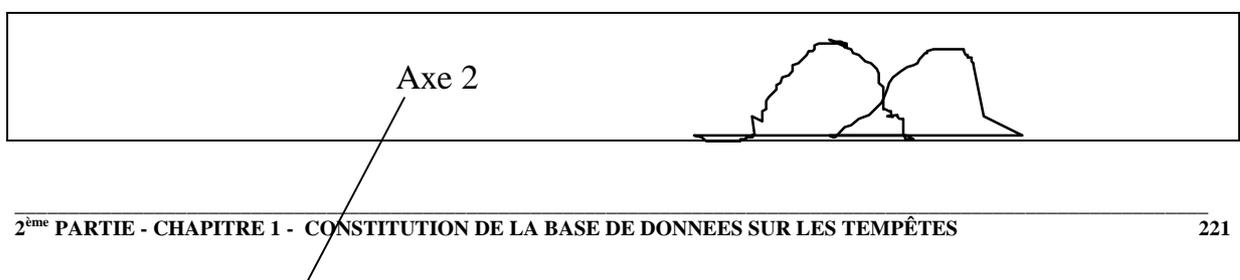
- la nature de l'événement : à travers cet indicateur, qui qualifie une catastrophe naturelle, il est possible de qualifier aussi un sinistre,
- la date de survenance : une date ne correspondant pas à des périodes de tempête qualifiées de catastrophes naturelles exclurait immédiatement le sinistre,
- le lieu de survenance : celui-ci permet d'une part de sélectionner les seuls sinistres survenus dans des zones qualifiées de catastrophes naturelles et, d'autre part, de mesurer la densité locale de la tempête; la zone postale, qui est retenue comme subdivision géographique pour préciser le lieu de survenance, fait l'objet d'un traitement particulier dans le paragraphe II.3.1,
- la *densité locale de la tempête* : cette variable rapporte, au niveau d'une zone postale et à une date donnée, le nombre de sinistres possédant une V.S. 12, au portefeuille estimé dans la zone postale; (on se reportera à la 1ère partie de ce chapitre pour obtenir de plus amples précisions sur l'estimation du portefeuille); une forte densité renforcera la décision de classer le sinistre dans la famille des sinistres dus à la tempête,

- la *part locale de la catastrophe naturelle* : cet indicateur rapporte le nombre de sinistres ayant une V.S. 30 au nombre total de sinistres possédant une V.S. 12 ou une V.S. 30 : à mesure que ce taux croît, diminue la probabilité que le sinistre soit consécutif à une tempête,
- le code *circonstance* : cette variable qualitative renvoie à un cas précisant les circonstances dans lesquelles le sinistre est apparu; en particulier, deux de ses modalités désignent la tempête, mais sa fiabilité est douteuse,
- la ou les *ventilations sinistrées choisies pour le règlement* : ce renseignement peut se révéler très précieux. En effet, il arrive qu'un règlement soit effectué sous forme d'avance à l'assuré sinistré, avant que la tempête ait fait l'objet d'un arrêté de catastrophe naturelle. Le règlement est donc enregistré sous une V.S. différente de la V.S. 30, précisant bien souvent la garantie affectée, qui peut être la tempête. Par la suite, à la publication de l'arrêté, une annulation d'écriture est opérée conduisant à l'enregistrement des sommes sous la V.S. 30. Ces écritures seront ensuite transmises par l'UAP à la C.C.R. à des fins de justificatifs,
- la *présence ou l'absence de la garantie tempête sur le contrat sinistré* : la carence de cette garantie explique le règlement du sinistre dans le cadre légal; à l'inverse, l'existence d'une garantie tempête conduit logiquement à la conclusion que la tempête n'est pas à l'origine du sinistre.

L'analyse discriminante répond à l'objectif que nous nous sommes fixé, qui est de séparer les deux types de sinistres, selon qu'ils sont ou non dus à la tempête. L'analyse factorielle discriminante consiste en effet à rechercher des axes a - dits discriminants - qui séparent le mieux les k catégories (ou sous-nuages) d'un échantillon (ou nuage) de n individus décrits par p variables numériques. Ces axes, résultat de la projection, selon une métrique donnée, des observations, sont choisis de telle manière que :

- les k centres de gravité g_j des sous-nuages soient aussi éloignés que possible les uns des autres, maximisant ainsi la *variance interclasse* B ,
- et que chaque sous-nuage se projette de manière aussi groupée que possible autour de son centre de gravité, minimisant ainsi la *variance intraclasse* W .

L'axe 1 de la figure 2.1.7 possède un bon pouvoir discriminant tandis que l'axe 2 (qui est l'axe principal usuel d'une analyse en composantes principales) ne permet pas de séparer en projection les deux groupes.



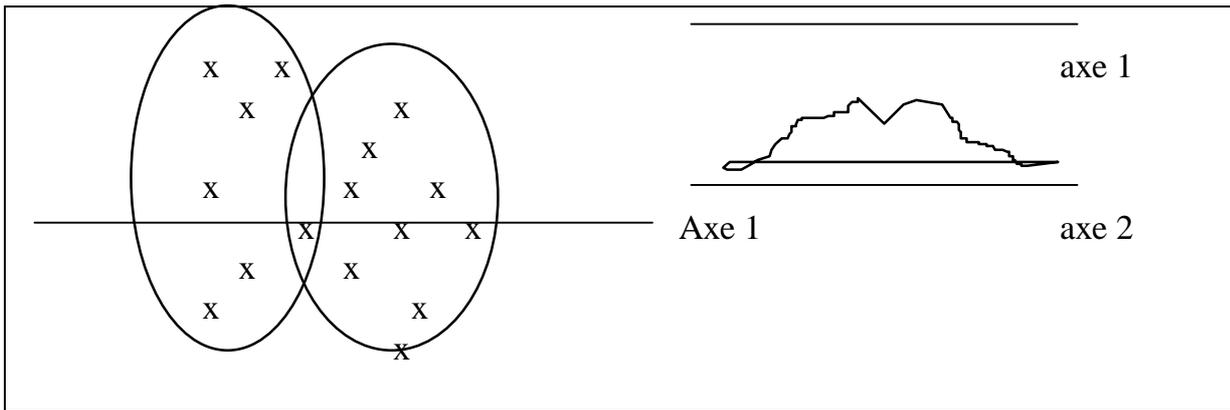


Figure 2.1.7 : Présentation d'axes aux pouvoirs discriminants différents

Cette méthode conduit donc à rechercher l'axe a maximisant le rapport de l'inertie interclasse à l'inertie totale, selon la formule matricielle suivante :

$$\max_a \frac{a' MBMa}{a' MVMa}, \text{ où :}$$

- M est la métrique associée à l'espace \mathfrak{R}^p des individus;
- B est la matrice (p,p) de variances interclasse, c'est-à-dire la matrice de variances des k centres de gravité des sous-nuages affectés des poids q_j :

$$B = \sum_{j=1}^k q_j (g_j - g)(g_j - g)', \text{ où } g \text{ est le vecteur } (p,1) \text{ centre d'inertie du nuage;}$$

si l'on centre les variables explicatives X_p , c'est-à-dire si $g = 0$, et que l'on considère chaque observation avec un poids identique, l'expression de B se simplifie :

$$B = \frac{1}{n} \sum_{j=1}^k n_j g_j g_j', n_j \text{ étant l'effectif du sous-nuage } E_j;$$

- V est la matrice (p,p) de variances du nuage E , c'est la somme des matrices de variances interclasse (B) et intraclasse (W), cette dernière étant la moyenne des matrices de variances des sous-nuages :

$$W = \frac{1}{n} \sum_{j=1}^k n_j V_j$$

La théorie nous enseigne que le maximum est atteint si a est vecteur propre de : $(MVM)^{-1}MBM$ associé à sa plus grande valeur propre λ_1 :

$$M^{-1}V^{-1}BMa = \lambda_1 a$$

λ_1 prend des valeurs comprises entre 0 et 1. Le cas $\lambda_1 = 1$ correspond à des variances intraclasse nulles et les sous-nuages appartenant chacun à un hyperplan orthogonal à a . Il y a évidemment discrimination parfaite si les centres de gravité se projettent en des points différents. $\lambda_1 = 0$ correspond au cas où le meilleur axe ne permet pas de séparer les centres de gravité qui sont confondus. Le nombre des valeurs propres non nulles, donc d'axes discriminants, est égal à $k-1$ dans le cas habituel où $n > p > k$ et où les variables ne sont pas liées par des relations linéaires.

Dans le cas présent, il n'y a qu'une seule variable discriminante puisque $k - 1 = 1$. L'axe discriminant est alors nécessairement la droite reliant les deux centres de gravité g_1 et g_2 . Par la suite, on peut considérer comme une variable qualitative à deux niveaux le fait que le sinistre ait été causé ou non par une tempête. Notre objectif étant de trouver des critères permettant de classer chaque sinistre selon son origine, il est alors logique de considérer que cette variable est à expliquer à travers un modèle de régression multiple impliquant ces "prédicteurs". En effet, l'on sait que, lorsque la variable à expliquer est qualitative, l'analyse discriminante n'est qu'une généralisation du modèle de régression.

Concrètement, on a le modèle :

$$y = XB$$

où y est la variable à expliquer, centrée et donc ne prenant que deux valeurs a et b sur les groupes de sinistres 1 et 2 respectivement ($n_1.a + n_2.b = 0$). X est la matrice (n, p) des critères et B est le vecteur discriminant que l'on calcule comme suit :

$$B = (X'X)^{-1}X'y.$$

$$\text{Le choix } a = \frac{n}{n_1}, b = -\frac{n}{n_2} \text{ conduit alors à } B = V^l(g_1 - g_2).$$

Les hypothèses habituelles de la régression ne sont pas vérifiées. Au contraire, y est non aléatoire et X l'est. On ne peut donc utiliser, autrement qu'à titre indicatif, les statistiques usuelles fournies par le logiciel SAS.

L'emploi de ce modèle nécessite bien évidemment la connaissance de la variable à expliquer. *Or, c'est justement ce que nous cherchons à connaître.* Pour pouvoir se servir de ce modèle comme d'un procédé d'affectation des sinistres en deux groupes, on va donc rechercher la véritable nature d'un certain nombre de sinistres, composés d'au moins 50 sinistres "cat. nat." suite à une tempête, et de 50 sinistres "cat. nat." ayant une origine différente. Ceci constitue notre vecteur y .

Bien évidemment, ce n'est qu'*a posteriori* que les sinistres étudiés nous révèlent leur qualité et il est donc exclu que le tirage aléatoire d'un certain nombre de ces sinistres conduise à une proportion de 50 % pour chacun des groupes. Le tirage au sort des numéros de sinistres s'effectue donc selon un schéma hypergéométrique dans lequel le groupe le moins nombreux doit atteindre 50 unités pour que le processus s'interrompe : l'effectif n de l'échantillon est alors une variable aléatoire. Sans

information complémentaire sur cette population, il est difficile de procéder à une stratification de celle-ci : chaque individu de la population des sinistres “ cat. nat. ” collectés, d’effectif N , est donc supposé avoir une probabilité égale à $1/N$ d’être tiré au sort et ne peut l’être qu’une fois. L’utilisation d’une fonction sous le logiciel SAS permet de se remettre dans les conditions du tirage P.E.S.R. (à probabilités égales sans remise).

Indirectement, la constitution d’un échantillon, destiné à déterminer la meilleure fonction linéaire discriminante de deux groupes, nous permet d’estimer la proportion p du nombre de sinistres “ cat. nat. ” suite à une tempête, relativement au nombre total de sinistres “ cat. nat. ” : ceci n’est qu’une évaluation réalisée sur plus de 100 sinistres, mais elle donne un “ ordre d’idée ” du poids de ces sinistres “ cat. nat. ” imputables à une tempête.

Le nombre de sinistres finalement tirés au sort est de 115. La catégorie des sinistres “ cat. nat. ” suite à une tempête contient 50 éléments. La proportion \hat{p} est donc égale à :

$$\hat{p} = \frac{50}{115} = 43,5\% .$$

L’étude des sinistres tirés au sort n’est pas toujours une sinécure : si l’analyse de certains d’entre eux n’offre pas de difficulté, en revanche, pour d’autres, il est nécessaire d’avoir recours au dossier sur support papier en raison de la double nature de la catastrophe naturelle. La lecture de la fiche informatique du sinistre ne permet pas en effet de distinguer :

- le règlement d’un dégât des eaux provoqué par l’inondation résultant des fortes pluies accompagnant la tempête,
- et le règlement des dommages entraînés par les infiltrations dues à l’arrachage d’éléments de couverture provoqué par le vent et aux précipitations qui l’ont accompagné.

Dans le 1er cas, seule la base de la construction a subi des dommages et il n’y a pas lieu de sélectionner ce sinistre. Dans le 2ème cas, en revanche, le caractère mixte de la catastrophe naturelle s’affirme à travers la double nature des dommages. Les termes de la garantie tempête précisent bien que les conséquences de l’arrachage de la toiture (i.e. les dégâts dus aux infiltrations) sont pris en compte durant les 48 heures suivant l’événement générateur, et il y a bien lieu de retenir ce sinistre, qui plus est pour la totalité du montant réglé.

Il se peut que le dossier du sinistre soit indisponible ou pilonné : les dizaines de kilomètres de galeries des deux lieux de stockage de l’UAP ne suffiraient pas en effet

pour conserver les millions de dossiers sinistres ayant existé. Il faut alors procéder à un tirage complémentaire.

Quant aux variables qualitatives composant la matrice X , elles sont transformées en variables numériques, booléennes et centrées.

Une analyse factorielle discriminante entre les sinistres dus à une tempête ou dus à une autre cause est alors réalisée sous le logiciel SAS. Elle fournit les coefficients des différents « prédictors » retenus dans le modèle. donne alors les résultats suivants.

Fort de la meilleure représentation de la séparation en 2 classes des n individus, on peut alors chercher à affecter une observation e à l'un des groupes. La règle naturelle d'affectation consiste à calculer les distances de l'observation à classer à chacun des 2 centres de gravité et à affecter selon la distance $d^2(e ; g_i)_{i=1,2}$ la plus faible. En d'autres termes, en utilisant la métrique W^{-1} :

$d^2(e; g_i) = (e - g_i)'W^{-1}(e - g_i)$, e étant le vecteur (p, l) des coordonnées de l'observation, ce qui, en développant, conduit à :

$$d^2(e; g_i) = e'W^{-1}e + g_i'W^{-1}g_i - 2e'W^{-1}g_i.$$

Comme $e'W^{-1}e$ ne dépend pas du groupe g_i , cette règle, dite de Mahalanobis - Fisher, règle consiste donc à chercher le minimum de l'expression linéaire par rapport à e : $g_i'W^{-1}g_i - 2e'W^{-1}g_i$, ou le maximum de : $e'W^{-1}g_i - (g_i'W^{-1}g_i)/2$. C'est en fait cette dernière expression que l'on conserve.

Il convient donc, pour chaque observation, de calculer 2 fonctions discriminantes et d'en chercher la valeur maximale. La décision finale conduit à affecter l'observation au groupe 1 si l'on a :

$$e'W^{-1}(g_1 - g_2) > \frac{1}{2} (g_1 + g_2)' \cdot W^{-1}(g_1 - g_2).$$

L'intérêt de ce score est d'analyser chaque sinistre selon les mêmes critères. Une étude détaillée d'un petit nombre de sinistres choisis selon une méthode de sondage à tirage aléatoire fournit des règles discriminantes que l'ensemble des sinistres " cat. nat. " pour les années 1982, 1983, 1986 à 1989 respecte. Cette méthode discriminante présente l'avantage d'automatiser et de simplifier le processus d'identification.

Il aurait été possible de recourir à la méthode dite du " rouleau compresseur " et consistant dans les étapes suivantes :

- recherche des dates et des lieux relatifs à des catastrophes naturelles dues à une tempête; on a vu que la nature n'est pas toujours d'une grande précision,
- sélection des sinistres survenus à ces dates et à ces lieux,

- enquête sur la nature précise de la cause de ce sinistre.

Mais, outre sa longueur et les gros moyens qu'elle nécessite, cette méthode présente le handicap d'exclure les éventuels sinistres dus réellement à la tempête mais dont le libellé officiel n'en ferait pas mention. Cette situation est pourtant assez fréquente pour les événements mixtes.

II.2.2.4 - Résultats de la méthode d'identification

Le tableau 2.1.16 présente les résultats de la méthode d'identification des sinistres "cat. nat." qui sont d'origine tempête. Le taux correspondant y figure par exercice de survenance.

Tableau 2.1.16 : Répartition du nombre de sinistres "cat. nat." d'origine tempête par exercice de survenance

Exercice de survenance	Nombre de sinistres "cat. nat."	Nombre de sinistres "cat. nat." d'origine tempête	Taux de sinistres "cat. nat." d'origine tempête
1982	7 116	4 160	58,5 %
1983	961	696	72,4 %
1986	92	27	29,3 %
1987	688	402	58,4 %
1988	230	157	68,3 %
1989	224	163	72,8 %
TOTAL	11 506	5 605	48,7 %

Au total, ce sont donc plus de 5 600 sinistres qui viennent rejoindre la base de sinistres extraits par les transactions mises à notre disposition par le département de l'informatique. Au total, la base sinistres contient près de 300.000 sinistres.

II.3 - Enrichissement de l'information sur les sinistres tempête

Mais l'on ne saurait se satisfaire de la seule connaissance du nombre de sinistres tempête et du montant des règlements occasionnés par chacun d'entre eux : si ces éléments sont indispensables pour le calcul d'une prime pure, ils sont néanmoins insuffisants pour l'édification d'un tarif, le prix du risque tempête devant en outre intégrer les dimensions temporelle et géographique de ces sinistres.

II.3.1 - Redressement du code postal indéterminé

II.3.1.1 - Constat chiffré

Si la date de survenance est toujours indiquée sur la fiche signalétique d'un sinistre, en revanche, le code postal, qui est la subdivision géographique retenue dans notre étude, ne se situe pas à ce niveau de qualité d'information. En effet, l'encodage du lieu de survenance du sinistre s'est pendant très longtemps limité aux deux caractères numériques du département. Le code postal, bien qu'apparu en France en 1972, n'a été adopté par l'UAP qu'en 1982, à la suite de la parution de la loi du 13 juillet. Les catastrophes naturelles pouvant avoir un caractère très localisé, les arrêtés interministériels précisent le nom des communes ou des cantons sinistrés. L'encodage du code postal sur la fiche signalétique permet alors un contrôle de validité des renseignements inscrits, dans la mesure où seuls les sinistres dont le lieu de survenance se situe dans les zones postales sinistrées peuvent accueillir une V.S. 30.

L'indétermination du code postal est un phénomène d'ampleur pouvant remettre en cause l'affectation de chaque zone postale à la station météorologique la plus proche.

En effet, cette imprécision touche plus de 44.000 sinistres dits " classiques ", extraits des transactions " sinistres tempête " mises à notre disposition par le département de l'informatique. C'est sans compter les zones postales extérieures à la France métropolitaine. Si l'on rapporte ce nombre au volume de sinistres tempête, le taux d'imprécision est donc proche de 15 %. Ce taux atteint même 93 % pour les exercices antérieurs à 1982. L'année 1982 constitue une rupture et, très vite, ce taux s'établit à environ 1 % pour les exercices postérieurs à 1982.

A ce groupe de sinistres " délocalisés ", il convient d'ajouter les sinistres d'origine " cat. nat. " et dont on veut connaître la nature pour opérer l'analyse discriminante, soit près de 8.000. Au total, ce sont environ 52.000 sinistres susceptibles d'être dus à la tempête qui se trouvent touchés par cette imprécision.

II.3.1.2 - Recueil d'information complémentaire

Si les renseignements saisis au niveau de la fiche signalétique du sinistre sont absents ou erronés, il est néanmoins parfois possible de les retrouver ou de les corriger grâce à la fiche signalétique du contrat sinistré.

C'est le cas pour le lieu de survenance du sinistre, auquel on peut substituer avec un relatif bonheur le lieu du risque, présent sur la fiche signalétique du contrat. En effet, si l'on excepte les risques industriels, souvent caractérisés par une tarification appliquée à une entreprise dans sa globalité, et donc aux établissements géographiquement distincts qui la composent, le lieu du risque est généralement identifié puisqu'il constitue une variable de tarification dont la saisie est obligatoire. Comme les biens assurés sont fixes, il ne peut y avoir de différence entre le lieu de

Tableau 2.1.17 : Répartition du nombre de contrats sinistrés “ délocalisés ” par réseau

type de contrat	contrats sinistrés en 1980	contrats sinistrés après 1980			
		réseau S	réseau C	réseau A	total
entité	tous réseaux				
nombre de contrats	3174	3202	1103	27667	31972
nombre de contrats sans fiche	1369	750	282	5316	6348
taux de contrats sans fiche	43,3 %	23,4 %	25,6 %	19,2 %	19,9 %

Cette collecte fastidieuse de renseignements présente des résultats encourageants, même si près de 20 % des contrats sinistrés “ délocalisés ” ne peuvent toujours pas être situés à l’intérieur d’un département.

Un autre recueil d’informations sur le portefeuille des contrats “ cat. nat. ” présumés tempête est aussi organisé dans le but de mettre en œuvre l’analyse discriminante décrite au paragraphe II.2.2.3 : plus de 7.000 écrans “ portefeuille ” sont ainsi recherchés : le résultat de l’investigation conduit à un taux de contrats sinistrés “ délocalisés ” voisins de celui des contrats sinistrés tempête.

Le 2ème axe de recherche est le retour du centre d’archivage d’un certain nombre de contrats dépourvus de fiche à l’issue de la 1ère phase de recherche. La quête est très décevante puisque le nombre de dossiers réceptionnés atteint seulement environ 3 % du nombre de ceux commandés.

Cette quête ne s’arrête pas là : en effet, il est envisageable qu’un contrat ait été sinistré à plusieurs reprises. C’est plus fréquemment le cas des risques d’entreprise ou agricole que du risque d’habitation. Il existe donc une probabilité de rencontrer des numéros de contrat possédant à la fois un code postal du lieu de survenance du sinistre se terminant par ‘999’ (signe d’indécision), et parfaitement valide : ce dernier code postal peut alors très bien figurer parmi les codes postaux de survenance les plus probables. On définit alors pour ces numéros de contrat ce que l’on appelle le *code postal de survenance le plus courant*. On peut aussi imaginer que, parmi ces numéros de contrat, il y en ait quelques uns dont la fiche portefeuille n’ait pas été trouvée au cours des étapes précédentes : le choix de ce code postal s’imposerait alors.

On trouvera dans l’annexe E le détail du programme TEMP8C, écrit sous SAS, de détermination du code postal le plus courant. Au final, une recherche lancée sur les

41.500 contrats tempête conduit à une liste de près de 5.000 numéros de contrats différents. Bien sûr, la proportion, parmi ces derniers, que les transactions d'interrogation du portefeuille historique n'aurait pas identifiés est bien plus marginale : par exemple, pour les contrats sinistrés en 1980, cette démarche nous a permis d'avoir un code postal bien identifié pour 57 contrats sur un total de 3.339 !

Enfin, une autre idée consiste à rechercher la zone postale la plus sinistrée pour une date et un département donnés : les tempêtes présentent parfois un caractère très localisé et peuvent ne toucher qu'une bande étroite du territoire d'un département. La difficulté de cette méthode réside dans sa mise en œuvre puisqu'elle repose sur le code postal que justement nous recherchons, mais qui fait en partie défaut et qui ne sera parfaitement connu qu'a posteriori. Néanmoins, différents indices, tels :

- le code postal du lieu de survenance quand il est parfaitement défini,
- le code INSEE du lieu du risque du contrat sinistré " délocalisé ",
- l'éventuel code postal le plus courant d'un contrat,

peuvent être réunis pour constituer ce que l'on appellera le *code postal le plus fréquent*.

II.3.1.3 - Algorithme de redressement du code postal indéterminé

A l'issue de cette phase préliminaire de collecte de renseignements relatifs à la situation du risque, il est permis d'écrire l'algorithme de redressement du code postal indéterminé. Certaines informations, extraites de la fiche synthétique d'un contrat, seront pleinement exploitées, à savoir :

- le numéro de contrat (NUCON),
- le code branche-risque (BRRI),
- le code postal de l'adresse (COPOS),
- la ville de l'adresse (VILLE),
- le code INSEE du lieu du risque (INSEE),
- l'adresse du risque (ADRRISQ),
- la ville du risque (VILLRISQ),
- le complément éventuel à apporter à l'adresse du risque (COMPRISQ),

auxquelles on peut ajouter le code postal du lieu de survenance du sinistre le plus courant (COPOSCOU) et le code postal du lieu de survenance du sinistre (COPOSIN).

Muni de ces différents indicateurs, l'algorithme de redressement du code postal va déterminer pour chaque contrat sinistré " délocalisé " son code postal du lieu du sinistre *le plus vraisemblable*. Le terme de " délocalisé " prend ici une acception large puisqu'il concerne tous les sinistres dont le code postal ne figure pas dans la table de correspondance entre le code INSEE et le code postal de novembre 1997 (cf. le paragraphe I.2.2.1 de ce présent chapitre). Il faut en effet faire preuve de méfiance vis à vis de la codification de la zone postale, quand bien même un contrôle de cohérence

empêche la saisie de codes postaux CEDEX ou absents de la table CPOSINS *au moment du sinistre* : en effet, la création ou la disparition de codes postaux est un phénomène relativement fréquent. Il est donc probable que quelques codes postaux ne figurent plus *aujourd'hui* dans la table CPOSINS.

L'algorithme contient deux phases : la 1ère réside dans le décryptage d'éventuelles informations intéressantes obtenues lors de l'étape précédente présentée au paragraphe II.3.1.2. Cette mise en valeur a pour résultat la génération de quelques codes postaux. La 2ème phase consiste en leur hiérarchisation, conduisant finalement à l'élection d'un seul.

La 1ère phase de l'algorithme conduit à la formation du maximum de codes postaux que les 3 zones ADRRISQ, VILLRISQ et COMPRISQ, de l'écran "portefeuille" à notre disposition, peuvent nous permettre de détecter. Cette analyse se fait sans conséquence sur le produit de notre recherche : ainsi, s'il se trouve un code postal généré correspondant au profil des codes postaux acceptables, il ne sera pas pour autant mis fin à l'investigation pour ce contrat, cette entreprise étant systématiquement étendue aux 3 zones d'un contrat.

L'une des difficultés pratiques présentées par cette 1ère phase est la programmation, sous le logiciel SAS, de *l'extraction de données numériques* d'une zone de texte libre normalement utilisée pour l'adresse, et dont la structure s'identifie à celle d'un code postal : 4 ou 5 chiffres. Les éventuels renseignements numériques qu'on peut y trouver, tel le numéro de rue, de boîte postale ou de bâtiment, ne peuvent en général, de par leur structure numérique dépassant très rarement 3 chiffres, provoquer de confusion. La consultation de la table CPOSINS nous renseigne sur la validité de ce code postal.

Le programme pousse la minutie jusqu'à analyser les nombres à deux chiffres présents dans l'une de ces 3 zones et leur égalité avec le département du lieu de survenance du sinistre (DEPSIN) : il n'est pas rare en effet de voir figurer ce dernier dans la zone de texte libre, suivi d'un nom de localité située dans ce département. La coïncidence est trop forte pour ne pas en extraire le contenu alphabétique et en déduire le code postal correspondant.

C'est ici qu'apparaît l'autre difficulté de programmation de l'algorithme : l'extraction des mots issus des zones alphabétiques et leur interprétation en tant que noms propres empruntés à la toponymie. Une table informatique, mise à notre disposition par UAP ASSISTANCE et propriété de l'Institut Géographique National (I.G.N.), offre la relation entre le nom de la commune ou d'un lieu dit et son code postal et nous permet de générer alors le code postal correspondant.

En termes plus concrets, cette traduction d'un nom en un code postal suppose au préalable une comparaison de deux chaînes de caractères, l'une empruntée au portefeuille de l'UAP et l'autre issue du fichier de l'I.G.N.. Il va sans dire que les chances de succès dans leur rapprochement dépendent dans une large mesure du

respect, par les personnes en charge à l'UAP ou en agence de la saisie des affaires nouvelles ou remplacements, des règles de syntaxe adoptées par l'I.G.N.. Ainsi, par exemple, empêcheront tout rapprochement :

- une lettre minuscule,
- les saints non orthographiés 'ST',
- l'absence de trait d'union entre les différentes parties d'un toponyme (à l'exception de l'article),
- l'absence de l'apostrophe derrière un article élidé et, bien sûr,
- une faute d'orthographe.

Ces règles ne sont bien sûr pas respectées, d'autres consignes semblant avoir été données et être utilisées avec plus ou moins de zèle. Il est par conséquent nécessaire de retraduire dans la syntaxe de l'I.G.N. les mots ou expressions citées ci-dessus.

En outre, il peut être opportun de coller des mots constituant un même toponyme mais appartenant à deux des trois zones alphabétiques. L'algorithme réalise le "collage" de ces mots séparés par un article ou un mot de liaison tel que " DE ", " SUR " ou " EN ".

Ainsi, une adresse étant généralement constituée de nombres et de mots, une chaîne de caractères sera extraite d'une des trois zones de texte libre si cette dernière ne contient pas de code postal. elle correspondra :

- au 1er mot si cette zone ne contient pas de chiffre,
- au dernier mot si cette zone contient un nombre dont la structure est distincte de celle d'un code postal.

On constate en effet une plus grande fréquence de zones d'expression libre constituées de nombres suivis de mots pour celles contenant les deux natures : en principe, un numéro de rue ou un code postal précèdent respectivement la rue et la ville. De même, la raison conduisant au choix du dernier mot est la présence quasi systématique d'un nom de rue avant le nom de la ville : nous en avons l'illustration dans l'exemple suivant où la zone libre VILLRISQ comprend l'expression suivante :

“ 2 RUE DE BELLOY SAINT FLOUR ”

Le résultat de ces extractions est couronné de succès si la concaténation du département du sinistre avec les huit premiers caractères de la chaîne extraite trouve un équivalent dans la table de l'I.G.N.. Si tel est le cas, le code INSEE que l'on y trouve conduit au code postal correspondant, grâce à l'utilisation de la table CPOSINS.

Puis vient la deuxième phase de l'algorithme, grâce à laquelle est désigné un code postal, quelle que soit l'issue de la phase de création précédente. Celui-ci doit nécessairement appartenir à la liste, fournie par la table CPOSINS, des codes postaux du département où le sinistre a eu lieu.

Si cette condition nécessaire est remplie, les éventuels codes postaux sont choisis en fonction :

- en premier lieu de leur fréquence d'apparition : c'est le code postal revenant le plus souvent qui aura notre faveur,
- en second lieu de leur nature; ainsi, on respectera la hiérarchie décroissante suivante, si les code postaux les plus fréquents ont un nombre d'apparition identique :
 - lieu du risque (INSEE),
 - adresse du risque (ADRRISQ, VILLRISQ ou COMPRISQ) fournissant un toponyme,
 - adresse du risque fournissant un code postal,
 - lieu le plus souvent sinistré (COPOSCOU),
 - adresse du souscripteur (COPOS).

En revanche, on se trouve dans une situation moins favorable si :

- aucun des codes postaux générés ne se situe dans le département du sinistre,
- ou bien, ce qui revient au même, aucun code postal n'a été généré.

Une alternative réside alors, pour le réseau agents, dans la recherche du code postal du lieu d'implantation de l'agence dont est issu le contrat sinistré. Cette information nous est donnée par l'exploitation de la table des agences où figure leur adresse. En revanche, ce redressement par le code postal de l'agence ne peut s'appliquer aux contrats du réseau S et C, dont le portefeuille n'est pas subdivisé en circonscriptions territoriales.

Le sort de tous les contrats sinistrés restant finalement sans code postal ne sera pas réglé par cet algorithme mais par les résultats qu'a posteriori il fournira : en effet, il sera possible de publier, de manière plus précise qu'avant le redressement du code postal, le palmarès des codes postaux les plus fréquents pour une date et un département donnés. C'est pour ce code postal qu'on optera alors pour tous les contrats sinistrés restés orphelins après l'application directe de la méthode de redressement.

On trouvera en annexe F le détail de l’algorithme de redressement au travers du programme SAS qui lui a été consacré.

II.3.1.4 - Résultats et validation de l’algorithme de redressement

L’objectif de cette méthode est d’utiliser au mieux toute l’information dont nous disposons, quelle que soit la difficulté pour mener à bien notre entreprise, dans la mesure où la cible est atteinte.

Est-il néanmoins possible de porter un jugement impartial sur les résultats obtenus ? Il faut pour cela donner une définition précise de ce que l’on souhaite mesurer et les instruments de mesure associés. En clair, il s’agit de mesurer :

- l’efficacité et l’ampleur du redressement, grâce au taux de contrats dotés d’un code postal et qui n’en avaient pas avant le redressement,
- la qualité du redressement, grâce à la comparaison des distributions par zone postale du nombre de contrats sinistrés dotés d’un code postal et du nombre de contrats sinistrés sans code postal, pour une tempête donnée.

Le tableau 2.1.18 présente l’ampleur prise par le redressement du code postal du lieu de survenance du sinistre. Il précise par ailleurs quelle a été la contribution de chaque type de renseignement à la détermination du code postal (grâce à la colonne « Nombre de codes postaux générés ») et la qualité de cette contribution (grâce aux colonnes « Nombre de codes postaux retenus » et « Taux de succès »). On voit que le code INSEE du lieu du risque présente une fiabilité très élevée, puisque 96,8 % des codes postaux générés ont été finalement considérés comme le code postal valide. Il en est de même pour le code postal de l’adresse du souscripteur (93,6 %), alors que la pertinence du code postal du lieu d’implantation de l’agence laisse plutôt à désirer (57,5 %).

Tableau 2.1.18 : Contribution de chaque type de renseignement au redressement du code postal du lieu du sinistre (en nombre de sinistres)

Type de renseignement	Nombre de codes postaux retenus	Nombre de codes postaux générés	Taux de succès
code INSEE	24 464	25 285	96,8 %
COPOSCOU	4 625	5 152	89,8 %
COPOS	23 320	24 914	93,6 %
code postal zone alpha.	1 573	1 905	82,6 %
toponyme zone alpha.	3 690	5 038	73,2 %
code postal d’agence	22 327	38 799	57,5 %
Tous types	48 323	48 323	100 %

L'efficacité de l'algorithme peut se mesurer au travers du taux de sinistres "relocalisés", qui rapporte le nombre de sinistres qui se sont vus affecter un code postal valide au nombre total de sinistres "délocalisés" avant cette opération. On pourrait adopter un autre taux, dont le dénominateur est égal au nombre de sinistres pour lesquels on a pu glaner quelque renseignement, arguant du fait que l'algorithme est inopérant si on ne lui soumet pas la moindre parcelle d'information pour son analyse.

Toujours est-il que le taux de sinistres dont on a dû localiser la survenance atteint près de 93 %, ce qui démontre l'ampleur de la méthode de redressement. Ce succès est d'autant plus manifeste qu'il concerne plus de 48.000 sinistres.

Peut-on dire pour autant que l'algorithme redresse "correctement" ? Rien ne permet de l'affirmer si ce n'est plusieurs indications :

- la contribution importante apportée par le lieu du risque, dont on connaît la forte corrélation avec le lieu de survenance du sinistre,
- le fait que le choix final du code postal soit fréquemment confirmé par plusieurs sources d'informations (on estime en effet à 58 % la part de codes postaux dont le choix a été établi sur la base de renseignements convergents).

Pour autant, on ne peut se satisfaire de ces seuls indices : aussi est-il souhaitable de comparer deux distributions de nombre de sinistres par zone postale, qui diffèrent par la connaissance que l'on a ou pas du lieu du sinistre. Pour ce faire, on applique la méthode de redressement précédente à la série pour laquelle le code postal est inconnu. On mesure alors la distance, mesurée selon la formule du χ^2 , entre les deux distributions pour un ou plusieurs événements, donc à des dates et des lieux identiques.

Notre choix se porte sur l'événement des 26 et 27 novembre 1983, qui a entraîné la déclaration de quelque 6.000 sinistres dans les régions du Nord-Pas-de-Calais, de Picardie et Normandie. A la fin de l'année 1983, la proportion de sinistres non localisés est d'environ 5 %, ce qui explique la présence significative de 232 sinistres survenus au cours de cet épisode venteux, par rapport à 5.728 sinistres localisés. L'agrégation par zone postale et le classement de ces deux distributions dans une séquence décroissante selon le nombre de sinistres sont ensuite effectués.

L'examen des 20 zones postales les plus sinistrées de chaque distribution montre que seulement 5 zones postales sont communes, dont 4 figurent parmi les 6 zones postales les plus sinistrées. Il est vrai que le nombre peu significatif de sinistres non localisés par zone postale à partir de la onzième zone postale classée dans un ordre décroissant du nombre de sinistres ne se prête pas à ce genre de comparaison.

En l'absence d'indices contraires, tout semble indiquer que l'algorithme de redressement ne génère pas de biais.

II.3.2 - Autres enrichissements de la base de données "Sinistres"

Il faut signaler pour être complet la nécessaire collecte de renseignements dont le fichier des catastrophes naturelles ne dispose pas mais que l'on retrouve en revanche dans la base constituée des transactions 2STEMP et 2STEMPQ : outre le code postal, on doit retrouver le montant des honoraires versés lors de l'expertise des biens endommagés pour chaque sinistre. Ce poste vient s'ajouter à la charge des sinistres supportée par l'UAP, mais doit être retiré du montant des dommages réellement subis par l'assuré car ce n'en est pas un élément constitutif.

La collecte se fait par la procédure déjà utilisée lors de la collecte des écrans "portefeuille" du paragraphe II.3.1.2. La seule différence tient à la transaction consultée : ici, ce sont les transactions de détail de paiements par sinistre (" 2SPH " et " 2SDP ") qu'il faut interroger. Puis, une agrégation par sinistre des seules écritures comptables relatives aux honoraires doit être réalisée.

III - Base de données météorologiques

La base de données météorologiques est une source d'informations nécessaire à la réalisation de nos travaux. Sa mise en relation avec les deux bases de données d'assurance, relatives au portefeuille et aux sinistres, s'avère même essentielle dans la perspective de l'objectif que nous nous sommes fixé d'une mesure des liens existant entre les dommages occasionnés par les tempêtes et les relevés météorologiques les caractérisant.

III.1 - Présentation de la base de données météorologiques

III.1.1 - Historique de la constitution de la base de données

Le chapitre 2 de la 1ère partie, consacré aux connaissances météorologiques du risque tempête, a été l'occasion de présenter les principales conclusions de l'étude de Météo-France, publiée en juillet 1995. Cette très intéressante publication constitue le document de synthèse du travail commandé par un groupe d'assureurs et de réassureurs auprès de l'établissement public.

La base de données créée pour l'occasion a été mise à la disposition de toutes les sociétés membres du groupe, dont l'UAP faisait partie. Elle contient les éléments suivants :

- les *valeurs des variables explicatives* retenues pour la classification des 524 situations du fichier élargi recensées entre 1962 et 1993. Leur liste se trouve en annexe 4 de l'étude de Météo-France. Citons simplement :
 - la date de la situation,
 - le type de la situation (N, W, SW, S, OR, E, NE; pour plus de précision se reporter au paragraphe I.4 du chapitre 2 de la 1ère partie),
 - sa localisation à travers la latitude du maximum de pression et du minimum de pression,
 - son étendue par la mesure du vent moyen sur les quatre parties du quadrant de la France,
 - la pression moyenne sur le quart sud-est,
 - la direction moyenne du vent géostrophique¹,
 - la survenance d'autres phénomènes atmosphériques l'accompagnant, tels que la neige, l'orage et la grêle, et leur importance à travers le nombre de stations touchées,
- les *résultats de la classification en 7 types* des tempêtes recensées entre 1962 et 1993, selon que leur durée est considérée égale à 1, 2 ou 3 jours,
- les *caractéristiques des 91 stations* retenues dans l'étude de Météo-France, à savoir :
 - département,
 - coordonnées physiques (latitude, longitude, altitude),
 - nom de la station,
- les listes et les *coordonnées des stations* touchées par des vents forts pour chaque type de tempête, et celles figurant dans le *panel de stations représentatives de chaque type* (cf. les paragraphes III et IV du chapitre 2 de la 1ère partie, consacrés à ce sujet),

¹ le vent géostrophique est le vent obtenu en négligeant les forces de frottement, c'est-à-dire en supposant que les forces de pression sont en équilibre avec la force de Coriolis due à la rotation de la terre; les composantes u et v du vent géostrophiques s'écrivent :

$$u = -\frac{1}{rf} \frac{\partial p}{\partial y}$$

$$v = \frac{1}{rf} \frac{\partial p}{\partial x}$$

, où p est la pression, f le paramètre de Coriolis et ρ la masse volumique de l'air.

Le vent géostrophique n'a pas de composante verticale; il est parallèle aux isobares; son intensité est d'autant plus forte que le gradient de pression est important.

- les **matrices de propagation** permettant d'étendre les valeurs du panel vers les 91 stations de l'étude pour les quatre paramètres descriptifs sélectionnés par Météo-France et le groupe d'assureurs, et dont la définition est donnée au paragraphe III.1 :

- intensité du vent maximal instantané,
- durée,
- direction,
- précipitations;

on y trouve les n coefficients estimés $A_1, A_2, \dots, A_n - n$, allant de 5 à 18, selon les types, représentant le nombre de stations du panel - B , la constante, et le coefficient de corrélation multiple,

- pour chaque type de tempête et chaque station de la zone touchée par les vents forts, **les valeurs du vent maximum instantané correspondant à des durées de retour égales à 5, 10, 15, 20, 25, 50 et 100 ans**, ainsi que les valeurs a et b de l'ajustement des relevés météorologiques par une loi de Gumbel et la demi-largeur de l'intervalle de confiance à 90 %,
- pour les 91 stations de l'étude et pour chaque type de tempête, les **tables de contingence de l'intensité du vent maximal et le deuxième paramètre descriptif** choisi selon le type de tempête,
- pour les 91 stations de l'étude et pour 108 situations fournies par les assureurs entre 1978 et 1993, les **valeurs des quatre paramètres descriptifs**,
- pour les 91 stations de l'étude et pour chaque type de tempête, **les valeurs des intensités moyennes** sur l'ensemble des situations d'un type donné.

A l'origine, le calcul de l'intensité moyenne n'avait pas été évoqué dans le cahier des charges adressé par le groupe d'assureurs et de réassureurs à Météo-France. Il s'est avéré nécessaire de l'inclure dans le rapport final pour être en mesure d'appréhender un phénomène dans sa *globalité*; le panel de stations représentatives, contenant entre 5 et 18 stations, ne permettant pas aux assureurs de synthétiser suffisamment les résultats pour quantifier les dommages sur la France.

III.1.2 - Evolutions de la base de données originelle

En 1996, suite à la demande de certains membres du groupe, des compléments ont été apportés à la base de données originelle. En particulier, Météo-France a porté à 168 le nombre de stations météorologiques présentes dans le fichier des paramètres descriptifs, ce dernier contenant toujours 108 situations entre 1978 et 1993.

Cependant, pour nombre d'entre elles, les données manquantes sont très souvent légion. On compte en effet, sur un total attendu de 8 316 nouveaux relevés, égal au nombre de stations ajoutées (77) multiplié par le nombre de situations (108) :

- près de 22 % des relevés qui font état de l'absence de l'intensité du vent et de la direction,
- un pourcentage identique de 17 % des relevés ne mentionnant pas la valeur de la durée et du niveau des précipitations.

La médiocre qualité de ces données explique la raison pour laquelle M étéo-France n'avait pas initialement retenu ces stations. Souvent, l'absence de relevés constatés pour certaines stations n'est pas due à un défaut d'instrumentation, mais à leur mise en place trop récente.

Cette situation présente des contrastes saisissants puisque, parmi les 77 nouvelles stations, on en recense :

- 9 présentant un historique des relevés complet sur la période 1980 - 1993 qui nous intéresse, égal à 91,
- 32 dont le nombre de situations postérieures ou égales à 1980 est strictement supérieur à 84,
- et 15 dont l'historique sur cette même période n'atteint pas 40 relevés.

III.1.3 - Derniers enrichissements apportés à la base de données

Comme on vient de le voir, la base de données de M étéo-France permet deux types d'approches pour l'analyse d'un événement :

- l'approche globale, pour laquelle on définit l'intensité d'un événement. Cette optique synthétique rend possible la mesure approximative des dommages provoqués par un événement de grande ampleur : en effet, l'intensité globale d'un événement croît à mesure que le nombre de stations enregistrant des vents forts augmente,
- l'approche locale, pour laquelle on dispose de relevés par station des quatre paramètres descriptifs considérés par les assureurs comme les plus significatifs des dommages. Le chiffrage des dommages occasionnés par l'événement requiert au préalable d'avoir mis en évidence une relation économétrique entre les dommages et les paramètres. C'est cette méthode, qui requiert une grande quantité de données locales, que l'on se propose par la suite de suivre.

Ayant à notre disposition les données d'assurances sur la période 1980 - 1996, nous avons donc cherché à obtenir les relevés météorologiques sur une période identique, donc à compléter la base de données locales pour les années suivant 1993.

Nous avons alors rencontré M. Duby, directeur scientifique de l'UAP, et lui avons exposé notre projet de mise en évidence d'une relation entre les données d'assurances et météorologiques. **M. Duby**, qui avait une très bonne connaissance des travaux effectués par Météo-France, **nous a soutenus dans notre entreprise en finançant l'achat de données météorologiques**. Nous l'en remercions vivement.

Concrètement, Météo-France a actualisé pour nous la précédente étude de 1995. Ce rafraîchissement a conduit à la sélection de 43 tempêtes supplémentaires sur la période de janvier 1994 à juin 1996.

Ces informations actualisées, disponibles pour les 168 stations de Météo-France et pour les 43 nouvelles situations, sont les suivantes :

- date de la situation,
- type de la tempête,
- valeurs des quatre paramètres descriptifs mesurés au niveau de la station.

Au total, 134 événements recensés par Météo-France constituent la base des données météorologiques utilisables pour une approche locale.

III.2 - Travaux préliminaires à l'utilisation de la base de données

Préalablement à l'exploitation de la base données météorologiques, un certain nombre de travaux sont à entreprendre :

- pour chaque zone postale, la détermination de la station météorologique située à la distance la plus faible,
- le choix des critères conduisant à la sélection des stations retenues pour mettre en exergue la relation économétrique (présentée au chapitre 4 de cette partie),
- éventuellement, le redressement des historiques de relevés des stations sélectionnées pour la suite.

III.2.1 - Détermination de la station météorologique la plus proche

Il convient, lors d'une étape préliminaire au rapprochement des données d'assurances et météorologiques, de déterminer la donnée météorologique à comparer à la donnée d'assurances. La logique prévoit de retenir le relevé appartenant à la

station située à la distance minimale de la zone postale considérée : en effet, il est de coutume d'estimer que les conditions prévalant à quelques kilomètres d'une zone postale frappée par une tempête sont peu différentes des conditions de cette zone, et qu'elles s'en éloignent à mesure que la distance augmente.

Pour ce faire, il suffit de créer un tableau contenant les distances de chaque zone postale à l'ensemble des stations météorologiques, d'en extraire pour chaque zone postale la distance minimale, et de noter la station météorologique correspondante.

La réalisation concrète de ce travail est plus délicate puisque le distancier possède une taille respectable : environ 6.000 zones postales pour 168 stations, soit plus d'un million de distances à calculer !

La réalisation du distancier requiert la connaissance de coordonnées (X, Y) du centre géographique de chaque zone postale et chaque station : cette source d'information est la propriété de l'I.G.N. dont une licence a été accordée à l'UAP. Elle donne pour chaque commune ou principal lieu dit de France métropolitaine les coordonnées Lambert (i.e. dans la représentation définie par Lambert).

On trouvera dans l'annexe G le programme APL MINDISTZPSM calculant cette distance minimale et affectant à chaque zone météorologique la station la plus proche.

Du travail précédent, il est possible de tirer quelques enseignements intéressants : on constate ainsi ***qu'une majorité de zones postales disposent d'une station météorologique située à moins de 22 kilomètres.*** Le maillage est certes plus dense dans les régions venteuses de l'ouest de la France que dans l'est, où l'on rencontre les distances minimales les plus élevées : E vian-les-Bains est, avec 85 kilomètres à vol d'oiseau de Lons-le-Saulnier, la ville la plus éloignée d'une station météorologique.

Il est commode de vérifier a posteriori la plus grande précision que le choix de la station météorologique la plus proche de chaque zone postale offre par rapport à la méthode simplificatrice consistant à retenir systématiquement l'une des stations situées dans le département de la zone postale ou, à défaut, dans l'un des départements voisins : en effet, une statistique tirée de l'affectation de chaque zone postale à une station météorologique indique que la ***proportion de stations choisies dans un département distinct de celui de la zone postale est de 27 %.***

On peut alors illustrer la méthode précédente d'affectation des zones postales à la station météorologique la plus proche par la figure 2.1.9, représentant la carte de France des frontières géographiques des 168 zones météorologiques ainsi définies.

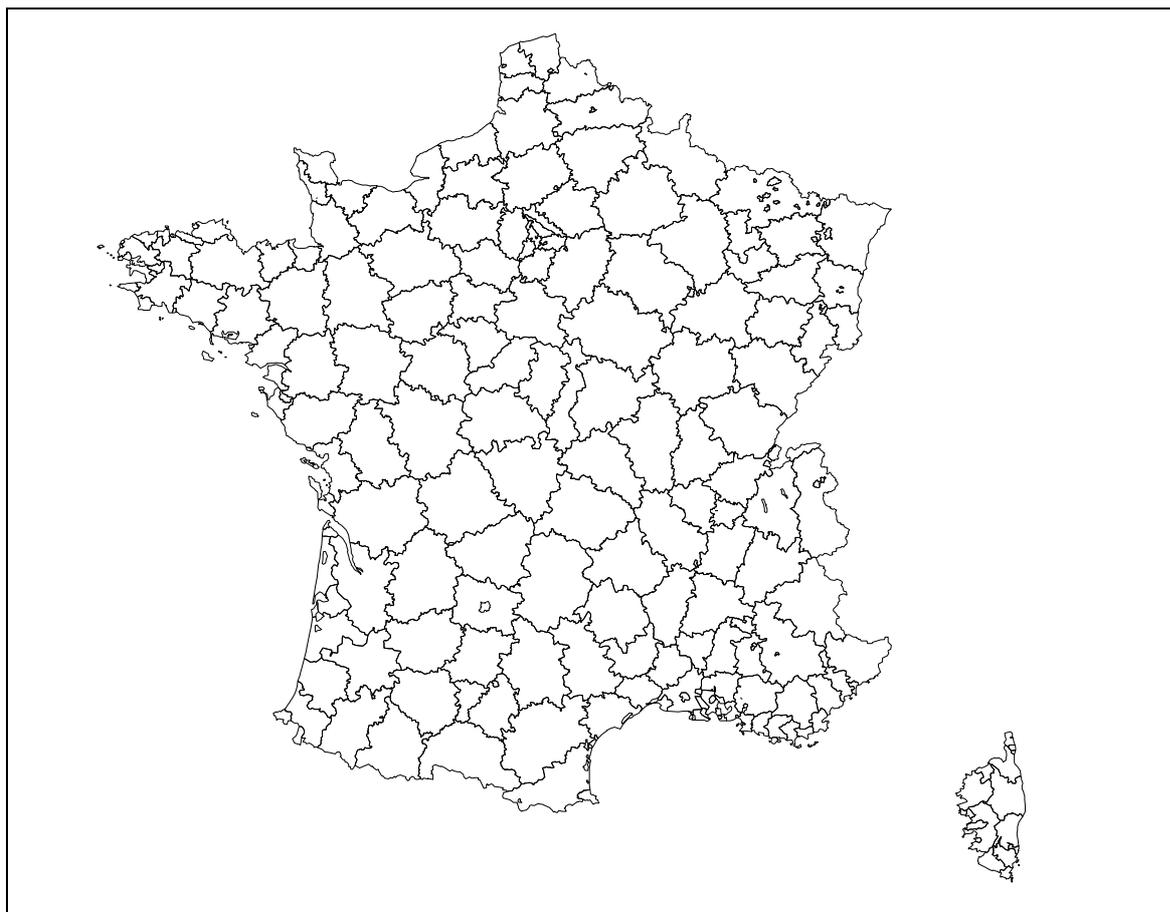


Figure 2.1.9 : Carte de France des 168 zones météorologiques

III.2.2 - Critères retenus lors du choix des stations météorologiques

L'actualisation de l'étude de Météo-France, grâce à un quasi doublement du nombre de stations disponibles, présente l'avantage d'une plus grande latitude dans le choix des stations finalement retenues pour la modélisation du risque tempête : il devient possible d'en rajouter certaines à la liste des 91 de l'étude de Météo-France, par exemple dans des régions éloignées des stations météorologiques, ou d'en retirer dans des zones où Météo-France est sur-représenté.

Mais quels critères nous dictent cette conduite ?

L'affectation de la station météorologique la plus proche à chaque zone postale conduit au découpage de la France en *zones météorologiques* dans lesquelles les quatre paramètres descriptifs auront une même valeur pour un événement donné. Il convient maintenant de fixer les critères permettant de savoir si une zone météorologique doit ou non être conservée. Voici ces principes de conservation, présentés par ordre d'importance décroissante :

- une **population de la zone météorologique dépassant 250.000 habitants**, de manière à pouvoir interpréter les résultats obtenus dans les modèles

économétriques : le fait qu'une station météorologique soit installée sur la pointe du Raz n'est intéressant que si la population alentour est importante,

- une *superficie supérieure à 2000 km² lorsque la population est inférieure à 250.000 habitants*; on introduit ainsi la notion d'*utilité* de la zone météorologique qui couvre une grande zone du territoire,
- la *disponibilité des informations* : l'absence trop courante de relevés (i.e. plus de 25 relevés absents entre 1980 et 1996), pour les stations trop récentes par exemple, est rédhibitoire.

De même, il nous a paru utile de calculer la distance séparant chaque station de celle qui lui est la plus proche. On trouvera en annexe G le programme APL nommé MINDISTSMSM conduisant à cette statistique. Cette distance permet de mieux évaluer les conséquences d'un rapprochement sur la taille de la zone météorologique résultante. Il va sans dire que la décision de conserver une station augmente avec cette distance.

Il peut être intéressant aussi de calculer la distance moyenne de chaque station aux zones postales qui lui sont rattachées : on verra que plus cette distance est grande, plus la zone météorologique comble un vaste territoire vide de stations, et plus elle se révèle utile.

A ce sujet, on prendra soin de pondérer cette distance moyenne par le nombre d'habitants de chaque zone postale rattachée à la zone météorologique : en effet, la population étant une variable corrélée très positivement avec le nombre de contrats assurés contre les dommages dus à la tempête, c'est cette distance qui reflète le mieux la vraie distance du bien assuré à la station météorologique.

Un calcul très simple montre que la *distance pondérée moyenne* entre une station et chacune de ces zones postales de rattachement est inférieure à *la distance équipondérée moyenne* (18,4 km contre 21,3 km), ce qui souligne la plus grande présence de stations dans des zones postales densément peuplées, dont les superficies sont moindres que les celles des zones postales rurales.

Il n'a pas échappé au lecteur le risque d'une importante divergence entre les relevés de la station et les caractéristiques réelles d'un événement frappant cette zone météorologique : ce phénomène ne peut que croître avec la taille de la zone, et l'on doit donc veiller à ne pas trop agréger les stations.

Il faut donc trouver un juste équilibre entre des zones météorologiques trop vastes, réduisant la pertinence du rapprochement entre données d'assurances et météorologiques, et des zones trop peu peuplées, qui ôtent tout crédit au calcul statistique.

La mise en application des règles précédentes nous permet d'obtenir la répartition géographique optimale des stations météorologiques, dont le nombre finalement retenu est de 130.

Une carte de France, où figurent les frontières de ces zones météorologiques ainsi définies, peut alors être dressée sur la figure 2.1.10. De plus, on trouvera les caractéristiques en terme de démographie et de distance dans l'annexe H.

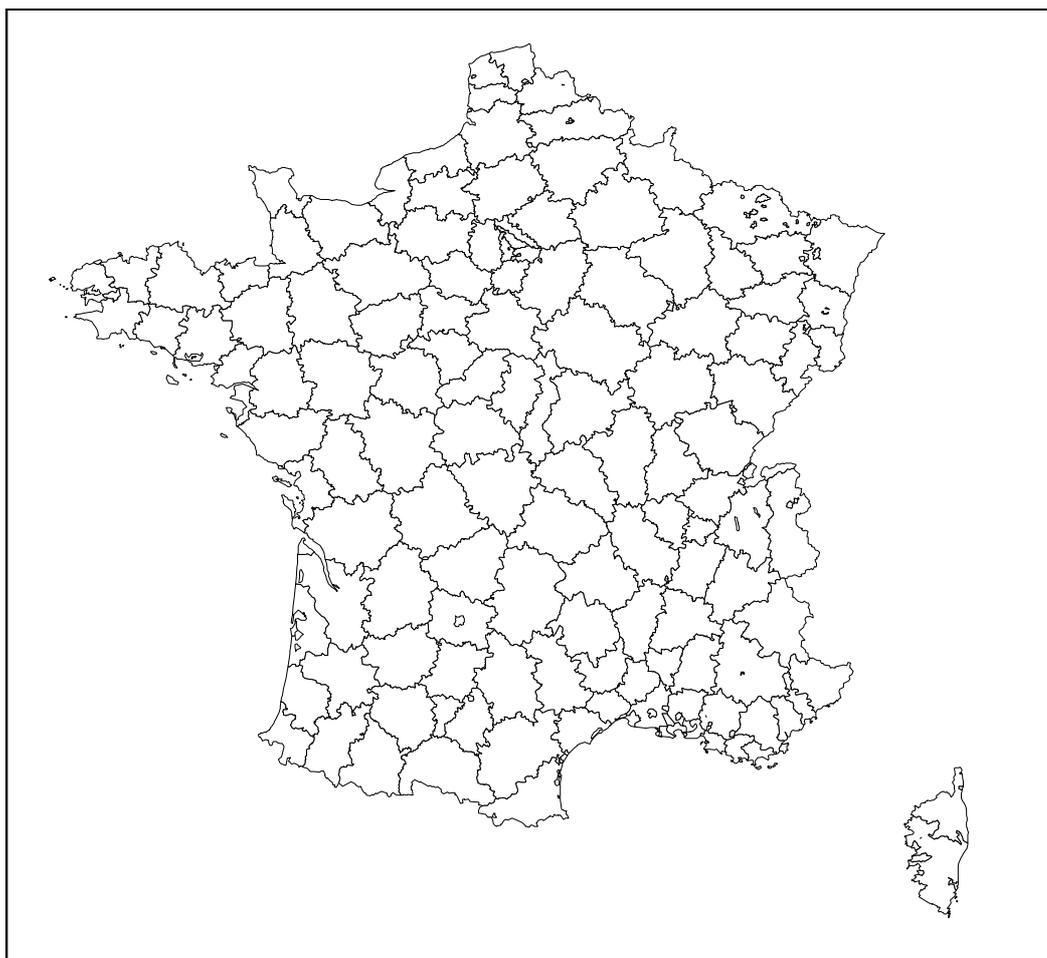


Figure 2.1.10 : Carte de France des 130 zones météorologiques

III.2.3 - Méthode de redressement des historiques des stations météorologiques

Au chapitre 2 de la 1ère partie, paragraphe III.4, nous nous sommes intéressés au calcul des corrélations qui pouvaient exister entre les stations. M étéo-France a alors présenté des matrices de propagation, permettant de déduire les paramètres descriptifs d'une station dont le relevé fait défaut grâce à la connaissance de relevés issus d'un panel de stations dites représentatives.

Cette méthode ne peut être reconduite ici pour l'estimation des valeurs absentes. En effet, les matrices de propagation sont élaborées *pour l'ensemble des stations* retenues

par Météo-France, et les *stations* du *panel* présentent de grandes disparités. Au contraire, nous souhaitons plus simplement remplacer la valeur manquante y_{i0} d'une station par la valeur prédite par la régression linéaire liant les valeurs y_i de la station en cause avec celles x_i de l'une des 3 stations géographiquement les plus proches, ce qui s'exprime par la formule suivante :

$$\hat{y}_{i0} = a \cdot x_{i0} + b,$$

où a est le paramètre estimé par la méthode des moindres carrés, et b une constante. On aura pris soin de vérifier que l'ensemble des valeurs de la station en cause constitue un échantillon gaussien.

Le choix de la station fournissant les valeurs x_i sera influencé par :

- la qualité de la reconstitution, c'est-à-dire par une valeur du carré du coefficient de corrélation (R^2) avoisinant 1,
- le nombre de données en commun entre les deux stations.

IV - Constitution de la base de données “Construction”

IV.1 - Importance de la constitution de la base de données “Construction”

Pour bien comprendre l'importance que revêt à nos yeux la constitution de la base de données “Construction”, il suffit d'imaginer l'étendue des dommages occasionnés à un chalet en préfabriqué par une tempête à la pointe du Raz, alors que la même tourmente ne provoquerait aucun dégât à une maison de style breton construite en granit !

Cette assertion veut souligner par son aspect excessif une vérité première : la vulnérabilité des biens assurés au risque tempête est très disparate. Le chapitre 3 de la 1ère partie l'a bien montré, dans l'analyse qui y est faite des dégâts dus au vent : la vulnérabilité est maximale pour les éléments de couverture d'une construction, par le jeu des forces exercées. Il a rappelé l'importance d'un bon dimensionnement d'un bâtiment et les règles de construction auxquelles ce dernier est soumis.

IV.2 - Constitution de la base de données relative aux normes de construction

Le chapitre 3 de la 1ère partie nous rappelle dans son paragraphe IV le champ d'application des normes de construction qu'il convient de respecter en France. En ce qui concerne le vent, les règles NV 65, toujours en vigueur, définissent une pression q dite “normale”, correspondant à une vitesse V normale du vent, définie comme la

vitesse de pointe instantanée atteinte ou dépassée seulement 3 jours par an, de par la relation qui lie ces deux valeurs :

$$q = \frac{V^2}{16,3} .$$

On en déduit alors la pression dynamique extrême, dans un rapport de 1,75 avec la pression dynamique normale. Ces règles NV 65 conduisent finalement à la définition de 3 régions selon les valeurs minimales requises des pressions dynamiques de base (cf. le paragraphe IV.3), et dont la carte figure en annexe A. On a vu, au paragraphe IV.4, que ces valeurs peuvent être modifiées par l'existence de phénomènes locaux tels que la hauteur du bâtiment par rapport au sol, le type du site, selon qu'il est protégé ou exposé, et l'effet de masque.

Ces règles NV 65 sont l'objet de critiques. Leur rédaction est ancienne, car elle repose sur les considérations faites il y a quarante ans. L'insuffisant historique des relevés météorologiques de l'époque et le maillage moins dense qu'aujourd'hui expliquent les quelques anomalies constatées dans l'affectation de certains cantons à l'une des 3 régions : ainsi, la région de Millau dans l'Aveyron, où M étéo-France compte depuis longtemps une station ayant enregistré de fortes tempêtes, est à juste titre classée dans la région " exposée ", alors que les cantons environnants se trouvent dans la région moyennement exposée ". A l'inverse, des zones battues par les vents, telle une partie du Pas-de-Calais, sont classées dans la région dite " faiblement exposée ", car à l'époque non pourvues de stations.

Ces cartes ne sont donc pas toujours d'une grande fiabilité, et il en résulte qu'un coefficient de sécurité est ajouté aux normes requises dans des lieux considérés à tort comme n'étant pas exposés.

L'existence de conflits d'intérêts entre les fabricants de matériaux de construction, qui souhaitent que les règles soient plus strictes, et les promoteurs, pour qui une telle mesure conduirait à un relèvement de leurs prix de vente, expliquent la difficulté et la lenteur à gommer les anomalies de ces cartes administratives. Malgré la mise en place en 1983 d'une commission chargée de revoir la rédaction de ces règles, celle-ci n'a pas, à notre connaissance, déposé ses conclusions.

D'autre part, nos collectes d'informations locales, auprès de différents organismes, tel le Centre Scientifique et Technique du Bâtiment, sur la répartition du parc immobilier, des établissements industriels ou des bâtiments d'exploitation agricole selon :

- les normes réellement appliquées ou la robustesse des bâtiments,
- la vétusté ou, à défaut, l'âge des constructions,

sont restées vaines. Il n'existe pas en France d'organisme professionnel disposant de cette volumineuse base de données.

Il faut donc se contenter des enseignements précieux fournis par la carte des vents de l'annexe A, même s'il n'est pas concevable de prendre en compte tous les facteurs locaux qui agissent sur le niveau de vulnérabilité des bâtiments assurés. Néanmoins, peut-il être envisagé la reconstitution d'une base de données indiquant dans quelle région chacune des zones postales se trouvent.

La saisie de ces informations se fait de la manière suivante :

- pour les départements appartenant entièrement à une région de vent, toutes leurs zones postales se voient affecter le même code,
- pour les autres départements, une transcription des noms de cantons en codes postaux par l'intermédiaire du nom de la zone postale s'avère nécessaire, même si certains obstacles se font jour, tels que :
 - l'évolution de la toponymie entre 1965 et 1997,
 - et surtout la difficulté du passage du canton à la zone postale, dont le découpage pour un même nom n'est pas toujours identique; on compte en France environ 3.500 cantons pour un peu plus de 6.000 zones postales, dont le découpage n'est pas du tout le même, en particulier en milieu urbain.

Quand plusieurs cantons appartenant à une même zone postale ne figurent pas dans la même région de vent, il est décidé d'attribuer à ces cantons le code du plus peuplé des deux.

L'ancienneté de ces normes peut constituer un atout sur le plan statistique : en effet, toutes les mises en construction depuis 35 ans répondent à ces règles ce qui contribue à homogénéiser le parc immobilier en terme de vulnérabilité. Une statistique de l'INSEE de 1994, reproduite sur la figure 2.1.11, nous éclaire sur la répartition du nombre de ménages en fonction de la date d'achèvement de l'immeuble qu'ils occupent :

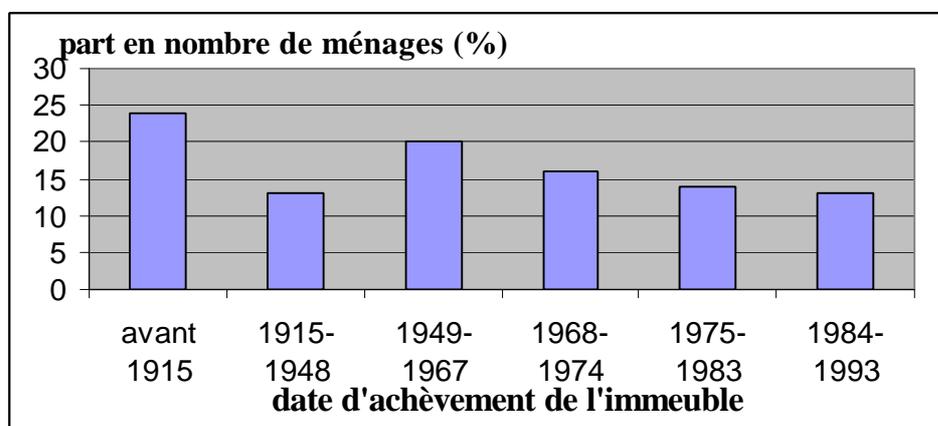


Figure 2.1.11 : Répartition des ménages selon la date d'achèvement de l'immeuble occupé

Il ressort de cette statistique que près de 50 % des ménages habitent en 1999 un immeuble répondant aux normes NV 65. Pour l'autre partie, il n'est pas possible de se prononcer. Mais il n'est pas hasardeux de pronostiquer que la construction d'une grande majorité de bâtiments a respecté les normes NV 65, même avant leur rédaction.

IV.3 - Constitution de la base de données relative à la vulnérabilité des biens assurés

Le chapitre 4 de la 1^{ère} partie s'est attaché, dans les paragraphes II.2 à II.4, à analyser le danger potentiel que constitue la tempête pour les risques d'habitation, professionnel et agricole. L'accent a été mis en particulier sur la distinction qu'il convient de faire entre le niveau d'exposition au risque tempête d'un bien et sa vulnérabilité, c'est-à-dire son incapacité à y faire face.

Ce travail qualitatif sera bonifié et même validé s'il peut être poursuivi et complété par sa composante quantitative. Cela passe par l'exploitation de la base des contrats de la branche " incendie " de l'UAP, contenue dans le fichier de synthèse déjà cité.

Concrètement, l'opération se compose de 4 étapes :

- la première étape consiste à rechercher et mettre en place une segmentation des contrats basée sur leur niveau différent d'exposition au risque,
- la deuxième étape a pour objectif la répartition, par code INSEE et par type d'assuré, du nombre de contrats de la branche incendie en portefeuille en cours au 31 décembre 1993,
- la troisième étape est la mise à disposition de la table de correspondances entre les codes INSEE et postaux,
- la dernière étape se résume dans la répartition, par code postal et par type d'assuré, du nombre de contrats précédents.

La première étape consiste, d'une part à sélectionner les variables, de la base des contrats, présentant une forte sensibilité aux effets du vent et, d'autre part, à regrouper leurs modalités selon leur niveau de sensibilité.

Trois variables répondent à ce profil :

- la *qualité de l'assuré*, dont les modalités sont :

- propriétaire, locataire, copropriétaire, pour les particuliers,
 - Société Civile Immobilière, syndic, nu-propriétaire pour les entreprises,
- le *type de locaux*, pour lequel on distingue :
- maison en agglomération,
 - immeuble,
 - maison hors d'une agglomération,
- le *type de résidence*, c'est-à-dire :
- résidence principale,
 - résidence secondaire.

En effet, les forces de pression et de succion ne s'exercent pas de la même manière et avec la même vigueur selon que le type de la résidence est un immeuble collectif, ou une maison en ville ou à la campagne. De même, les dégâts occasionnés à une résidence secondaire peuvent être aggravés en l'absence des propriétaires. Quant à la première variable, il convient de se souvenir que le champ d'application de la garantie tempête varie selon la qualité de l'assuré. La garantie tempête couvre en effet :

- les bâtiments pour le propriétaire non occupant,
- le mobilier intérieur et les embellissements pour le locataire,
- tous ces éléments pour le propriétaire occupant.

Pour le risque d'habitation, la construction d'un arbre pyramidal obtenu par le croisement des modalités des variables précédentes permet de mettre en évidence 20 branches, au niveau d'exposition au risque tempête variable. On peut citer, par exemple, le propriétaire non occupant d'une maison en agglomération.

Malheureusement, ces branches ne correspondent pas forcément à des cases tarifaires, qui sont les seules à pouvoir être dénombrées. Finalement, il ressort après analyse que la segmentation peut comporter 13 rubriques, présentes dans le tableau 2.1.19.

Tableau 2.1.19 : Segmentation des contrats d'habitation selon l'exposition au risque tempête

sigle	définition
LOA	Locataire occupant d'appartement
POA	Propriétaire occupant d'appartement

LOM	Locataire occupant de maison
POM	Propriétaire occupant de maison
PNOM	Propriétaire non occupant de maison
PNOX	Propriétaire non occupant
PNOI	Propriétaire non occupant d'immeuble
PNOA	Propriétaire non occupant d'appartement
LOX	Locataire occupant
POX	Propriétaire occupant
OCX	Occupant
OCA	Occupant d'appartement
OCM	Occupant de maison

La segmentation proposée supra est le résultat d'une démarche visant à utiliser toutes les données présentes dans la base de contrats et de nature à nous éclairer sur le niveau d'exposition au risque théorique de chaque contrat. Elle prend en compte évidemment certaines contraintes, telle l'indisponibilité de certaines informations sur le fichier de synthèse, particulièrement pour les contrats non "mécanisés". (Ceci explique d'ailleurs la présence d'un " X " dans certains sigles de modalités, et l'abandon de la variable " type de résidence ", trop peu fréquente).

En outre, la réalisation informatique de la segmentation désirée se révèle délicate car :

- certains codes, présents dans le " fichier de synthèse " ne figurent pas dans la nomenclature officielle : il faut alors souvent procéder par assimilation (suite d'une série réservée à un certain type de contrats, traduction numérique d'un codage alphabétique, etc...),
- le nombre de cas à considérer est important, en raison du volume des branches-risques.

On aura dans l'annexe I le détail du programme d'extraction REPZPHAB et de table d'encodage des différentes modalités de la segmentation précédente, PYRAMHAB.

IV.4 - Constitution de la base de données relative au parc immobilier

L'exploitation de base de données de l'INSEE peut aussi être mise à profit pour mieux cerner le niveau d'exposition local au risque tempête du parc immobilier : le recensement de la population contient en effet un volet relatif aux conditions d'hébergement des ménages. Ces informations ne concerneront donc plus le parc immobilier des assurés de l'UAP, mais de la France.

Les variables présentes dans le fichier issu du recensement de la population effectué en 1990 et présentant un intérêt à être incluses dans notre étude sont les suivantes :

- le *nombre d'immeubles*, comprenant au moins deux logements,
- le *nombre de résidences principales* (on y trouve les logements ordinaires, les logements-foyers pour personnes âgées, les pièces indépendantes louées, sous-louées ou prêtées à des particuliers, les chambres meublées dans les hôtels ou les garnis, les constructions provisoires et les habitations de fortune. Les habitations mobiles et les logements en collectivité ne sont pas compris dans les résidences principales),
- le *nombre de résidences secondaires*, (logements utilisés pour les week-ends, les loisirs ou les vacances, notamment les maisons et logements loués (ou à louer) pour les saisons touristiques dans les stations balnéaires, de sports d'hiver, etc.),
- le *nombre de logements occasionnels* (i.e. utilisés une partie de l'année pour des raisons professionnelles),
- le *nombre de logements vacants* (comprennent à la fois les logements disponibles pour la vente et la location, les logements neufs achevés mais non encore occupés à la date du recensement, ainsi que les logements, autres que les résidences secondaires, que leurs titulaires n'occupent pas à la date du recensement),
- la *population des résidences principales*,
- la *population sans double compte* (ce concept est utilisé pour calculer la population d'un ensemble de communes, chaque personne étant alors prise en compte une seule fois).
- la *taille de l'agglomération* à laquelle appartient la commune.

Ces variables, disponibles pour les recensements successifs de 1982 et 1990, nous apportent de précieux renseignements sur l'environnement urbain de chaque commune et sur son niveau d'exposition à la tempête.

Le passage du code INSEE au code postal s'effectue selon le processus déjà présenté dans le paragraphe I de ce présent chapitre, grâce à l'utilisation de la table de correspondances CPOSINS.

CHAPITRE 2 : DÉTERMINATION ET ACTUALISATION DU PRÉJUDICE DÛ À LA TEMPÊTE

Après avoir complètement décrit les conditions dans lesquelles la base de données relative aux tempêtes a été constituée, il est temps d'aborder une autre étape importante : l'actualisation de cette base de données, préliminaire indispensable à tout calcul statistique ou toute modélisation économétrique.

En effet, la période relativement longue (1980 - 1997) sur laquelle repose l'historique des tempêtes nous oblige à réaliser une étude détaillée de tous les facteurs ayant pu influencer, par leur évolution, le règlement des dommages résultant des tempêtes.

Dans ce chapitre, seront ainsi abordés deux thèmes primordiaux conditionnant la validité de tous les calculs qui vont suivre :

- d'une part, *le calcul du préjudice réellement subi par l'assuré* suite à une tempête,
- d'autre part, *l'actualisation de ce préjudice à sa valeur de fin 1997*.

On pourrait à juste titre se demander pour quelle raison notre intérêt se porte aussi sur le préjudice de l'assuré, c'est-à-dire le coût réel des dommages, et non pas seulement sur le coût supporté par la compagnie d'assurances qui, seul, conduit au calcul d'un tarif.

Deux raisons justifient ce choix :

- en premier lieu, c'est le coût réel des dommages, et non le coût supporté par la compagnie qu'il convient de comparer aux données météorologiques, lors de la modélisation économétrique,
- en second lieu, il est nécessaire de connaître la courbe de distribution des dommages dans son intégralité, et non sa version amputée du montant restant à la charge de l'assuré, pour être en mesure d'établir en toute connaissance de cause le niveau de la franchise acquitté par l'assuré sinistré.

L'objectif du présent chapitre est de pouvoir répondre au type de question suivant : combien aurait coûté la tempête du 6 novembre 1982 si elle s'était produite 15 ans après ?

I - Principes de détermination et d'actualisation du préjudice dû à la tempête

I.1 - Eléments constitutifs du préjudice dû à la tempête

Il convient avant tout de définir la notion de préjudice : sous ce vocable, on retrouve la somme des éléments suivants :

$$\begin{aligned} \text{préjudice} &= \text{coût du sinistre supporté par l'UAP} \\ &- \text{montant des honoraires d'expertise versés par l'UAP} \\ &+ \text{franchise} \\ &+ \text{vétusté} \\ &+ \text{application de la règle proportionnelle} \\ &+ \text{application du taux de coassurance} \\ &+ \text{montant des exclusions liées à la tempête} \\ &+ \text{part du montant des dommages dépassant la limite contractuelle} \\ &\quad \text{d'indemnité} \end{aligned}$$

Le *préjudice* est le montant des dommages occasionnés par la tempête. Le calcul du préjudice se fait, selon les cas, en prenant la valeur de remplacement du bien assuré, ou le montant des réparations nécessaires à une utilisation identique à celle qui avait cours avant l'événement déclencheur.

Le préjudice ne peut être inférieur au *coût du sinistre supporté par la compagnie d'assurances*, puisque la survenance d'un sinistre ne peut conduire à l'enrichissement du sinistré. Le coût du sinistre est le montant enregistré dans le compte de sinistres de la compagnie. C'est ce montant que nous avons extrait. Lorsque la gestion du dossier est achevée par le rédacteur de la compagnie, le sinistre est clos, et le coût du dossier est égal au total des paiements effectués. L'évaluation restante est alors mise à zéro. Dans la base " sinistres ", les derniers sinistres étant survenus en 1997, il n'y qu'une infime minorité de sinistres dont le montant des paiements ne soit pas forcément égal au coût final du sinistre.

De ce coût supporté par la société, il faut défalquer les coûts induits par la tempête mais qui ne figurent pas parmi les éléments constitutifs des dommages occasionnés par elle : il s'agit principalement du *montant des honoraires versés aux experts* de la société en charge de l'examen des biens endommagés.

En revanche, tous les autres éléments non encore cités viennent se greffer au coût supporté par la société, telle la *franchise*. Celle-ci a longtemps été appelée *franchise d'avarie* pour bien signifier à l'assuré que le sinistre aurait pu être évité si les parties endommagées d'un bien avaient fait l'objet d'un entretien régulier.

La *vétusté* figure aussi parmi les éléments pouvant rester à la charge de l'assuré. D'après le Larousse, elle correspond à un " *état de détérioration résultant de l'ancienneté* ". Par ce biais, l'assureur ne règle que la valeur vénale d'un bien, et non

sa valeur actuelle. Il se prémunit ainsi contre la tentation que pourrait avoir l'assuré d'attendre la prochaine tempête pour changer une partie usagée de sa toiture...

L'article L 121-5 du Code des Assurances relatif à la sous-assurance est ainsi rédigé :

“ S’il résulte des estimations que la valeur de chose assurée excède au jour du sinistre la somme garantie, l’assuré est considéré comme restant son propre assureur pour l’excédent, et supporte, en conséquence, une part proportionnelle du dommage, sauf convention contraire. ”

Cette règle vise à responsabiliser l'assuré, qui peut être parfois tenté de minorer le capital garanti, dans l'espoir d'une réduction équivalente de la prime d'assurance. Son application peut être levée si une clause du contrat le spécifie. Cette réduction proportionnelle des capitaux, appliquée ou non selon les contrats, ne doit pas être confondue avec la réduction proportionnelle des primes, toujours applicable s'agissant d'une sanction, qui survient lorsqu'il apparaît que le risque sinistré ne correspond pas à la description qu'en avait faite l'assuré à la souscription.

Le *taux de coassurance*, ou part du risque prise par la compagnie d'assurances, s'il n'augmente pas la part du dommage à la charge de l'assuré, diminue en revanche celle prise par cette compagnie.

Les *exclusions liées à la tempête* comprennent principalement les biens non couverts par la garantie contre les effets du vent. Le recours aux exclusions n'est pas seulement dicté par la perspective d'économies réalisées sur des biens particulièrement exposés au vent : cette pratique offre en outre un certain confort dans la rédaction des contrats puisqu'elle permet l'utilisation de la formule du “ *tout sauf* ”, plus synthétique qu'une rédaction indiquant tout ce qui est garanti, et protégeant l'assureur des conséquences d'un texte prêtant à l'interprétation.

Enfin, il arrive qu'un contrat mentionne l'existence d'une *limite contractuelle d'indemnité*, c'est-à-dire d'un montant maximal de règlement à l'assuré. Ce dernier devient alors son propre assureur pour le montant du préjudice dépassant cette limite.

Une liste exhaustive des éléments constitutifs du préjudice ayant été dressée, il faudra alors rechercher leur influence et leur évolution au cours de la période d'étude 1980 - 1997.

I.2 - Facteurs d'actualisation

Avant cela, il convient de rechercher d'éventuels autres facteurs ayant eu une incidence non négligeable sur l'évolution de la série du montant des dommages provoqués par les tempêtes. Puis, après les avoir identifiés, il restera à déterminer leur

degré d'influence, à travers un *indice d'actualisation* qui les caractérise, et dont la valeur annuelle est à multiplier par la valeur de la série étudiée. Un facteur ayant une influence nulle sera donc crédité d'un indice d'actualisation égal à 1 sur tout l'historique. Après cette opération, la série sera corrigée de ces effets biaisants et considérée comme homogène sur toute la période de l'étude.

Un des plus importants facteurs d'actualisation est *le taux de souscription de la garantie tempête*, c'est-à-dire la part de contrats du portefeuille de la branche incendie garantissant les effets du vent. Le suivi précis de son évolution est essentiel car ce taux, ainsi que le niveau de franchise, vient affecter le nombre total de sinistres tempête.

Les autres facteurs viennent modifier le coût moyen des sinistres, et non leur fréquence. Il faut évoquer en particulier :

- *l'indice de réparation des dommages aux bâtiments*,
- *le taux d'inflation*,
- *l'évolution démographique*,
- *la transformation du parc immobilier*,
- *la modification de la composition du portefeuille de l'UAP* (entre locataires et propriétaires, entre maisons et immeubles, etc...),
- *l'apparition de nouvelles techniques de construction et de matériaux plus robustes mais aussi plus onéreuses*,
- *et bien sûr l'évolution des couvertures d'assurance*.

Ces facteurs ne seront évoqués que dans la mesure où leur influence peut être clairement identifiée. Dans le cas où cette preuve ne pourrait être apportée, leur influence ne serait pas retenue.

I.3 - Modalités de calcul

Il faut bien veiller, lors de la réalisation pratique des calculs d'estimation des éléments du préjudice et des indices d'actualisation, au respect des points ayant trait :

- à l'ordre de réalisation,
- au degré de finesse des calculs.

I.3.1 - Ordre de réalisation des calculs

L'ordre de réalisation des calculs a en effet son importance. Lorsque l'on analyse le rapport d'expertise d'un bien endommagé par la tempête, on constate en effet que l'expert part d'une estimation du préjudice pour en déduire une indemnité. Il s'agit ici de refaire le chemin dans l'autre sens, c'est-à-dire de partir de l'indemnité pour en déduire le préjudice, dans l'ordre inverse de l'expert :

- rajouter d'abord la franchise au montant indemnitaire,
- puis y inclure éventuellement la vétusté,
- y appliquer si nécessaire l'inverse du taux de la règle proportionnelle,
- enfin y joindre les possibles exclusions pour retrouver le montant du préjudice.

Les esprits sceptiques pourront se convaincre que l'arrangement de ces calculs n'est pas neutre par l'exemple suivant. Ainsi, considérons un sinistré dont le préjudice atteint une valeur de 10.000 Frs. Malheureusement pour lui, la somme garantie ne représente que 90 % du montant de la valeur du bien assuré, et son contrat ne stipule pas l'abrogation de la règle proportionnelle sur les capitaux : une réduction proportionnelle de l'indemnité lui sera alors appliquée. De même, son contrat ne garantit pas les dommages aux clôtures et aux persiennes, qui atteignent 1.500 Frs. Il limite d'autre part la vétusté à un taux de 25 %, inférieur au taux réel de vétusté, égal à 35 %. Enfin, une franchise d'avarie de 10 % est prévue à son contrat.

Deux questions se posent : premièrement, quel ordre l'expert doit-il respecter pour calculer la vraie indemnité ? Deuxièmement, l'ordre importe-t-il si tous les calculs conduisent au même résultat ?

Le tableau 2.2.1 répond à ces deux questions, à partir de 3 arrangements différents des calculs.

Tableau 2.2.1 : Comparaison de 3 calculs d'indemnités à partir d'un même préjudice dû à la tempête

cas n° 1		cas n° 2		cas n° 3	
opération	somme restante	opération	somme restante	opération	somme restante
préjudice	10.000	préjudice	10.000	préjudice	10.000
règle proportionnelle	9.000	vétusté	9.000	exclusions	8.500
exclusions	7.500	règle proportionnelle	8.100	règle proportionnelle	7.650
vétusté	6.750	exclusions	6.600	vétusté	6.885
franchise	6.075	franchise	5.940	franchise	6196,5
différence	3.925	différence	4.060	différence	3.803,5

C'est bien évidemment le cas n° 3 qui correspond au bon cheminement de calcul. Les autres arrangements, sélectionnés parmi bien d'autres calculs à ne pas suivre, montrent qu'une erreur dans l'ordre conduit à des valeurs indemnitaires sensiblement différentes.

Ce classement partiel des éléments de préjudice sera bien sûr à respecter, mais il constituera une troisième étape, car il faudra auparavant avoir mené :

- dans un premier temps, les calculs d'actualisation tenant compte de l'érosion monétaire ou, ce qui revient au même, de l'inflation,
- puis, aussitôt après, les calculs conduisant à la génération de nouveaux sinistres (taux de souscription de la garantie tempête, niveau de franchise).

En effet, il ne serait pas logique de considérer comme équivalentes deux franchises exprimées en Frs et d'un montant identique mais appliquées à des époques distinctes : en période d'inflation, la franchise appliquée à la date la plus ancienne a eu un impact plus important sur le nombre de sinistres non indemnisés car d'un montant inférieur à la franchise, que celle appliquée à la date la plus proche. L'indice d'actualisation tenant compte de l'inflation conduira donc à ce que les deux franchises d'un même montant ne génèrent pas le même nombre de sinistres à niveau de franchise nul.

Ceci justifie que l'indice d'actualisation dû à l'inflation soit appliqué au montant de chaque sinistre - déduction faite des éventuels montants d'honoraires - *avant* le calcul du surcoût généré dans l'hypothèse d'une franchise nulle et d'une garantie tempête systématiquement présente dans les contrats de l'étude.

De même, il est logique de procéder au calcul du nombre de sinistres " fictifs " et du coût généré par eux *avant* de leur ajouter les éléments de préjudice restant à la charge de l'assuré.

Connaissant alors le *nombre de sinistres qu'il aurait dû y avoir à taux de souscription de la garantie tempête égale à 100 % et à niveau de franchise nul*, on est en mesure ensuite d'estimer la part non indemnisée pour chacun de ces sinistres.

I.3.2 - Degré de finesse des calculs

Parallèlement, il est primordial de connaître la structure croisée selon laquelle les calculs sont à effectuer. Le degré de finesse le plus fin d'un calcul est bien sûr le sinistre : cette situation est celle où l'on disposerait du temps nécessaire à l'analyse de chacun des 300.000 dossiers de sinistres pour en déduire à chaque fois les éléments constitutifs du préjudice ! Cette solution peu réaliste nous amène à définir un niveau plus agrégé pour le calcul des éléments de préjudice et des différents indices d'actualisation.

Sans dévoiler le contenu des paragraphes suivants, qui explicitent en détail les modalités de chaque calcul (de préjudice, ou d'estimation d'indice), on peut néanmoins dire que le résultat de ce calcul est toujours ventilé par *exercice de survenance* des tempêtes - c'est la raison d'être d'un indice d'actualisation - et par *risque* (d'habitation, professionnel, agricole), même si, dans certains cas, les valeurs sont identiques.

II - Actualisation liée à l'évolution des prix

Au cours de la période 1980 - 1997, l'économie française a connu deux phases :

- la première (1980 - 1986) est très inflationniste, conséquence de la hausse des produits pétroliers décidée en 1979 par les pays producteurs de l'Organisation des Pays Exportateurs de Pétrole (OPEP), avec néanmoins une tendance baissière affirmée : la fin de l'indexation des salaires sur le rythme d'inflation décidée en 1983 a pu juguler en partie cette croissance des prix, qui se situe entre 5 et 15 % par an,
- la seconde (de 1986 à 1997) présente une période de faible inflation, que l'on attribue principalement au rôle modérateur joué par la grande distribution, et au fort développement de la concurrence entre grandes marques. L'indice des prix est alors compris entre 0 et 5 %.

Parallèlement, l'*indice de réparation des dommages aux bâtiments* suit une courbe ascendante : les techniques de construction et les matériaux utilisés gagnent en qualité et en résistance, ce qui se traduit bien évidemment par une hausse du coût de la construction. La décélération est néanmoins moins prononcée que pour l'indice des prix à la consommation.

Notre préférence va à l'indice F.N.B. (ou indice du coût de la construction) et l'indice R.I. (risques industriels). En effet, chaque contrat est rattaché à une " branche-risque " représentant un type de contrat particulier. Or, la fiche signalétique d'une " branche-risque " de la branche incendie contient toujours un élément d'indexation des primes qui se trouve être choisi parmi l'un de ces deux indices. Nous les adoptons donc tous les deux, et ne tenons pas compte de l'inflation.

Par la suite, il suffit d'appliquer à chaque montant de sinistre de la base " sinistres " un indice d'actualisation fonction de :

- la date de règlement par rapport à la date correspondant à un indice égal à 100 (décembre 1997); plus la date de règlement est éloignée, plus l'indice d'actualisation est important,

- l'élément d'indexation caractéristique de la b ranche-risque (F.N.B. ou R.I.).

On considère que le règlement de l'indemnité de sinistre s'effectue à une date unique et qu'il suit d'environ trois mois la date de survenance. Ces hypothèses très réalistes permettent de constituer une première table d'actualisation ventilée selon la structure croisée par branche-risque et nombre de trimestres séparant la date de règlement de la date de référence.

III - Actualisation liée à l'évolution de la souscription de la garantie tempête

III.1 - Bref historique

L'intérêt de ce paragraphe résulte d'un constat indiscutable : la présence de la garantie tempête dans les contrats incendie - multirisque n'a pas toujours été au niveau actuel. Il en résulte que la charge des sinistres dus aux effets du vent doit faire l'objet d'une réactualisation, sur la base d'un *taux de souscription de la garantie tempête* égal à 100 %, particulièrement pour la période comprise entre 1980 et 1990.

En effet, comme l'a exposé en détail le chapitre 4 de la première partie, le niveau de souscription de cette garantie n'a pas été étale :

- avant 1970, les trois compagnies, qui allaient former en 1968 l'UAP, proposent déjà, soit systématiquement, soit en option, une garantie tempête,
- en 1970, le nouveau produit commun appelé MH (Multirisque Habitation) contient une garantie intégrée qu'on ne peut exclure, à l'exception de la formule MH à option; les souscripteurs de ce produit, ainsi que du contrat IH (Incendie Habitation) se verront attribuer obligatoirement la garantie contre les effets du vent à partir de 1975. Les contrats agricoles et professionnels, ainsi que ceux relevant du tarif industriel, disposent seulement d'une garantie facultative,
- en 1980, en risque d'habitation, le contrat Globassur remplace le contrat MH : une possibilité est offerte aux assurés de ne pas souscrire la garantie tempête,
- en mars 1983, à la demande des pouvoirs publics, une pollicitation est réalisée pour proposer *systématiquement* une garantie tempête améliorée aux détenteurs de contrats incendie - multirisque, à l'exception de ceux relevant du tarif industriel. (A ce sujet, on pourra se reporter pour mémoire au chapitre 4 de la première partie, paragraphe II.1). Sans notification écrite de leur refus, celle-ci leur sera attribuée moyennant 2 Frs pour le risque d'habitation et 20 Frs pour les risques professionnels et agricoles,

- parallèlement, à partir de cette date, il est *systématiquement proposé, mais non imposé*, cette garantie lors de chaque acte de production, que ce soit lors d'une affaire nouvelle ou d'un remplacement (cf. le chapitre 3 de cette partie, paragraphe I, pour le détail des conditions tarifaires).
- en 1990, la loi du 25 juin 1990 fait obligation aux assureurs d'intégrer dans leur contrat relevant de la branche incendie une garantie contre les effets du vent : avant cette date, le taux de souscription de cette garantie est estimé à :
 - 99 % pour la multirisque habitation,
 - 90 % pour la multirisque professionnelle,
 - 99 % pour la multirisque agricole,
 - 85 % pour la multirisque industrielle.
- après cette date, la souscription de cette garantie est maximale.

III.2 - Difficultés de calcul du taux de présence de la garantie tempête

Si la période 1990 - 1997 ne recèle aucun secret quant à la présence ou l'absence d'une garantie contre les effets du vent, en revanche, il n'en est pas de même pour la période 1980 - 1989. Il est en effet impossible de connaître avec précision, pour chacune de ces années, l'exacte valeur du taux de présence de la garantie tempête (appelé par la suite taux de présence).

Il faut en effet savoir que jamais le risque lié à la tempête n'a fait l'objet d'une tarification particulière, avant 1990 comme après. On se référera au chapitre 3 de cette partie, paragraphe I, pour disposer d'un éclairage précis sur cette question.

Il en résulte que les informations présentes sur la fiche signalétique d'un contrat de la branche incendie peuvent parfois, mais hélas pas systématiquement, déceler la présence d'une garantie tempête. Voici ce que disait à ce sujet, dans le compte rendu de la réunion du 16 février 1983, la commission chargée à l'UAP de l'organisation de cette pollicitation et de son suivi statistique et comptable :

“ Il est estimé que 90 % des contrats des particuliers (simples ou agricoles) comportent une garantie tempête, et qu'il n'est pas possible d'identifier sans consultation du dossier ceux qui ne la comportent pas. (...) Pour les risques industriels, seuls les contrats comportant l'intercalaire P 13 pourront être détectés, l'intercalaire P 15 n'étant pas codifié. ”

S'il est vrai que le calcul exact du taux de présence de la garantie tempête est impossible, il n'est pas interdit d'en faire une estimation pour les années précédant 1990. D'ailleurs, comment peut-on à la fois donner une estimation, certes peu précise, de ce taux, et dire qu'il n'est pas possible de le faire ? Lorsque l'enjeu est

d'importance, il est parfois nécessaire de vérifier le fondement de certaines affirmations, quitte à y consacrer un peu de temps.

III.3 - Estimation du taux de présence de la garantie tempête

Le calcul du taux de présence de la garantie tempête procède selon trois étapes dont les deux premières sont successives :

- la première étape consiste dans *l'étude du caractère facultatif ou obligatoire de la présence de cette garantie, ou de son absence*, au niveau de la subdivision administrative déjà citée plus haut, la b ranche-risque,
- la seconde étape réside dans *l'estimation du taux de présence de la garantie tempête pour les seules branches-risques où son caractère facultatif est avéré*; le recours à des méthodes statistiques est alors nécessaire,
- enfin, la troisième étape est la *collecte de données relatives au nombre de contrats que compte chaque branche-risque aux inventaires des exercices 1979 à 1989* : l'agrégation de ces données conduit alors à l'estimation d'un taux de présence global, que l'on étalonnera à celui qui était communément avancé sans justification.

III.3.1 - Nature de la présence de la garantie tempête pour chaque branche-risque

La première étape, consistant à déterminer, pour chacune des quelque 340 b ranches-risques relevant de la branche incendie, la présence, obligatoire ou facultative, ou l'absence de la garantie tempête, a pris différentes formes.

III.3.1.1 - Exploitation des bases informatiques

Une première piste de recherche consiste à *tirer le meilleur parti des informations, exploitables par l'informatique, et relatives aux caractéristiques des contrats et des sinistres*. Il s'agit plus précisément :

- de la *base des sinistres*, présentée en détail au paragraphe I du premier chapitre de cette partie,
- des *fiches signalétiques des contrats touchés par un sinistre tempête* dont le lieu de survenance ne pouvait pas être connu au niveau de la zone postale; ceci requerrait une extraction de quelque 40.000 de ces fiches. (On pourra se référer au chapitre 1 de cette partie, paragraphe II.3, pour plus de précisions),

- des *fiches signalétiques des branches-risques*, déjà utilisées pour y trouver l'indice d'indexation de référence.

On peut alors se demander à juste titre pour quelle raison les branches-risques des contrats sinistrés par la tempête ne posséderaient pas systématiquement la garantie tempête. Une telle conclusion ne peut être tirée que si l'on a l'assurance que l'assuré sinistré a perçu une indemnité.

Il faut en effet vérifier, *dans le fichier "sinistres"*, que des règlements indemnitaires, dans lesquels n'entrent pas les règlements d'honoraires, ont bien été effectués sur les sinistres qui peuvent comporter la garantie tempête. Ce contrôle s'explique par le fait que la déclaration d'un dommage subi par un assuré de l'UAP entraîne toujours l'ouverture d'un sinistre "déclaré provisoire". C'est seulement après qu'intervient la vérification de la présence sur le contrat de la garantie sinistrée. S'il s'avère que le contrat ne comporte pas la garantie qui permettrait de régler l'assuré sinistré, le sinistre est considéré comme *sans suite*.

La présence d'un règlement indemnitaire certifie donc que le contrat sinistré, et par suite la branche-risque correspondante, comporte la garantie tempête. En revanche, il n'est pas possible d'en déterminer le caractère obligatoire ou facultatif.

A l'inverse, l'absence d'une branche-risque de la liste des branches-risques touchées par les tempêtes survenues entre 1980 et 1990 permet de conclure à l'absence de garantie couvrant les effets du vent.

De même, *l'analyse des fiches signalétiques* des contrats touchés par un sinistre tempête dont le lieu de survenance n'est pas déterminé avec précision peut conduire à des résultats intéressants puisqu'on y trouve les variables suivantes :

- les "*ventilations de primes*" (V.P.), qui correspondent à la transcription numérique des garanties, de base ou facultatives, présentes sur le contrat,
- les *numéros de garanties annexes*,
- la *date d'entrée en risque du contrat*, c'est-à-dire la date à partir de laquelle le bien, objet de l'assurance, a été couvert,
- le *code branche-risque*.

S'il ne peut être question de remettre en cause l'affirmation citée au paragraphe II.2, sur l'impossibilité de prouver l'absence de la garantie tempête sur une branche-risque, en revanche, il est possible de démontrer sa présence. Cette situation apparaît en présence :

- d'un code V.P. 12 (le " 12 " symbolisant la tempête),
- de certains codes de garanties annexes.

Il est à noter que la présence d'un code V.P. 12 s'interprète différemment selon que la date d'entrée en risque est antérieure ou postérieure à la loi du 25 juin 1990 : si cette date est antérieure à 1990, on conclut au caractère facultatif de la présence de la garantie, et à son caractère obligatoire dans le cas inverse.

La troisième source d'information exploitable par l'informatique est la base des *fiches signalétiques des branches-risques*, dont le principal mérite est de fournir la liste exhaustive des ventilations de primes acceptées sur chaque b ranche-risque : il est alors possible de statuer sur la présence de la garantie tempête, mais pas sur son caractère systématique.

Très souvent, les indices extraits des bases informatiques précédentes n'ont permis que d'établir la présence ou l'absence de la garantie tempête. Il n'a pas été possible d'en déterminer le caractère obligatoire ou facultatif. C'est pourquoi un deuxième type de recherche doit être poursuivi pour les branches-risques pour lesquelles nous n'avons pu statuer.

III.3.1.2 - Exploitation de la base documentaire

Ainsi, il a fallu souvent *se mettre en quête des documents officiels présentant chaque produit*, et en premier lieu des *conditions générales d'assurances* des générations de produits créés depuis 1945, que ce soit pour l'habitation ou les risques agricole, professionnel et industriel. Il peut paraître surprenant de remonter aussi loin dans le temps. Cependant, il ne faut pas oublier que le portefeuille incendie recelait, au début des années 1980, une grande proportion de contrats qui n'avaient jamais été modifiés. Par la suite, dans un souci de rationalisation, l'informatique prenant une part croissante dans la gestion des contrats, de nombreuses opérations de migration de ces anciennes formules vers de nouveaux produits ont été réalisées.

Il n'en demeure pas moins une très significative présence de ces anciens contrats au début des années 1980, comme le démontre le résultat de notre collecte de données, objet de notre troisième étape. Cela nécessite la recherche des conditions générales ou, si celles-ci s'avèrent insuffisantes, du *tableau des garanties* du produit. Dans ce document, la garantie tempête peut figurer soit dans les garanties de base, soit être citée parmi les garanties facultatives, soit au contraire ne pas apparaître. Cette dernière situation ne doit pas conduire à la conclusion d'une absence de la garantie tempête sans la consultation préalable des *conditions particulières*. Cette pièce donne en effet un détail précis du calcul de la prime dont le souscripteur devra s'acquitter. Cependant, le rappel des garanties ne concerne souvent que celles constituant la prime de base et les garanties dites annexes ne sont pas citées.

Une autre source riche en information est l'ensemble des *notes d'instructions* délivrées par la direction Générale aux réseaux des agents ou des salariés, les courtiers ne pouvant en être destinataires. Il a été possible de retrouver au secteur des produits des particuliers de l'UAP toutes les notes d'instruction depuis leur création, en 1970, et, parmi celles-ci, les notes de lancement des nouveaux produits dans lesquelles figure une présentation détaillée des garanties.

D'autres documents officiels ont pu être exploités, tels que les répertoires de tarifs, le texte d'annexes, les notes administratives, etc...

Ainsi, lorsque les documents officiels ont livré tous leurs secrets ou ont montré leurs limites, la recherche d'éléments déterminants sur la présence ou l'absence de la garantie tempête dans les produits de la branche incendie prend alors une nouvelle forme : ***le rapatriement et l'exploitation des contrats archivés***. On y trouve, outre les conditions générales et le tableau des garanties, les conditions générales remplies, ce qui permet en général de conclure. La transcription numérique ou alphabétique des garanties annexes, qui est un art consommé du monde des assurances, cache parfois un texte de garantie contre les effets du vent. Cela ne facilite pas notre approche du sujet puisqu'il faut alors trouver la signification de ce code. Or, de nombreuses annexes ont vu le jour bien avant l'informatisation de la gestion des produits, ce qui fait que le texte auquel elles font référence n'est pas disponible à l'écran, et ce qui les rend presque plus rares que des pièces de musée ! En revanche, le texte des autres annexes, plus récentes, peut être aisément consulté.

Il est à noter une ***difficulté particulière*** dans l'exploitation des documents officiels : ***l'établissement d'une correspondance entre le document et la branche-risque à laquelle il se réfère***. Dans la plupart des cas, les conditions générales ou particulières, de même que le tableau des garanties ne font nulle part allusion à la codification de la branche-risque correspondante. Ce n'est pas le cas des notes d'instruction datant d'avant 1985 car, à l'époque, la branche-risque n'était pas encore déterminée automatiquement à partir des informations présentes dans la note de couverture, mais saisie elle aussi. Il fallait donc connaître la valeur de la b branche-risque, sous peine de ne pouvoir achever l'acte de production. Ce cas n'est nullement envisageable lors de l'exploitation des contrats archivés, puisque le numéro de contrat désarchivé est forcément lié à un code b branche-risque.

Cependant, les contrats désarchivés mais souscrits avant 1971 ne comportent pas, sauf en cas de remplacement, de code b branche-risque, puisque la date d'apparition de cette subdivision administrative, postérieure à la fusion de l'UAP, remonte à 1971. Les codifications que l'on y trouve sont sans rapport avec les codes b branche-risque.

Dans la plupart des cas, l'établissement d'un lien entre le texte et le code b branche-risque résulte, non pas de la similitude des codes b branches-risques, mais de la ressemblance du libellé de cette b branche-risque et de la dénomination du produit apparaissant sur le document.

On a pu aussi procéder par *assimilation*, lorsque le libellé, la date d'apparition et les caractéristiques d'une branche-risque donnée, dont on connaît le statut de la garantie tempête, coïncident avec ceux relevés pour d'autres branches-risques : on considère alors que ces dernières partagent le statut de la première branche-risque. Cette technique d'assimilation est très employée car un produit nouveau se présente sous différentes facettes. Par exemple, le produit " Incendie Professionnelle ", lancé en 1985, regroupe 5 branches-risques qui permettent de distinguer le marché auquel il s'adresse (particulier, PME-PMI, entreprise), et le caractère automatique ou pas de la tarification appliquée.

D'une manière générale, la progression de nos travaux se heurte à de nombreuses difficultés : le taux de retour des contrats désarchivés est inférieur à 10 %, les dossiers ne contiennent pas les pièces intéressantes ou ont changé de code b ranche-risque, les documents officiels ne sont plus disponibles, et les fiches signalétiques sont rarement interprétables relativement à la nature de la garantie tempête...

Le tableau 2.2.2 fournit les résultats finalement obtenus.

Tableau 2.2.2 : Répartition des branches-risques selon la nature de la présence de la garantie tempête

Nature de la présence de la garantie tempête	Nombre de branches-risques	Pourcentage de branches-risques
présence obligatoire	186	54,7 %
présence facultative	118	34,7 %
absence systématique	19	5,6 %
indétermination entre la présence obligatoire ou facultative	17	5,0 %
total	340	100 %

Malgré le faible pourcentage (5 %) de branches-risques au sort indéterminé, il faut mettre l'accent sur la difficulté d'identification de la véritable nature de la présence de la garantie tempête. Au total, près de 55 % des branches-risques accueillent toujours la garantie tempête. Il faut ajouter que certaines branches-risques regroupent des produits de l'Union et de l'Urbaine de nature différente : les b ranches-risques en question sont considérées comme ne comportant pas systématiquement la garantie tempête.

III.3.2 - Estimation du taux de présence de la garantie tempête pour certaines branches-risques

III.3.2.1 - Contraintes informatiques

Après avoir déterminé les branches-risques admettant en option ou en garantie de base la tempête, il convient de ne retenir que la première catégorie pour le calcul du taux de présence de la garantie tempête, pour les inventaires des exercices 1979 à 1989, le taux de présence étant par définition de 100 % pour la deuxième catégorie.

Les contrats dont la garantie tempête est facultative disposent en général d'une fiche signalétique dont la présence du code V.P. 12 atteste de la souscription par l'assuré de la garantie tempête. A l'inverse, l'absence du code V.P. 12 est synonyme de non garantie contre les effets du vent.

Il existe malheureusement une partie importante de ces contrats dont la garantie tempête n'est repérable que par la consultation du dossier papier. Ainsi, le taux calculé t_c à partir de la part de contrats dont la fiche signalétique présente la V.P.12 est inférieur au taux de présence réel t_r . Pour retrouver une estimation sans biais, une idée consiste à mesurer le taux de V.P. 12 pour des contrats sinistrés par la V.P. 12 ($t_c(t)$) et à faire l'hypothèse que la proportion de contrats avec de la V.P. 12 est la même entre les contrats sinistrés par la tempête et ceux qui ne le sont pas. (Cela revient à dire que les contrats disposant de la garantie tempête ne sont pas plus affectés par la tempête que les contrats n'en disposant pas). La valeur estimée du taux de présence pour l'ensemble des contrats d'une branche-risque est donc égale à :

$$\hat{t}_r = t_r(t) \cdot \frac{t_c}{t_c(t)}$$

Si, par exemple, le taux calculé de présence de la V.P. 12 pour un ensemble de contrats sinistrés par la tempête est de 70 %, alors que le taux réel est égal à 100 % (par définition pour des contrats ayant fait l'objet d'un règlement suite à une tempête), et que l'on trouve une proportion de V.P. 12 égale à 55 % pour l'ensemble des contrats de la branche-risque, alors on estime que le taux de présence réel pour la b ranche-risque est égal à :

$$\hat{t}_r = 100\% \cdot \frac{55\%}{70\%} = 79 \%$$

III.3.2.2 - Réalisation d'un sondage

Pour arriver à ce genre de résultat, il faut par conséquent procéder à l'exploitation de deux bases informatiques :

- l'une contenant les fiches signalétiques des contrats sinistrés à la suite d'une tempête,
- l'autre, à constituer, contenant les fiches signalétiques des contrats sinistrés ou non.

Le rapport des taux de présence de la V.P. 12 permet de se ramener à un taux allant de 0 à 100 %.

Pour ce faire, l'idée consiste à *réaliser un sondage* à partir de l'exploitation informatique d'un nombre suffisant de contrats relevant d'une des branches-risques en question. Etant donné la création récente de ce fichier " historique des contrats " (on y trouve les contrats résiliés depuis 1983 et les affaires nouvelles depuis 1987), la répartition du nombre de contrats en cours par exercice d'inventaire décroît très rapidement à mesure que l'on remonte le temps. La présence de contrats en cours avant 1983 y est réduite. Il est donc vain de vouloir réaliser un sondage par année, où seuls les contrats en cours pourraient faire partie de la population sondée.

On considère donc qu'une branche-risque conserve un taux de présence de la garantie tempête équivalent tout au long de sa présence sur le portefeuille précédant la pollicitation de 1983 : en effet, la décision d'un assuré de se garantir contre les effets du vent peut être l'occasion pour le producteur de procéder à un remplacement du contrat ou - mieux - à une affaire nouvelle, pour laquelle il touche une commission. Mais, en ce qui nous concerne, l'important est de constater que cette opération aboutit à un changement de branche-risque.

La pollicitation de mars 1983, visant à généraliser la présence d'une garantie tempête, a certes connu un grand succès en augmentant le taux de présence puisque, globalement, la nouvelle garantie a été acceptée par près de 98 % des assurés en raison de son prix modique. En revanche, la tarification proposée pour les affaires nouvelles et les remplacements étant ramenée à un niveau normal, le taux de refus de cette garantie n'a pas été aussi faible entre 1983 et 1990.

Les branches-risques sont nombreuses mais il est possible de regrouper ces dernières par produit : ne sont retenus alors dans le sondage que les groupes dont le nombre de contrats dépasse un pourcentage significatif du risque auquel ils appartiennent. Le reste des autres branches-risques est regroupé dans un même ensemble.

Nous utilisons le principe de l'échantillon issu d'un tirage au sort à P.E.S.R. (probabilités égales sans remise). Le tirage au sort doit assurer une chance équivalente à l'ensemble des contrats. Chaque contrat ne peut être tiré au sort qu'une fois.

On montre que le taux de V.P. 12 obtenu à partir d'un échantillon de $n = 500$ unités est un estimateur sans biais du taux de V.P. 12 étendu à l'ensemble de la b ranche-risque ou du regroupement de branches-risques.

En effet, si y est la variable aléatoire mesurant les valeurs des contrats de l'échantillon et prenant les valeurs 1 (i.e. la V.P. 12 est présente sur la fiche signalétique du contrat) ou 0 (i.e. la V.P. 12 est absente), alors l'estimateur \bar{y} vaut :

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

et l'espérance de cet estimateur $E(\bar{y})$ vaut :

$$E(\bar{y}) = \frac{1}{n} \sum_{i=1}^n E(y_i).$$

Or y_i vaut Y_a avec la probabilité $1/N$, où Y est la variable booléenne ayant la même définition que y mais appliquée à l'ensemble des N contrats de la b ranche-risque et α un de ces contrats, d'où :

$$E(y_i) = \frac{1}{N} \sum_a Y_a = \bar{Y} \text{ et donc : } E(\bar{y}) = \frac{1}{n} \sum_{i=1}^n \bar{Y} = \bar{Y}$$

Ensuite, il ne reste plus qu'à renouveler ce sondage pour chacun des regroupements de branches-risques et ce pour les contrats sinistrés en tempête ou pour l'ensemble des contrats : le rapport de ces taux donne une estimation de la présence de garantie tempête entre 1980 et 1982 pour chaque branche-risque.

Pour 1983, année de la pollicitation, la relation entre le taux de présence après pollicitation t_{ap} et le taux de présence avant pollicitation t_{av} s'établit à partir du tableau de passage 2.2.3 suivant.

Tableau 2.2.3 : Relation entre les taux de présence de la garantie tempête avant et après la pollicitation

	Après	acceptation de la nouvelle garantie	refus de la nouvelle garantie	total
Avant				
contrats avec garantie tempête		$n_{av} \cdot t_a$	$n_{av} \cdot (1-t_a)$	n_{av}
contrats sans garantie tempête		$(n-n_{av}) \cdot t_s$	$(n-n_{av}) \cdot (1-t_s)$	$n-n_{av}$
tous contrats		$n \cdot t$	$n \cdot (1-t)$	n

On note t_a et t_s respectivement les taux d'acceptation de la nouvelle garantie tempête pour les contrats qui en étaient dotés avant et après la pollicitation, t étant le taux

d'acceptation global (et estimé, selon les décomptes faits par le secteur Etudes et produits, à 98 %). n_{av} et n_{ap} étant le nombre respectif de contrats disposant de la garantie tempête avant ou après la pollicitation, la relation qui lie ces deux grandeurs est la suivante :

$$n_{ap} = n_{av} + (n - n_{av}) \cdot t_s \text{ soit encore : } t_{ap} = t_{av} + (1 - t_{av}) \cdot t_s$$

En fait, ce n'est pas t_s que l'on connaît mais t , grâce au décompte fait à partir du code d'une annexe signifiant " refus de la nouvelle garantie ".

Si l'on considère que t_s est peu différent de t , alors t_{ap} devient calculable.

Rappelons que le portefeuille des risques industriels est resté à l'écart de cette pollicitation et que seuls les affaires nouvelles et les remplacements ont été sollicités.

On trouvera dans l'annexe J le programme SAS (TPGTEMP) des calculs du taux de présence. Nous reproduisons ici le tableau 2.2.4 donnant ce taux pour les 4 risques sur la période 1980 – 1983.

Tableau 2.2.4 : Taux de présence de la garantie tempête par risque entre 1980 et 1989

Risque	Période	
	1980 - 1982	1983-1989
Habitation	93 %	99 %
Agricole	80 %	99 %
Entreprise	67 %	99 %
Industriel	67 %	70 %
Branche Incendie	90 %	98 %

III.3.3 - Précision du calcul du taux de présence de la garantie tempête

La précision de notre calcul peut être aisément trouvée grâce à l'utilisation d'un sondage à PESR : dans ces conditions, on sait qu'il est possible de construire un intervalle de confiance dans lequel se trouve le taux de présence de la garantie tempête de chaque groupe, pour un seuil de confiance donné. En effet, l'application du théorème central limite garantit, sous des conditions peu restrictives, la normalité asymptotique de \bar{y} .

Au seuil de confiance de 95 %, l'intervalle de confiance de \bar{y} est : $\bar{y} \pm 2s \sqrt{\frac{(1-f)}{n}}$, où f est le taux de sondage retenu dans notre étude.

Si l'on suppose que nos différents sondages sont réalisés indépendamment les uns des autres, il est alors possible de considérer les différents estimateurs comme des

variables indépendantes suivant des lois normales, de même que toute combinaison linéaire de ces variables. On peut donc donner un intervalle de confiance pour chacun des 4 risques retenus dans cette question.

Si l'on considère en outre que le taux de présence de la garantie tempête des branches-risques :

- à présence obligatoire suit une loi $N(100,0)$,
- ou à absence systématique est une constante,

il est alors envisageable d'établir un intervalle de confiance pour *toutes* les branches-risques d'un risque donné. C'est ce que l'on retrouve dans le tableau 2.2.5 suivant.

Tableau 2.2.5 : Intervalle de confiance lié au taux de présence de la garantie tempête par risque entre 1980 et 1989

Risque	Période	
	1980 - 1982	1983-1989
Habitation	93 % ± 0,4 %	99 % ± e %
Agricole	80 % ± 2,1 %	99 % ± e %
Entreprise	67 % ± 3,0 %	99 % ± e %
Industriel	67 % ± 3,0 %	70 % ± 3,0 %
Branche Incendie	90 % ± 0,7 %	98 % ± 0,2 %

III.4 - Calcul de la charge de sinistres à taux de présence de la garantie tempête égale à 100%

Après avoir procédé à l'estimation du taux de présence de la garantie tempête entre 1980 et 1989 pour les différents risques considérés (habitation, professionnel, agricole), il convient de conclure notre étude en utilisant ces coefficients d'actualisation pour déterminer le montant de la charge de sinistres qui aurait été réglé si le taux de présence avait toujours été égal à 100 %.

Nous allons pour ce faire établir une relation entre la charge de sinistres avec un taux de présence maximum et égal à 100 % (C_T), et celle correspondant à un taux de présence réellement constaté (C_G). On a la relation classique suivante :

$$C_T = c_T \cdot n_T \cdot \frac{n_{ST}}{n_T} = c_T \cdot n_T \cdot f_T \quad \text{où :}$$

- c_T est le coût moyen calculé sur l'ensemble des sinistres survenus pour un taux de présence égal à 100 %,
- n_T est le nombre de contrats total,

- n_{ST} et f_T sont respectivement le nombre et la fréquence de sinistres survenus pour un taux de présence égal à 100 %.

Or, la prime pure π_T est égale au rapport de la charge au nombre de contrats soit :

$$P_T = \frac{C_T}{n_T} = c_T \cdot f_T,$$

ce qui conduit au résultat suivant, si l'on dresse le rapport des primes pures π_T et π_G :

$$C_T = \frac{P_T}{P_G} \cdot \frac{1}{t_G} \cdot C_G$$

Ce résultat, somme toute logique, montre que la charge totale sera d'autant plus importante par rapport à la charge réellement constatée que :

- le rapport de la prime pure calculée sur l'ensemble des contrats sur celle calculée sur les seuls contrats dotés de la garantie tempête sera grand,
- et que le taux de présence sera faible.

Mais qu'en est-il exactement des contrats sans garantie tempête ? Quels types de sinistres génèrent-ils ? Et de quelle importance ? Et selon quelle fréquence par rapport aux contrats dotés ?

Pour pouvoir apporter une réponse à ces interrogations, il peut être intéressant d'utiliser la *prime nette d'annulation de la garantie incendie*, information que l'on peut extraire de la fiche signalétique des contrats ayant participé au sondage pour l'estimation du taux de présence de la garantie tempête. (En effet, nous avons pris bien soin de ne pas faire disparaître cette intéressante source d'information). Cette prime, hors taxe mais alourdie par les différentes charges ou commissions, présente sans doute une bonne corrélation avec la prime pure de la garantie tempête.

Si les conditions du sondage sont strictement identiques à celles du précédent, néanmoins la variable étudiée diffère en ce sens qu'elle n'est pas booléenne mais prend des valeurs réelles. La constitution d'une variable discrète, à partir des valeurs des primes pures constatées dans les échantillons, en k classes d'un poids relativement équivalent et dépassant un cardinal de 5, permet de se retrouver dans un schéma classique de comparaison de deux échantillons (pour les calculs, on pourra se reporter au paragraphe I.3.3.2 du premier chapitre de cette partie). La distance du χ^2 associée est alors égale à :

$$d_0^2 = \sum_{i=1}^k \frac{(n_i^t - n_i^g)^2}{n_i^t + n_i^g}$$

où n_i^g et n_i^t représentent respectivement l'effectif dans la classe i ($i=1, \dots, k$) pour les échantillons extraits de la population totale et de la population des contrats avec la garantie tempête.

Dans l'hypothèse H_0 , le nombre de degrés de liberté est égal à $k-1$. Si le nombre de classes est égal à 10, au niveau de confiance de 95 %, la valeur critique du χ^2 est égal à 16,92. Or, le calcul numérique de d_0^2 donne : 15,2. On accepte donc H_0 .

On en conclut que la prime nette incendie des contrats dotés de la garantie tempête n'est pas significativement différente de celle de l'ensemble des contrats. Par extension, on en déduit que la prime pure tempête ne serait pas significativement différente si elle était calculée pour un portefeuille où chaque contrat serait doté de la garantie contre les effets du vent. *L'indice d'actualisation dû à un taux de présence de la garantie tempête en croissance entre 1980 et 1989 ne concerne que le nombre, et non pas le coût moyen, des sinistres.*

III.5 - Validation des calculs précédents

Il peut être riche d'enseignements d'utiliser les données de la base " sinistres " pour valider le résultat du calcul du taux de présence de la garantie tempête. En effet, il faut se souvenir que, à l'occasion de grandes et puissantes tempêtes (1982 et 1987), des arrêtés interministériels ont assuré une indemnisation des assurés non garantis contre les effets du vent. Dans la mesure où la base " sinistres " contient ces sinistres réglés en catastrophe naturelle, il peut être utile de dresser le rapport, pour ces grosses tempêtes :

$$\frac{nbVS12}{nbvs30 + nbVS12} \quad \text{où :}$$

- $nbVS30$ représente le nombre de sinistres tempête réglés selon le dispositif de la loi du 13 juillet 1982,
- $nbVS12$ représente le nombre de sinistres tempête réglés selon les termes du contrat.

Ce rapport peut être considéré comme le taux de présence de la garantie tempête mesuré sur une grande échelle. En effet, on considère que, pour ces tempêtes, tous les sinistrés ont été indemnisés ce qui revient à considérer le taux de souscription égal à 100 %. Il permet de mieux situer les estimations précédentes réalisées sur un échantillon de taille relativement modeste ($n=500$), compte tenu de la proximité à 100 % de ce taux.

Pour réaliser ce calcul, trois événements tempête ont été retenus, compte tenu de leur taille (i.e. nombre de sinistres supérieur à 1000 sur une période de 72 heures), et du règlement d'une partie des dommages selon le dispositif légal (i.e. présence de la V.S. 30).

Il en ressort le tableau 2.2.6 suivant.

Tableau 2.2.6 : Taux de présence de la garantie tempête (TPGT) pour les tempêtes déclarées catastrophes naturelles

Tempête	Nombre de sinistres « cat. nat. »	Nombre total de sinistres	TPGT
6 - 10 novembre 1982	3 185	27 280	88,3 %
18 - 20 juillet 1983	401	3 918	89,8 %
15 - 16 octobre 1987	301	22 557	98,7 %

Les résultats du tableau corroborent donc tout à fait les estimations du taux de présence de la garantie tempête effectuées au paragraphe III.3.2.

III.6 - Modalités pratiques de l'actualisation

Sur le plan théorique, nous avons démontré la nécessité d'une actualisation de la charge des sinistres supportée par la compagnie, de même qu'apporté les éléments permettant l'estimation de cette actualisation. En revanche, il reste à préciser les modalités pratiques de cette actualisation en terme de :

- degré de finesse du calcul,
- redistribution du volume supplémentaire de sinistres dans l'optique d'une présence systématique de la garantie tempête.

Les calculs précédents ont été réalisés sur les variables dont les modalités offrent une grande disparité de la valeur du taux de souscription : *risque et exercice de survenance* en font partie. Il faut y ajouter la variable qualitative " *événement* ", qui identifie chaque tempête à partir de la sélection des sinistres selon leurs lieu et date de survenance. Il faut en effet se rappeler que certaines de ces tempêtes ont été qualifiées de catastrophes naturelles, entre 1982 et 1989, ce qui a permis aux assurés non garantis contre la tempête d'être indemnisés : l'indice d'actualisation doit donc être de 1 pour ces tempêtes.

Un *niveau de finesse de calcul* plus grand que celui décrit précédemment ne peut pourtant pas être apporté, même s'il paraît souhaitable : il n'est en effet pas possible de donner quelque précision supplémentaire sur le taux de souscription de la tempête entre les zones exposées et les zones non exposées car les développements informatiques sont trop lourds.

Outre la structure croisée de l'indice d'actualisation, il convient de définir la *structure de redistribution* dans la base " sinistres " de ce surcoût engendré par ces sinistres " fictifs ". On ne peut se satisfaire de connaître ce simple montant sans y apporter quelques précisions relatives au nombre de sinistres générés et à la distribution de leur montant.

Faut-il pour autant définir le sinistre selon toutes les caractéristiques le concernant, telles que :

- code branche-risque,
- code postal du lieu du sinistre,
- entité d'appartenance,
- date de survenance,
- montant de l'indemnité ?

Bien qu'il soit possible par des programmes informatiques sous SAS de générer des nombres appartenant à des lois de probabilités classiques, il n'est pas nécessaire de caractériser chaque sinistre. Il faut savoir en effet que le modèle économétrique établit un lien entre les données d'assurances et les données météorologiques à partir d'informations agrégées par *risque, station météorologique et événement tempête*. Cette structure de redistribution du nombre et de la charge supplémentaires de sinistres peut donc se limiter à une rupture par risque et événement, dans la mesure où l'on fait l'hypothèse, faute d'informations complémentaires, d'un taux de souscription égal dans chaque zone météorologique.

Tableau 2.2.7 : Taux de présence de la garantie tempête (TPGT): niveaux de rupture et hypothèse retenus dans les calculs

Niveau de finesse du calcul du taux de présence (1)	Structure de redistribution des sinistres " fictifs " (en nombre et en montant) (2)	Niveau d'agrégation des données du modèle économétrique (3)
- risque - exercice (événement)	- risque - événement tempête	- risque - événement tempête - zone météorologique
hypothèse pour passer de (1) à (2)		hypothèse pour passer de (2) à (3)
TPGT identique pour chaque tempête survenue dans la même année		TPGT identique pour chaque zone météorologique, à risque et événement donnés

De même, si l'on suppose que ce taux est le même pour chaque événement venteux survenu au cours d'une même année, cette structure est compatible avec le niveau de finesse du calcul de l'indice d'actualisation qui est, rappelons-le, par *risque* et *exercice*, à l'exception de certaines tempêtes pour lesquelles l'indice est inopérant.

Le tableau 2.2.7 précédent présente le niveau de rupture auquel les différentes informations sont calculées et les hypothèses faites lors du passage d'une structure à une autre.

IV - Prise en compte de la franchise

IV.1 - Intérêt du calcul de l'indemnisation à niveau de franchise nul

Nous allons maintenant examiner un des éléments les plus importants entrant dans la composition du préjudice : la franchise. L'institution de la franchise visait officiellement à responsabiliser les assurés afin que les travaux d'entretien soient régulièrement entrepris et de diminuer les très petits sinistres coûteux en charge de gestion. Mais on s'est très vite rendu compte que sa véritable raison d'être était d'alléger sensiblement la charge des sinistres supportée par les compagnies d'assurances.

Au cours de la période de l'étude 1980 - 1997, la franchise a globalement vu son montant décroître, et cela pour les trois risques. Il est donc nécessaire de procéder, sur cette période, au calcul de l'indemnisation qui aurait été versée à l'assuré si le montant de la franchise avait été nul.

Les franchises ont été au préalable actualisées selon l'indice d'évolution des prix présenté au paragraphe II et qui prend pour référence décembre 1997. Quant à savoir dans quel ordre doivent s'effectuer les calculs de l'indemnisation à franchise nulle et du surcoût occasionné par un taux de souscription de la garantie tempête maximal, nous considérons ces deux opérations comme indépendantes ce qui nous permet d'en réaliser une sans avoir au préalable à prendre en considération les résultats de l'autre.

IV.2 - Aspects théoriques

IV.2.1 - Adéquation de la distribution du montant des sinistres à une loi log-normale tronquée

Par essence, les sinistres enregistrés dans les fichiers présentent la caractéristique d'avoir tous fait l'objet d'une indemnisation - à l'exception des sinistres ouverts *sans suite* - amputée d'un montant restant à la charge de l'assuré. Mais, en l'absence de franchise, quel aurait été le montant de l'indemnisation globale ?

Outre la somme des règlements effectués, deux composantes supplémentaires se greffent à ce montant, c'est-à-dire :

- d'une part, la charge des sinistres non indemnisés car d'un montant inférieur à la franchise,
- d'autre part, la part de l'indemnisation restant à la charge de l'assuré sinistré.

Pour reconstituer la charge globale d'indemnisation, il est d'usage de considérer que la courbe de distribution du montant de l'indemnisation globale des sinistres suit une loi log-normale. En effet, cette loi correspond bien au profil de distribution du montant des sinistres réellement constaté : des petits sinistres connaissent une forte augmentation de leur fréquence à mesure de l'augmentation de l'indemnisation, puis, pour les tranches de coût supérieur, la fréquence décroît aussi très vite.

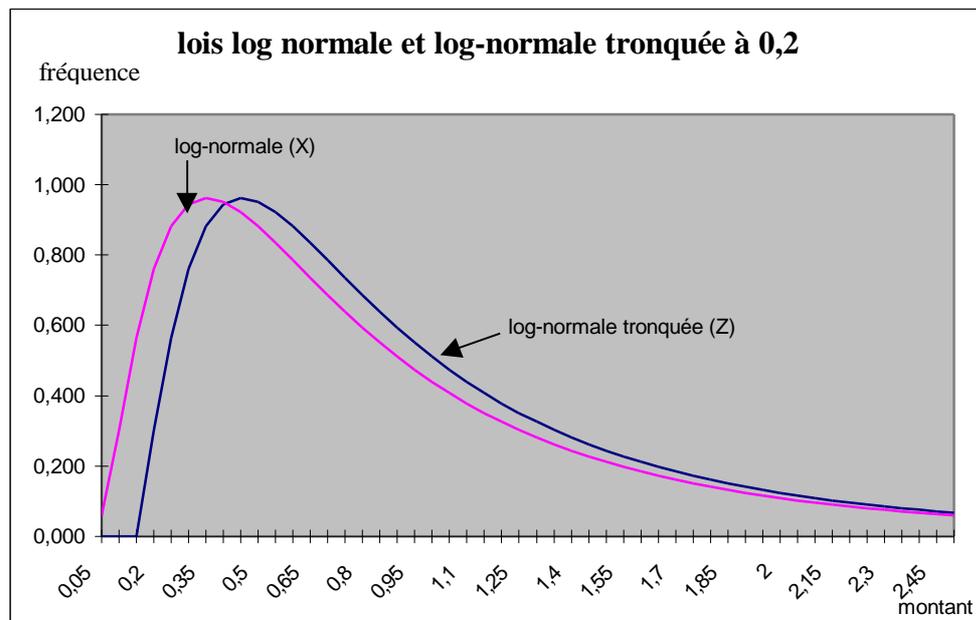


Figure 2.2.1 : Représentation d'une loi log-normale et d'une loi log-normale tronquée

Il est possible d'établir une relation entre la charge globale d'indemnisation, objet de nos recherches, et le montant des règlements de sinistres présents dans la base de données, que nous estimons suivre une loi log-normale tronquée, du fait de l'application d'une franchise. Pour cela, un rappel sur la loi log-normale s'impose : c'est la loi d'une variable positive X telle que son logarithme népérien suive une loi normale. Alors :

$$\ln X \sim N(m; S^2)$$

où m et s^2 sont les deux premiers moments de la loi normale N . Par un simple changement de variable, nous en tirons la fonction de densité $f_{ln}(x)$ de la loi log-normale :

$$f_{ln}(x) = \frac{1}{\sqrt{2p.s.x}} \exp\left(-\frac{1}{2}\left(\frac{\ln x - m}{s}\right)^2\right)$$

Intéressons-nous maintenant à la densité d'une loi log-normale tronquée, nommée Z , telle que $Z = X - \gamma$, où γ est le seuil de troncature. La figure 2.2.1 représente sur un même graphe ces deux lois.

Nous avons la relation suivante entre les deux lois de probabilité :

$$P(Z \leq x) = P(X - \gamma \leq x) = P(X \leq x + \gamma) = F(x + \gamma)$$

Soit g la fonction de densité de la loi log-normale tronquée. Alors :

$$g(x) = f_{ln}(x + \gamma) = \frac{1}{\sqrt{2p.s.(x + \gamma)}} \exp\left(-\frac{1}{2}\left(\frac{\ln(x + \gamma) - m}{s}\right)^2\right).$$

IV.2.1 - Recherche des moments de la loi log-normale

Si l'on dispose d'un n -échantillon (x_1, x_2, \dots, x_n) issu de la loi de X , grâce aux travaux de Cohen, on a pu montrer qu'il est alors possible d'estimer les deux premiers moments de la loi de X , m et s^2 (cf. le mémoire de B. Valleaux) :

$$\hat{m} = \bar{y} - Q(w)(\bar{y} - \ln \gamma)$$

$$\hat{s}^2 = S_Y^2 + Q(w)(\bar{y} - \ln \gamma)^2$$

où :

- Y , qui suit la loi normale, est telle que $Y = \ln(X)$,

- $y_i = \ln(x_i)$,

- $\bar{y} = \frac{1}{n} \sum_{i=1}^n \ln x_i$,

- $S_Y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$,

- $w = \frac{S_Y^2}{(\bar{y} - \ln \gamma)^2}$,

- et $Q(w) = \frac{Z(\hat{u})}{Z(\hat{u}) - \hat{u}}$, où :

- $Z(x) = \frac{f(x)}{1-F(x)}$, $f(x)$ et $F(x)$ désignant respectivement la fonction de densité et la fonction de répartition de la loi normale centrée réduite,

- $\hat{u} = \frac{\ln \hat{\xi} - \hat{m}}{s^2}$ est solution de l'équation :

$$\frac{1 - Z(u)(Z(u) - u)}{(Z(u) - u)^2} = w .$$

ω est un coefficient qui va nous permettre de résoudre numériquement les équations du maximum de vraisemblance.

Un programme de résolution de cette équation se trouve dans le mémoire de B. Valleaux et est mis à profit pour calculer les paramètres de la loi log-normale et pour procéder à des tests d'adéquation de la distribution du montant des sinistres à la loi log-normale tronquée, grâce au calcul de fréquences théoriques par tranche de coût.

Nous en déduisons alors l'estimateur de l'espérance d'une loi log-normale de paramètres (m, σ^2) , égal à :

$$\hat{E}(X) = \exp\left(\hat{m} + \frac{\hat{s}^2}{2}\right)$$

Comme les z_i , $i=1, \dots, n_z$, sont les montants des n_z sinistres présents dans la base "sinistres", on a la relation suivante liant les z_i et les x_i :

$$x_i = z_i + g$$

Si l'on désigne par C_z le montant total des indemnisations :

$$C_z = \sum_{i=1}^{n_z} z_i$$

alors on peut établir la relation suivante entre la charge totale des indemnisations à franchise nulle (C) et la charge des indemnisations réellement effectuées (C_z) :

$$C = C_z + n_z \cdot g + n \frac{\xi}{2}, \text{ où :}$$

- $n_z \cdot \xi$ est le total du montant des sinistres restant à la charge de l'assuré,
- $n \frac{\xi}{2}$ est l'estimation de la charge des n sinistres non réglés car d'un montant inférieur à la franchise, le coût moyen des sinistres étant estimé à la moitié d'une franchise ($\frac{\xi}{2}$).

De même, puisque l'on a :

$$\hat{E}(X) = \frac{C_z + n_z \cdot g + n_g \cdot \frac{\varepsilon}{2}}{n_z + n_g}$$

on en tire une estimation du nombre n_g de sinistres générés par le passage d'une loi tronquée (Z) à une loi non tronquée (X) :

$$\hat{n}_g = \frac{C_z + n_z \cdot (g - \hat{E}(X))}{\hat{E}(X) - \frac{g}{2}}$$

On dispose donc d'une relation nous précisant le surcoût occasionné par la suppression de la franchise, et le nombre de sinistres générés en résultant.

Néanmoins, cette relation n'a de sens que pour certaines valeurs de $E(X)$, ce que le tableau 2.2.8 présente, à travers les variations de la fonction h :

$$h : E(X) \rightarrow n_g$$

Tableau 2.2.8 : Etude des variations de la fonction $h : E(X) \rightarrow n_g$

$E(X)$	0	$\frac{\varepsilon}{2}$	$\frac{C_z}{n_z} + g$
n_g	$-2(\frac{C_z}{g} + n_z) < 0$	$+\infty$	0

Comme on le voit, cette fonction a un point de non continuité lorsque l'espérance de X est égale à la moitié de la valeur de la franchise. On voit d'autre part que n_g n'est positive que si l'espérance $E(X)$ fait partie de l'intervalle formé par le montant de la demi-franchise ($\frac{g}{2}$) et le montant moyen du règlement d'un sinistre augmenté du montant de la franchise ($\frac{C_z}{n_z} + g$). En outre, il faut ajouter que la pertinence de l'estimation de n_g s'estompe au fur et à mesure que $E(X)$ se rapproche de $\frac{g}{2}$, le nombre estimé s'envolant démesurément.

Dans le cas où l'estimation précédente diverge ou présente un nombre négatif, l'évaluation du nombre de sinistres d'un montant de préjudice inférieur à la

franchise repose sur l'échantillon des sinistres de montant z_i dont on estime le nombre présentant un montant inférieur à la franchise ?.

IV.3 - Modalités pratiques du calcul

Après avoir présenté les aspects théoriques de l'estimation du montant global de l'indemnisation des assurés dans l'hypothèse d'une annulation de la franchise, il convient ensuite d'en assurer la réalisation pratique. Celle-ci requiert la mise en œuvre d'un certain nombre d'opérations successives :

- la recherche de la formule du montant des franchises,
- le regroupement des branches-risques selon le montant des franchises,
- la détermination du niveau de finesse auquel doit s'effectuer le calcul,
- et, enfin, la détermination de la structure de redistribution de la charge supplémentaire.

IV.3.1 - Recherche de la formule du montant des franchises

La première étape réside dans la collecte d'informations relatives à la formule permettant de retrouver le montant des franchises tempête. Cette recherche utilise les mêmes filons que ceux retenus lors de l'enquête relative à la présence, facultative ou obligatoire, ou à l'absence de la garantie tempête (cf. le paragraphe III.3), à savoir :

- l'exploitation des bases informatiques,
- l'exploitation de la base documentaire.

IV.3.1.1 - Exploitation des bases informatiques

Une première piste de recherche réside dans *l'exploitation des bases informatiques* mises à notre disposition, où se trouvent consignées des informations relatives à l'application de la franchise lors du règlement d'un sinistre tempête. (Ces bases se distinguent de notre base " sinistres ", où la franchise ne figure pas).

Il est à noter que ces montants résultent de l'application pratique d'une formule théorique, ce qui explique l'existence d'éventuelles divergences entre le montant de franchise constaté et celui que l'on est en droit d'attendre par l'application d'une formule. Ces divergences tiennent à plusieurs facteurs :

- l'application réitérée d'une franchise à plusieurs biens sinistrés couverts par une même garantie tempête,
- l'absence de retranchement de la franchise,
- la mauvaise application de la formule de la franchise.

En dehors de ces cas, il est toujours possible de retrouver le montant de la franchise qui a été appliquée lors du règlement des sinistres de la base " sinistres ". Ceci requiert au préalable de retrouver les formules de chaque garantie tempête. Pour cela, il convient de tirer partie des informations contenues dans la base documentaire des contrats de l'UAP.

IV.3.1.2 - Exploitation de la base documentaire

Comme nous l'avons déjà vu, la base documentaire de l'UAP relative aux contrats est très importante, ce qui explique souvent que l'édition de certains documents anciens ne soit plus possible. Néanmoins, il faut se mettre en quête des documents qui ont régi le règlement des tempêtes depuis 1980 jusqu'à aujourd'hui. Comme nous l'avons vu dans le chapitre 4 de la première partie, paragraphe II.1, il est classique de distinguer trois périodes dans l'évolution du montant des franchises :

- avant 1983, époque pendant laquelle les franchises représentaient près de 20 % des dommages indemnisés,
- entre 1983, date de pollicitation relative à une nouvelle garantie tempête, et 1990, période durant laquelle les franchises représentent en général un plus faible pourcentage,
- en 1990, une des conséquences de la nouvelle garantie tempête universelle est le remplacement de franchises auparavant *proportionnelles* en franchises à *montant fixe*.

A ce sujet, il convient peut-être de préciser la nature des franchises rencontrées. L'on note :

- *F* le montant de la franchise,
- *R* le règlement effectué par la compagnie d'assurances,
- *P* le préjudice subi par l'assuré.

On trouve ainsi deux types de franchises à l'UAP, la franchise *fixe* et la franchise *proportionnelle*.

La franchise fixe se caractérise par le programme suivant :

$$\begin{cases} R = 0 & \text{si } P < F \\ R = P - F & \text{si } P \geq F \end{cases}$$

La franchise proportionnelle fait intervenir trois paramètres : un minimum (*Min*), un maximum (*Max*) et un taux *t* :

$$\begin{cases} R = 0 & \text{si } C \leq \text{Min} \\ R = P - \text{Min} & \text{si } \text{Min} \leq P \leq \frac{\text{Min}}{t} \\ R = (1-t)P & \text{si } \frac{\text{Min}}{t} \leq P \leq \frac{\text{Max}}{t} \\ R = P - \text{Max} & \text{si } P \geq \frac{\text{Max}}{t} \end{cases}$$

D'autres types de franchises existent (franchise fixe relative, franchise mixte par exemple) mais elles n'ont pas cours à l'UAP.

La liste des documents collectés figure dans le paragraphe précédent : on pourra donc s'y reporter pour connaître leur nature. Il est à noter que certains documents, telles les notes d'instructions, peuvent annuler et remplacer d'autres documents plus anciens, à l'occasion par exemple de grandes réformes comme la pollicitation de 1983, les nouveaux textes étant généralement plus favorables à l'assuré. Le succès remporté par cette opération - plus de 98 % des contractants ont opté pour cette nouvelle garantie - nous conduit à choisir donc systématiquement la formule la plus récente. La date du règlement du sinistre nous permet de déterminer la formule de garantie appliquée, qui l'emportera donc sur celle du type de contrat sinistré.

Mais l'examen de dossiers de sinistres tempête, destiné à évaluer le montant des différents éléments d'exclusion et évoquer au paragraphe V, nous a permis de remarquer des erreurs dans l'application des franchises, certaines instructions ayant été oubliées : il en résulte quelquefois le choix d'une formule de franchise effectivement présente dans le contrat d'origine, et qu'une garantie annexée au contrat vient contredire. On peut aussi rencontrer des situations où l'assuré, pour ne pas payer de surprime, a refusé toutes les améliorations des conditions d'application de la garantie tempête, mais se voit malgré tout accorder la formule la plus avantageuse.

Comme précédemment, les divers documents collectés font référence à un produit souvent désigné par deux ou trois lettres : MH désigne par exemple le produit Multirisque Habitation lancé en 1971. Ce sigle se retrouve en général dans les fichiers informatiques, ce qui permet d'établir une table de correspondances entre le type de produit et la subdivision très fine des produits IARD de l'UAP, la b ranche-risque, et d'affecter à cette dernière la formule qui lui revient.

Au final, une dizaine de textes précisant les dispositions en matière de franchise sont recensés : ils sont souvent communs à plusieurs produits ou peuvent même provenir d'instances professionnelles, telles les intercalaires P13 et P15.

IV.3.2 - Regroupement des produits selon le montant minimal des franchises

Après avoir déterminé la formule de la franchise de chacune des 340 branches-risques rattachées à la branche incendie, il convient de réaliser l'étape suivante : réunir ces branches-risques en quelques groupes homogènes du point de vue du niveau de la franchise. *Il est possible de réunir des produits ayant des formules de franchise différentes à la condition que le montant minimal restant à la charge de l'assuré soit commun.* La difficulté la plus sérieuse résulte de l'existence de nombreuses franchises ayant une valeur minimum différente.

Considérons une courbe de distribution du montant des sinistres avant l'application de la franchise, à laquelle on applique l'un des deux types de franchises présents à l'UAP (franchises fixe et proportionnelle) et dont le montant minimal de la franchise proportionnelle (*Min* selon les notations précédentes) est identique au montant de la franchise fixe (*F*). Alors, lorsque l'on compare les deux courbes de distribution du montant des sinistres déduction faite de la franchise, on peut constater que la première courbe est plus incurvée vers la droite, ce qui signifie que le montant des indemnités versées aux assurés sinistrés est plus conséquent. *Pour autant, nous ferons l'hypothèse que le mélange de ces sinistres différents du point de vue de la franchise est admissible.* D'ailleurs, ce n'est que lorsque le montant du sinistre avant application de la franchise dépasse $\frac{Min}{t}$ que les deux courbes se séparent.

Il est à noter que, lors du classement de certaines tempêtes en catastrophe naturelle, un débat s'est instauré pour déterminer quelle franchise (tempête ou catastrophe naturelle) était applicable. En général, la franchise prévue par la loi s'applique, même si le bien assuré est couvert avec une franchise plus favorable pour l'assuré par la garantie tempête. Une particularité peut toutefois se présenter lorsque l'arrêté dispose que, "à titre exceptionnel", la garantie contractuelle souscrite (i.e. la garantie tempête) s'applique en premier lieu, la clause légale n'intervenant qu'en complément. Ainsi, dans le cas de dommages à des habitations consécutifs à une tempête, la prise en charge de ces dommages étant supposée identique en garantie contractuelle et en garantie légale, et la franchise légale étant de 800 Frs (à l'époque), la franchise supportée par l'assuré sera la suivante :

- 800 Frs si la franchise contractuelle est d'un montant supérieur,
- la franchise contractuelle si celle-ci est inférieure à 800 Frs.

IV.3.3 - Degré de finesse des calculs

Comme il a été suggéré plus haut, une branche-risque peut changer de formule de franchise, au gré des réformes ou de l'évolution des couvertures d'assurances. Il est donc nécessaire de calculer le surcoût et le surnombre occasionnés par la mise à zéro de la franchise selon une rupture permettant de repérer ces grandes évolutions : le choix de l'exercice de survenance du sinistre paraît ainsi tout indiqué.

De même, les groupes de branches-risques partageant un même montant minimal de franchise ont été constitués. Compte tenu de la prise en compte de la dimension temporelle, la composition de ces ensembles peut tout à fait évoluer d'un exercice à l'autre, même si l'on a préféré, autant que faire se peut, privilégier leur homogénéité tout au long de leur existence.

Une remarque importante est à apporter : ces ensembles doivent toujours représenter une subdivision des trois risques d'habitation, professionnel et agricole, sous peine de ne pouvoir ensuite redistribuer les surcoût et surnombre pour cette rupture.

Enfin, il n'a pas été jugé utile de rechercher une adéquation des distributions du montant des sinistres à une loi log-normale au niveau de chaque station météorologique : le nombre élevé de stations retenues, près de 130, et l'absence a priori de disparité régionale en terme de politique de franchise expliquent notre choix.

En résumé, il est décidé que le calcul du surcoût et du surnombre générés dans l'hypothèse de la suppression des franchises s'effectue selon la structure croisée de l'exercice de survenance et d'une segmentation des produits permettant une agrégation par risque.

IV.3.4 - Structure de redistribution de l'indemnité à niveau de franchise nul

Comme cela a été souligné dans le paragraphe III, la structure de redistribution de l'indemnité supplémentaire résultant de l'abolition des franchises dépend étroitement :

- du degré de finesse de calcul de ce surcoût (par exercice de survenance et regroupement de branche-risque),
- et du niveau d'agrégation des variables présentes dans le modèle liant les différentes bases (par risque, événement et station météorologique).

La structure de redistribution de l'indemnité supplémentaire se fait donc par exercice de survenance et par risque, ce qui sous-entend la validité de deux hypothèses :

- d'une part, l'égalité entre la somme des surcoûts indemnitaires, calculés selon la théorie présentée au paragraphe IV.2, pour chaque tempête d'un exercice de survenance, et le surcoût calculé selon la même théorie mais en regroupant l'ensemble des sinistres d'un exercice de survenance,
- d'autre part, l'équivalence des caractéristiques en matière de franchise entre les différentes zones météorologiques auxquelles sont rattachés les sinistres.

En outre, il est souhaitable d'apporter une modification à la méthode, présentée au paragraphe IV.2, présidant à l'estimation de la part de l'indemnisation restant à la charge de l'assuré sinistré. Dans ce procédé, cette surcharge est estimée à $n_z \cdot \gamma$. Or, si cette évaluation est exacte pour une formule de franchise fixe, il n'en est pas de même lorsque la franchise relève d'une formule proportionnelle puisque alors son montant est égal à $t.P$ tout en variant entre *Min* et *Max* (cf. le paragraphe IV.3.1.2). Puisque chaque branche-risque se voit affecter une formule de franchise, un calcul plus réaliste du montant de la franchise appliquée à chaque sinistre est tout à fait envisageable.

Le tableau 2.2.9 donne, pour chaque risque, une idée du montant représenté par les franchises relativement au niveau d'indemnisation avant l'application de la franchise.

Tableau 2.2.9 : Taux de franchise, par risque, relativement au montant de l'indemnisation avant l'application de la franchise

Risque	Agricole	Habitation	Entreprise	Total
Taux de franchise	24,6 %	8,2 %	10,8 %	12,2 %

V - Actualisation liée à l'évolution du niveau d'exclusion

V.1 - Importance du calcul d'un indice d'exclusion

Le paragraphe II du chapitre 4 de la première partie a présenté de manière détaillée l'évolution des garanties contractuelles contre les effets du vent au cours des 30 dernières années. Et force est de constater que cette évolution a été particulièrement marquée par la volonté affichée par l'assureur d'améliorer ses relations avec son client lors d'un sinistre. Ce souci d'offrir un meilleur service s'est traduit par des actes concrets et réitérés, d'amélioration de la couverture offerte par la garantie contre les effets du vent. Ainsi :

- en 1983, la pollicitation est l'occasion de proposer une garantie tempête renouvelée, où la limite de garantie et la double vétusté disparaissent, de même que les franchises diminuent,
- en 1990, une nouvelle garantie met fin à l'application de la vétusté à hauteur de 25 % et surtout aux nombreuses exclusions qui subsistaient.

Une évidence apparaît nettement : on a assisté entre 1980 et 1997 à une réduction sensible du *niveau moyen d'exclusion*, c'est-à-dire du rapport du montant moyen des éléments constitutifs du préjudice au montant du préjudice. A cet effet, 1990 constitue une année charnière avec l'opération volontaire menée par l'UAP en octobre 1990 d'amélioration de la garantie tempête, et dont une des principales mesures consiste à intégrer dans le champ de la garantie des éléments autrefois exclus (stores, clôtures, antennes de télévision, etc...).

Une interrogation apparaît tout naturellement, à savoir le niveau et l'impact sur la charge des sinistres de ces mesures. En effet, les exclusions constituent un des éléments du préjudice occasionné par la tempête aux biens assurés.

V.2 - Méthode de calcul de l'indice d'exclusion

V.2.1 - Echantillonnage des sinistres tempête

Pour essayer d'estimer l'impact, sur le niveau moyen d'exclusion, de la suppression de nombreuses exclusions sur la garantie tempête, une solution réside dans l'examen de dossiers sinistres réglés en 1997 : en effet, les sinistrés qui ont fait l'objet d'une indemnisation récente suite à une tempête bénéficient tous d'une garantie très favorable, leur permettant de se voir rembourser des éléments exclus du champ de la garantie avant 1990. Le chiffrage de ces exclusions, dont le détail nous est fourni par le rapport d'expertise, autorise ensuite à en voir l'importance à travers leur fréquence et leur coût moyen.

L'hypothèse, consistant à considérer que, sur la période 1980 - 1997, la composition des éléments sinistrés n'a pas subi d'évolution marquée doit être ici faite. Ceci suppose de considérer que :

- les éléments exclus hier ne sont pas disparus aujourd'hui (par exemple, cela revient à faire l'hypothèse que les antennes de télévision restent toujours aussi présentes sur les toitures de nos jours qu'en 1980),
- les tempêtes observées entre 1995 et 1997 étaient « représentatives » de l'ensemble des tempêtes de l'étude, relativement à la part représentée par les éléments d'exclusion sur le montant total du préjudice.

Cette conjecture étant posée, il convient ensuite de procéder à un tirage au sort de 200 numéros de dossiers de sinistres présentant les particularités suivantes :

- ils ont donné lieu à une indemnisation à l'assuré,
- un expert de compagnie a rédigé un rapport d'expertise, document dont l'analyse est très enrichissante,
- le sinistre a été réglé récemment.

En l'absence de renseignements précis, le sondage est choisi à probabilités égales sans remise. Le nombre de dossiers répondant aux caractéristiques précédentes est d'environ 4200 ce qui porte le taux de sondage à environ 5 %. L'utilisation sous SAS de la fonction UNIFORM() dans l'opération sélective suivante :

« IF UNIFORM(43724) < 200/4200 ; »

permet d'obtenir un échantillon de numéros de 200 sinistres normalement archivés, c'est-à-dire stockés dans l'un des centres prévus à cet effet avant l'envoi au pilon.

V.2.2 - Grille de lecture des dossiers sinistres

La sélection aléatoire des dossiers tempête donne ensuite lieu à l'analyse du contenu de ces dossiers dans l'optique de déterminer la part des anciennes exclusions sur le montant du préjudice.

Pour cela une grille de lecture de chaque dossier est mise en place, de manière à guider la personne en charge de l'analyse de leur contenu, et faire en sorte qu'un dossier soit décrypté de manière similaire par deux personnes différentes. Par exemple, le thème « type d'exclusion » contient un ensemble de rubriques, symbolisées par une ou deux lettres, qui constituent une partition des éléments anciennement exclus, à savoir :

- antenne (A),
- clôture (C),
- enseigne publicitaire (E),
- fil aérien (F),
- gouttière - chéneau (G),
- mobilier (M),
- panneau solaire (P),
- store (S),
- vitrage (VI),
- volet (V).

La catégorie « Autres » est proscrite car elle est trop souvent utilisée par facilité lorsque le rapport d'expertise utilise un synonyme d'un des termes précédents mais non repris dans cette liste, alors que l'assimilation à l'une de ces catégories aurait pu convenir.

Il est alors porté sur la grille de lecture, pour chaque sinistre, l'éventuelle « exclusion » symbolisée par sa (ou ses) lettre(s) ainsi que son montant.

V.2.3 - Constitution d'un tableau des exclusions par produit

Puis vient la phase de détermination des exclusions caractéristiques de chaque produit. En effet, il faut se souvenir que les produits d'un même risque qui se succèdent diffèrent souvent par la composition de la liste des exclusions de la garantie tempête. Cette différence existe bien sûr entre des produits appartenant à des risques différents. Leur niveau d'exclusion est donc forcément différent.

Le tableau 2.2.10 donne à titre d'illustration les dispositions retenues pour quelques produits standards et quelques types d'exclusions.

Tableau 2.2.10 : Exemples des dispositions prises en matière d'exclusions pour quelques produits UAP

Produit	clôture (C)	antenne (A)	volet persienne (VO)	chéneau gouttière (G)	élément vitré (VI)
C.H.LA-LM-PA-PM	couverte	exclue	couvert	couvert	exclu
M.H.	exclue	exclue	exclu sauf si bâtiments détériorés	exclu sauf si bâtiments détériorés	exclu sauf si bâtiments détériorés
H.M.	exclue	exclue	exclu sauf si bâtiments détériorés	exclu sauf si bâtiments détériorés	exclu sauf si bâtiments détériorés
M.P. 85	couverte	exclue	exclu sauf si bâtiments détériorés	exclu sauf si bâtiments détériorés	exclu sauf si bâtiments détériorés
R.I.	exclue	exclue	exclu sauf si bâtiments détériorés	exclu sauf si bâtiments détériorés	exclu sauf si bâtiments détériorés
M.C.	exclue	exclue	exclu sauf si bâtiments détériorés	exclu sauf si bâtiments détériorés	exclu sauf si bâtiments détériorés
C.A.	exclue	exclue	exclu sauf si bâtiments détériorés	exclu sauf si bâtiments détériorés	exclu sauf si bâtiments détériorés

Ensuite, il convient, à partir des résultats contenus dans ce tableau et du poids estimé de chaque exclusion, d'évaluer le niveau d'exclusion de l'ensemble des produits. Le poids estimé de chaque élément d'exclusion s'obtient en rapportant le montant de cette exclusion au montant total du préjudice subi, sur l'ensemble des sinistres sondés. Le niveau d'exclusion moyen d'un produit est alors la somme des poids effectuée sur les seuls éléments d'exclusion, à l'exclusion des éléments garantis. Par suite, le calcul d'un niveau d'exclusion par risque s'effectue en pondérant le niveau d'exclusion de chaque produit par le nombre de contrats qu'il représente relativement au total du risque considéré.

L'hypothèse sous-jacente sur laquelle reposent nos calculs est l'équivalence des poids des éléments d'exclusion, quel que risque que l'on considère. Cette supposition, qui consiste à établir une seule hiérarchie des éléments d'exclusion, valable pour chacun des risques, peut paraître osée. Elle est néanmoins dictée par la réalité du sondage, vu le faible nombre de sinistres de l'échantillon, qui rend impossible le calcul du niveau d'exclusion à un niveau de finesse plus grand.

Le tableau 2.2.11 reprend ces chiffres tirés du sondage aléatoire, et ventilés par risque.

Tableau 2.2.11 : Estimation du poids de chaque élément d'exclusion extrapolé d'un sondage

Type d'exclusion	Agricole	Habitation	Entreprise	aucun
Antenne	-	1,1 %	-	
Clôture	2,8 %	2,8 %	2,8 %	
Enseigne		2,6 %	2,6 %	
Fil aérien				x
Gouttière	1,3 %	1,3 %	1,3 %	
Mobilier	1,7 %	1,7 %	1,7 %	
Panneau solaire				x
Store				x
Vitrage		0,1 %		
Volet		3,8 %		

V.3 - Structure de redistribution du montant induit par l'indice d'exclusion

L'étape précédente nous a permis d'estimer le taux d'exclusion pour chacun des trois risques considérés. De même, la déclinaison de ce taux selon l'exercice de survenance du sinistre est souhaitable dans la mesure où la quasi totalité du portefeuille incendie se voit attribuer au cours de 1990 une nouvelle garantie.

En revanche, il ne semble pas utile de décliner le taux d'exclusion selon un axe supplémentaire car cela aboutirait à une structure croisée très fine comportant un grand nombre de croisements d'effectifs trop faibles. En outre, cela nécessiterait alors des moyens trop importants pour un résultat très aléatoire : en effet, il ne semble pas à première vue exister de disparités en terme de taux d'exclusion entre les différentes régions de France, ou entre les tempêtes d'un même exercice de survenance, qui nécessitent une recherche complémentaire.

Le calcul du taux d'exclusion se limite donc à une ventilation par exercice de survenance et risque. La structure de redistribution du surcoût généré par l'abolition des exclusions reste identique à celle déjà décrite au paragraphe précédent, c'est-à-dire qu'elle est par exercice de survenance et par risque. Les agrégats utilisés dans le modèle économétrique du chapitre 4 de cette partie étant sommés par zone météorologique, risque et tempête, on supposera donc que le niveau d'exclusion ne diffère pas selon les tempêtes d'un exercice et selon les zones météorologiques considérées.

VI - Actualisation liée à l'évolution du taux de vétusté appliquée

VI.1 - Intérêt du calcul d'un indice de vétusté

Au cours de la première partie, au paragraphe II du chapitre 4, nous avons évoqué les deux grandes dates qui ont marqué l'évolution dans l'application d'un taux de vétusté (i.e. rapport des montants de la vétusté au préjudice) sur le montant des dommages dus à la tempête :

- d'une part, en mars 1983, où une nouvelle garantie massivement souscrite met un terme à la règle de la double vétusté, lorsque la vétusté dépasse 40 %,
- d'autre part, en octobre 1990, où la nouvelle garantie tempête, proposée à l'ensemble du portefeuille - à l'exception de la branche agricole - à l'occasion d'une opération de révision, permet le remboursement différé, à hauteur de 25 %, du montant de la vétusté après la réalisation effective des travaux de réparation du bien endommagé.

Pour la branche agricole, il a fallu attendre le mois de juin 1993 pour que les dégâts causés aux bâtiments d'habitation sous toiture distincte des bâtiments d'exploitation soient indemnisés sans déduction de la vétusté. En revanche, tous les autres types de bâtiments (agricoles ou pour lesquels on a sous un même toit les bâtiments agricoles et d'habitation) se voient encore appliquer aujourd'hui un éventuel taux de vétusté.

Le *paiement différé* du montant de la vétusté, plafonné à un taux de 25 %, peut être versé dans les deux ans suivant le premier règlement. Son institution vise à affecter

réellement le montant des indemnités à l'entretien des biens endommagés. Cette nouvelle disposition du règlement de la tempête, figurant depuis longtemps dans le règlement des sinistres incendie, a pour but d'assainir le marché et d'éviter que les zones sinistrées par la tempête restent à jamais défigurées si d'aventure les bâtiments endommagés n'étaient pas restaurés.

Si l'on fait l'inventaire des différentes situations de règlement d'un sinistre tempête prévalant après 1990, en fonction de la valeur du taux de vétusté et de l'éventualité de la survenance d'un règlement différé, on aboutit au tableau 2.2.12 qui suit. Ce dernier fait référence aux numéros des cas de règlements chiffrés explicités dans le tableau 2.2.13.

Tableau 2.2.12 : Présentation des situations de règlements en tempête après 1990 selon le niveau de la vétusté

Règlement différé	Vétusté		
	nulle	inférieure à 25 %	supérieure à 25 %
effectué	cas 5	cas 1	cas 2
non effectué	sans objet	cas 3	cas 4

Même si les cas 3 et 4 prévalaient avant 1990, l'analyse des dossiers sondés, menée au paragraphe suivant, prouve qu'on peut encore rencontrer ces situations de règlements dans les dossiers tempête actuels.

Tableau 2.2.13 : Exemple chiffré des situations de règlements en tempête après 1990 selon le niveau de la vétusté

cas	Montant du préjudice	Montant de la vétusté	Montant de l'indemnité avant abattement dû à la franchise
1	100	20	100
2	100	30	95
3	100	20	80
4	100	30	70
5	100	0	100

De même, il importe de calculer un indice correctif lié à l'application, avant 1983, de la règle de la double vétusté. Une note du 27 août 1975, émanant du Secrétariat Technique du Département Incendie - Dommages de l'UAP et adressée au directeur de la délégation du sud-ouest, précise les modalités de règlements lors de l'application de cette règle :

« Lorsque le taux réel de vétusté est supérieur à 40 %, cette indemnité doit être calculée (...) après l'abattement correspondant au taux réel de vétusté, en appliquant au chiffre ainsi obtenu, le rapport existant entre 40 % et le taux réel de vétusté, soit :

Dommages.....	10.247,48
Vétusté 60 %.....	<u>6.148,48</u>
Reste.....	4.099,--
 Indemnité après abattement :	
$4.099 \times \frac{40}{60} =$	<u>2.732,--</u> »

Le même sinistre, indemnisé aujourd'hui, l'aurait été à un montant de 6.660,81 Frs de l'époque, soit près de 2,5 fois plus qu'avant 1983 ! Ce simple exemple montre l'intérêt du calcul d'un indice correctif.

VI.2 - Etude de la distribution du taux de vétusté

Le sondage présenté au paragraphe V n'a pas livré tous ses secrets : il peut encore être exploité pour le calcul du *taux de vétusté* dans les conditions actuelles de paiement des sinistres tempête qui servira d'estimation pour l'ensemble de la période de l'étude. Il faut en effet savoir que cette information n'est pas disponible dans les fichiers informatiques centraux.

Contrairement au paragraphe précédent pour le taux moyen d'exclusion, on ne peut se satisfaire de la seule connaissance de la moyenne du taux de vétusté. Il est nécessaire aussi de connaître sa dispersion, de manière à avoir une juste vision de la part des sinistres pour lesquels la vétusté dépasse le seuil de 25 % de la valeur vénale. Il est intéressant aussi de mesurer la part de sinistrés pour lesquels un règlement différé a été effectué, en raison de la présentation de factures prouvant la réalité des réparations : une situation où la part serait proche de zéro nous ramènerait au contexte qui prévalait avant 1990. La combinaison de ces deux effets (limitation du versement de la vétusté à 25 % et absence de règlement différé) nous conduit à définir le *taux de règlement*, c'est-à-dire le rapport du montant de l'indemnité *réglée* avant abattement dû à la franchise (*MTREG*) au montant du préjudice (*MTPREJ*),

La figure 2.2.2 représente l'histogramme de distribution du taux de vétusté, tous risques confondus compte tenu du faible nombre de sinistres sondés (84).

L'histogramme de distribution du taux de vétusté présente plusieurs modes :

- un, très grand, correspondant à un taux de vétusté appliquée de 0 %, ce qui représente environ 17 % des sinistres sondés; il s'agit principalement de dommages sur des bâtiments neufs ou restaurés , ou encore de petits sinistres.

Mais il n'est pas à exclure qu'une mauvaise application des textes soit à l'origine de ce taux nul : l'annulation du taux de vétusté peut être un moyen de les détourner en ajoutant au premier paiement le règlement différé, sans attendre la réparation effective des travaux. Mais cette pratique condamnable ne peut être qu'exceptionnelle,

- d'autres, plus modestes, correspondant à des nombres multiples de 5 : 10 %, 15 %, 20 % et 25 %. C'est pour l'expert de la compagnie une échelle de mesure suffisante, car l'estimation du taux de vétusté au pourcentage près n'est pas réaliste. D'ailleurs, à part le taux à 12 %, il n'y a pratiquement pas de valeur comprise entre les multiples de 5 % qui résulte presque toujours d'une combinaison de plusieurs taux.

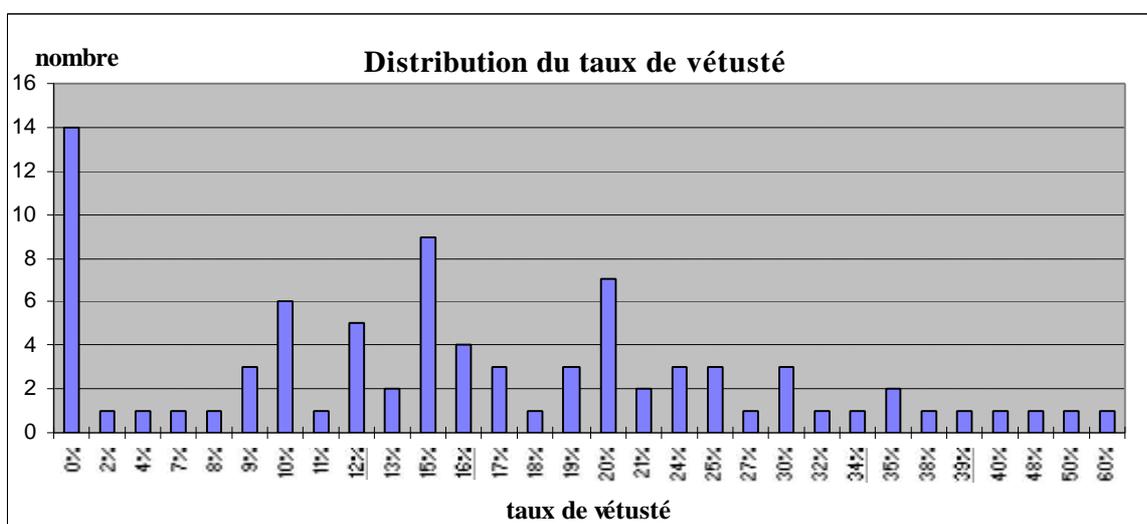


Figure 2.2.2 : Histogramme de distribution du taux de vétusté appliquée

La part de dossiers dont la vétusté dépasse 25 % est d'environ 17 %. Pour ces sinistrés, un abattement égal au pourcentage de différence entre le taux de vétusté réel et 25 % sera appliqué à leur préjudice.

Quant à la moyenne du taux de vétusté réelle, son estimation résulte du rapport entre le montant total de la vétusté et de la somme des préjudices des dossiers sondés. L'intervalle de confiance à 95 % de cet estimateur \bar{y} s'obtient, dans l'hypothèse de sa normalité asymptotique :

- en estimant la variance de cet estimateur par la quantité calculée sur l'échantillon :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \text{ qui estime sans biais la variance corrigée de la population}$$

de référence d'effectif N :

$$S^2 = \frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})^2,$$

- en formant la quantité $\bar{y} \pm 2.s.\sqrt{\frac{1-f}{n}}$, où f est le taux de sondage.

Le calcul numérique donne : 16,0 % ± 2,6 %, tous risques confondus. Le tableau 2.2.14 nous renseigne sur le taux de vétusté réelle par risque.

Tableau 2.2.14 : Estimation du taux de vétusté par risque extrapolé d'un sondage

Risque	Agricole	Habitation	Entreprise	Total
Nombre de dossiers	16	53	15	84
Taux de vétusté	18,8 % ± 6,7%	15,9 % ± 3,3 %	13,4 % ± 4,9 %	16,0 % ± 2,6 %

VI.3 - Etude du taux de règlement

Compte tenu du taux de vétusté précédent, et du règlement de la vétusté pour les risques professionnels et d'habitation et pour les dommages aux bâtiments d'habitation pour le risque agricole, on pourrait en déduire que la conséquence d'une telle mesure est le renchérissement, à hauteur du règlement de la vétusté, du coût des sinistres.

Or, contrairement à ce que l'on pourrait penser, le rapport du montant effectif de l'indemnisation réglée à l'assuré avant abattement dû à la franchise (*MTREG*), sur le montant du préjudice (*MTPREJ*), mesurant *le taux de règlement*, s'établit à 87,2 %. Le tableau 2.2.15 détaille les résultats selon les risques.

Tableau 2.2.15 : Estimation du taux de règlement par risque extrapolé d'un sondage

Risque	Agricole	Habitation	Entreprise	Total
Nombre de dossiers	16	53	15	84
Taux de vétusté	79,1 % ± 8,8 %	89,9 % ± 3,6 %	86,6 % ± 6,0 %	87,2 % ± 3,1 %

VI.3.1 - Cas des risques professionnel et d'habitation

La faiblesse de ce taux par rapport au taux maximum d'indemnisation possible, égal à 100 % pour les risques professionnel et d'habitation, s'explique par la combinaison de deux effets :

- d'une part, l'effet de plafond agissant lorsque le taux de vétusté dépasse 25 %,
- d'autre part, l'absence éventuelle du règlement différé.

Ainsi, pour le risque habitation, le montant de la vétusté, qui représente 15,9 % du préjudice, n'est réglé qu'à hauteur de 5,8 % du préjudice. En effet, si l'on désigne par $TXVET$ le taux de vétusté réelle et par $TXVETREG$ le taux de vétusté réglée, alors l'on a :

$$TXVETREG = TXREG - (1 - TXVET) = 89,9 \% - (100 \% - 15,9 \%) = 5,8 \%$$

Il se trouve donc que, à partir de 1990, la compagnie ne règle dans les faits qu'un montant édulcoré de la vétusté, égal à :

$$MTVETREG = TXREG \cdot MTPREJ - MTINDSUP,$$

inférieur à l'abattement effectué avant cette date et calculé sur la base de la vétusté réelle :

$$MTVET = MTPREJ - MTINDSUP$$

$MTINDSUP$ étant le niveau d'indemnisation prévalant avant 1990 et avant abattement dû à la franchise.

Si l'on veut maintenant connaître le supplément de charge que le paiement de cette vétusté atténuée représente, on forme le rapport :

$$INDVET = \frac{MTVETREG}{MTINDSUP} = \frac{5,8}{100 - 15,9}.$$

Ce taux, égal à 6,9 % en habitation et à 0 % pour le risque professionnel, s'interprète comme le supplément de charge de sinistres consécutif à la prise en compte de la vétusté à hauteur de 25 % et l'institution d'un règlement différé dans l'indemnisation des contrats sinistrés.

VI.3.2 - Cas du risque agricole

D'autre part, il faut apporter quelques explications concernant les modalités de règlement des sinistres tempête pour le risque agricole. En effet, ces règles diffèrent sensiblement de celles qui prévalent depuis 1990 pour les deux autres risques. Le risque agricole, qui n'entraîne pas dans le champ d'application de l'opération de révision

de la garantie tempête de 1990, n'a connu sa véritable réforme qu'en juin 1993, lorsque les mesures suivantes ont été étendues à l'ensemble du portefeuille :

- les bâtiments d'habitation sous toiture distincte des bâtiments d'exploitation font l'objet d'une indemnisation identique à celle prévue en incendie : le remboursement de l'éventuelle vétusté des habitations est prévu,
- en revanche, le montant de réparation de tous les autres bâtiments, qu'ils soient agricoles ou d'habitation mais sous la même toiture, se voit amputé du montant de la vétusté,
- enfin, le règlement de l'indemnité globale du sinistre s'effectue d'abord par un premier versement pour un montant de 70 % de l'indemnité, le deuxième versement de 30 % intervenant uniquement sur présentation de factures ou de justificatifs. (Notons que cette mesure a été appliquée dès 1988 aux affaires nouvelles).

Nous sommes en droit de nous demander l'impact sur la charge des sinistres de cette réforme et en premier lieu son signe : celle-ci est-elle coûteuse ou paradoxalement allège-t-elle la charge ?

En effet, le tableau 2.2.16 nous fait bien comprendre que le paiement de la vétusté peut être largement financé par l'institution de la règle, valable en incendie, des « 70-30 », ce qui rend la mesure bénéfique.

Tableau 2.2.16 : Règles d'indemnisation de la tempête en risque agricole avant et après 1993

Période		partie habitation du sinistre	partie exploitation du sinistre
Avant 1993		$P_H \cdot RP_H \cdot (1 - TV_H) - F_H$	$P_E \cdot RP_E \cdot (1 - TV_E) - F_E$
Après 1993	Premier règlement	$p1 \cdot (P_H \cdot RP_H - F_H)$	$p1 \cdot (P_E \cdot RP_E \cdot (1 - TV_E) - F_E)$
	Second règlement	$p \cdot P2 \cdot (P_H \cdot RP_H - F_H)$	$p \cdot P2 \cdot (P_E \cdot RP_E \cdot (1 - TV_E) - F_E)$

Les notations suivantes ont été adoptées : les indices *E* et *H* désignent respectivement les bâtiments d'exploitation et d'habitation. De même, on définit par :

- P_E (P_H) le montant des dommages des bâtiments de la catégorie E (H),
- RP_E (RP_H) le taux de la règle proportionnelle appliquée,
- TV_E (TV_H) le taux de vétusté des bâtiments de la catégorie E (H),
- F_E (F_H) le montant la franchise appliquée aux bâtiments de la catégorie E (H),

- $P1$ et $P2$ les pourcentages de l'indemnisation représentés par les 2 versements,
- p la probabilité de survenance du deuxième versement.

Ainsi, il se peut que la faible fréquence du deuxième versement mène à une situation où la charge de sinistres résultante après la réforme soit inférieure à celle qui aurait été supportée avant celle-ci.

C'est d'ailleurs ce que montre l'exploitation des résultats du sondage : *la réforme mise en application en 1993 pour le risque agricole a eu pour conséquence la baisse de – 24,5 % de la charge de sinistres, le paiement de la vétusté de la partie habitation ne compensant qu'à hauteur de 5,5 % du montant du préjudice l'institution du règlement différé.*

VI.4 - Etude de la double vétusté

Enfin, il faut se remettre dans la situation qui prévalait avant 1983, avec la règle de la double vétusté. L'histogramme de distribution du taux de vétusté réel indique la proportion de sinistres sondés dont le taux de vétusté dépasse strictement 40 % : environ 4 %. Pour mesurer l'impact de l'abrogation de cette règle sur la charge des sinistres, il convient de calculer l'indemnisation avant abattement dû à la franchise qui aurait été versée pour ces sinistres en application de cette règle. (Un exemple de calcul de la double vétusté a été donné au paragraphe VI.1).

En termes plus concrets, il convient de rapprocher, sur l'ensemble des sinistres sondés, ces deux montants d'indemnisation dont l'expression mathématique est :

- $MTINDSUP = MTPREJ - MTVET$ pour les règles en vigueur entre 1983 et 1990,
- $MTINDSUP2 = (MTPREJ - MTVET) \cdot \min\left(\frac{TXVET_0}{TXVET}, 1\right)$, où $TXVET_0$ désigne le taux seuil de 40 %, pour les règles en vigueur avant 1983.

Leur comparaison fournit l'indice correctif lié à l'évolution des règles d'indemnisation relatives au taux de vétusté.

Globalement, la suppression de la règle de la double vétusté conduit à une augmentation de la charge équivalente à 1,0 % de celle-ci.

VI.3 - Hypothèses de redistribution du surcoût lié au paiement de la vétusté

Compte tenu des informations à notre disposition, nous ferons l'hypothèse que la vétusté réelle des bâtiments endommagés par la tempête n'a pas connu de grandes évolutions au cours de la période d'étude considérée. Cette supposition se justifie par le fait que le secteur du bâtiment en France intervient sur un marché de renouvellement, arrivé à saturation, en raison de la stagnation de la population. L'indice correctif lié à l'application de la vétusté réellement appliquée prend quant à lui des valeurs différentes selon que la date de règlement est ou non postérieure à 1990.

La structure de redistribution du surcoût lié à l'application de la vétusté réellement appliquée s'identifie donc à la ventilation par risque et exercice.

Certaines personnes ont émis l'idée que la vétusté pouvait diminuer de manière très sensible dans une région frappée par une tempête importante, en raison des réparations qui y sont effectuées; cette théorie s'appuie sur le témoignage des experts et l'exploitation de rapports d'expertises de bâtiments frappés de nouveau par une deuxième tempête, quelques mois seulement après la première. Cette thèse, connue sous le nom d'«*effet de mémoire* » montre alors non seulement que le règlement de la vétusté est moindre, mais en outre que les bâtiments résistent mieux aux assauts du vent, ce qui a pour effet de diminuer le taux de destruction. Malheureusement, aucun élément chiffré ne vient en appui de cette thèse, qu'il pourrait néanmoins être intéressant de compléter par une analyse statistique. Dans l'attente, nous considérons que le niveau de vétusté de chaque événement tempête est identique. Nous émettons la même hypothèse concernant la région, en considérant que chaque zone admet le même taux de vétusté.

VII - Autres opérations d'actualisation

Il existe bien d'autres facteurs d'actualisation à examiner. Nous présentons maintenant ceux dont l'influence sur la charge des sinistres n'est pas nulle, et le calcul de l'indice correctif qu'il importe d'appliquer à cette charge pour la rendre homogène dans le temps.

VII.1 - Calcul de l'indice lié à la limite de garantie

A l'UAP avant 1983, l'usage était de limiter le montant de l'indemnisation versée à l'assuré pour les dommages occasionnés aux bâtiments, quel que fût le montant du préjudice, le montant d'indemnisation maximal pour les dégâts subis par le mobilier ou les objets précieux étant quant à lui défini contractuellement.

On recense ainsi 5 garanties tempête, en cours entre 1980 et 1983 :

- la garantie annexe T.G.A. (Tempête, Grêle - toiture, chute d'appareils de navigation aérienne) issue de l'Urbaine, dont le pouvoir d'indemnisation est limité, par risque et par sinistre, à 600 fois la valeur en francs de l'indice, ce qui représente entre 120.000 et 180.000 francs entre 1980 et 1983,
- la garantie annexe n° 7103 de l'Union, dont la limite d'indemnisation s'élève à seulement 40.000 Frs,
- l'annexe T de la Séquanaise, qui prend en charge les dommages causés aux bâtiments à concurrence de 1.000 indices,
- les garanties tempête proposées après la fusion de 1968 pour les risques agricole, professionnel et d'habitation, et dont les textes très proches indiquent une limite de garantie égale à 900 fois la valeur de l'indice, soit un montant compris entre 185.000 et 275.000 francs,
- enfin, un texte de l'APSAD, daté de 1983, dont la limite de garantie s'élève à 630 fois la valeur de l'indice, ce qui représente entre 130.000 et 190.000 francs.

L'abrogation de cette limitation de garantie est stipulée dans le texte de la garantie tempête proposée par pollicitation en 1983. Elle concerne toutes les composantes de la branche incendie, hormis le risque industriel pour lequel cette limite de garantie reste contractuelle. Pour déterminer l'influence que peut avoir cette mesure sur la charge des sinistres, il importe de procéder en deux étapes successives :

- dans un premier temps, le regroupement des branches-risques selon le niveau maximal d'indemnisation prévu par la garantie,
- dans un deuxième temps, l'étude des courbes de distribution, selon ces 5 niveaux de limites de garantie, du montant des sinistres survenus au cours des exercices 1980 à 1982, et l'analyse de leur queue de distribution.

Le regroupement des branches-risques par type de garantie tempête, s'il ne présente pas de difficulté sur un plan théorique, est moins aisé du point de vue pratique. En effet, aucun élément ne permet d'identifier simplement le type de garantie. C'est de l'analyse :

- du code branche-risque, auquel est rattaché un type de produit et donc un type de garantie,

- ou du code entité d'appartenance (réseau S pour la Séquanaise, parties Urbaine ou Union du réseau Agent),

que peut être identifié le type de garantie auquel est rattaché chacun des sinistres.

La deuxième étape réside dans l'estimation, pour chacune des garanties et chacun des risques, de la part du préjudice restant à la charge de l'assuré du fait du dépassement de la limite de garantie. Pour ce faire, la méthode la plus simple consiste à mesurer la part que représentait en moyenne ce préjudice pour les années postérieures à 1982, suite à l'abrogation du plafond maximal d'indemnisation. Il est donc fait l'hypothèse que la part à la charge de l'assuré reste du même ordre de grandeur sur les exercices 1980, 1981 et 1982 pris individuellement que sur la période 1983 à 1992. La vraisemblance de cette hypothèse importe moins que sa contribution à effacer l'impact fluctuant, sur la charge globale, des sinistres de très grande importance.

D'autre part, il convient de traiter avec soin le cas particulier de la garantie n° 7103 de l'Union, dont la limite de garantie est de 40.000 Frs. En effet, en période de forte inflation (comme c'était le cas pour les années 1980 à 1982), la valeur actualisée à la fin 1992 de cette somme est sensiblement plus forte pour 1980 que pour 1982.

Finalement, le tableau 2.2.17 donne l'estimation de la part du préjudice restée à la charge de l'assuré compte tenu de la limite de garantie, pour chacun des risques et pour les cinq garanties. Du fait de la proximité de leur plafond, les garanties T.G.A. et A.P.S.A.D. sont regroupées et la garantie annexe n° 7103 de l'Union est considérée pour chacun des trois exercices 1980 à 1982.

La combinaison de ces proportions avec le poids, exprimé en nombre de contrats, de chacune de ces garanties conduit finalement à la valeur de l'indice correctif décliné pour chacun des risques.

Tableau 2.2.17 : Estimation de la part de préjudice à la charge de l'assuré compte tenu de la limite de garantie

risque	Type de garantie					
	900 FNB	600 FNB	1000 FNB	40 kF 1980	40 kF 1981	40 kF 1982
habitation	0,8 %	1,4 %	0,8 %	12,1 %	14,0 %	16,3 %
professionnel	3,4 %	6,0 %	2,9 %	30,4 %	33,3 %	36,3 %
industriel	11,3 %	13,9 %	11,3 %	44,2 %	46,8 %	50,6 %
agricole	3,0 %	4,1 %	2,7 %	25,7 %	28,9 %	32,9 %
total	2,9 %	4,1 %	2,7 %	21,6 %	24,1 %	27,1 %

VII.2 - Calcul de l'indice lié au taux de T.V.A.

D'autres facteurs peuvent avoir eu une influence, certes relativement plus modeste que les facteurs précédents, sur la charge des sinistres au cours de la période 1980 - 1997. C'est le cas de la taxe à la valeur ajoutée (T.V.A.), impôt appliqué au montant hors taxe du sinistre lorsque le statut fiscal du sinistré ne permet pas à ce dernier de récupérer cette taxe : les particuliers et toutes les entreprises non productrices de T.V.A. sont à classer dans cette catégorie.

Le taux de la T.V.A. a crû à deux reprises au cours de la période d'étude :

- d'une part, en juillet 1982, de 17,6 % à 18,6 %,
- d'autre part, en juillet 1995, de 18,6 % à 20,6 %,

ce qui contribue au renchérissement du montant des sinistres les plus récents par rapport aux plus anciens. L'application d'un indice correctif au montant des sinistres s'impose donc.

Il convient pour cela de déterminer l'usage, professionnel ou privé, qui est fait du bien. Le risque professionnel, qui comporte uniquement des contrats couvrant des biens à usage professionnel, est à l'abri de ce réajustement. En revanche, les risques d'habitation et agricole peuvent contenir des contrats garantissant des biens ayant les deux usages. Pour le risque agricole, on rencontre d'ailleurs souvent, à l'intérieur du même contrat, des bâtiments d'exploitation et d'habitation dont le statut fiscal diffère.

Pour le risque d'habitation, le code branche-risque permet généralement de déterminer l'usage auquel est destiné le bien assuré. Pour le risque agricole, le sondage effectué sur des dossiers de sinistres nous a permis d'estimer la part représentée par l'indemnisation des dommages causés aux biens à usage privé : 6,8 %. L'hypothèse de stabilité de la partie « privée » de ce règlement sur la période d'étude permet de conclure.

Tableau 2.2.18 : Estimation de l'indice correctif lié à l'évolution de la T.V.A.

période	Risque agricole	Risque d'habitation	Risque d'entreprise
Avant juillet 1982	$(1 + \frac{20}{1186} \cdot \frac{68}{1000}) \cdot (1 + \frac{10}{1176} \cdot \frac{68}{1000}) = 1,002$	$(1 + \frac{20}{1186}) \cdot (1 + \frac{10}{1176}) = 1,026$	1
Juillet 82- juin 95	$1 + \frac{20}{1186} \cdot \frac{68}{1000} = 1,001$	$1 + \frac{20}{1186} = 1,017$	1
Après juin 95	1	1	1

Le tableau 2.2.18 reprend l'évolution, par risque, de la valeur du terme correctif lié à l'évolution du taux de T.V.A..

L'indice correctif lié à l'évolution du taux de T.V.A. est appliqué au montant de chaque sinistre, en fonction de sa date de règlement et de la branche-risque de rattachement.

VII.3 - Autres facteurs

Il existe bien d'autres facteurs perturbant l'homogénéité de l'historique du montant des sinistres. Il serait vain d'en dresser une liste complète et plus encore d'en mesurer les effets sur cette série, tant les autres facteurs présentés supra ont un effet incomparablement plus grand.

C'est le cas, notamment :

- du remboursement, à partir de 1990, des frais de déblais, à hauteur de 5 % du montant des dommages,
- de la définition de la tempête, dont un critère est la vitesse du vent puis l'importance des dégâts alentour,
- et d'une manière générale, de l'évolution des couvertures d'assurances.

Citons simplement les plus importants facteurs dont l'influence non négligeable sur l'indemnisation n'a pas à notre connaissance connu d'évolution marquée dans le sens d'un renchérissement ou d'un allègement du coût de la tempête :

- l'application de la règle proportionnelle des primes,
- l'évolution du taux de coassurance de l'UAP,

D'autres facteurs, dont l'impact sur le niveau des réparations en tempête n'est pas à établir, n'ont pas fait l'objet d'un développement spécifique. Ce redressement interviendra lors de la construction du modèle économétrique. Il s'agit en particulier :

- des normes de construction,
- de la composition du portefeuille habitation de l'UAP entre locataires et propriétaires, maisons individuelles et appartements en immeubles,
- de la concentration urbaine,

- du portefeuille.

Il paraît acceptable de considérer que les deux premiers effets n'ont pas connu de modification notable dans leur composition. En effet, les normes de construction n'ont pas fait l'objet de mesures visant à leur renforcement au cours de la période d'étude. De même, la structure du portefeuille ne semble pas avoir connu de grandes évolutions. Ces deux facteurs « non évolutifs » du point de vue temporel ne sont donc retenus dans le modèle économétrique que pour leurs disparités du point de vue géographique.

En revanche, les deux autres facteurs sont dits « évolutifs », c'est-à-dire que leurs variations prennent à la fois une dimension spatiale et temporelle et influencent donc celles de la charge de sinistres. La concentration urbaine, ou la part des sinistres entre la ville et la campagne, est appréhendée dans le modèle économétrique par l'introduction de la variable « taille de l'agglomération » fournie par l'INSEE.

Enfin, le degré d'influence du portefeuille sur le nombre de sinistres se mesure à travers ce que l'on appelle la fréquence de sinistres, notion que l'on retrouvera aussi dans le modèle au chapitre 4 de cette partie.

On pourra se reporter à l'annexe K, qui présente le programme SAS (PREJUDICE) où sont consignées toutes les opérations d'actualisation et de calcul du préjudice.

VIII - Validation des calculs et de l'actualisation du préjudice

La correction de la série historique des sinistres par les indices calculés aux paragraphes précédents rend normalement possible la comparaison des données sur toute la période de l'étude.

Il reste néanmoins à en mesurer la validité. Pour ce faire, on utilise le résultat obtenu au chapitre 2 de la première partie, et illustré par la figure 1.2.8 : le nombre de tempêtes par an est aléatoire et les trois périodes identifiées entre 1962 et 1993 ne constituent donc que des fluctuations climatiques sans tendance significative.

En revanche, rien n'indique que l'intensité des tempêtes soit stationnaire. C'est pourquoi une comparaison du montant actualisé des tempêtes d'exercices de survenance différents et occupant un rang identique du point de vue de leur coût est préférable à la comparaison des coûts moyens annuels des tempêtes : on sait en effet que les statistiques de rang offrent une plus grande robustesse que les moyennes aux valeurs extrêmes. Et notre série en comporte.

Nous retenons pour chaque exercice la neuvième tempête par le coût qu'elle a généré : en effet, il convient de retenir une tempête de taille moyenne, se distinguant

nettement des épisodes venteux les plus destructeurs. Pour cela, il convient de définir les règles de sélection des tempêtes en termes de taille et d'identité de lieu et de date.

Si l'on considère qu'une tempête peut durer au maximum trois jours consécutifs et doit rassembler au minimum 30 sinistres quotidiennement, alors cette définition conduit à un nombre d'événements égal à 431.

La figure 2.2.3 donne l'évolution du coût actualisé de la neuvième tempête sur la période 1980 - 1997.

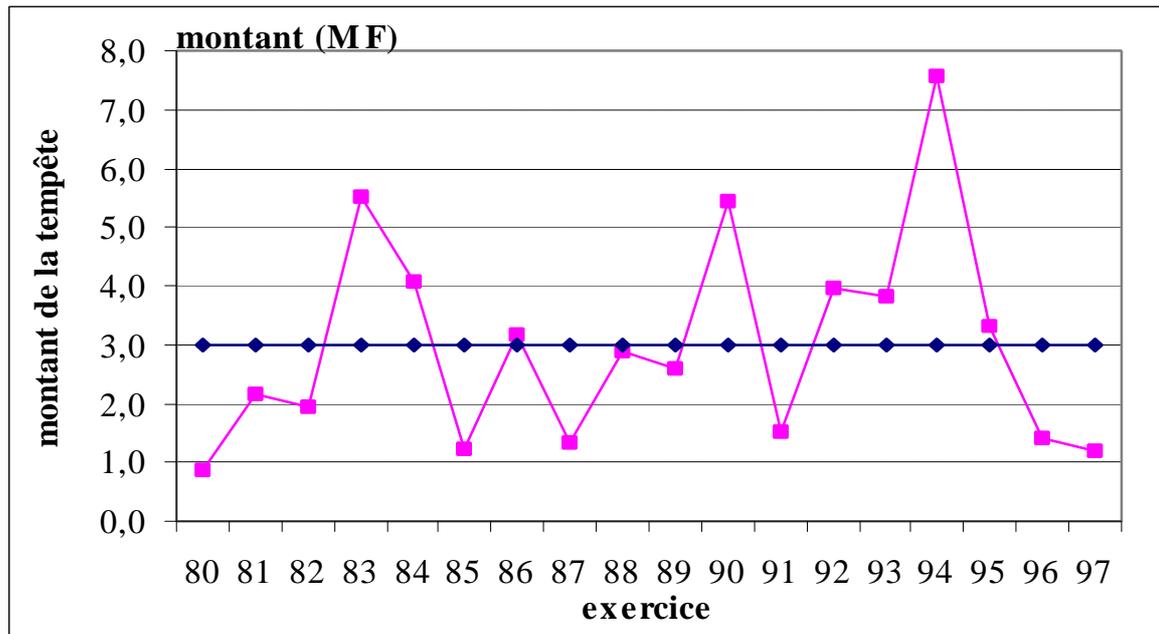


Figure 2.2.3 : Evolution du coût actualisé de la neuvième tempête (1980 - 1997).

Sur ce graphique, il n'apparaît pas de tendance bien définie : à part trois années qualifiées de venteuses (1983, 1990, 1994), le reste de l'historique se situe à un niveau proche de la moyenne, matérialisée par la droite à losanges d'équation « $y = 3,00$ », ou en dessous.

Si l'on recherche par la méthode des moindres carrés la droite de régression s'ajustant au mieux avec les données du graphique, on trouve :
 $y = 0,057.x - 2,06$.

Si l'on calcule le rapport entre la valeur estimée du coefficient de régression (0,057) et l'erreur-type qui y est rattachée, fournie par les logiciels, (1,85), on trouve :

$$t = \frac{0,057}{1,85} = 0,031.$$

Cette valeur est à comparer au fractile d'ordre 0,975 $t_{16}(0,975)$ d'une loi de Student à 16 degrés de liberté (2,12), correspondant à une probabilité de 5 % que la valeur de t soit dépassée si l'hypothèse d'une nullité du coefficient directeur de la droite est exacte. Comme : $t < t_{16}(0,975)$, on ne peut considérer le coefficient directeur de la droite comme significativement différent de zéro.

Les considérations précédentes permettent donc de valider l'actualisation et les corrections apportées aux données de la base « Sinistres ».

CHAPITRE 3 : LA TARIFICATION DU RISQUE TEMPETE

Il est frappant de constater que, durant ces trente dernières années, les bases de la tarification du risque tempête ont peu évolué. Cette évolution a été, de plus, différente selon la société constituant le groupe UAP.

L'évolution de la législation en la matière, rendant obligatoire la présence de la garantie tempête, a notamment contribué à modifier l'approche que l'assureur pouvait en avoir.

Après avoir étudié dans une première partie ce que nous appellerons la tarification classique, nous proposerons dans la seconde, grâce aux travaux d'actualisation et aux opérations de correction menées au cours du deuxième chapitre, une nouvelle méthode de tarification.

I - La tarification classique

Il est paradoxal de constater que le contenu de la garantie tempête a évolué, notamment après la loi de 1990, bien qu'il n'y eût jamais de paramètres spécifiques de tarification pour ce risque, pourtant si particulier.

Les critères de tarification de la garantie tempête furent toujours, soit ceux de la garantie Incendie, soit ceux du contrat de base auquel la garantie tempête était rattachée.

L'importance de la loi du 25 juin 1990 nous conduit tout naturellement à distinguer la tarification avant ou après 1990.

I.1 - La tarification avant la loi du 25 juin 1990

I.1.1 - Avant la création de l'UAP

A la veille de la création de l'UAP, en 1968, trois sociétés distinctes, l'Union, l'Urbaine et la Séquanaise qui devaient fusionner par la suite pour former le premier assureur français, proposaient cette garantie :

- soit en tant qu'élément couvert par la garantie de base (l'incendie),
- soit en tant qu'annexe à la police incendie.

Dans tous les cas, les variables de tarification étaient identiques pour le risque Incendie et la garantie Tempête.

Lorsqu'elle était souscrite en option, le tarif était alors exprimé en pourcentage du tarif Incendie et faisait l'objet d'une surprime.

Il n'y avait jamais de tarification spécifique du risque tempête.

I.1.1.1 - Pour l'Union

La majoration applicable au tarif incendie était de 25% pour toutes les situations avec un minimum de 10 F. Cette garantie tempêtes, ouragans et grêle - couvertures était proposée aux seuls propriétaires assurés en garantie illimitée.

Le montant de la garantie totale était fixé à 30 000 F. Cette garantie fut supprimée en 1971.

Pour les propriétaires d'immeuble, la majoration pour la même garantie était de 30%.

I.1.1.2 - Pour l'Urbaine

La garantie tempêtes, grêle – toitures, chute d'appareils de navigation aérienne pouvait être souscrite à titre facultatif (avenant T.G.A.), mais d'autres formules, tel le produit Globale Habitation, prévoyait systématiquement une couverture contre les effets du vent.

La garantie était accordée dans la limite des garanties Incendie avec une limite contractuelle d'indemnité (L.C.I.) de 600 fois la valeur de l'indice en francs.

La surprime variait de 10% à 40% de la prime Incendie selon la nature du risque, et valait plus précisément :

- 10% pour le locataire simple,
- 30% pour le propriétaire,
- 40% pour le propriétaire non exploitant agricole.

Cette surprime était calculée sur la prime totale Incendie avec un minimum de 6 F, à l'exclusion, lorsque cette garantie figurait au contrat, de la portion de prime " Pertes indirectes ", ladite garantie ne devant jamais être accordée sur T.G.A.

Il est intéressant de noter que ces tarifs ne pouvaient s'appliquer lorsque les risques étaient situés dans des cantons côtiers de l'atlantique, de la Manche et de la mer du Nord, ou à une altitude supérieure à 1 000 mètres.

De plus un zonage était prévu puisque les départements étaient classés en deux groupes selon l'exposition aux risques. Ainsi la surprime d'un locataire en zone 1 était de 10%, contre 15% pour un locataire situé en zone 2.

De façon générale la surprime appliquée pour le département de la zone 2 était supérieure de 5 points par rapport aux départements situés en zone 1.

Cependant, la garantie tempête pouvait être également incluse dans les garanties principales, comme dans le cas de la police “ Globale Habitation ”. Quelle que fût la qualité du souscripteur, locataire ou propriétaire, la prime était basée sur la catégorie dans laquelle étaient classés les locaux de l’immeuble suivant le décret du 10 décembre 1948 sur les loyers.

En outre, pour l’assurance des occupants, il était tenu compte du nombre de pièces principales, des locaux loués ou occupés.

Dans le cas du propriétaire non occupant, la tarification était effectuée au m².

Enfin, pour bénéficier de l’extension de la garantie tempête à l’assurance “ Grêle - toitures ”, le souscripteur devait acquitter une surprime uniforme (pour locataire ou propriétaire) de 25%. Cette surprime était calculée sur la prime nette totale à l’exclusion toutefois de la portion de prime afférente à l’assurance des véhicules automobiles.

I.1.2 - Après la création de l’UAP

A partir de 1970, les produits des trois sociétés devinrent communs.

I.1.2.1 - Risque d’habitation

Le contrat Multirisque - Habitation ne prévoyait pas une tarification spécifique du risque tempête.

Cette garantie pouvait être accordée comme précédemment dans le cadre d’une option avec surprime jusqu’en 1975 (contrats Incendie - Habitation ou Multirisque Habitation).

A partir de 1975, la garantie tempête et grêle-couvertures fut obligatoirement souscrite, il n’y avait pas de tarification spécifique prévue, donc pas de montant particulier de prime; celle-ci était toujours incluse dans la prime incendie.

Cependant la L.C.I. était au maximum de 900 fois l’indice (dont 600 fois au titre de la garantie “ grêle - couvertures ”).

Les variables de tarification du contrat “ Assurance Incendie Habitation ” étaient toujours fonction du nombre de pièces assurées et du montant des capitaux sur mobilier.

I.1.2.2 - Risque professionnel

Pour le risque professionnel, la garantie tempête était proposée en option moyennant une surprime forfaitaire.

Ainsi lorsque cette garantie était souscrite par un professionnel en vue d'assurer le bâtiment, le contenu professionnel et le mobilier personnel contenu dans les locaux accessoires en communication intérieure et privée avec les locaux professionnels, le prix était alors fonction de la surface des bâtiments :

- 55 F pour une surface comprise entre 0 et 300 m²,
- 120 F pour une surface comprise entre 301 et 500m²,
- 155 F pour une surface comprise entre 501 et 600m²,

une tarification sur mesure étant prévue au-delà.

Lorsque le propriétaire était non occupant ou propriétaire occupant partiel, le montant de la prime était porté à 0,31 F/m² ou 0,40 F/m² en fonction de la catégorie de l'immeuble.

D'autres types de contrats garantissent les risques professionnels; ainsi :

- le contrat Assurance *Hôtel - Multirisque*, dont la prime était établie en fonction du nombre d'étoiles de l'hôtel et de la qualité de l'assuré selon le tableau 2.3.1 suivant (1975),
- l'assurance multirisque des communes de moins de 5 000 habitants, dont la garantie tempêtes et grêle - couvertures était accordée également à titre facultatif, et dont la surprime était de 15% du montant de la prime nette "Incendie – Dégâts des eaux – Vol – Risques divers ". Cette surprime était ramenée à 10% si le souscripteur acceptait l'exclusion des dommages causés aux clochers. La L.C.I. était globalement fixée à 630 fois l'indice.

Tableau 2.3.1 : Tarification du risque tempête pour le contrat Hôtel- Multirisque (1975)

		1 étoile	2 étoiles	3 étoiles
☞ Propriétaire occupant total	Prime en francs calculée par	8	10	14
☞ Locataire et copropriétaire		Chambre	4	5
☞ Propriétaire non occupant	Prime calculée par m ² de la surface au sol	0,25F / m ²		

Rappelons également que le montant de la taxe applicable à cette garantie tempête et grêle - couvertures était de 9%.

L'APSAD a réalisé en juillet 1990 une étude tarifaire sur la période 1982 - 1988, reprise dans le tableau 2.3.2, portant sur le contrat Multirisque Habitation et Multirisque Artisan Commerçant.

Cette étude mentionne notamment le montant de la prime d'équilibre, pour chacun de ces 2 contrats, des garanties Multirisque, Incendie et T.G.N.

Tableau 2.3.2 : Primes d'équilibre de l'incendie et de la tempête des contrats Multirisque Habitation et artisan - commerçant

MULTIRISQUE HABITATION	MULTIRISQUE ARTISAN COMMERCANT
Prime équilibre multirisque : 1 090 F (y compris T.G.N.)	Prime équilibre multirisque : 4 158 F (y compris T.G.N.)
Prime équilibre Incendie : 337 F	Prime équilibre Incendie : 1 821 F
Prime équilibre TGN : 78 F (soit 23% de la prime Incendie)	Prime équilibre TGN : 250 F (soit 13,7% de la de la prime Incendie)
ou 7,7% de la prime MRH (hors TGN)	ou 6,4% de la prime MRA/C (hors TGN)

I.1.2.3 - Risque agricole

Pour le risque agricole, la garantie tempête et grêle - couvertures faisait l'objet d'une surprime, fonction de la qualité de l'assuré, et égale à :

- +25% pour un propriétaire non exploitant,
- +25% pour un propriétaire exploitant,
- +15% pour un fermier ou un métayer.

Cette surprime était applicable à la somme de la prime de base exploitation et habitation, préalablement majorée dans les cas suivants :

- ♦ si la surface au sol cumulée des bâtiments d'exploitation et des bâtiments d'habitation occupés par les préposés logés était supérieure à un nombre de m² occupés, lui-même déterminé par rapport à la superficie réelle d'exploitation. Celle-ci ne devait pas excéder pour un tarif préétabli une surface de :
 - 60 ha pour les exploitations de cultures,
 - 80 ha pour les exploitations herbagées ou mixtes.

Ainsi, à titre d'exemple, cette surface ne devait pas excéder :

- 400 m² pour une superficie réelle exploitable de 10 ha,
- 800 m² pour une superficie réelle exploitable de 30 ha.

Au delà, l'assuré était passible d'une majoration de la prime de base de 20%,

◆ si les bâtiments comportaient dans leur ensemble :

- plus de 10% de matériaux légers pour la couverture,
- entre 30% et 50% de matériaux légers pour la construction.

Dans ce cas, l'assuré subissait une majoration de la prime de base de +15%. Cette même majoration était de + 20% si :

- plus de 10% de matériaux légers pour la couverture,
- plus de 50% de matériaux légers pour la construction.

D'autres majorations venaient s'ajouter encore à la prime de base lorsque des extensions de garanties étaient accordées du type :

- utilisation d'une moissonneuse-batteuse automotrice,
- entrepôt de liquide inflammable et / ou vins, alcools,

mais il serait fastidieux d'en dresser une liste exhaustive.

Retenons simplement que la surprime initiale de 25% de la prime de base Exploitation plus Habitation, pour un propriétaire exploitant, atteignait près de 35% du montant de la prime de base majorée dans les cas envisagés ci-dessus, pour bénéficier de la garantie Tempêtes et Grêle - couvertures.

I.1.3 - La pollicitation de 1983

Suite à la loi de 1983, l'UAP proposa à tous les assurés en incendie la garantie tempête par avenant moyennant une surprime de :

- 2 F par contrat pour le risque Habitation,
- 20 F pour les risques professionnels et agricole.

Le tarif des affaires nouvelles était alors le suivant :

- risque d'habitation : 15 % de la prime Incendie
ou 3 % de la prime Multirisque en zone 1
ou 5 % de la prime en zone 2

- risques commerciaux : 15% de la prime Incendie
- risques agricoles : 30% de la prime Incendie
- risques industriels : 0,15% des capitaux assurés contre l'Incendie

Pour information, on trouvera dans le tableau 2.3.3 une synthèse de la tarification de la garantie tempête appliquée vers 1985 pour l'ensemble des risques de la branche incendie en 1985.

Tableau 2.3.3 : Tarification de la garantie tempête HT en proportion de la prime Incendie ou des capitaux garantie

CATEGORIE DE CONTRAT	TYPE	TARIFICATION TEMPETES (HT)
HABITATION	CH (1983)	Selon les zones (7,5,4,3,2,1 %) de la prime formule
	CHLA/CHLM/CHPM/CHPA (1983)	prime globale
	IH non occupant (1971)	12 % de la prime globale incendie tempête – attentats - catastrophes naturelles
	MH non occupant et immeuble en copropriété (1971)	5 % (14 ,29 si extension dégâts des eaux) de la prime globale
	Contrat classique (avant 1970)	15 % de la prime incendie
PROFESSIONNEL	MP 85 (1985)	Prime globale pour les garanties de base
	Ancienne MP (1974)	15 % de la prime incendie
	MP simplifiée (1985)	Prime globale pour les garanties de base
	HOTELS (1985)	15 % de la prime incendie
	Contrat classique (avant 1970)	15 % de la prime incendie
AGRICOLE	LA (1975)	30 % de la prime incendie
	Contrat classique (avant 1970)	30 % de la prime incendie
	CA (1985)	30 % de la prime incendie mais comprise dans la prime globale
RISQUES INDUSTRIELS	MP RIC (1985)	0 ,15 % des capitaux incendie
	MP ME (1985)	Compris dans la garantie de base : - 0,06 ‰ au titre du bâtiment - 0,10 ‰ au titre du contenu
	Contrat classique (avant 1970)	0,15 ‰ des capitaux incendie

I.2 - La tarification après la loi du 25 juin 1990

La loi du 25 juin 1990 (cf. la 1^{ère} partie), généralisant la couverture des tempêtes à tous les contrats d'assurances garantissant les dommages contre l'incendie ou les pertes d'exploitation après incendie, à des biens situés en France, ainsi qu'aux véhicules terrestres à moteur, obligea les assureurs à revoir la tarification de cette garantie.

Dans un premier temps, l'UAP effectua les modifications réglementaires puis, très rapidement, proposa une opération volontaire d'amélioration de la garantie Tempête dans les contrats Incendie, Multirisque – Habitation et professionnels, le risque agricole en étant exclu.

I.2.1 - La conséquence de cette nouvelle loi à l'UAP

Tous les contrats concernés :

- Incendie,
- Multirisque Incendie,
- Pertes d'exploitation après Incendie,
- Automobile avec garantie Incendie,
- tous les autres contrats de dommages aux biens comportant une garantie Incendie (notamment R.C. Bâtiments),

virent leurs effets automatiquement étendus à la couverture des tempêtes à partir du 1^{er} août 1990, même si l'assuré n'avait pas souscrit cette assurance.

Les dispositions prises par l'UAP en date du 1^{er} août 1990 furent différentes selon le type du contrat :

◆ *Contrats en portefeuille*

- *le contrat “ d'assurance des véhicules automobile ”* comportait la garantie “ action des forces de la nature ”, et couvrait donc déjà de ce fait le risque de Tempêtes,
- *contrats R.C. bâtiments* : lorsque le contrat comportait l'assuré “ dommages d'incendie subis par les biens sur chantiers ”, il fut décidé d'accorder gratuitement la garantie Tempête à hauteur de 25 000 F avec une franchise de 2 500 F,
- *contrats Incendie et Multirisque “ Risques industriels ”* : Les assurés ayant refusé la garantie Tempêtes, lors de la pollicitation de 1983, ou après, lors de la souscription d'une affaire nouvelle, se virent accorder la seule assurance des dommages dus au vent, gratuitement mais limitée à 200 000 F,

- *contrats Incendie et multirisque “ Risque simple ”* : très peu de contrats ne comportaient pas la garantie Tempête. Cette garantie fut donc proposée gratuitement :

- sur les bases des garanties contractuelles Tempêtes prévues dans les contrats,
- à défaut, sur la base des garanties Incendie et / ou pertes d’exploitation, mais dans ces deux cas, dans les strictes limites édictées par la loi, c’est-à-dire seulement pour les dommages dus au vent.

◆ *Affaires nouvelles*

Tous les contrats Incendie, multirisque ou pertes d’exploitation devaient comporter la garantie Tempêtes.

Le tarif et les conditions de cette garantie étaient ceux appliqués en août 1990.

I.2.2 - Amélioration de la garantie Tempête

Très rapidement, c’est-à-dire en octobre 1990, l’UAP décida d’effectuer une opération volontaire d’amélioration des garanties Tempêtes dans les contrats Incendie et Multirisque.

En effet, la disparition du mécanisme d’anti-sélection permit d’améliorer sensiblement le contenu de la garantie et surtout de supprimer les exclusions traditionnelles de cette garantie.

Cette opération fut réalisée en automne 1990. Elle concernait les différents types de contrats d’habitation et professionnels (sauf agricoles, pertes d’exploitation, valeur vénale et risques industriels).

La tarification fut alors adaptée dans les conditions suivantes :

I.2.2.1 - Risque d’habitation

Pour les *affaires nouvelles*, les dispositions suivantes furent adoptées :

- pour le contrat Habitation (CH), le tarif fut fixé à 6F par pièce,
- pour le contrat Habitation simple (CH1), la délivrance de cette garantie entraîna le rétablissement du jeu de l’indice.

Le barème des primes se trouva modifié, le nouveau tarif tenant compte de 3 paramètres :

- le nombre de pièces principales,
- le capital mobilier garanti,
- la zone du risque assuré.

Les affaires en portefeuille firent l'objet des mesures suivantes :

- pour le contrat Habitation simple (CH1), les contrats subirent le jeu normal de l'indice,
- pour les autres contrats, la majoration de prime fut fixée à 3% de la prime nette annuelle avec un maximum de 60 F à l'indice 431,40 F.

I.2.2.2 - Risque professionnel

Les clients garantis par un contrat Incendie ou Multirisque professionnelle (occupant ou non occupant) et hôtel, bénéficièrent immédiatement de la couverture Tempête améliorée (suppression des exclusions traditionnelles,).

Celle-ci fut accordée aux clients en portefeuille et proposée systématiquement avec chaque affaire nouvelle (ou remplacement) M. P. ou Hôtel sur la base d'un taux de révision de 2% en sus de l'indice, avec un maximum de 160 F H.T..

I.2.2.3 - Cas particulier des contrats agricoles

Contrairement aux contrats précédemment étudiés, l'amélioration du contenu de la garantie tempête ne fit son apparition qu'en juin 1993 pour les seuls affaires nouvelles et remplacements du contrat C.A.. Sa forme diffère ainsi de l'opération de révision d'octobre 1990 pour les risques habitation et professionnel.

Les anciennes formules conservèrent leurs anciennes dispositions et, à moins d'un remplacement, les exclusions liées à la tempête.

L'amélioration de la garantie tempête se fit selon les quatre directions, présentées par la suite.

- **La modification de la définition des événements garantis** : dorénavant la grêle n'est plus limitée aux dommages sur les toitures. L'exclusion des " inondations, engorgements et refoulement des égouts et canalisation souterraines, débordements des sources, cours d'eau et étendues d'eau, ainsi que les masses de neige ou de glace en mouvement " disparaît.

- **La nouvelle définition des biens assurés** : certaines exclusions spécifiques à la tempête disparaissent des conditions générales tels que volets, persiennes, gouttières, chéneaux, vitrages, glaces et toute matière translucide ou transparente. Les antennes de

radio et de télévision, les fils aériens et leurs supports, les stores, les enseignes et panneaux publicitaires, les panneaux solaires sont garantis. Seuls restent exclus :

- les bâtiments en cours de réfection,
- les bâtiments clos au moyen de bâches,
- les cribs et leur contenu,
- les bâtiments dont les murs et la couverture comportent en quelque proportion que ce soit :
 - des plaques de toute nature non posées et non fixées selon les procédés préconisés par le fabricant,
 - des matériaux tels que carton, feutre, toile ou papier goudronné, etc...

Enfin, une définition plus précise de ce qu'il est convenu d'appeler un hangar fut adoptée. Ce point est important car les hangars représentent à eux seuls 40% du parc des bâtiments agricoles. Le but étant d'exclure les hangars dont l'édification avait été faite par l'agriculteur lui-même et par conséquent dont la construction pouvait présenter des points flagrants de non conformité aux règles de l'art.

Ainsi, seuls furent garantis les hangars construits sur piliers en bois, fer ou maçonnerie scellés et fixés par des ferrures d'ancrage dans des fondations, soubassements ou de maçonnerie enterrés d'au moins 40 cm, si les ferrures d'ancrage faisaient corps avec les assises sur les piliers sur lesquels elles étaient boulonnées et tirefonnées, les simples goujons ne pouvant être considérés comme des ferrures.

• ***Un nouveau mode d'indemnisation*** : avant cette proposition d'amélioration de la garantie tempête, l'indemnisation se faisait toujours en valeur de reconstruction, déduction faite d'une vétusté calculée de manière indépendante pour les diverses parties sinistrées du bâtiment (couverture, maçonnerie, charpente).

Par la suite, il fut proposé d'aligner le mode d'indemnisation des *bâtiments d'habitation sous toiture distincte* sur celui de la garantie Incendie, c'est-à-dire de régler le montant de la réparation du bâtiment, y compris la vétusté dans la limite de 25 %.

En revanche, l'indemnisation des dommages des autres bâtiments, comprenant une partie agricole, ne donne toujours pas lieu au règlement de la vétusté. Cette indemnisation est égale :

- d'une part pour *le bâtiment ancien*, à la valeur de remplacement au jour du sinistre par un bâtiment moderne sans que celle-ci puisse excéder la valeur de reconstruction vétusté déduite,

- d'autre part pour *le bâtiment moderne*, à la valeur de remplacement par un bâtiment neuf avec prise en charge de la vétusté à hauteur de 25% si celle-ci était inférieure ou égale à 50%. Dans le cas où la vétusté était supérieure à 50%, l'indemnisation était calculée sur la base de matériaux évalués comme matériaux de démolition.

Le paiement de l'indemnité globale (pour les bâtiments d'habitation et d'exploitation) était, dès 1988, effectué en deux fois : 70% dans le mois suivant l'accord des parties, le solde sur production de facture justifiant la réparation.

• **La réduction de la franchise** : le niveau de la franchise fut fixé, en cas de sinis tre touchant seulement :

- l'habitation sous toiture distincte : à 900 F (comme le contrat habitation),
- les bâtiments à usage mixte ou d'exploitation : à 20% des dommages avec minimum de 7,5 fois l'indice F.N.B. et un maximum de 40 fois l'indice F.N. B.

I.3 - La tarification actuelle

Il s'agit dans ce paragraphe d'étudier la tarification établie par l'UAP pour la garantie Tempête avant la fusion avec AXA, puis celle pratiquée aujourd'hui par ce groupe d'assurances.

I.3.1 - La tarification du risque Tempête à l'UAP avant la fusion avec AXA

I.3.1.1 - Risque d'habitation

En 1995, quelle que fût la qualité de l'assuré, le tarif du contrat Habitation, et donc de la garantie Tempête, grêle, neige qui faisait partie des garanties de base, était fonction des paramètres suivants :

- la *zone géographique du lieu d'assurance* : 3 zones avaient été créées à cet effet,
- le *type de l'habitation* à savoir : appartements, maisons en agglomération, maisons hors agglomération,
- le nombre de pièces principales,
- le *capital mobilier choisi par l'assuré* (ce capital pouvant être " renforcé "),

- la présence de mesures de sécurité renforcée, si la garantie vol avait été souscrite,
- la présence ou non d'une franchise générale, d'exclusions de garanties, de garanties facultatives et de clauses spécifiques.

Le montant minimum de prime était de 260 F HT. Il n'y avait pas de tarification particulière pour le risque Tempête; cependant, le tarif était présenté TTC.

La taxe applicable au risque Incendie était de 30% mais de 9% pour les autres risques. La fiscalité nécessitait donc d'être en mesure de déterminer une prime pure pour l'Incendie et une prime pure pour les autres risques.

La prime afférente au risque tempête fut estimée à 15% du montant de la prime totale HT du contrat à laquelle elle était rattachée.

Enfin, à ce tarif de base TTC, était appliqué un coefficient de majoration de 0,10 pour les maisons situées hors agglomération ; des réductions des primes étaient accordées dans le cas d'une franchise générale égale à 2 fois l'indice (-10%) ou si des mesures de sécurité renforcée étaient prises pour les appartements situés en étage (-3,21%) (Indice janvier 1995 : 506,10 F).

I.3.1.2 - Risque professionnel

De façon générale la tarification du risque Tempête suivait le principe de la tarification du contrat principal auquel elle était rattachée; la garantie Tempête faisant partie des garanties de base :

- incendie et événements assimilés,
- catastrophes naturelles,
- tempête, grêle et neige ou dégâts des eaux.

Le tarif était toujours présenté TTC sans qu'il y ait de distinction entre la taxe sur l'Incendie et la taxe sur les autres risques.

- Contrat Multirisque professionnelle - Formule générale
 - Bâtiments ou risques locatifs : la prime était déterminée en fonction de :
 - la superficie développée des locaux à usage professionnel,
 - la classe de tarification résultant de l'activité exercée, selon une nomenclature pré-établie,
 - L'assurance du contenu professionnel (matériel, stocks, aménagements) : la prime était fonction de :

- la valeur assurée sur le matériel, les stocks et les aménagements,
 - la classe de tarification résultant de l'activité exercée.
- Multirisque immeuble :
 - Bâtiments et recours : la prime était déterminée en fonction de :
 - la surface totale développée,
 - la classe de tarification résultant de l'activité exercée,
 - la surface réservée à l'usage professionnel et à l'usage d'habitation.
 - Multirisque Hôtel :
 - Bâtiment : la prime était déterminée en F/m² en fonction de :
 - la surface développée,
 - nombre d'étoiles.
 - Assurance du contenu : la prime était forfaitaire et fonction :
 - du nombre d'étoiles,
 - du nombre de chambres,
 - du capital assuré sur contenu par chambre.
 - Contrat Multirisque professionnel des stations-service
 - Bâtiment ou risques locatifs : la prime était déterminée en fonction de la surface développée des locaux à usage professionnel.
 - Assurance du contenu professionnel (Matériel, stocks, aménagements) : la prime était déterminée en fonction de la valeur assurée sur le matériel, les stocks et les aménagements (à l'exclusion des carburants en cuves enterrées).

I.3.1.3 - Risque agricole

De façon générale, le tarif était décomposé en 3 sous-tarifs :

- Habitation : tarif par zone géographique,
- Exploitation : tarif par activité agricole (polyculture, élevage traditionnel, élevage industriel, viticulture),
- Protection juridique.

- **Habitation**

La tarification était effectuée, quelles que fussent la qualité de l'assuré (sauf propriétaire non occupant) et de l'activité agricole exercée, en fonction :

- de la zone géographique où était situé le bâtiment d'habitation (7 jours),
- du nombre de pièces principales,
- du capital mobilier choisi par l'assuré.

Les primes étaient proposées TTC et correspondaient aux garanties de base suivantes : Incendie et événements assimilés, Catastrophes naturelles, Tempête, grêle et neige, dégâts des eaux, vol, bris de glace.

Cependant, une extension de la garantie Tempête, grêle et neige “ améliorée ” était possible, lorsque le bâtiment d’habitation était sous toiture distincte des bâtiments d’exploitation.

La majoration des primes, pour bénéficier de cette garantie T.G.N. améliorée, était de 16 F HT par pièce principale.

- ***Exploitation***

La prime était déterminée en fonction :

- du type d’activité agricole exercée par l’assuré : polyculture, élevage traditionnel, élevage industriel, viticulture,
- de la zone géographique où était située l’exploitation, à l’exception de l’élevage industriel.

Les départements étaient répartis en zone (assurance des bâtiments) et en classes (assurance du contenu).

Le montant de la prime était déterminé de la façon suivante :

- Bâtiment : produit de la surface au sol des bâtiments assurés par le prix en francs du m² ; celui-ci étant fonction de :
 - la classification du département (zones géographiques),
 - la qualité de l’assuré (propriétaire, fermier, métayer).
- Contenu agricole : produit du capital choisi par le taux correspondant à la formule de garantie souhaitée par l’assuré.

Une option était possible :

- formule Multirisque regroupant la totalité des événements suivants : incendie et événements assimilés, Tempête, grêle et neige, vol, bris de machine et perte de liquide,
- formule réduite regroupant, outre l’incendie et la Tempête, grêle et neige (obligatoires), une ou plusieurs des garanties prévues en multirisque.

La répartition de la prime Incendie et de la prime Tempête pour les bâtiments d'exploitation était la suivante :

- ***Elevage traditionnel***

La prime Tempête représentait :

- 32% de la prime totale TTC,
- 47% de la prime Incendie TTC.

Si l'on tient compte des taxes sur les bâtiments d'exploitation, à savoir :

- 15% sur la garantie Incendie (contribution au fond de garantie contre les calamités agricoles),
- et 7% sur la garantie Tempête,

on aboutit à une tarification Tempête H.T. égale à 50,6% de la prime Incendie H.T. Le taux moyen des taxes étant de 12,3%.

- ***Viticulture***

La prime Tempête représentait :

- le 1/3 de la prime totale TTC (Incendie + Tempête),
- la moitié de la prime TTC Incendie. Pour le contenu, cette répartition de la prime Incendie et de la prime Tempête était identique.

I.3.2 - La tarification du risque Tempête après la fusion de l'UAP avec AXA

En mars 1999, le principe général de la tarification des contrats habitation, incluant la garantie Tempête dans la garantie de base est la suivante :

- le tarif est la multiplication d'une constante par des coefficients correspondant aux critères tarifaires (zone, nombre de pièces...),
- à ce produit sont ajoutés les frais de gestion et la taxe attentat.

Chacune des formules a sa constante et sa propre liste de critères tarifaires pertinents :

- nombre de pièces,
- capitaux mobiliers,
- taux des objets de valeur,
- zonage (18 zones),
- type de logement (maison, appartement, rez de chaussée),
- qualité de l'occupant (locataire, propriétaire, copropriétaire),
- l'inhabitation inférieure ou supérieure à 90 jours,
- le rachat de la franchise.

Cependant, les demandes d'autorisation préalables à la souscription sont nombreuses :

- capitaux importants assurés,
- nombre de pièces principales supérieur à 12,
- surface des dépendances supérieure à 300 m²,
- dépendances se trouvant à une autre adresse que celle du risque,
- bâtiments classés monuments historiques,
- moyens de protection ne répondant pas aux normes imposées,
- contrat proposé faisant l'objet durant les 24 derniers mois soit de sinistres, soit d'une résiliation par le précédent assureur,
- dérogation au montant ou à l'étendue des garanties,
- bâtiments situés en Corse.

L'ensemble de ces demandes d'autorisation préalables conduisent fréquemment à :

- des limitations contractuelles de garanties,
- des limitations ou exclusions de garanties,
- des demandes de protection particulières.

II - Proposition de tarification à partir des données d'assurances

Après avoir procédé au chapitre 2 à l'actualisation et au calcul du préjudice dû à la tempête, il est dorénavant possible d'étudier la série des sinistres.

II.1 - Analyse statistique des données d'assurances

II.1.1 - Etude de la série chronologique du nombre de sinistres

II.1.1.1 - Nombre et fréquence annuels des sinistres

La figure 2.3.1 présente la répartition annuelle du nombre de sinistres tempête au cours de l'historique d'étude 1980-1997, représentée par une courbe à motifs

constitués de ronds. La courbe comportant des triangles est la moyenne mobile d'ordre 3 de la série du nombre annuel de sinistres.

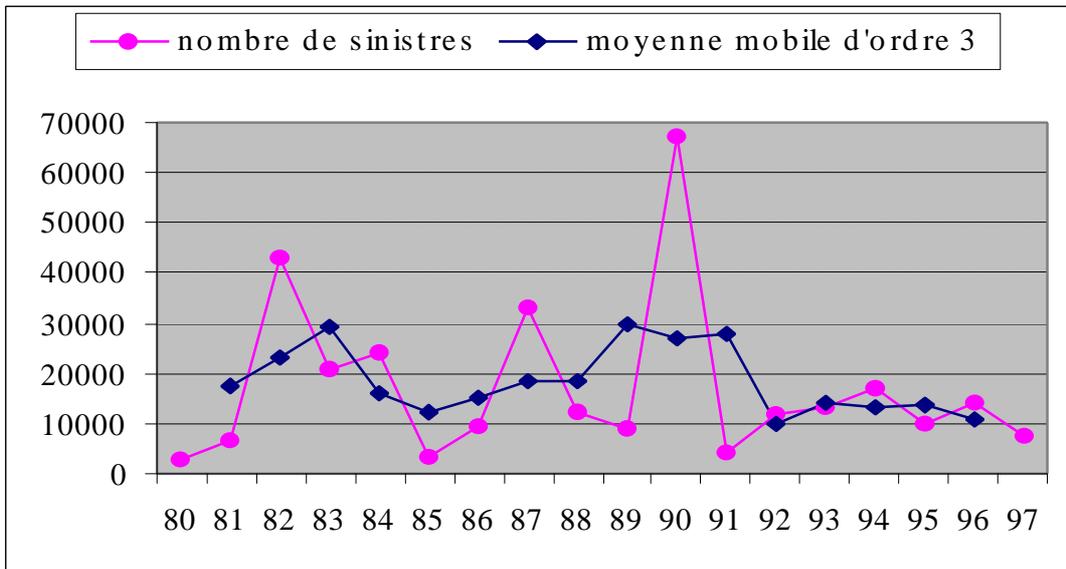


Figure 2.3.1 : Evolution annuelle du nombre de sinistres tempête (1980 - 1997)

Ce graphe illustre la très grande variation du nombre de sinistres d'une année sur l'autre, dans un rapport de 1 (en 1985) à 21 (en 1990). C'est en 1982 et 1990 que l'on recense le plus grand nombre de sinistres, dont l'origine est la survenance d'épisodes venteux particulièrement dévastateurs. La courbe de la moyenne mobile d'ordre 3 semble indiquer que ces années venteuses constituent des points isolés qui ne s'inscrivent pas dans une quelconque tendance.

Pour vérifier cette assertion, on réalise donc un test du χ^2 . Il consiste à comparer les valeurs observées chaque année à la valeur théorique moyenne calculée sur les 18 années d'une loi uniforme.

$$c^2 = \sum_{i=1}^N \frac{(n_{i\text{observé}} - n_{i\text{théorique}})^2}{n_{i\text{théorique}}}$$

où N est le nombre d'années (18 dans notre cas) et les n_i sont en milliers d'unités.

On compare la valeur χ^2 calculée à une valeur critique c_0^2 fournie par une table et fonction du nombre de degrés de liberté $v = N-1$ et du degré de confiance du test (ici égal à 95 %) : comme la valeur calculée (309,9) est très supérieure à la valeur

théorique (27,58), on rejette l'hypothèse d'une adéquation de la courbe du nombre de sinistres à la loi uniforme.

De même, si l'on remplace la valeur des années exceptionnelles 1982 et 1990 par la moyenne des valeurs qui les entourent, le même test du χ^2 conduit à une valeur du χ^2 calculée égale à 63,2. En revanche, si ce résultat montre que la distribution empirique ne peut être rattachée à une loi uniforme, cela ne prouve pas l'existence d'une quelconque tendance.

Pour en savoir plus, il convient de rechercher la droite de régression linéaire :

$$y = Tb$$

s'ajustant au mieux avec notre distribution annuelle du nombre de sinistres. Cette droite a pour équation :

$$y = -243,5t + 38811$$

où y et t désignent respectivement la série du nombre de sinistres et l'année de survenance correspondante.

Ces coefficients correspondent aux estimateurs des moindres carrés ordinaires. Il convient de tester le caractère significatif de l'estimateur \hat{b}_1 du coefficient directeur de la droite de régression. On forme alors le test suivant :

$$\begin{cases} H_0 : \hat{b}_1 = 0 \\ H_1 : \hat{b}_1 \neq 0 \end{cases}$$

Dans l'hypothèse H_0 , b_1 suit la loi de Laplace Gauss $N(\hat{b}_1, \sigma \sqrt{(T'T)^{-1}})$, et la variable t_{n-2} , définie par :

$$t_{n-2} = \frac{b_1 - \hat{b}_1}{\sqrt{\frac{\|y - Tb\|^2}{n-2} (T'T)^{-1}}}$$

suit une loi de Student à $n-2$ degrés de liberté. Sa valeur empirique est égale à :

$$t_{n-2} = -\frac{243,5}{16741} = -0,0146.$$

Or, si l'on consulte une table, on constate que le fractile d'ordre 0,975 d'une loi de Student, correspondant à une probabilité de 5 % que cette valeur soit dépassée si H_0 est vraie, est égal à 2,12. On ne peut donc considérer le coefficient \hat{b}_1 comme significativement différent de zéro.

En l'absence d'une série plus longue, il n'est pas possible de conclure dans le sens d'une augmentation ou d'une baisse du nombre annuel de sinistres.

Il est alors tentant de comparer cette courbe à celle de la fréquence annuelle, qui rapporte le nombre annuel de sinistres au portefeuille d'assurances mesuré en début de période. Le portefeuille ayant connu en volume une décroissance assez sensible sur la période 1980 - 1997, le dénominateur des exercices les plus récents est de moindre influence.

II.1.1.2 - Répartition mensuelle des sinistres

La figure 2.3.2 donne la distribution mensuelle des sinistres tempête. Elle indique clairement que la tempête est un phénomène particulièrement fréquent en automne et en hiver : les mois de novembre et surtout de février ressortent avec une part dépassant respectivement 20 et 25 % du nombre de sinistres totaux.

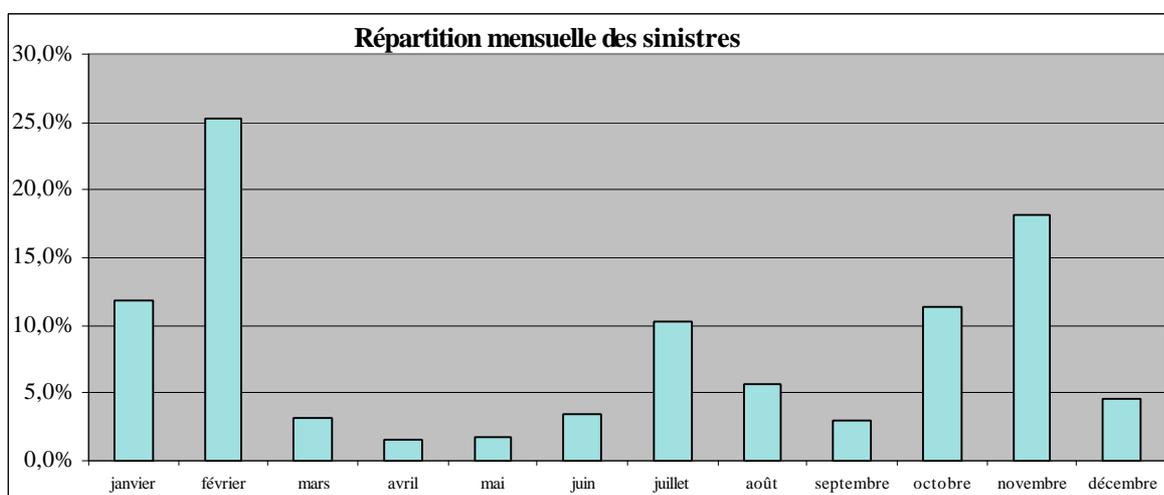


Figure 2.3.2 : Répartition mensuelle des sinistres tempête

Il est à noter que cette distribution des sinistres tempête diffère quelque peu de celle des tempêtes présentée sur la figure 1.2.9 et issue des archives de Météo-France, où les mois de décembre et mars présentent un poids nettement plus important.

II.1.2 - Répartition géographique de la fréquence

Il est possible de fournir, à partir de l'estimation, d'une part du nombre moyen de *contrats du portefeuille UAP ayant subi un préjudice* au cours de la période d'étude, et, d'autre part, du nombre moyen de *contrats en portefeuille*, une évaluation de la *fréquence des sinistres tempête*. Les résultats révèlent une différence sensible de niveau selon les trois composantes de la branche incendie (agricole, habitation, professionnel), ce qui justifie a posteriori cette segmentation. De même, leur visualisation sous la forme de cartes départementales permet une meilleure appréhension de leur répartition géographique.

II.1.2.1 - Risque agricole

La figure 2.3.3 fournit la répartition départementale de la fréquence de sinistres pour le risque agricole. Cette distribution présente une grande dispersion puisqu'il existe un rapport de 1 à 6 entre le niveau moyen de la fréquence des 2 classes de départements extrêmes (51 ‰ - 149 ‰ et 6 ‰ - 20 ‰). Les départements du nord de la France et du centre présentent en général une fréquence plus élevée que ceux situés dans le Sud-Est, le Poitou, la Franche-Comté, le Roussillon, et le l'est et le sud du Massif Central. Ce constat peut justifier une tarification intégrant la dimension géographique.

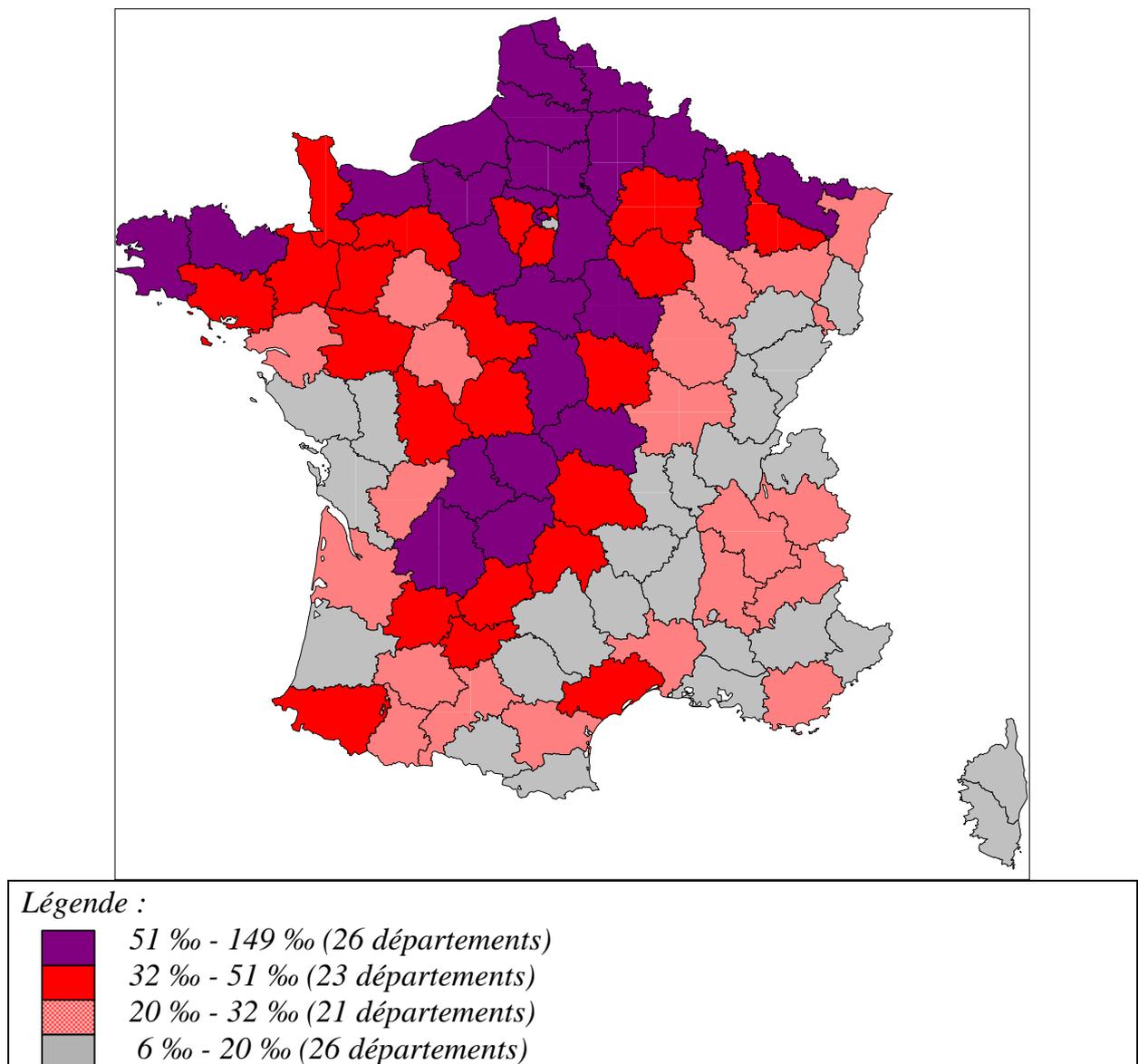
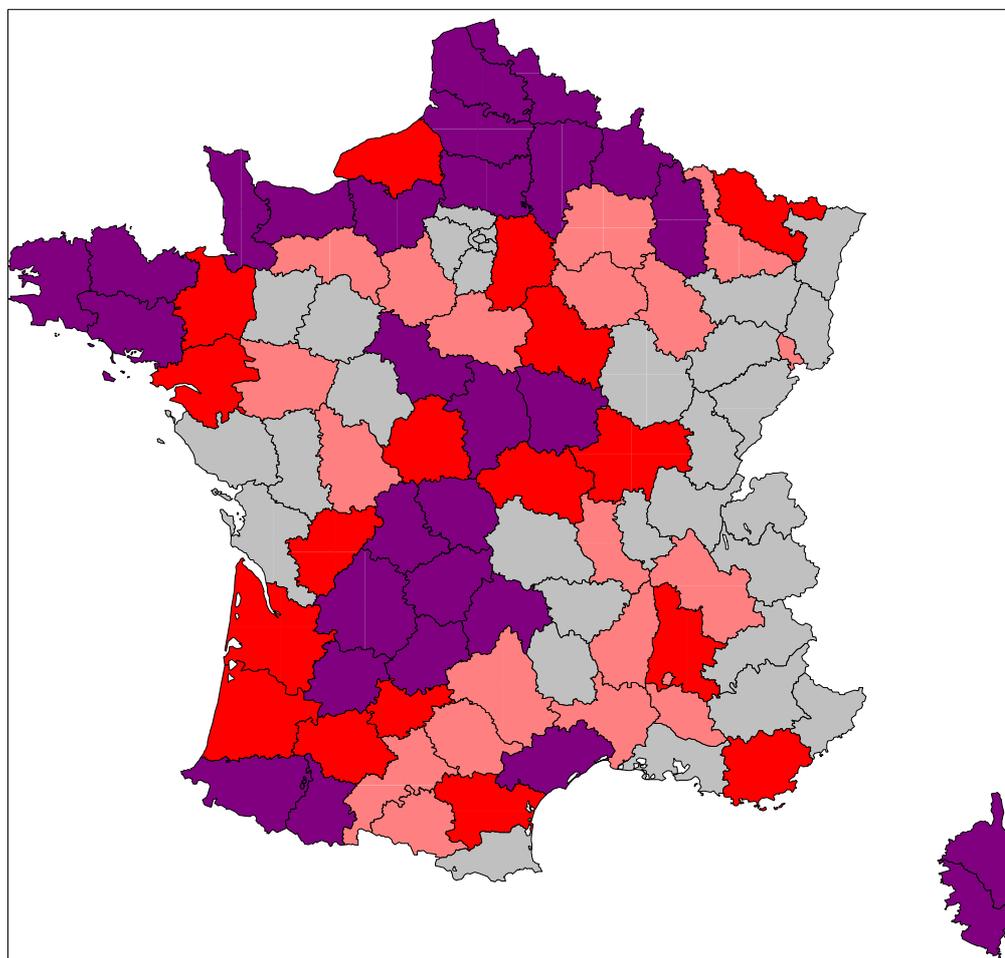


Figure 2.3.3 : Répartition départementale de la fréquence de sinistres pour le risque agricole

II.1.2.2 - Risque d'habitation

La figure 2.3.4 fournit la répartition départementale de la fréquence de sinistres pour le risque d'habitation.



Légende :

	8 % - 16 % (28 départements)
	6 % - 8 % (17 départements)
	5 % - 6 % (19 départements)
	1 % - 5 % (32 départements)

Figure 2.3.4 : Répartition départementale de la fréquence de sinistres pour le risque d'habitation

Contrairement au risque agricole, cette distribution présente une moindre dispersion puisque le rapport des niveaux moyens de fréquence des deux classes de départements les plus éloignées (8 % - 16 % d'une part, 1 % - 5 % d'autre part) n'est que d'environ 4. En revanche, il existe la même opposition entre le centre, le nord de la France et la Bretagne, où la fréquence annuelle se situe entre 8 % et 16 %, et l'est et la région parisienne, où elle ne dépasse pas les 5 %.

II.1.2.3 - Risque professionnel

La figure 2.3.5 fournit la répartition départementale de la fréquence de sinistres pour le risque professionnel.

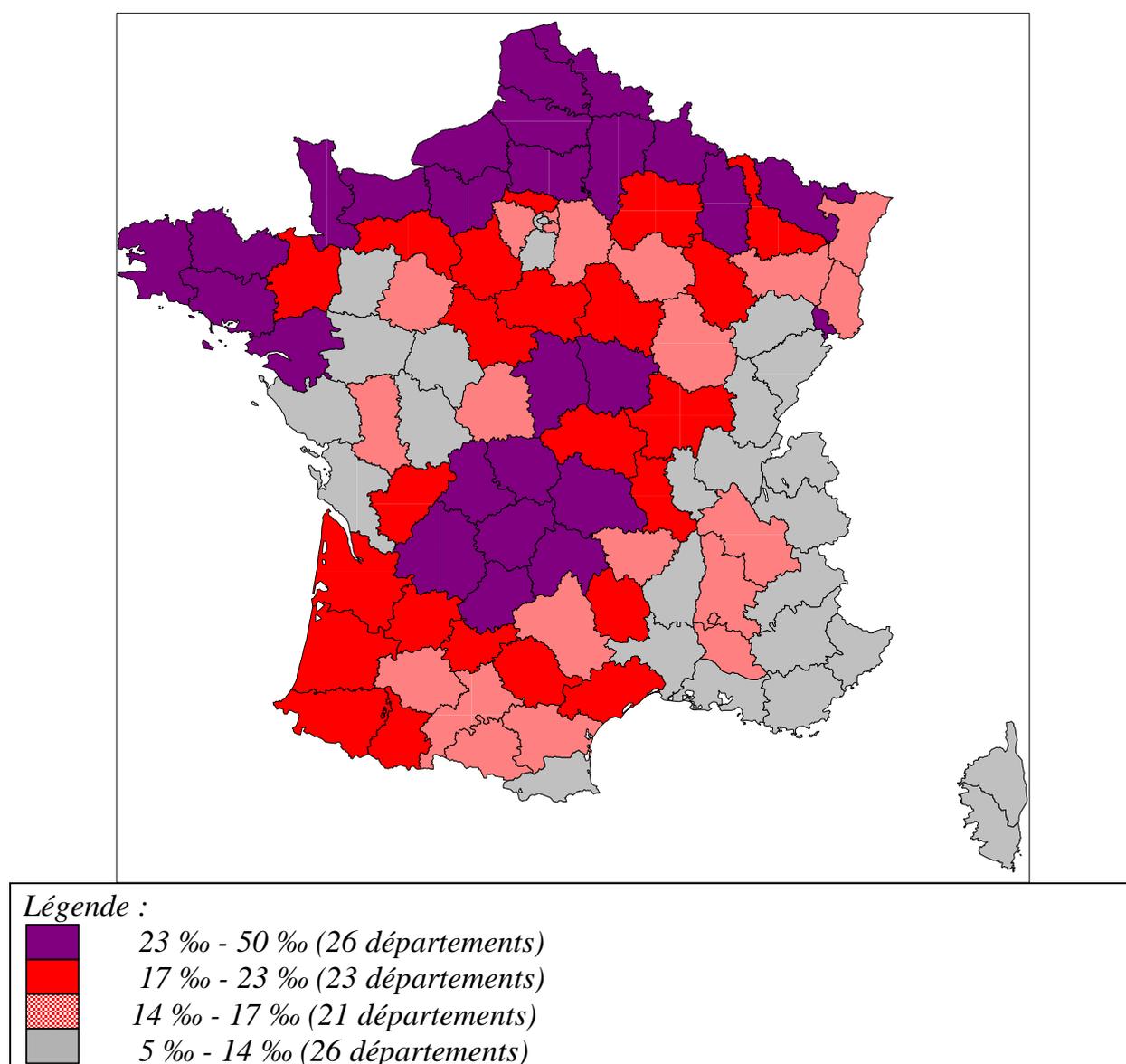


Figure 2.3.5 : Répartition départementale de la fréquence de sinistres pour le risque professionnel

Le risque professionnel présente une fréquence sensiblement supérieure à celle du risque d'habitation, car un contrat incendie garantit souvent plusieurs établissements distincts. La fréquence reste néanmoins inférieure à celle enregistrée pour le risque agricole. L'opposition centre – nord contre est – centre ouest décrite précédemment reste d'actualité.

II.1.3 - Répartition géographique de la prime pure

Pour compléter cette approche tarifaire, il convient d'indiquer le niveau de la *prime pure relative au risque tempête*. Cette grandeur est ici équivalente au montant total actualisé du préjudice dû à la tempête, en annulant la franchise, rapporté au nombre total de contrats appartenant au portefeuille incendie : cela revient en fait à évaluer ce que coûte en moyenne annuellement à chaque assuré le risque tempête.

La tempête étant par définition très dépendante du relief sur lequel elle souffle, il a été choisi de visualiser la disparité du niveau de la prime pure à partir de cartes de France. Enfin, comme pour la fréquence, la prime pure diffère fortement selon le risque considéré, ce qui renforce le choix de cette segmentation.

II.1.3.1 - Risque agricole

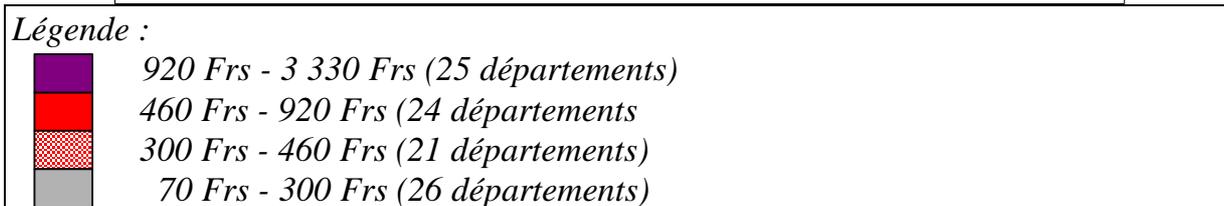
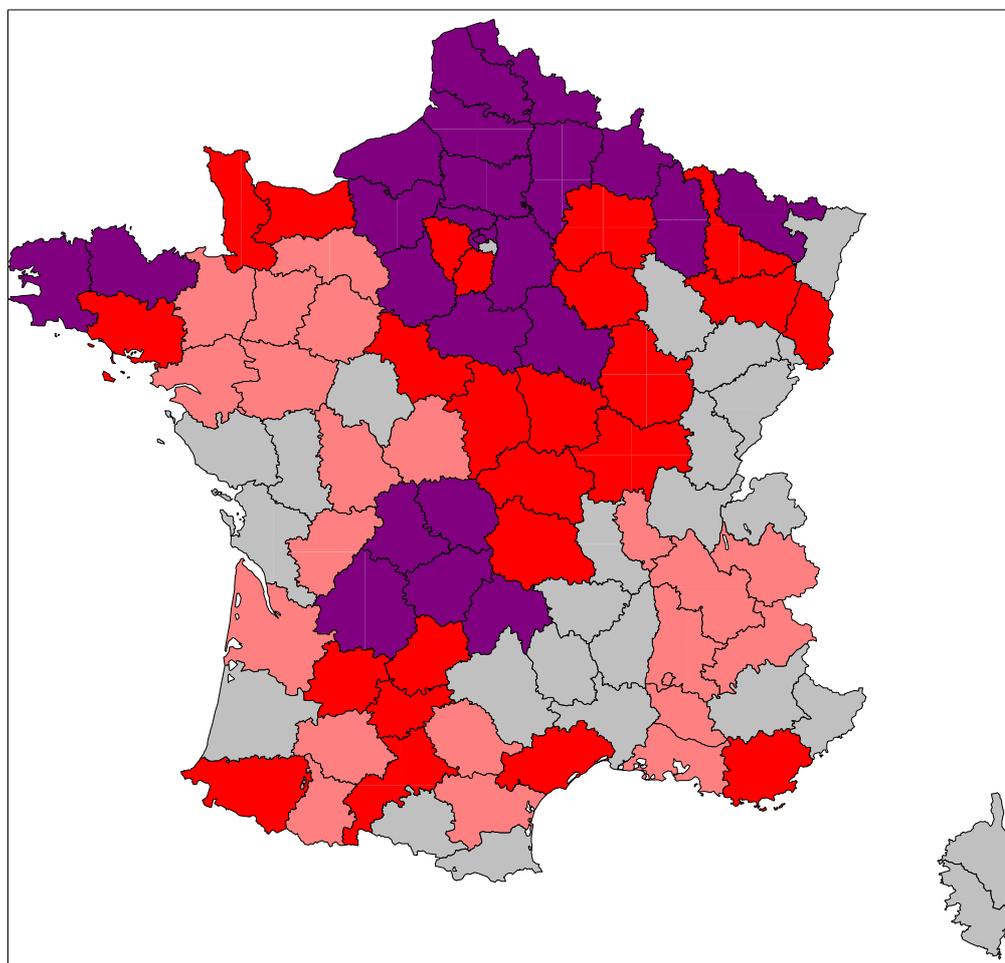


Figure 2.3.6 : Répartition départementale de la prime pure pour le risque agricole

La figure 2.3.6 représente le niveau, dans chaque département, de la prime pure tempête du risque agricole.

Cette carte montre une grande disparité dans l'évaluation de la prime pure puisque cette dernière se situe entre 920 F et 3.330 F dans les départements en violet, concentrés dans le nord (Ile-de-France, Nord-Picardie, Haute-Normandie, Champagne-Ardennes, Bretagne Nord) et dans le centre (Limousin, Dordogne, Cantal) de la France, et à moins de 250 F pour les départements en gris, plutôt situés à l'ouest (Poitou) ou à l'est (Franche-Comté, Cévennes, Roussillon). La palme de la prime pure la moins élevée revient à la Lozère avec 73 F, le département le plus cher étant l'Aisne.

II.1.3.2 - Risque d'habitation

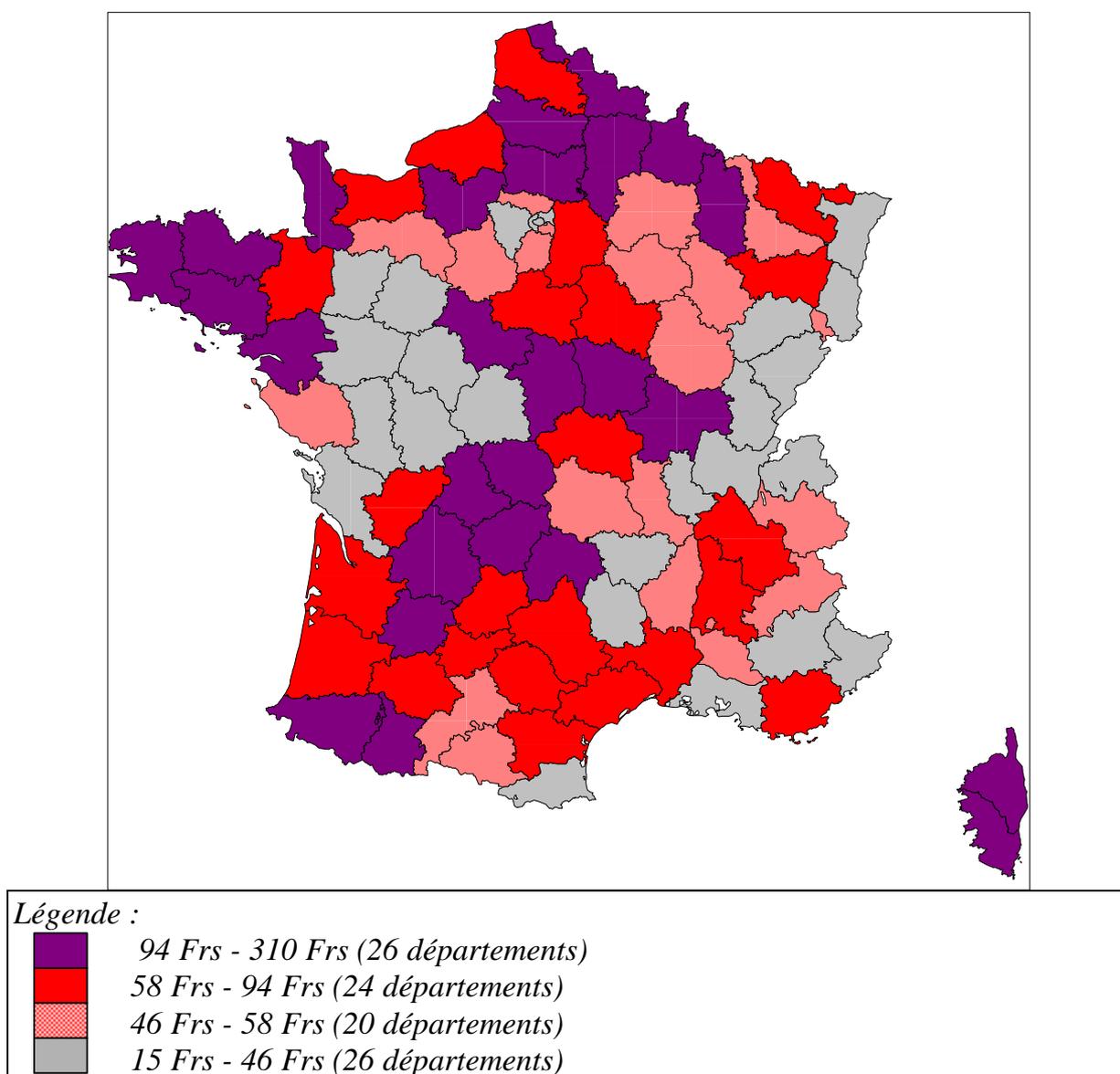


Figure 2.3.7 : Répartition départementale de la prime pure pour le risque habitation

La figure 2.3.7 donne la répartition par département de la prime pure tempête du risque d'habitation. La dispersion du niveau de la prime pure y est bien moins importante que pour le risque agricole : là où le rapport de la prime pure la plus élevée à celle la moins élevée est d'environ 45 pour le risque agricole, cet indicateur n'est que de 21 pour le risque d'habitation. Ceci peut s'expliquer en partie par le poids plus important de ce risque, par sa moindre vulnérabilité face à la tempête et à sa plus grande homogénéité. Là encore, les régions du nord et du centre de la France présentent un niveau de prime pure plus élevé que celui des régions du centre ouest et de l'est (Pays de la Loire, sud-est, Franche-Comté, Alsace).

II.1.3.3 - Risque professionnel

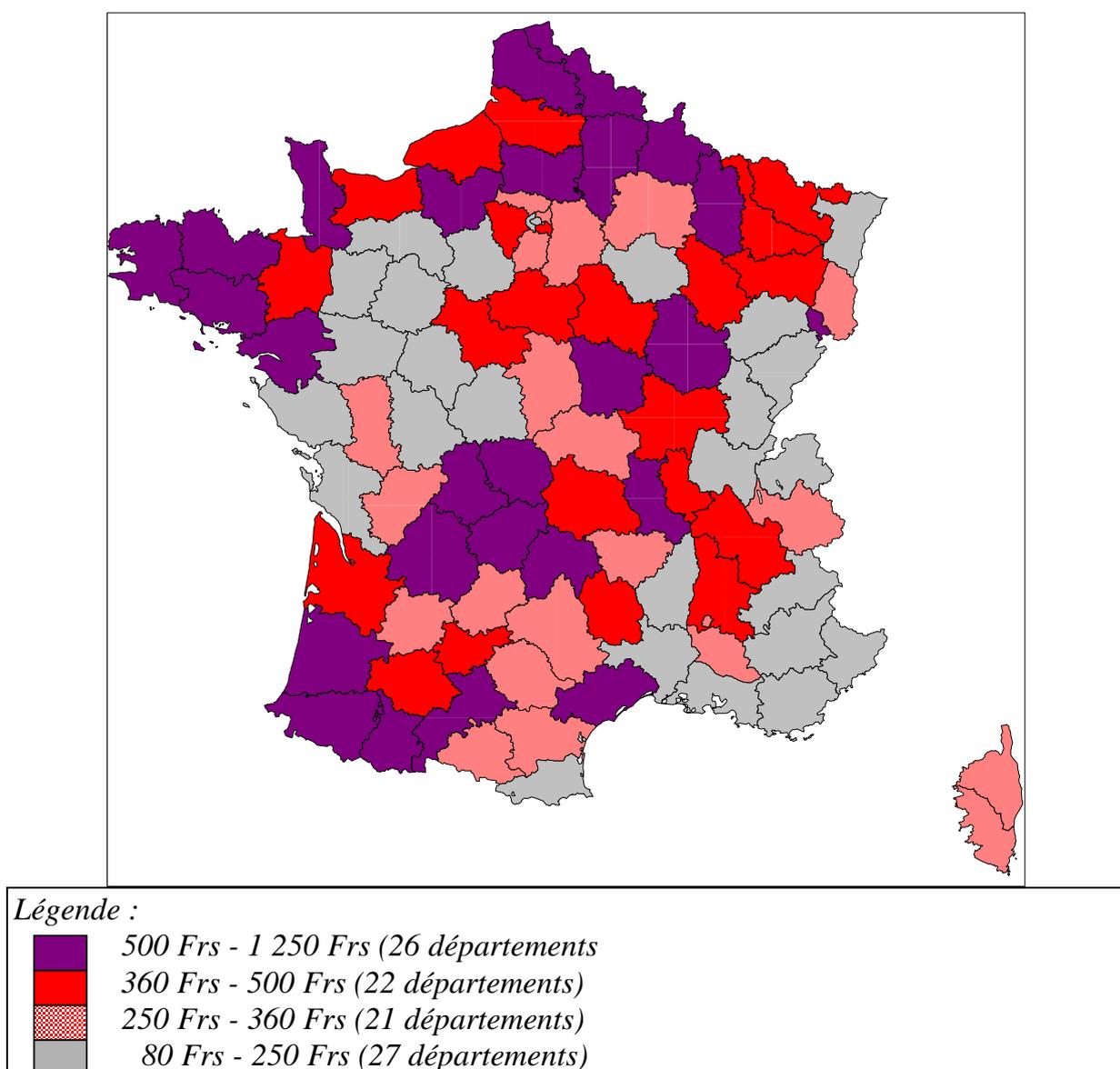


Figure 2.3.8 : Répartition départementale de la prime pure pour le risque professionnel

La figure 2.3.8 donne la répartition par département de la prime pure tempête du risque professionnel.

La dispersion de la prime pure du risque professionnel est relativement peu marquée puisque le rapport des primes pures extrêmes est d'environ 15. Là encore, on relève une opposition entre les régions de Bretagne, du Nord et du Limousin, de la Dordogne et des Pyrénées-Atlantiques, où la prime pure est souvent élevée, et les régions du centre-ouest, du sud-est et de Franche-Comté, dont les contrats professionnels sont peu exposés à ce risque. Le département dont la prime pure est la plus haute est l'Eure (1249 F), alors que Paris se situe à l'extrémité de ce classement avec 82 F.

II.1.4 - Analyse de la distribution du montant des sinistres

Au chapitre 2 ont été menés le calcul et l'actualisation du préjudice causé par le vent aux différentes composantes de la branche incendie. Ainsi, chaque sinistre ayant fait l'objet d'un règlement s'est vu attribuer :

- un indice correctif intégrant les différents éléments venant en déduction du préjudice réel,
- un indice d'actualisation tenant compte de l'évolution des prix des matériaux utilisés lors de la réparations des biens endommagés par la tempête.

En outre, il a été procédé à une évaluation du nombre et de la charge de sinistres qui auraient été réglés en cas de souscription systématique de la garantie tempête ou d'absence de franchise. Pour la franchise, le coût moyen de ces sinistres a été supposé égal à la moitié du niveau de la franchise minimum et leur distribution équivalente à celle d'une loi uniforme. Pour le taux de souscription de la garantie tempête, le coût moyen de ces sinistres « générés » est considéré comme équivalent aux sinistres réglés et la distribution de leur montant épouse donc celle des vrais sinistres.

Il est alors possible, muni de ces informations, de représenter la courbe de distribution du montant du préjudice dû à la tempête.

II.1.4.1 - Risque agricole

La figure 2.3.9 représente la distribution du montant du préjudice causé aux biens agricoles par le risque tempête.

Cette distribution est caractérisée par une moyenne égale à 16.528 F. Si l'on excepte les petites valeurs de préjudice, pour lesquelles on note un nombre de sinistres relativement faible, sa forme se rapproche de celle d'une loi log-normale. Il convient de vérifier la portée de cette affirmation en prenant le logarithme népérien de cette série ($y = \ln(TRC)$), où TRC est le montant du préjudice en hecto francs) et en utilisant

l'option « NORMAL PLOT » de la procédure UNIVARIATE de SAS. Le résultat de ces actions est consigné en annexe L où l'on trouve :

- d'une part l'ajustement graphique, sous la forme de la droite de Henry, de la distribution de y relativement à la distribution normale de moyenne et d'écart-type ceux de y ,
- et, d'autre part, la valeur du test de normalité de Kolmogorov-Smirnov. Celui-ci vaut 0,0717, ce qui signifie que la probabilité d'accepter à tort l'hypothèse de normalité de la variable y est inférieure à 0,01.

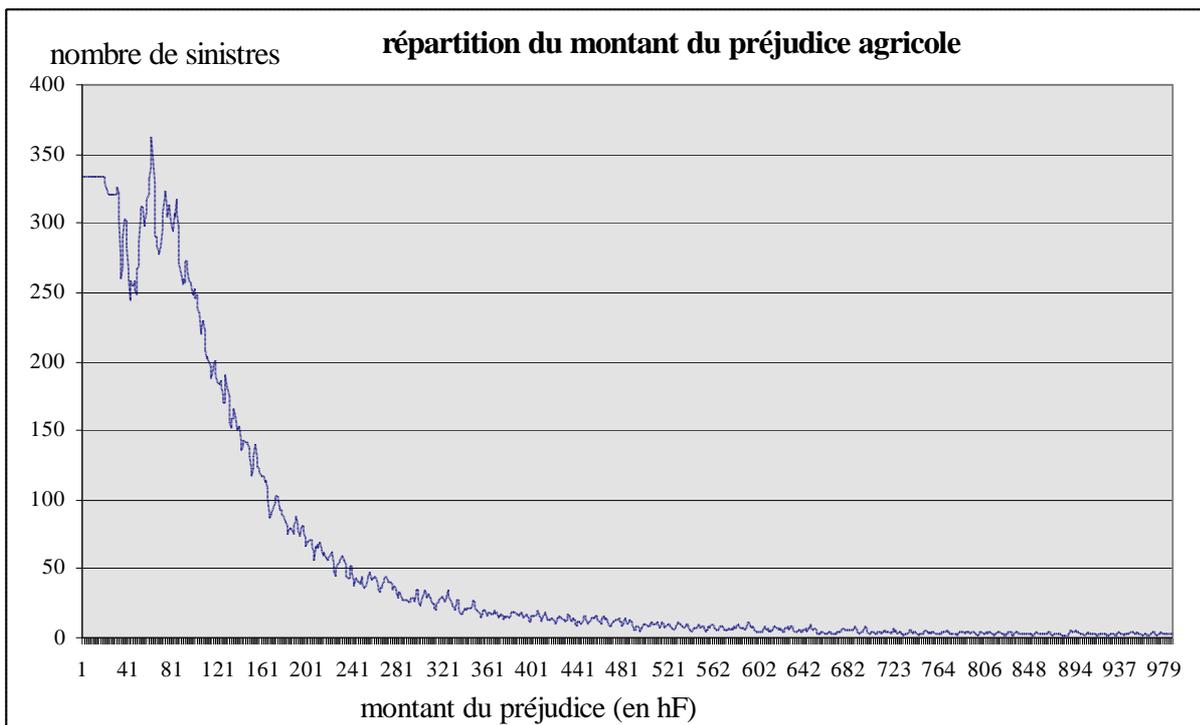


Figure 2.3.9 : Distribution du montant du préjudice dû à la tempête pour le risque agricole

Les sinistres importants de plus d'un million de F sont au nombre de 10 : ils représentent 0,16 % du nombre de sinistres, mais 3,0 % du préjudice total.

II.1.4.2 - Risque d'habitation

La figure 2.3.10 représente la distribution du montant du préjudice causé aux biens d'habitation par le risque tempête.

Cette distribution est caractérisée par une moyenne égale à 12.361 F. Comme pour le risque agricole, sa forme peut être ajustée par celle d'une loi log-normale : l'annexe L précise que la valeur du test de normalité de Kolmogorov-Smirnov, appliqué au

logarithme népérien du montant de préjudice d'habitation, est égale à 0,0704 : le risque de première espèce est, pour cette valeur, inférieur à 1 %. On peut donc accepter l'hypothèse de normalité pour y .

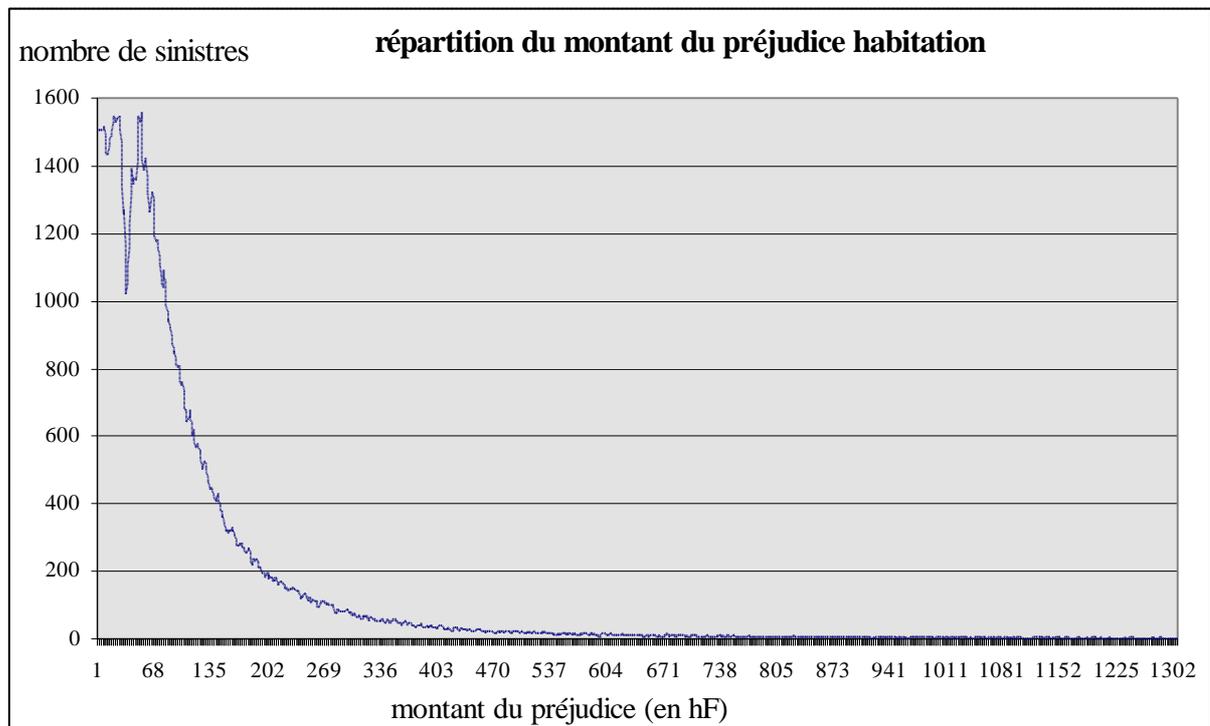


Figure 2.3.10 : Distribution du montant du préjudice dû à la tempête pour le risque d'habitation

Les sinistres importants de plus d'un million de F sont au nombre de 15, ce qui situe leur probabilité de survenance à 0,08 ‰. Leur charge représente néanmoins 2,0 % du préjudice total.

II.1.4.3 - Risque professionnel

La figure 2.3.11 représente la distribution du montant du préjudice causé aux biens professionnels par le risque tempête.

Cette distribution est caractérisée par une moyenne égale à 24.136 F. Comme pour le risque d'habitation, l'ajustement à une loi log-normale est satisfaisant. Aussi bien l'ajustement graphique (i.e. la droite de Henry) que le test statistique montrent la justesse du choix d'une loi log-normale. L'annexe L précise que la valeur du test de normalité de Kolmogorov-Smirnov, appliqué au logarithme népérien du montant de préjudice d'habitation, est égale à 0,05814 : le risque de première espèce est, pour cette valeur, inférieur à 1 %. On peut donc accepter l'hypothèse de normalité pour y .

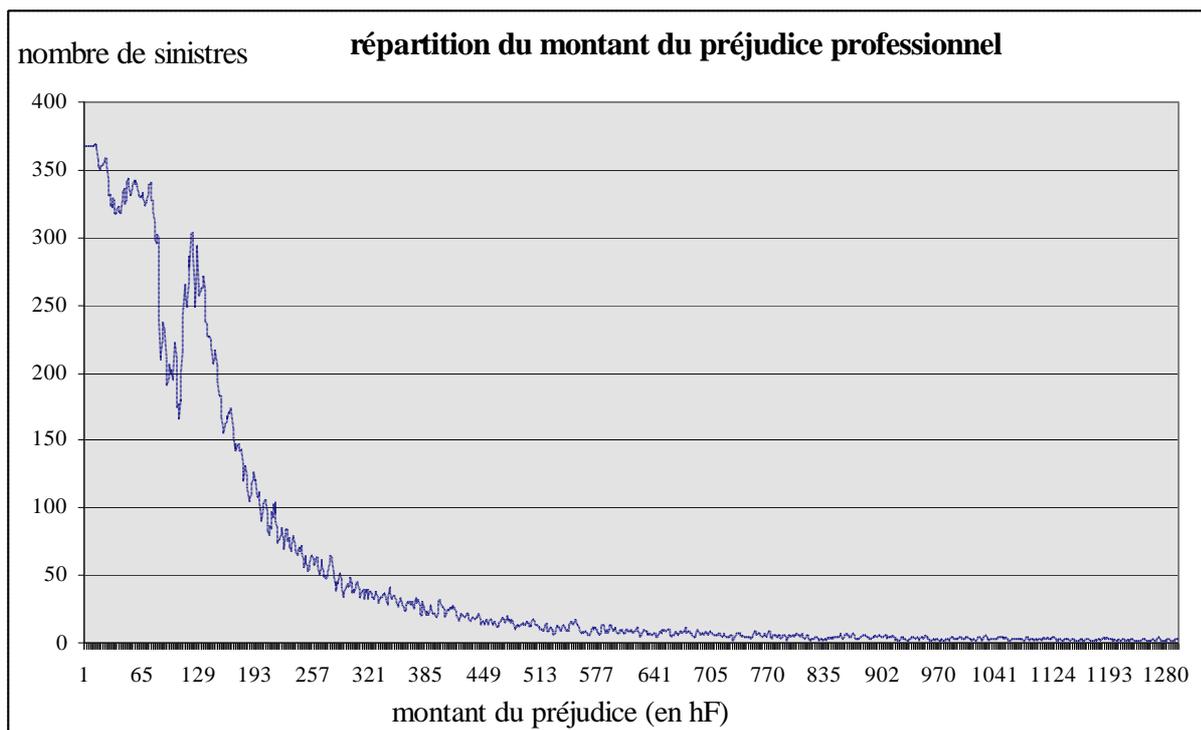


Figure 2.3.11 : Distribution du montant du préjudice dû à la tempête pour le risque professionnel

Les sinistres importants de plus d'un million de F sont au nombre de 150; ces derniers peuvent faire l'objet d'une réassurance dans le cas où ils appartiennent à un événement qui l'est. Ils représentent 0,21 % du nombre de sinistres total mais atteignent 21,5 % du préjudice total.

II.1.5 - Réduction de la prime pure par l'application d'une franchise

S'il est possible à l'assureur de recourir à la réassurance pour le règlement des tempêtes très coûteuses, il peut en outre fixer contractuellement un niveau de franchise restant à la charge de l'assuré. Il peut être intéressant de voir l'impact de l'application d'une franchise sur le montant de la prime pure tempête, et ce par risque. Une comparaison de la réduction de la franchise suite à l'utilisation de deux types de franchises (mixte ou forfaitaire) est ainsi menée successivement.

Le tableau 2.3.4 permet de voir l'impact de l'application de la franchise, mixte ou absolue, sur le niveau de la prime pure d'une des trois risques.

L'application d'une franchise mixte égale à 10 % du montant des dommages et d'un montant minimum de 5 fois l'indice FNB a un impact plus important, relativement à la réduction de la prime pure, que l'application d'une franchise absolue de 5 fois l'indice FNB : en effet, pour les montants de dommages supérieurs à 50 fois l'indice, la franchise appliquée est alors d'un montant inférieur à 10 % des dommages. On

préfèrera donc la franchise mixte à la franchise absolue, pour sa plus grande efficacité et son caractère plus équitable.

Tableau 2.3.4 : Impact de l'application d'une franchise sur le montant de la prime pure tempête

Risque	prime pure à franchise nulle	prime pure après application d'une franchise			
	montant (en F)	mixte (10 % - 5 FNB) montant (en F)	évolution	absolue (5 FNB) montant (en F)	évolution
agricole	729	600	-17,7 %	625	- 14,2 %
habitation	71	55	-22,0 %	57	- 19,5 %
professionnel	429	368	- 14,3 %	392	- 8,7 %

II.2 - Etude des événements tempête

Les sinistres précédents s'inscrivant dans le temps, il est donc possible de les regrouper selon cette dimension pour accéder à la notion d'*événement*, qui est la seule à être utilisée dans le domaine de la réassurance.

II.2.1 - Définition d'un événement tempête

La notion d'événement doit être définie avec précision. Il nous a semblé que sa définition devait retenir les deux aspects suivants :

- l'*intensité destructrice*, c'est-à-dire le nombre de sinistres survenus à une date donnée,
- la *durée* : le nombre de jours consécutifs pour lesquels l'intensité dépasse un certain seuil.

De manière à donner une certaine consistance à la tempête, il nous a semblé nécessaire d'adopter *un seuil minimal d'intensité destructrice journalière. Celui-ci a été fixé à 30 sinistres*. A l'opposé, compte tenu de l'existence de clauses dans les traités de réassurance fixant à *72 heures la durée maximale d'une tempête*, il nous a semblé logique de conserver cette disposition dans notre définition des tempêtes. Ainsi, quatre jours de tempête consécutifs (i.e. pour lesquels le nombre de sinistres survenus dépasse 30) constitueront au final deux tempêtes, respectivement de trois et un jour.

Cette définition, basée sur *l'importance des dommages*, s'écarte bien évidemment de celle retenue par Météo-France, dans laquelle il est mis l'accent sur *l'intensité du*

vent (devant dépasser 100 Km/h) et l'étendue de l'événement (devant toucher au moins 5 % des stations), sans qu'il soit fait référence à son caractère destructeur.

II.2.2 - Etude de la série chronologique des tempêtes

II.2.2.1 - Nombre et fréquence

La figure 2.3.12 représente la répartition annuelle du nombre de tempêtes au cours de l'historique d'étude 1980 - 1997, la notion de tempête étant celle définie au paragraphe précédent. La courbe comportant des triangles est la moyenne mobile d'ordre 3 de la série du nombre annuel de tempêtes, représentée quant à elle par une courbe à motifs constitués de ronds.

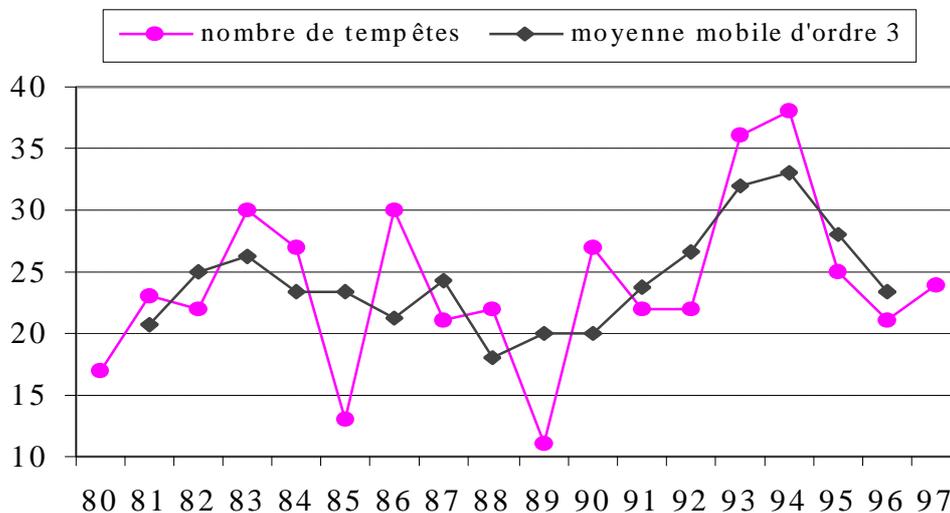


Figure 2.3.12 : Evolution annuelle du nombre de tempêtes (1980 – 1997)

Le nombre de tempêtes sur la période est de 431 ce qui représente une moyenne annuelle d'environ 24. L'année 1989 a été particulièrement peu venteuse avec 11 tempêtes, alors que 1994 s'est particulièrement distinguée dans l'autre sens, avec 38 tempêtes recensées.

On peut s'interroger sur le caractère stationnaire de cette série. Pour cela, on forme un test du χ^2 . Il consiste à comparer les valeurs observées chaque année à la valeur théorique moyenne calculée sur les 18 années d'une loi uniforme.

$$c^2 = \sum_{i=1}^N \frac{(n_{i\text{observé}} - n_{i\text{théorique}})^2}{n_{i\text{théorique}}}$$

où N est le nombre d'années (18 dans notre cas) et les n_i le nombre de tempêtes.

On compare la valeur χ^2 calculée à une valeur critique c_0^2 fournie par une table et fonction du nombre de degrés de liberté $\nu = N-1$ et du degré de confiance du test (ici égal à 95 %) : comme la valeur calculée (32,3) est supérieure à la valeur théorique (27,6), on rejette l'hypothèse d'une adéquation de la courbe du nombre de sinistres à la loi uniforme.

Si l'on cherche maintenant la droite de régression linéaire :

$$y = Tb$$

s'ajustant au mieux avec notre distribution annuelle du nombre de tempêtes, on trouve la droite d'équation :

$$y = 0,338.t - 5,92$$

où y et t désignent respectivement la série du nombre de tempêtes et l'année de survenance correspondante.

Il convient de vérifier le caractère significatif du coefficient directeur de la droite de régression linéaire, \hat{b}_1 . On forme alors le test suivant :

$$\begin{cases} H_0 : \hat{b}_1 = 0 \\ H_1 : \hat{b}_1 \neq 0 \end{cases}$$

Dans l'hypothèse H_0 , b_1 suit la loi de Laplace Gauss $N(\hat{b}_1, \sigma \sqrt{(T'T)^{-1}})$, et la variable t_{n-2} , définie par :

$$t_{n-2} = \frac{b_1 - \hat{b}_1}{\sqrt{\frac{\|y - Tb\|^2}{n-2} (T'T)^{-1}}}$$

suit une loi de Student à $n-2$ degrés de liberté, soit ici 16. Sa valeur empirique est égale à :

$$t_{n-2} = \frac{0,338}{6,85} = 0,049.$$

Or, le fractile d'ordre 0,975 d'une loi de Student, correspondant à une probabilité de 5 % que cette valeur soit dépassée si H_0 est vraie, est égal à : 2,12. On ne peut donc considérer le coefficient \hat{b}_1 comme significativement différent de zéro.

La série chronologique de 18 ans ne permet pas de conclure quant à une éventuelle évolution du nombre de tempêtes en France métropolitaine.

II.2.2.2 - Répartition mensuelle des tempêtes

La figure 2.3.13 représente la répartition mensuelle du nombre de tempêtes au cours de l'historique d'étude 1980 - 1997.

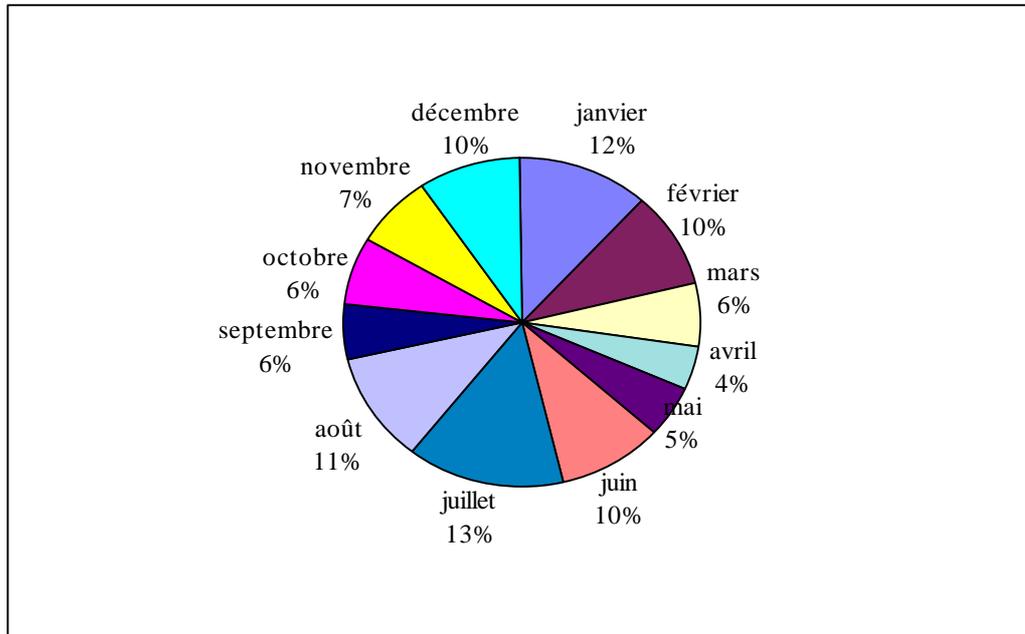


Figure 2.3.13 : Répartition mensuelle du nombre de tempêtes

Ce camembert montre une relative prédominance des mois d'été (31 % du nombre de tempêtes) et d'hiver (28 %) pour la survenance de tempêtes. Les saisons de l'automne (23 %) et du printemps (18 %) montrent une moindre propension à leur manifestation. Cette vision tranche avec celle de la figure 1.2.9 et issue des archives de Météo-France, où les mois d'été ne représentent que 10 % du nombre total de tempêtes. Il est probable que la manifestation de phénomènes météorologiques destructeurs et localisés, tels les orages, que Météo-France ne considère pas forcément comme des tempêtes, est une des explications principales de cette différence.

II.2.3 - Analyse de la distribution du nombre de sinistres des tempêtes

Une autre question intéressante est la distribution du nombre de sinistres des tempêtes : est-il possible de l'approcher par une loi de densité connue ? Un éventuel ajustement de la distribution empirique à une loi à un paramètre pourrait en effet permettre de déterminer la période de retour d'un événement rare.

La figure 2.3.14 représente le début de la distribution empirique du nombre de sinistres des tempêtes, les plus grandes tempêtes n'étant pas représentées; l'échelle utilisée est la dizaine de sinistres.

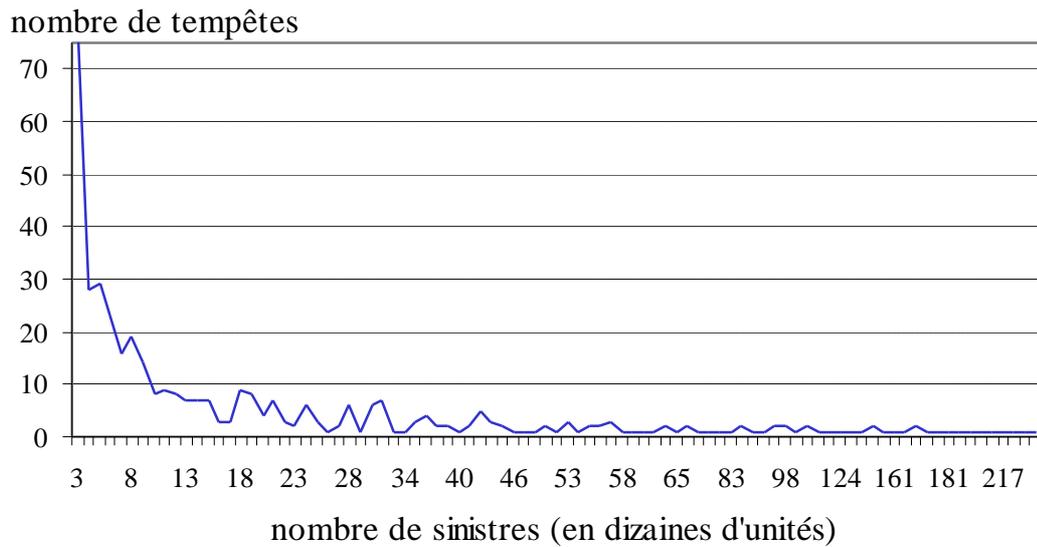


Figure 2.3.14 : Distribution du nombre de sinistres par tempête (1980 -1997)

Compte tenu de la forme de la courbe, il peut être tentant de tester l'adéquation de cette distribution empirique avec une loi de Pareto, dont l'écriture de la loi de densité est la suivante :

$$f(x) = \frac{a-1}{J} \left(\frac{J}{x}\right)^a, \text{ pour } x > \theta > 0 \text{ et } \alpha > 1$$

Par la méthode du maximum de vraisemblance, il est possible de déterminer le paramètre de la loi de Pareto α , θ étant la valeur à partir de laquelle la fonction est définie (ici, c'est égal au seuil d'intensité destructrice).

La fonction de vraisemblance s'écrit :

$$L(x_1, \dots, x_n; \mathbf{a}) = (a-1)^n J^{n \cdot (a-1)} \prod_{i=1}^n \left(\frac{1}{x_i}\right)^a,$$

et, si l'on prend le logarithme népérien de cette expression :

$$\ln L(x_1, \dots, x_n; \mathbf{a}) = n \cdot \ln(a-1) + n \cdot (a-1) \cdot \ln J - \sum_{i=1}^n a \ln x_i$$

$\hat{\mathbf{a}}$ est l'estimateur du maximum de vraisemblance s'il rend minimum la fonction en α :

$$n \cdot \ln(a-1) + n \cdot (a-1) \cdot \ln J - \sum_{i=1}^n a \ln x_i$$

c'est-à-dire si :

$$\frac{\partial \ln L(x_1, \dots, x_n, a)}{\partial a} = 0$$

$$\frac{\partial \ln L(x_1, \dots, x_n, a)}{\partial a} = \frac{n}{a-1} + n \cdot \ln J - \sum_{i=1}^n \ln x_i = 0$$

$$\text{D'où : } \hat{\alpha} = 1 + \frac{1}{\frac{1}{n} \sum_{i=1}^n \ln x_i - \ln J}$$

L'application numérique donne ici :

$$\hat{\alpha} = 1 + \frac{1}{2,69 - 1,09} = 1,62$$

Nous allons former le test du χ^2 pour voir si la décision de retenir la loi de Pareto de paramètres ($\alpha = 1,62$; $\theta = 3$) est valide. Le tableau 2.3.5 permet de comparer les distributions théorique et empirique après le partitionnement des observations en 22 classes, dont l'effectif dépasse le nombre d'unités requis (8).

Tableau 2.3.5 : Test du χ^2 d'ajustement du nombre de sinistres par tempête à une loi de Pareto

Tranche du nombre de sinistres (en dizaine d'unités)	effectifs réels	effectifs théoriques	$\frac{(n_i - n.p_i)^2}{n.p_i}$
3	81	71,0	1,40
4	28	46,9	7,64
5	29	33,8	0,67
6	23	25,7	0,28
7	16	20,3	0,92
8	19	16,6	0,35
9	14	13,8	0,00
10	8	11,7	1,20
11-12	17	19,0	0,21
13-15	21	21,0	0,00
16-18	15	15,4	0,01
19-22	22	15,3	2,93
23-26	12	11,5	0,02
27-30	15	9,0	3,96
31-40	21	16,0	1,53
41-50	17	10,7	3,70
51-60	15	7,8	6,77
61-80	8	10,6	0,65
81-100	10	7,1	1,22
101-150	10	10,6	0,04
151-300	15	13,0	0,31
301-	15	24,1	3,42
Total	431	431	37,2

Au seuil de 0,5 % de considérer à tort notre distribution empirique comme une loi de Pareto de paramètre 1,62, la table statistique du χ^2 donne, pour un nombre de degrés de liberté égal à 20 (i.e. nombre de classes (22) moins le nombre de relations entre les effectifs réels (1) et le nombre de paramètres estimés (1)), une valeur de 40 : on ne peut donc rejeter à ce seuil l'ajustement précédent. Notons en outre que c'est pour la valeur $\alpha = 1,60$ que la distance du χ^2 est la plus faible (36,8), ce qui renforce encore l'acceptation de l'ajustement.

La distribution du nombre de sinistres par tempête peut être ajustée par la loi de Pareto de paramètres (a = 1,62; q = 3).

La période de retour d'un événement rare peut alors être estimée comme l'inverse de la probabilité qui lui est attachée. Ainsi, si l'on considère la tempête la plus destructrice de l'étude, causant préjudice à plus de 37 790 assurés (à portefeuille constant), la probabilité rattachée à cet événement est donnée par :

$P_p(X > 3779)$

où P_p est la probabilité de la loi de Pareto de paramètre 1,62 et X est la variable aléatoire du nombre de sinistres par tempête. Comme l'on a :

$$P_p(X > 3779) = \left(\frac{J}{X}\right)^{a-1} = \left(\frac{3}{3779}\right)^{0,62} = 0,013 = 1,3\%$$

la période de retour d'un événement de cette intensité destructrice peut être estimée à environ 75 ans.

La tempête du 6 novembre 1982, la plus destructrice de la période 1980 - 1997, peut être considérée comme une tempête revenant tous les 75 ans.

II.3 - Proposition d'une tarification du risque tempête

Compte tenu des éléments précédents, la recherche d'une tarification du risque tempête pourrait paraître inutile : la faiblesse de la prime pure pour le risque d'habitation (71 F) n'encourage pas à mener plus avant l'étude.

Cependant, plusieurs raisons poussent à revoir cette position : d'une part, les cartes départementales ont souligné de fortes disparités entre les régions du centre et du nord et les régions périphériques. En outre, la prime pure tempête, qui représente entre 10 et 15 % de la prime incendie, fait partie de la prime de base, ce qui signifie qu'elle est payée par tous les assurés de la branche incendie et qu'elle représente un volume de chiffre d'affaires important.

Ainsi, la tarification du risque tempête peut s'appuyer sur les principes suivants :

- compte tenu des coûts de gestion supplémentaires qu'entraînerait la séparation complète des gestions incendie et tempête, ***la tarification de base reste celle de la garantie incendie;***

- une *surprime tempête* est appliquée à la prime incendie, et est calculée à partir de l'application d'un coefficient dépendant de variables tarifaires sélectionnées au préalable et dont les variations ont un impact sur le montant de la prime.

Ces variables, choisies parmi :

- le type de local (maison en agglomération ou en dehors, appartement),
- le type d'assuré (propriétaire occupant ou non, locataire),
- le niveau de norme de construction requise
- le type d'agglomération, etc..

font l'objet d'une analyse plus détaillée au prochain chapitre, relatif à la modélisation du risque tempête.

Le tarif comportera donc plusieurs cases tarifaires mais, au final, le chiffre d'affaires généré par cette surprime doit correspondre à la multiplication de la prime pure chargée par le nombre de contrats en portefeuille.

CHAPITRE 4 : MODELISATION DU RISQUE TEMPÊTE

Grâce aux travaux de création et d'actualisation des différentes bases de données relatives à la tempête, et dont l'exposé a été fait dans les chapitres 1 et 2 de cette partie, la modélisation du risque tempête est maintenant possible. Elle consiste en l'établissement d'un lien entre les données d'assurances (portefeuille, sinistres) et les données météorologiques mais aussi celles liées au type de construction. L'objectif étant la mesure des impacts sur le coût des sinistres des variations des facteurs présents dans le modèle, pour une nouvelle approche de la tarification.

I - Présentation du modèle

I.1 - Aspects théoriques

Comme explicité au chapitre 4 de la première partie, le risque tempête intrinsèque d'un bien (P) dépend intimement de l'*exposition* (E) à la tempête de ce bien et de sa *capacité* à y faire face. C'est à travers la prime pure, c'est-à-dire la charge actuelle probable par contrat, que le risque tempête de chaque bien garanti peut être mesuré. Cette relation physique qui lie la prime pure aux deux autres grandeurs s'écrit :

$$P = a. E. V. K \quad (1)$$

où V est l'inverse de la capacité à faire face à la tempête, c'est-à-dire la *vulnérabilité*, K est le *capital sous risque*, et a une constante positive.

Les grandeurs qui composent la notion d'exposition au risque tempête sont d'ordre météorologique ou géographique :

- l'intensité du vent maximal instantané,
- la direction principale du vent,
- la durée de la tempête,
- le niveau des précipitations,
- l'altitude (type de relief).

Les grandeurs qui composent la notion de vulnérabilité face au risque tempête sont liées à l'habitat et au type de construction :

- la norme de construction requise pour chaque région,
- la proportion de locataires et de propriétaires,
- le type d'habitat (appartement, maison, immeuble),
- l'importance de l'agglomération (ville, campagne).

Le capital sous risque est indépendant de l'exposition et de la vulnérabilité. Il est lié par la relation :

$$P = TxP \cdot K$$

où P et TxP représentent respectivement la prime pure et le taux de prime pure tempête.

La relation (1) peut être transformée et linéarisée en en prenant le logarithme népérien :

$$\ln P = A + \ln E + \ln V + \ln K \quad (2)$$

Il est donc tentant de rechercher une relation linéaire entre les données " sinistres ", dont on tente d'expliquer les variations, et les données des bases météorologique (E), " habitat " (V) ou " portefeuille " (K). Pour ce faire, le modèle linéaire suivant est adopté :

$$\begin{cases} Y = Xb + e \\ E(e) = 0 \quad V(e) = \mathbf{S}^2 I \\ V(e) = \mathbf{S}^2 I \end{cases} \quad (3)$$

On définit les grandeurs suivantes :

- $Y = (Y_1, \dots, Y_n)$ est le vecteur formé par l'ensemble des valeurs de la variable à expliquer, c'est-à-dire la valeur actuelle probable du préjudice moyen subi par chaque assuré du fait de la tempête. Y est de dimension $(n,1)$ où n représente le nombre d'individus étudiés. En l'occurrence, il s'agit ici du nombre d'événements tempête ayant frappé les biens d'un risque donné, pour une zone météorologique déterminée,

- X est la matrice $(n, p+1)$ formée par le vecteur unitaire et l'ensemble des valeurs mesurées des p Variables exogènes X_i :

$$X = (\underline{1}, X_1, \dots, X_p) = \begin{bmatrix} 1 & X_{11} & \dots & X_{1p} \\ 1 & X_{21} & \dots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \dots & X_{np} \end{bmatrix}.$$

En ce sens, on peut dire que les variables X_i ne sont pas aléatoires mais déterministes,

- e est le vecteur des résidus représentant les fluctuations non expliquées entre Y et $E(Y/X_1, \dots, X_p)$. On suppose en outre que chacun de ces résidus est une variable aléatoire d'espérance nulle, de même variance, indépendantes et identiquement distribuées,

- I est la matrice diagonale constituée de 1,
- $b = (b_0, b_1, \dots, b_p)$ est le vecteur constitué des coefficients du modèle.

Pour valider le modèle et en particulier ces coefficients par des tests, il est nécessaire d'intégrer une autre hypothèse relative aux résidus :

$$\forall i, 1 \leq i \leq n, e_i \sim N(0, \mathbf{S}^2)$$

L'estimation des coefficients b se fait par la méthode des moindres carrés, consistant à minimiser la distance entre une combinaison linéaire des variables X_i et la variable expliquée Y :

$$\hat{b} = (X'X)^{-1} X'Y$$

(Ceci est vrai sous réserve que X soit de plein rang).

La théorie nous enseigne de plus que cet estimateur est sans biais et, qui plus est, de variance minimale parmi tous les estimateurs de la forme bY .

Pour pouvoir juger du caractère significatif des composantes du vecteur \hat{b} , il est nécessaire de faire l'hypothèse de la normalité des résidus : celle-ci peut être facilement vérifiée par l'outil SAS. A cet effet, il peut être utilisé la procédure "UNIVARIATE" et l'option "NORMAL / PLOT".

I.2 - Contraintes liées au modèle

I.2.1 - Incohérence des dates et des lieux de tempête

La constitution du modèle nécessite la recherche des éléments des différentes bases de données relatives à chacune des tempêtes décrites par Météo-France. Ce rapprochement est réalisé avec le concours du code postal et de la date de mesure de la grandeur, deux variables présentes dans chacune des bases.

Néanmoins, deux écueils se présentent lors de cette opération :

- d'une part, la non concordance des dates de survenance des événements tempête,
- d'autre part, la valeur souvent approximative des dates de survenance dans la base "sinistres" (par exemple, la journée du 29 février 1990 apparaît à plusieurs reprises).

Une interrogation naît de la comparaison des dates de survenance des événements tempête puisqu'il n'est pas rare de trouver des événements ayant provoqué de gros dommages mais absents des relevés de Météo-France : on peut ainsi citer la " tempête " du 23 juillet 1988, qui a occasionné des dégâts à quelque 2.000 risques de l'Essonne, du Val-de-marne, de la Marne et des Ardennes, mais dont Météo-France ne dit mot dans le fichier recensant tous les événements d'une durée maximale de 72 heures.

Le caractère très isolé du phénomène climatique, comme peuvent l'avoir les tornades, conjugué à l'éloignement de la plus proche des stations, constitue une explication satisfaisante de l'absence de relevés météorologiques. D'ailleurs, il est vrai que le maillage des stations météorologiques dans le centre de la France, où ces manifestations atmosphériques sont courantes, est de moindre densité, ce qui rend plausible cet argument.

Mais il peut s'agir aussi d'autres phénomènes atmosphériques qui se distinguent de la tempête : orage, fortes pluies, grêle, chutes de neige, etc.... Le déchaînement des éléments et leur caractère brusque et inattendu font que, bien souvent, les régleurs des compagnies d'assurances les assimilent à des tempêtes. On sait que l'expression " tempête de neige " est souvent utilisée, non pour signaler la présence de vent, mais pour mettre l'accent sur l'intensité des chutes de neige. Cette difficulté à définir la tempête est encore renforcée par le fait que, dans les contrats, la référence à une vitesse minimale (100 km / h pour l'UAP) a disparu, seule une notion de dommages alentour étant retenue. La définition de la tempête par l'UAP prend ici une acception large, englobant la grêle et l'orage, alors que la définition des météorologues répond à des critères précis (on se reportera au paragraphe II.3 du chapitre 2 de la première partie).

En résumé, l'absence de relevés météorologiques pour certains événements présents dans la base " sinistres " est souvent la conséquence de la définition très large de la tempête retenue par l'UAP : le réseau d'instruments de mesure de Météo-France a décelé ce phénomène mais ne le définit pas comme une tempête.

Dans cette situation, il n'est pas possible de conserver dans l'étude ces sinistres dont l'origine ne peut être attribuée à la tempête.

Il se peut d'autre part que l'inverse se produise, une tempête ayant été décelée par Météo-France, mais celle-ci n'ayant pas causé de dommages au portefeuille : ces situations apparaissent dans des régions peu peuplées ou habituées aux vents forts. Ces données météorologiques sont à maintenir dans l'étude, même en l'absence de données issues de la base " sinistres ", car il s'agit d'observations pour lesquelles le vent n'est pas nul.

I.2.2 - Hétérogénéité des zones météorologiques

Toutes les données collectées étant ventilées par zone postale, le regroupement par zone météorologique ou affectée par les dommages conduit inévitablement à des associations de données hétérogènes. Cette difficulté se pose avec acuité lorsque le type d'agglomération ou la norme de construction diffèrent entre les différentes zones postales d'un même périmètre météorologique ou sinistré, ou, par exemple, lorsque la proportion de locataires y varie fortement.

Néanmoins, ces cas litigieux restent l'exception et c'est par l'utilisation de la pondération démographique de chaque zone postale constituant la zone sinistrée qu'est déterminé le niveau de norme de construction finalement retenu.

Supposons pour illustrer nos propos que les trois quarts de la population d'une zone sinistrée soient situés dans une région faiblement exposée aux vents, où les normes de construction requises sont peu strictes, et que le quart restant se trouve dans une région de vent moyennement exposée. L'affectation arbitraire d'une valeur numérique mesurant l'exposition au vent, égale à 1 pour les zones postales situées dans la région faiblement exposée et 2 pour les autres, conduit à la valeur 1,25. La zone météorologique est donc à rattacher *en globalité* à la région des vents dits faibles.

I.2.3 - Exposition au risque distincte au sein du portefeuille habitation

Une autre difficulté est la manière d'intégrer dans le modèle la segmentation des contrats du portefeuille d'habitation : cette introduction s'avère utile car chaque élément de cette segmentation est crédité d'un niveau d'exposition au risque théoriquement bien déterminé (cette question est l'objet du paragraphe IV.3 du premier chapitre de la deuxième partie). Comment peut-on rendre, à travers quelques indicateurs, cette richesse d'information de la manière la plus juste possible ?

La solution, consistant à attribuer un niveau théorique d'exposition au risque à chaque segment et, compte tenu du poids respectif de ces derniers, à en déterminer un niveau global d'exposition au risque, n'est pas satisfaisante. En effet, son caractère réducteur conduit à une échelle des valeurs très limitée reflétant imparfaitement le spectre des expositions entre les différentes zones météorologiques.

La solution qui a notre préférence est celle qui respecte un peu mieux la variété des situations à l'intérieur de la zone. Ainsi, il est utile de conserver les deux notions suivantes :

- la *qualité de l'assuré* (propriétaire occupant ou non occupant, locataire, copropriétaire),
- le *type de locaux* (maison en agglomération ou en dehors, immeuble).

Ces deux variables présentent en effet l'avantage d'être relativement indépendantes et d'admettre des contrats dans chacune des cases de leur tableau de contingence, résultat du croisement de leurs modalités. Ainsi, on peut très bien imaginer une zone météorologique présentant une grande proportion de propriétaires habitant dans une maison en dehors d'une agglomération, alors qu'une autre sera richement dotée en locataires habitant dans un immeuble.

Il reste alors à affecter chaque modalité des deux variables à une échelle d'exposition au risque, déterminée à partir de l'analyse des risques réalisée au cours du chapitre 4 de la première partie, paragraphe II.2. De la sorte, il sera considéré que, pour la variable " qualité de l'assuré ", le propriétaire (ou le copropriétaire) présente un niveau 2 sur l'échelle de l'exposition au risque, alors que le locataire ne présente qu'un moindre risque, de niveau 1. Il est fait la même opération pour la variable " type de locaux " : l'immeuble présente un degré faible (1) sur l'échelle de l'exposition au risque, alors que la maison en agglomération est à un niveau moyen (2) et la maison hors de l'agglomération présente une exposition au risque plus élevée (3).

Le choix de valeurs numériques arbitraires pour la mesure du degré d'exposition est dicté par la nécessité de calculer l'exposition au risque moyenne, présentée par les assurés et par le type de locaux, pour une zone météorologique donnée. En effet, connaissant les proportions respectives présentées par chacune des modalités d'une variable, on peut en déduire ainsi la valeur moyenne de chaque zone météorologique. La palette de ces valeurs permet de différencier les zones selon l'exposition au risque tempête des populations ou des locaux assurés.

I.2.4 - Suppression de l'effet de taille pour les phénomènes localisés

Il arrive qu'une tempête ne frappe qu'une partie réduite du territoire couvert par la station météorologique, le reste de ce territoire n'étant pas atteint. Cette situation apparaît dans le cas de tempêtes prenant un aspect orageux le plus souvent dans des régions escarpées, où le relief contrarie la trajectoire des flux atmosphériques. L'espérance de la charge actualisée par contrat prend alors une composante due à l'*effet de taille* - qui n'existe évidemment pas au niveau de la zone postale, en plus du taux de destruction, principal facteur explicatif de la charge par contrat.

Dans le cas de deux tempêtes (symbolisées par les lettres *a* et *b* sur la figure 2.4.1) provoquant un montant de dommages identique mais sur une zone respectivement réduite ou étendue, le taux de destruction global est le même mais l'existence d'un *effet de taille* explique de grandes disparités dans la distribution spatiale de ce taux de destruction. La tempête *a* a un pouvoir destructeur très fort, comme en témoigne la densité du gris de la zone sinistrée, et très limité, alors que la tempête *b* touche indifféremment et de façon homogène toute la zone météorologique.

Il est souhaitable que ces deux tempêtes disposent de composantes distinctes dans la matrice *X* du modèle linéaire adopté. Pour s'en convaincre, il suffit de reprendre la relation (2) du paragraphe I.1 : dans le cas présent, puisque les deux tempêtes

présentent une égalité du terme de gauche de la relation ($\ln P$), et que leur terme respectif lié à l'exposition ($\ln E$) est différent (i.e. les paramètres météorologiques ne sont pas identiques), alors la vulnérabilité des biens assurés ($\ln V$) étant la même, il importe donc que le terme relatif au montant des capitaux assurés ($\ln K$) diffère.

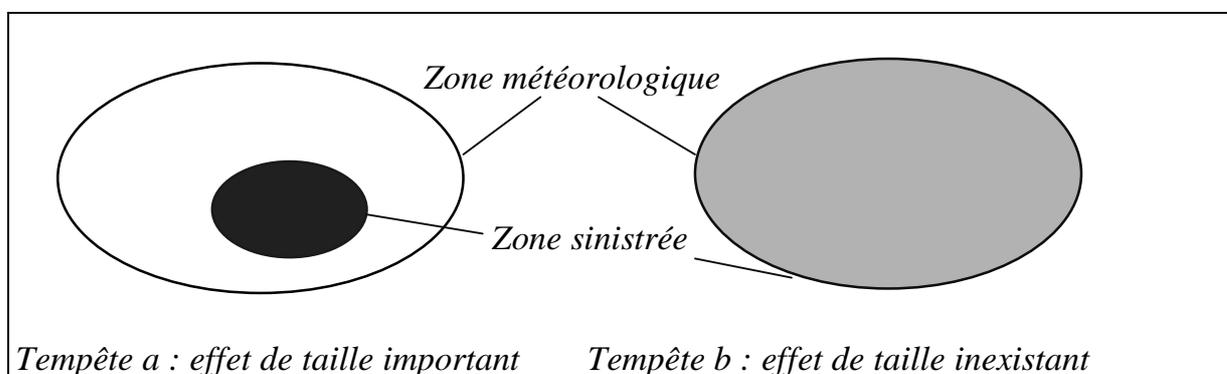


Figure 2.4.1 : Comparaison de deux tempêtes au coût identique mais touchant partiellement ou totalement une zone météorologique

Dans ce but, on forme, pour chaque tempête, le rapport du capital assuré sur les seules zones postales affectées par la tempête au capital total : cet indicateur permet d'évaluer le caractère localisé de la tempête et de mesurer l'importance de l'effet de taille.

Ces préliminaires étant posés, il convient dès lors de procéder à la modélisation du risque tempête pour pouvoir juger de la réalité, sur le plan quantitatif, de la typologie décrite par Météo-France.

I.2.5 - Détermination et actualisation du capital sous risque

Nous avons introduit dans ce chapitre, par l'intermédiaire de la relation (1), la notion de capital sous risque, pouvant être défini comme le montant des biens exposés au risque tempête. Sa présence dans la relation (1) n'est pas sans poser quelques problèmes. En effet, cette grandeur est absente des fichiers informatiques à notre disposition. Cependant, à un niveau d'agrégation suffisant, telle la zone météorologique, on peut considérer comme plausible l'hypothèse d'une proportionnalité du capital sous risque et du montant des primes émises nettes d'annulations : en effet, la majeure partie du chiffre d'affaires est constituée de la part afférente à l'incendie, à laquelle la prime tempête est proportionnelle.

Le capital sous risque étant donc supposé proportionnel au chiffre d'affaires de la branche incendie, la linéarisation de la relation (1), par l'intermédiaire du logarithme népérien conduit à la relation (2'), finalement opérationnelle :

$$\ln P = A + \ln E + \ln V + \ln P \quad (2')$$

où P désigne le montant des primes émises nettes d'annulations de la branche incendie.

D'autre part, il convient d'actualiser le terme " $\ln P$ " du fait de l'inflation, comme l'a été le terme " $\ln P$ ", pour que les primes de chaque exercice soient comparables. Pour cela, il est pris comme référentiel l'indice F.N.B. à la date de fin 1997. Les opérations de calcul du préjudice qui ont été menées pour l'estimation du terme " $\ln P$ " (et dont l'exposé a été fait au chapitre 2) n'ont bien sûr pas de sens pour le terme " $\ln P$ ".

II - Validation de la typologie des tempêtes définie par Météo-France

Au cours du chapitre 2 de la première partie a été présentée la méthode définie par Météo-France, conduisant à une typologie des tempêtes, et basée sur une analyse discriminante relativement aux valeurs prises par ces tempêtes pour certaines variables explicatives. Cette typologie a identifié sept types de tempêtes caractéristiques aux appellations suivantes :

- nord (N),
- ouest (W),
- sud (S),
- orage (OR),
- sud-ouest (SW),
- est (E),
- nord-est (NE).

Si cette typologie est riche d'enseignements sur le plan de la connaissance des caractéristiques qualitatives des tempêtes, elle ne répond en aucune manière à l'interrogation des assureurs sur l'intensité des dommages : rien ne permet de dire en effet qu'un type de tempête se distingue des autres pour l'intensité des dommages qu'il occasionne.

Pour tenter de répondre aux questions suivantes :

- quel type de tempête provoque les plus gros dégâts ?
- quel type de tempête est le plus dangereux ?
- quels sont les types de tempêtes les plus proches ?

il sera mené, dans un premier temps, une analyse descriptive de ces différents types puis, dans un deuxième temps, une analyse économétrique. Mais avant toute chose, pour pouvoir caractériser ces tempêtes, doit être menée l'opération d'appariement des

événements de Météo-France et des tempêtes du fichier “ sinistres ” reposant sur la comparaison des dates de survenance.

II.1 - Appariement des événements de Météo-France et des tempêtes du fichier “ sinistres ”

Météo-France a extrait de ses archives, sur la période de 1962 à 1993, la liste des 524 événements d’une durée maximale de 72 heures qui correspondent à la définition exposée au chapitre 2, paragraphe II.3 de la première partie, d’une tempête. A la demande des rédacteurs de ce mémoire, la série a pu être complétée pour les années 1994 à 1996.

Cependant, les bases de données étant limitées à la période 1980 - 1997, l’analyse descriptive des différents types de tempêtes se limite à 1980 - 1996, période sur laquelle Météo-France a recensé 293 événements.

Au paragraphe II.2 du chapitre 3, une définition de la tempête est proposée, intégrant en particulier la notion d’intensité destructrice journalière minimale, le seuil étant fixé à 30 sinistres par jour, et celle de durée maximale de l’épisode venteux (3 jours). Si l’on se conforme à cette définition, 394 tempêtes sont ainsi identifiées.

Cependant, le rapprochement des événements de Météo-France d’une part, et de ces tempêtes du fichier “ sinistres ” d’autre part, ne conduit à l’appariement que de 167 couples événements - tempêtes, les autres événements étant indétectables dans le fichier “ sinistres ”.

Deux raisons expliquent cette situation :

- d’une part, le décalage des dates de survenance retenues par Météo-France et par les assureurs,
- d’autre part, un niveau de destruction inférieur à 30 sinistres par jour.

Pour y remédier, il est procédé à l’abaissement à 15 sinistres par jour du seuil d’intensité destructrice et à une étude visant à repérer les décalages. Au total, 223 événements sont finalement identifiés dans le fichier “ sinistres ” entre 1980 et 1996, les autres restant indétectables.

Il existe aussi une situation pour laquelle ces règles d’appariement ne conduisent pas aux résultats escomptés, lorsque l’épisode venteux d’une durée supérieure à trois jours conduit à la génération de plusieurs tempêtes; or, Météo-France ne recense quant à lui qu’un événement. Ce cas peut être illustré par l’événement des 15 et 16 octobre 1987, réparti entre les tempêtes 266 et 267 dans le fichier “ Sinistres ”. Les règles de

l'appariement doivent donc être légèrement amendées en procédant au regroupement des deux tempêtes.

Il est à noter que la situation inverse, pour laquelle l'événement donnerait lieu à plusieurs relevés de Météo-France sur une période pendant laquelle une seule tempête aurait été recensée dans le fichier " Sinistres ", ne peut se produire : en effet, les relevés de Météo-France sont extraits du fichier des événements dont les dates de survenance sont au moins séparées de 72 heures.

II.2 - Analyse descriptive de la typologie des tempêtes

Le tableau 2.4.1 fournit les statistiques globales par type de tempête, permettant de mieux situer leurs caractéristiques premières.

Tableau 2.4.1 : Caractéristiques de la sinistralité des différents types de tempêtes définis par Météo-France

Rubrique	Type de tempête							Total
	E	N	NE	OR	S	SW	W	
Nombre de tempêtes	13	69	16	20	37	26	42	223
Nombre de sinistres	2 976	51 280	2 104	27 283	52 314	73 160	32 740	241 857
Nombre de sinistres par tempête	229	743	132	1 364	1 414	2 814	780	1 085
Nombre de zones postales sinistrées	24	77	29	74	79	95	107	78
Superficie (km ²)	2 149	7 905	2 760	9 645	8 849	9 182	11 116	8 267
Montant total des dommages (MF)	41,9	661,1	20,8	658,4	839,5	1 165,1	372,7	3 759,5
Montant moyen des dommages par tempête (MF)	3,2	9,6	1,3	32,9	22,7	44,8	8,9	16,9
Montant moyen des dommages par sinistre (F)	14 095	12 893	9 880	24 130	16 047	15 926	11 383	15 544
Indice de densité	259	202	259	130	131	168	203	169
Taux de destruction	4,9 %	7,4 %	2,4 %	13,5 %	10,8 %	22,4 %	6,2 %	10,1 %

Le tableau précédent est très instructif puisqu'il nous renseigne sur la fréquence de ces différents types de tempête et sur la forme qu'ils prennent en terme :

- d'importance, exprimée en nombre et en montant de sinistres,
- d'intensité destructrice,
- de superficie couverte.

Leur *fréquence d'apparition* est inégalement distribuée, puisque le type N est cinq fois plus fréquent que le type E. Le *nombre de sinistres* générés par ces 223 tempêtes est également très inégal selon les types, le type SW en comptant plus de 70.000 contre un peu plus de 2.000 pour le type NE.

C'est le type de tempête SW qui a causé sur la période d'étude, avec plus d'un milliard de F, le plus gros *montant de dommages*, les types S, N et OR se révélant aussi très onéreux. A l'inverse, les types E et NE présentent la facture la moins lourde.

L'importance de chaque type de tempête peut être définie par le *nombre de sinistres par tempête*, lui aussi très variable d'un type à l'autre : le type SW génère en moyenne environ 2.800 sinistres par événement, contre un peu moins de 140 pour le type NE. La taille de chaque type de tempête s'exprime aussi en valeur par l'intermédiaire du *montant moyen des dommages par tempête* : les types SW et OR sont les plus coûteux avec respectivement 44,8 et 32,9 millions de F.

Deux composantes contribuent à renforcer l'importance d'une tempête : d'une part, son intensité destructrice, et, d'autre part, la superficie de la région qu'elle frappe. L'intensité destructrice se définit ici par le *taux de destruction*, dont la définition est la suivante :

$$\text{Taux de destruction} = \frac{\text{nombre de sinistres}}{\text{nombre de contrats de la zone atteinte}}.$$

Le taux de destruction atteint un niveau bien plus élevé pour les tempêtes de type SW (22,4 %) et OR (13,5 %) que pour le type NE dont le niveau n'atteint que 2,4 % du nombre de contrats de la zone atteinte par la tempête.

Enfin, la *superficie* est l'autre axe selon lequel une tempête peut expliquer sa taille. Elle donne une idée de l'aire moyenne couverte par l'événement tout au long de sa manifestation. Il s'agit d'une moyenne quotidienne, et non de la superficie maximale atteinte par l'événement. L'étendue de cette tempête, exprimée en kilomètres carrés, peut aussi être définie en nombre moyen quotidien de zones postales atteintes. Comme pour les autres indicateurs, les types de tempête se distinguent par la superficie couverte : les types OR, SW et W se développent sur une superficie voisine de 10.000 km², ce qui représente environ l'étendue de deux zones météorologiques, alors que les types E et NE ne couvrent qu'environ 2.500 km², soit seulement la moitié d'une zone météorologique.

De l'examen du *montant moyen des dommages par sinistre*, il ressort une différence notable entre le type de tempête OR, dont la moyenne culmine à plus de 24.000 F, et les autres types, dont le montant reste compris entre 10.000 F et 16.000 F.

Enfin, le tableau précédent introduit *l'indice de densité*, grandeur précisant le niveau de densité démographique de la zone atteinte par la tempête et dont l'analyse peut se révéler riche d'enseignements : en effet, au chapitre 4 de la première partie, l'hypothèse a été faite que le type d'agglomération n'est pas un facteur sans influence sur le montant du préjudice provoqué par la tempête, les bâtiments étant moins exposés en agglomération qu'en dehors. Il n'est donc pas vain d'étudier cette grandeur dont les écarts constatés entre les différents types de tempêtes peuvent expliquer les différences de niveau pour la prime pure, le taux de destruction ou le coût moyen.

Il reste néanmoins à en donner une définition admissible, ce que le rapport du nombre d'habitants à la superficie des zones postales atteintes par la tempête n'est pas. En effet, cette densité privilégie les zones postales les plus peuplées au détriment de celles qui le sont moins, car il suffit d'un seul contrat sinistré issu de cette zone postale peuplée pour que cette dernière soit intégrée au calcul de la densité. Or, comme nous l'avons vu au paragraphe II.3.1 du chapitre 1 de cette présente partie, il n'est pas rare que le code postal d'un sinistre " délocalisé " trouve comme point final de localisation le chef-lieu d'un département, commune en général la plus peuplée d'un département.

Une définition acceptable de la densité doit donc tenir compte de la *fréquence* respective de chaque zone postale sinistrée. La formule retenue pour la densité d est la suivante :

$$d = \frac{\sum_{i=1}^{n_{zp}} f_i \cdot d_i}{f_i}, \text{ où :}$$

- n_{zp} désigne le nombre de zones postales atteintes par la tempête,
- f_i et d_i représentent respectivement la fréquence de sinistres et la densité de la zone postale n° i .

L'indice de densité étant supérieur de près de 50 % à la densité démographique de la France (169 contre environ 100 habitants par km²), il est probable que la pondération utilisée ne limite qu'en partie l'influence des zones postales densément peuplées : on ne peut en effet imaginer que les tempêtes frappent plus fréquemment les villes que les campagnes. Néanmoins, cet indicateur présente des différences très marquées selon les types de tempêtes : si les types E et NE ont frappé des zones plus densément peuplées (indice égal à 259), en revanche les types OR et S (indice proche de 130) ont atteint des zones plus rurales.

II.3 - Recherche d'une quantification de la typologie des tempêtes

Au travers des premières analyses du tableau 2.4.1, le paragraphe précédent nous a permis de comprendre la nécessité de “ décorrélérer ” la fréquence ou le coût moyen des tempêtes des influences de la densité, de la répartition du portefeuille entre locataires et propriétaires, du type d'habitat (maison individuelle ou immeuble collectif), ou des normes de construction. Ce paragraphe doit nous permettre d'aller au-delà des premières impressions et de répondre au style de question suivante : à vitesse et à niveau de concentration d'assurance identiques, quel type de tempête représente le péril le plus important ?

Pour tenter de répondre à la question relative à l'intensité des dommages, il convient d'intégrer dans le modèle :

- les 193 événements tempêtes identifiés dans le fichier “ sinistres ”, quel que soit leur type, entre 1980 et 1993,
- les variables explicatives météorologiques décrivant le phénomène de manière *globale*, et dont la liste figure au paragraphe III.1.1 du chapitre 1; rappelons pour mémoire que l'on y trouve la pression moyenne sur le quart sud-est, sa localisation à travers les latitudes du maximum et du minimum de pression, son intensité et son étendue par la mesure du vent moyen sur les quatre parties du quadrant de la France, la survenance d'autres phénomènes atmosphériques l'accompagnant, tels que l'orage, la neige ou la grêle, et leur importance à travers le nombre de stations touchées,
- de même qu'une variable qualitative, dont chaque modalité correspond à un type de tempête.

On retrouve alors un modèle d'analyse de variance, forme particulière du modèle linéaire général lorsque certaines variables explicatives, appelées *facteurs*, ne sont pas quantitatives, mais admettent au contraire des catégories connues sous le terme de *niveaux*. Notons que ce modèle est équivalent à celui dans lequel on intégrerait sept variables indicatrices des catégories. L'écriture (3) du modèle devient alors :

$$Y = Xb + \mathbf{a}_i + e \quad (4)$$

où \mathbf{a}_i est l'effet principal du niveau i du facteur A “ type de tempête ”, et X la matrice des autres “ prédicteurs ”.

Pour pouvoir identifier les paramètres du modèle (4), il est nécessaire d'imposer une contrainte sur les paramètres :

$$\sum_{i=1}^7 n_i \cdot \mathbf{a}_i = 0$$

où n_i représente l'effectif de l'échantillon des tempêtes classées dans le type i . Il est aussi d'usage de retenir la contrainte :

$$\mathbf{a}_7=0,$$

qui correspond à l'option par défaut retenue par la procédure GLM du logiciel SAS.

Ensuite, on montre que les différents types de tempêtes ont un impact réellement différent sur la charge actualisée des sinistres si les paramètres $\mathbf{a}_1, \dots, \mathbf{a}_7$ sont distincts deux à deux. Il convient en d'autres termes de former le test suivant :

$$\begin{cases} H_0 : \mathbf{a}_1 = \mathbf{a}_2 = \dots = \mathbf{a}_6 = \mathbf{a}_7 = 0 \\ H_1 : \exists i, j \mathbf{a}_i \neq \mathbf{a}_j \end{cases}$$

On suppose ici que les résidus suivent une loi normale $N(0, \sigma^2)$ pour pouvoir effectuer le test. Si H_0 est vraie, on sait qu'alors la relation suivante est vérifiée :

$$\frac{\frac{S_A^2}{k-1}}{\frac{S_R^2}{n-k}} \sim F(k-1; n-k), \text{ où :}$$

- S_A^2 et S_R^2 représentent les deux termes de l'équation de la variance, respectivement la variance due au facteur A et la variance résiduelle,

- F est la variable de Fisher - Snedecor, c'est-à-dire le rapport de deux lois du χ^2 à $k-1$ et $n-p-k$ degrés de liberté respectivement, k étant ici égal à 7 et p étant le nombre des autres variables explicatives.

On forme le rapport $\frac{S_A^2/k-1}{S_R^2/n-p-k}$, que l'on compare à la valeur critique de la variable de Fisher - Snedecor au degré de confiance de 95 % (1,60). Si l'on modélise le montant de la charge actualisée d'une tempête, l'application numérique donne :

$$\frac{S_A^2/k-1}{S_R^2/n-k} = \frac{5,62 \cdot 10^{16} \cdot 182}{4,26 \cdot 10^{17} \cdot 6} = 4,00$$

Le rapport des variances pondérées par les nombres est supérieur à la valeur critique, ce qui signifie que les coefficients $\mathbf{a}_1, \dots, \mathbf{a}_7$ sont globalement différents de zéro : ***on en déduit que le type de tempête influence effectivement le montant de la charge actualisée par tempête.***

On trouvera en annexe M les détails de l'analyse de la variance dont la réalisation a été faite grâce à la procédure GLM du logiciel SAS. On trouvera en outre le

programme CUMNBSIN écrit sous SAS et assurant le rapprochement des événements définis par Météo-France et des tempêtes définies par nos propres soins.

III - Utilisation du modèle pour une nouvelle approche tarifaire

L'objectif de ce paragraphe est l'établissement d'un lien, établi à un niveau local, entre les données d'assurances et les données météorologiques, de manière à évaluer l'impact réel des paramètres météorologiques sur le montant des dommages. De même, l'influence précise de variables présentes dans les bases " habitat " et " construction " constitue un point évoqué au chapitre 3 et resté en suspens : il peut trouver dans ce présent paragraphe une réponse plus quantitative.

III.1 - Etude de l'impact local de la tempête

Nous reprenons le modèle (3) et en proposons une nouvelle utilisation pour l'adapter à notre objectif : celui d'une tarification du risque tempête basée sur l'utilisation des distributions bidimensionnelles par station météorologique observées sur une longue période.

Cette approche locale de la tempête doit en outre permettre de juger sur pièces la manière dont les conditions particulières d'une zone météorologique interfèrent dans le résultat final : la concentration des biens assurés, le type d'agglomération, les normes requises de la construction, la proportion de locataires sont autant de facteurs locaux dont l'impact sur la prime pure ne peut être négligé. En adoptant cette vision de proximité, il est certain que les paramètres estimés du modèle en tiennent compte.

III.1.1 - Présentation du modèle d'analyse par station

III.1.1.1 - Détermination des événements

Le modèle d'analyse, procédant au découpage des tempêtes selon 130 zones météorologiques du territoire métropolitain, rassemble au total 134 événements :

- les 91 plus anciens (1980 - 1993) sont extraits du fichier que les assureurs et réassureurs avaient communiqué pour alimenter l'étude de Météo-France de 1995,
- les 43 plus récents (1994 - 1996) ont été commandés par les auteurs de ce mémoire à Météo-France.

Il est à noter que les 91 premiers événements sont plus facilement détectables dans le fichier “ sinistres ” que les 43 suivants, car ils font référence à des tempêtes ayant entraîné des dommages pour les assurés.

La technique d’identification des événements de Météo-France dans le fichier “ Sinistres ” utilise la même méthode d’appariement que celle présentée au paragraphe II : il est nécessaire que le seuil limite de 15 sinistres survenus par jour soit dépassé pour qu’une journée soit considérée comme venteuse et intégrée à une tempête, cette dernière ne pouvant durer plus de 72 heures.

Sur la période de janvier 1980 à juin 1996, 625 tempêtes sont ainsi extraites du fichier “ Sinistres ” mais, suite au regroupement de 10 d’entre elles, 615 seulement sont retenues.

L’étape suivante consiste en l’étude des dates retenues par Météo-France pour situer un événement dans le temps. Cependant, on note pour 9 événements une anticipation d’un ou de deux jours de ces dates par rapport à la date de survenance du “ pic ” de sinistres. Dans ce cas, le rapprochement d’un événement décalé et d’une tempête ne saurait se faire sans avoir au préalable apporté la preuve de la similitude de lieux frappés par l’épisode venteux pour Météo-France et la base “ sinistres ” : en d’autres termes, on vérifie que la station météorologique de la région la plus durement touchée par la tempête indique une vitesse de vent supérieure à la normale.

L’appariement des événements de Météo-France et des tempêtes précédentes donne finalement lieu à l’identification de 119 événements sur 134, caractérisés par une date de début et une date de fin d’événement. Une évaluation du montant local du préjudice occasionné par les tempêtes est alors possible.

Une alternative à cette méthode eût été d’établir *localement*, c’est-à-dire pour chaque zone météorologique, une liste de tempêtes dès lors que le seuil de sinistres quotidien, ramené à un niveau plus faible (de l’ordre de 5 sinistres par jour), eût été dépassé. Mais cette technique, en limitant la durée des tempêtes par rapport à la méthode *nationale*, renforce le problème de décalage de date et limite sensiblement la liste locale des tempêtes. Au contraire, les dates retenues étant communes à l’ensemble des zones météorologiques, la méthode nationale offre l’avantage d’une définition unique des tempêtes. En outre, elle présente l’intérêt de proposer à la modélisation des événements pour lesquels les dommages sont inexistantes, et donc d’enrichir de manière sensible les modèles, et de pouvoir établir une vitesse seuil de déclenchement des dommages.

III.1.1.2 - Choix final des stations météorologiques

Au paragraphe III du premier chapitre, la question du choix des stations météorologiques a été abordée, un juste équilibre devant être trouvé entre la précision des relevés météorologiques, renforcée par un nombre maximal de stations, et un poids démographique suffisant des zones retenues. La suppression des stations situées sur les

îles ou les côtes ou à l'historique trop récent a d'abord conduit à une première sélection de 130 stations météorologiques sur 168.

Puis, à la lumière de l'historique du fichier " sinistres ", certaines stations, ne disposant pas du relevé relatif à des événements importants, n'ont pu être conservées. Cette mesure s'appliquant si plus de 25 % du nombre de sinistres rattachés à une station se trouvent exclus du fait de l'absence de relevés météorologiques. Cette règle conduit à la suppression de 21 stations, dont une part importante se situent dans le centre de la France.

En revanche, les stations d'Istres et du Bec de l'Aigle (Bouches-du-Rhône), qui n'ont pas été présélectionnées parce qu'elles appartiennent à une région au quadrillage dense, mais dont l'historique est complet, ont finalement pu être repêchées en remplacement de stations alentour ne respectant pas les critères de sélection.

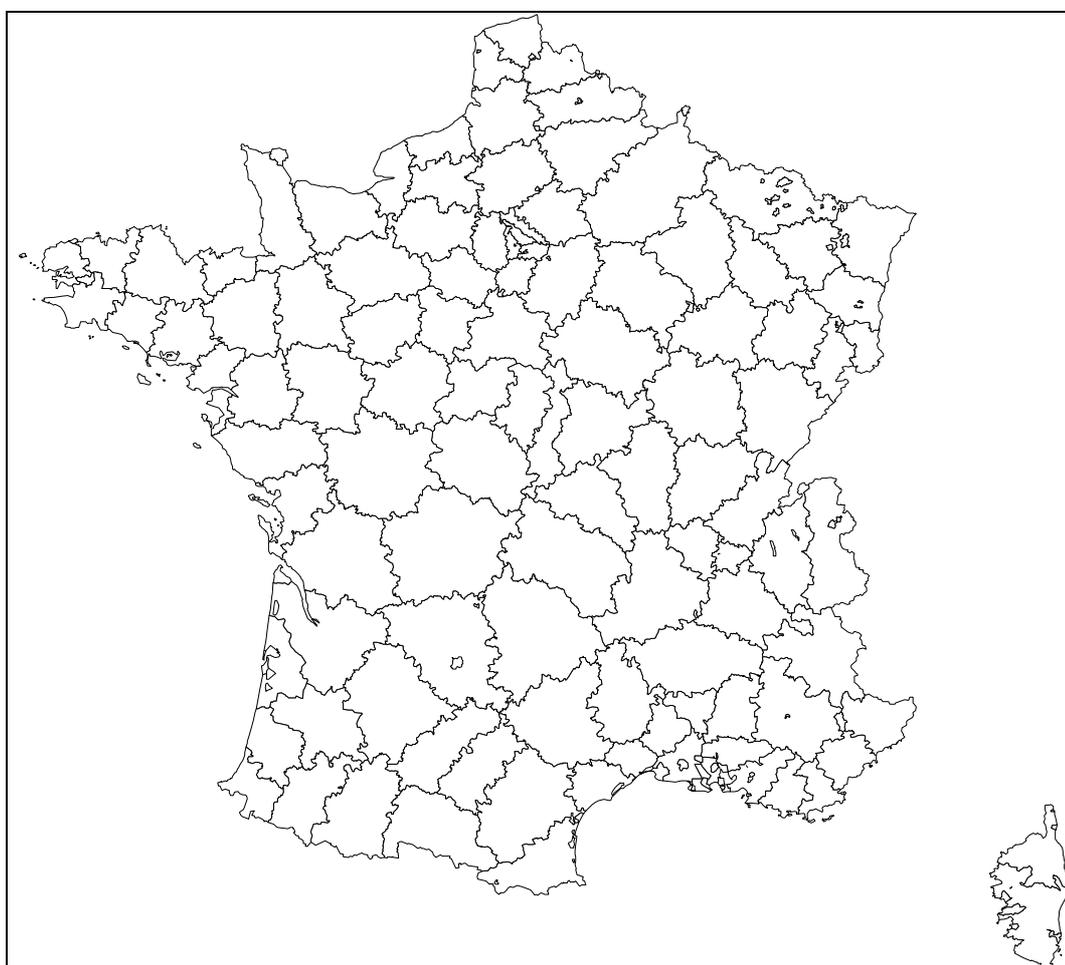


Figure 2.4.2 : Contour des 111 zones météorologiques finalement retenues

Le nombre de stations finalement retenu est de 111. La distance moyenne entre chaque zone postale et la station la plus proche, pondérée par le poids démographique

de la zone postale, passe de 18,3 km pour 130 stations à 21,1 km, alors que la distance équipondérée passe de 25,0 km à 28,1 km.

La figure 2.4.2 donne la carte de France des 111 zones météorologiques finalement retenues pour la modélisation.

III.1.1.3 - Construction des modèles de régression linéaire

Les 119 événements tempête représentent, tous risques confondus, plus de 216.000 sinistres et 3,2 milliards de francs actualisés à la fin 1997. On y trouve donc les trois-quarts des sinistres, parmi lesquels figurent les plus importants épisodes venteux.

Ces sinistres sont répartis par événement, mais aussi par zone météorologique et risque (agricole, habitation, professionnel), ce qui conduit à la constitution de 333 modèles. Chacun de ces modèles comporte au maximum 119 événements, certains pouvant parfois n'en contenir qu'une petite centaine.

La variable expliquée (Y) est l'un des indicateurs suivants :

- le montant de chaque tempête par contrat,
- le taux de destruction en volume,
- le montant de chaque tempête par sinistre.

Les variables explicatives introduites dans le modèle sont :

- le type de tempête (TYPTEMP)
- la vitesse maximale du vent (LVIT dans la modèle d'analyse de la variance),
- la durée de la tempête (LDUR),
- la variable booléenne indiquant le caractère normal ou anormal de la direction principale prise par la tempête (DIR),
- la variable booléenne indiquant une présence significative ou symbolique de précipitations (PREC),
- la part de capital sous risque touché par la tempête (EFFLOCM),
- la part du portefeuille touché par la tempête (EFFLOCN),
- le niveau moyen de la prime incendie par contrat (RICHESSSE).

En revanche, les variables dont la valeur prise sur une zone météorologique est constante ne sont pas retenues pour ce modèle local. Il s'agit plus précisément :

- de la densité démographique (DENSITE),
- de la superficie (SUPERF),
- du type de local (POITLOC),
- du type d'assuré (POITASS),
- du niveau des normes de construction requises (NORMCONS).

Ces variables trouveront leur raison d'être au paragraphe III.2, pour le modèle d'analyse de la tempête par risque.

Pour mener à bien la construction de chacun des 333 modèles de régression linéaire, trois étapes importantes sont à mener successivement. Il s'agit en effet de vérifier :

- le *respect des hypothèses de base sur les résidus*,
- la *qualité de l'ajustement et le caractère significatif des coefficients*,
- la *robustesse de la régression*, à travers la détection des observations influentes, la non colinéarité des variables explicatives lors du choix de celles-ci.

Les calculs statistiques reposent sur un certain nombre d'hypothèses relatives aux résidus. Il s'agit entre autres de vérifier :

- l'hypothèse d'*indépendance des résidus*; il existe en théorie plusieurs tests non paramétriques mais SAS ne pouvant les réaliser, nous supposons cette indépendance acquise afin de poursuivre nos validations,
- l'hypothèse d'*homoscédasticité*, c'est-à-dire d'égalité des variances des résidus. Le test, effectué sous SAS, est global et utilise pour tester l'hypothèse H_0 d'hétéroscédasticité une statistique suivant une loi du k hi-deux; en cas d'acceptation de H_0 , le caractère global de ce test ne permet pas de connaître la variable responsable de l'hétéroscédasticité,
- l'hypothèse de *normalité des résidus* (dans le cas de l'indépendance et de l'homoscédasticité des résidus); cette vérification s'avère nécessaire avant tout recours au test sur le caractère significatif des coefficients. La normalité des résidus peut être prouvée par l'utilisation de la statistique de Kolmogorov - Smirnov :

$D_n = \sup |F_n^*(x) - F(x)|$ où $F_n^*(x)$ est la fonction de répartition empirique d'un n-échantillon d'une variable aléatoire de distribution $F(x)$. Au seuil $\alpha = 0,05$, la région critique est $D_n > \frac{1,3581}{\sqrt{n}}$.

Puis, il est possible de mesurer la qualité de l'ajustement par l'intermédiaire du coefficient de détermination R^2 . Soient les grandeurs suivantes :

- $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ la moyenne des valeurs de Y,
- $SCT = \sum_{i=1}^n (Y_i - \bar{Y})^2$ la somme des carrés totale,

$$- SCR = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \text{ la somme des carrés résiduelle,}$$

$$- SCE = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \text{ la somme des carrés expliquée,}$$

alors R^2 mesure la corrélation entre Y et \hat{Y} et est défini par :

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT}.$$

Ainsi, plus sa valeur est proche de 1, plus le modèle est de bonne qualité. Cependant, pour tenir compte du nombre de variables (p) et d'observations (n), on introduit la notion de \bar{R}^2 ajusté :

$$\bar{R}^2 = 1 - \frac{n-1}{n-p-1} (1 - R^2) = \frac{(n-1) \cdot R^2 - p}{n-p-1}$$

D'autre part, le caractère significatif des paramètres estimés peut être vérifié de deux manières différentes :

- en ce qui concerne l'ensemble des paramètres B , grâce à l'utilisation de la statistique de Fisher F :

$$F = \frac{n-p-1}{p} \cdot \frac{SCE}{SCT} = \frac{n-p-1}{p} \cdot \frac{R^2}{1-R^2}.$$

Sous l'hypothèse H_0 , : " $B = 0$ ", $F \sim F(p, n-p-1)$. On considère généralement que le modèle est valide lorsque F est supérieur à 2,

- en ce qui concerne un paramètre b_j , grâce à l'utilisation de la statistique de

Student T . Sous l'hypothèse H_0 , : " $b_j = 0$ ", $T_j = \frac{\hat{b}_j}{\mathbf{s}_j^2} \sim T(n-p-1)$. Dès lors

que la valeur absolue de ce rapport dépasse une certaine valeur, alors l'apport de la variable dans le modèle est significatif.

S'il est indispensable que les variables à intégrer aient un pouvoir explicatif du phénomène à étudier, cette condition n'est pas suffisante : il faut en outre, pour confirmer leur intégration, avoir vérifié leur non-corrélation. C'est à ce prix que l'on assurera la robustesse du modèle.

Enfin, il peut être utile de détecter les observations influentes, pour pouvoir les éliminer. L'étude des résidus " studentisés " est une des nombreuses techniques de détection utilisables. Si l'on note :

$$- H = X(X'X)^{-1}X',$$

$$- \hat{Y} = HY,$$

$$- e = Y - \hat{Y} = (I - H)Y,$$

la diagonale de H comporte les n coefficients $h_{ii} = X'_i(X'X)^{-1}X_i$ appelés leviers de l'observation i . Un levier représente l'influence de l'observation i sur la valeur ajustée \hat{Y}_i , à cause des valeurs X_i prises par les variables en i . On a : $\text{Var}(e_i) = \sigma^2.(1 - h_{ii})$ et les

résidus “ studentisés ” : $r_i = \frac{e_i}{\hat{S}^2 \sqrt{1 - h_{ii}}}$. Bien que le numérateur et le dénominateur ne

soient pas indépendants, on considère que r_i suit une loi de Student à $n - p - 2$ degrés de liberté.

Lorsque n est grand, les r_i doivent rester compris entre -2 et 2 pour un seuil de 5 %.

III.1.1.4 - Présentation des résultats

On trouvera en annexe N le programme MODTATEMP écrit sous SAS et assurant le rapprochement des événements définis par M étéo-France et des tempêtes du fichier “ sinistres ”. Ce programme est aussi l'occasion de la création des 111 bases de données locales.

Figure en outre un exemple détaillé de la modélisation du logarithme népérien du montant actualisé du préjudice par contrat, réalisée grâce à la procédure GLM du logiciel SAS, pour la station de Dinard (code 3501) et ce par risque. La modalité 1 du risque correspond au risque agricole, la modalité 2 au risque d'habitation, et la 3 au risque professionnel.

L'ensemble des résultats ne peuvent être reproduits ici. Ils sont la propriété d'AXA dont l'UAP fait maintenant partie.

En règle générale, les R^2 sont compris entre 70 % et 90 % pour les trois risques et pour l'ensemble des zones météorologiques.

Les tests de normalité sont effectués sous SAS grâce à la procédure UNIVARIATE et l'option “ NORMAL PLOT ”. C'est le test de Shapiro - Wilk qui est ici utilisé, les échantillons comprenant moins de 119 observations. Les résultats rejettent parfois cette normalité, ce qui peut remettre en cause la validité du test de Student sur l'apport réel des variables dans le modèle.

Malgré le nombre important de modèles, il existe certains invariants :

- la *vitesse maximale du vent* est un paramètre présentant un apport toujours significatif; néanmoins, on relève un signe négatif du coefficient dans 17 des 333 modèles, ce qui se traduit par une réduction du montant des dommages à mesure de l'élévation des dommages; cette situation peut s'expliquer localement par la prédominance de facteurs non répertoriés, ou à l'excessive corrélation

entre cette variable et une autre présente dans le modèle, ce à quoi nous pouvons remédier,

- la variable *type de tempête* a, 9 fois sur 10, un impact net sur la variable expliquée,
- l'effet taille de la zone météorologique joue un rôle essentiel de sorte que la variable “ *part de capital sous risque sinistré* ” est très significative.

En revanche, les autres variables ne présentent pas cette constance : on ne peut faire l'économie d'une analyse détaillée des résultats de chaque modèle. Les paramètres descriptifs météorologiques ne peuvent pas toujours faire partie de la liste de variables retenues : on s'accorde généralement pour considérer qu'une variable a un pouvoir explicatif des variations de la variable modélisée dès que la statistique de Fisher dépasse le seuil de 2.

Pour les variables “ *direction atypique* ” et “ *caractère significatif des précipitations* ”, la rareté de l'une de leurs modalités est à l'origine d'un niveau d'inertie insuffisant pour certaines stations, et les rend sans objet.

La variable *durée* se compte le plus souvent parmi les facteurs retenus dans le modèle. On peut toutefois s'étonner de la valeur négative prise dans certains modèles par le paramètre estimé.

III.1.1.5 - Détermination de la vitesse d'apparition des dommages

Avant d'aller plus avant dans notre méthode de détermination d'une prime pure tempête, il est possible d'utiliser les résultats du modèle précédent pour déterminer la *vitesse d'apparition des dommages*, c'est-à-dire le seuil de vitesse à partir de laquelle les premiers dommages apparaissent. Pour cela, il suffit de reprendre l'écriture du modèle linéaire en considérant que la relation suivante est vérifiée, pour chacun des 333 modèles :

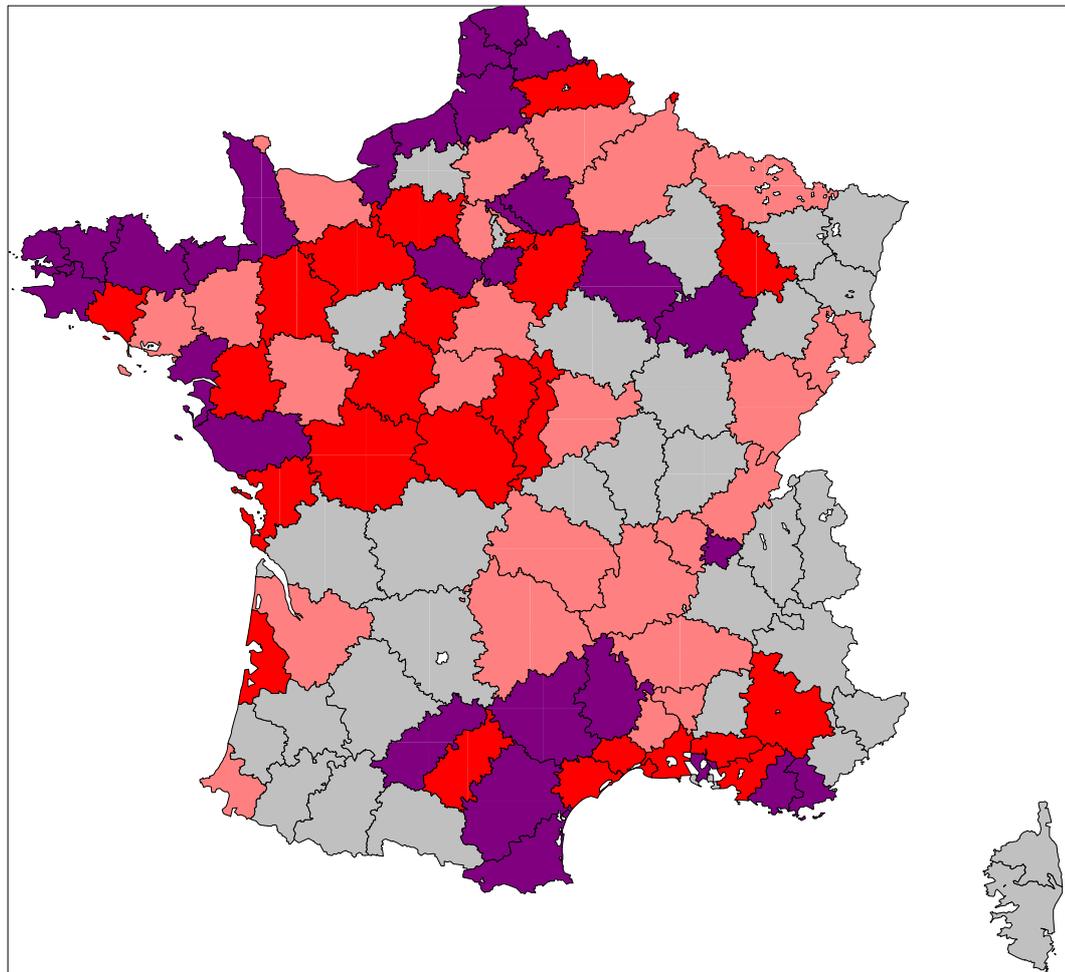
$$Y_e = a \cdot \ln(V) + b + e \quad (4)$$

L'estimation de la prime pure par événement est fournie par $Y_e = Xb$, combinaison linéaire des coefficients précédemment estimés et des valeurs des variables explicatives; mais d'autres variables, non prises en compte dans ce modèle, expliquent les variations du montant du préjudice par contrat (Y) : il nous semble donc que la relation entre l'intensité du vent V et Y_e doit être plus forte et moins sujette aux valeurs extrêmes (donc plus robuste), que celle entre V et Y .

Il ne reste plus alors, pour déterminer la vitesse *d'apparition des dommages* V_0 , qu'à estimer les coefficients a et b de la relation (4), Y_e devenant alors une variable aléatoire. V_0 est la vitesse qui annule le montant du préjudice, telle que :

$$\ln V_0 = -\frac{b}{a}$$

La figure 2.4.3 fournit une illustration géographique de la vitesse d'apparition des dommages pour le risque agricole.



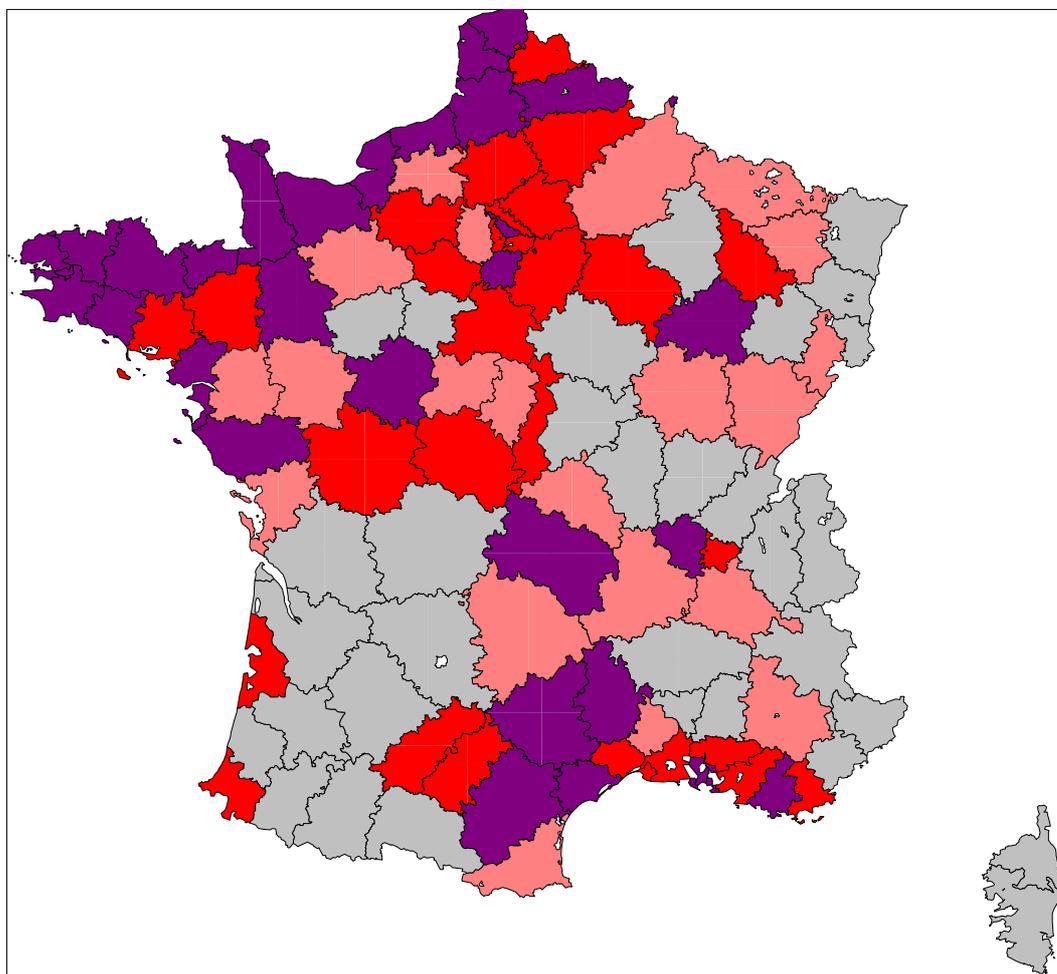
Légende :

	15,3 – 28,3 m/s (29 zones météorologiques)
	14,2 – 15,3 m/s (24 zones météorologiques)
	13,3 – 14,2 m/s (26 zones météorologiques)
	6,0 – 13,3 m/s (32 zones météorologiques)

Figure 2.4.3 : carte des vitesses de vents déclencheurs de dommages pour le risque agricole

La carte précédente montre une opposition assez nette entre les façades maritimes (côtes méditerranéenne et nord-atlantique) et l'intérieur des terres : les premières présentent des vitesses d'apparition des dommages plus élevées que les secondes, moins confrontées à la tempête, et dont les bâtiments agricoles offrent une moins grande résistance.

La figure 2.4.4 fournit une illustration géographique de la vitesse d'apparition des dommages pour le risque d'habitation.



Légende :

	15,9 – 26,6 m/s (30 zones météorologiques)
	14,7 – 15,9 m/s (28 zones météorologiques)
	13,5 – 14,7 m/s (21 zones météorologiques)
	7,6 – 13,5 m/s (32 zones météorologiques)

Figure 2.4.4 : carte des vitesses de vents déclencheurs de dommages pour le risque d'habitation

La carte précédente ressemble à s'y méprendre à celle du risque agricole, même si les régions septentrionales sont, dans une plus importante proportion que les régions méridionales, dans la zone offrant une résistance au vent plus forte. Il est à noter que la vitesse moyenne d'apparition des dommages est un peu supérieure pour ce risque que le risque agricole (53 km/h contre 51), ce qui peut dénoter une moindre exposition aux flux atmosphériques.

La figure 2.4.5 fournit une illustration géographique de la vitesse d'apparition des dommages pour le risque professionnel.

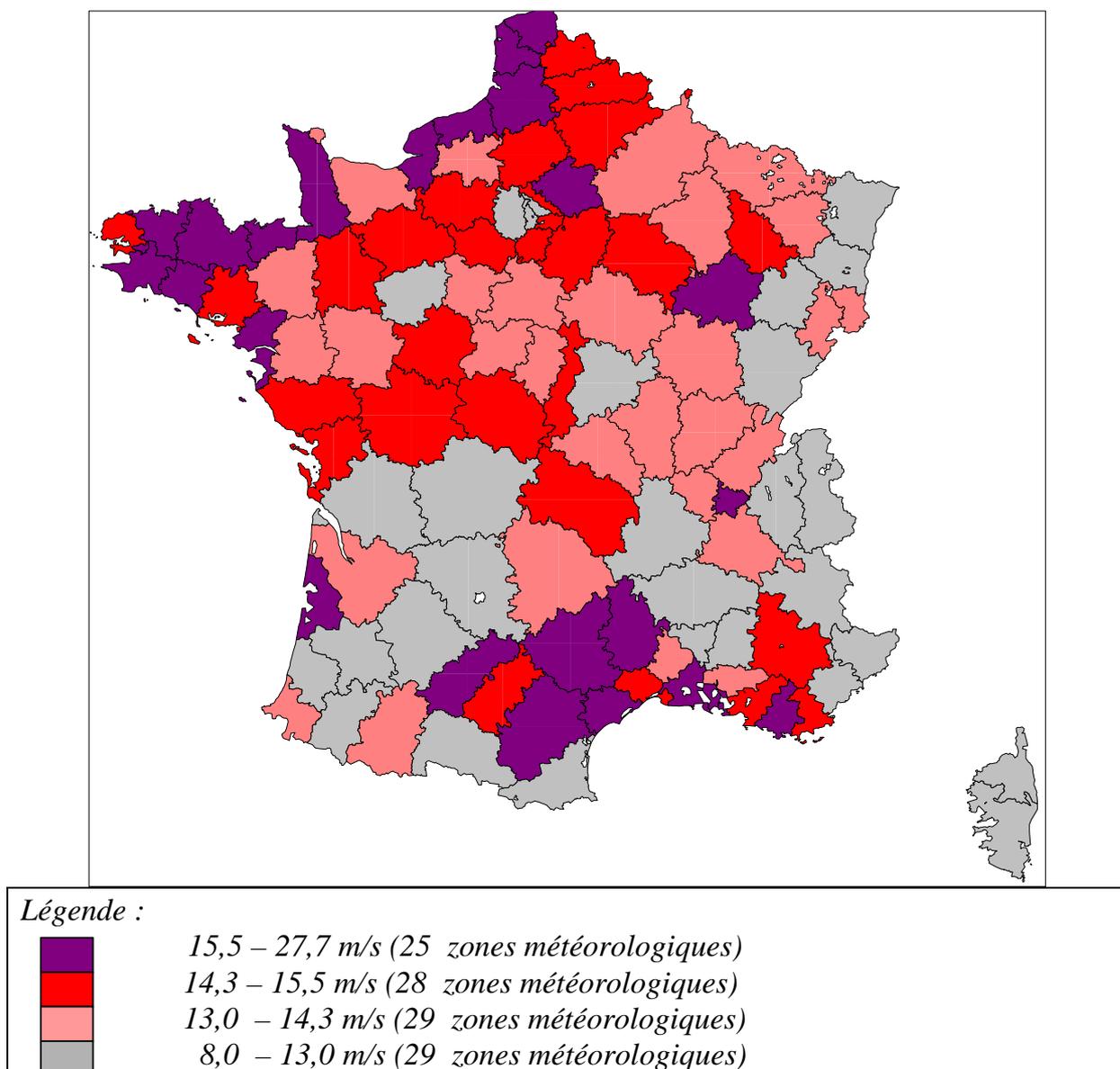


Figure 2.4.5 : carte des vitesses de vents déclencheurs de dommages pour le risque professionnel

Là encore, la même opposition entre les régions maritimes et l'est et le sud-ouest de la France est à souligner. L'appartenance d'une majeure partie de l'Aquitaine et du Midi-Pyrénées à la zone grisée (quart inférieur de la distribution des vents) nous rappelle l'importance des phénomènes atmosphériques locaux, que ne décèlent pas les stations météorologiques (la vitesse enregistrée étant faible). La vitesse moyenne d'apparition des dommages est égale à celle du risque agricole, 51 km/h.

D'une manière générale, on peut dire qu'il existe une importante corrélation entre ces cartes et la carte des vents, figurant en annexe. Elle souligne l'importance des

mesures de protection pour lutter avec efficacité contre les effets du vent, qui ont pour conséquence le relèvement des seuils à partir desquels les vitesses se révèlent destructrices.

III.1.2 - Utilisation des distributions bidimensionnelles par station

III.1.2.1 - Hypothèses simplificatrices

En outre, Météo-France a mis à notre disposition les distributions selon l'intensité du vent et une autre variable descriptive choisie parmi la durée, le niveau des précipitations et la direction du vent. Ces distributions, fournies pour 91 stations météorologiques comportant un historique complet sur la période 1962 – 1993 et pour chacun des sept types de tempêtes, ont fait l'objet d'une présentation détaillée au paragraphe III.5 du chapitre 2 de la première partie, consacré à l'étude de Météo-France.

Ainsi, partant du modèle (3), liant la charge actualisée des sinistres par contrat (Y) aux différents prédicteurs contenus dans la matrice X , dont deux paramètres météorologiques descriptifs, il doit être possible de “ probabiliser ” ces distributions empiriques et, par là même, de donner l'espérance de Y relativement aux valeurs prises par les deux paramètres descriptifs.

Malheureusement, il n'est pas possible de trouver une loi d'adéquation bidimensionnelle continue, s'ajustant aux probabilités empiriques de survenance d'une vitesse et d'un autre paramètre compris dans un intervalle de valeurs. De la même manière, l'ajustement du tableau de contingence sur une loi discrète à deux dimensions nécessite l'utilisation des probabilités conditionnelles.

D'autre part, compte tenu du nombre réduit d'événements (119) sur lequel repose le modèle (3), il n'a pas été jugé opportun de multiplier le nombre de modèles résultant de la multiplication des nombres de zones météorologiques (111) et de risques (3) par le nombre de types de tempêtes : cela aurait abouti à 2.441 modèles ! De ce fait, les distributions des intensités des vents par type de tempête sont agrégées au niveau de chaque zone météorologique.

Signalons en outre que, sur les 111 stations finalement retenues pour l'étude, 83 figurent parmi les 91 stations dotées de distributions des intensités de vent. Les 28 autres, qui ne disposent donc pas de distributions propres, sont rattachées à la station alliant proximité géographique et similitude de relief.

III.1.2.2 - Recherche d'une adéquation à une loi exponentielle

Nous venons de déterminer au paragraphe III.1.1.5, grâce au modèle liant les dommages et la vitesse du vent, l'intensité seuil du vent à partir de laquelle apparaissaient les premiers sinistres. Il est clair que ce n'est que sur la partie

supérieure de la distribution de la vitesse du vent, pour laquelle la vitesse est supérieure à ce seuil, que se concentrent les sinistres.

La figure 2.4.6 représente la fonction de distribution de l'intensité du vent (en m/s) pour la station de Poitiers (code 8601).

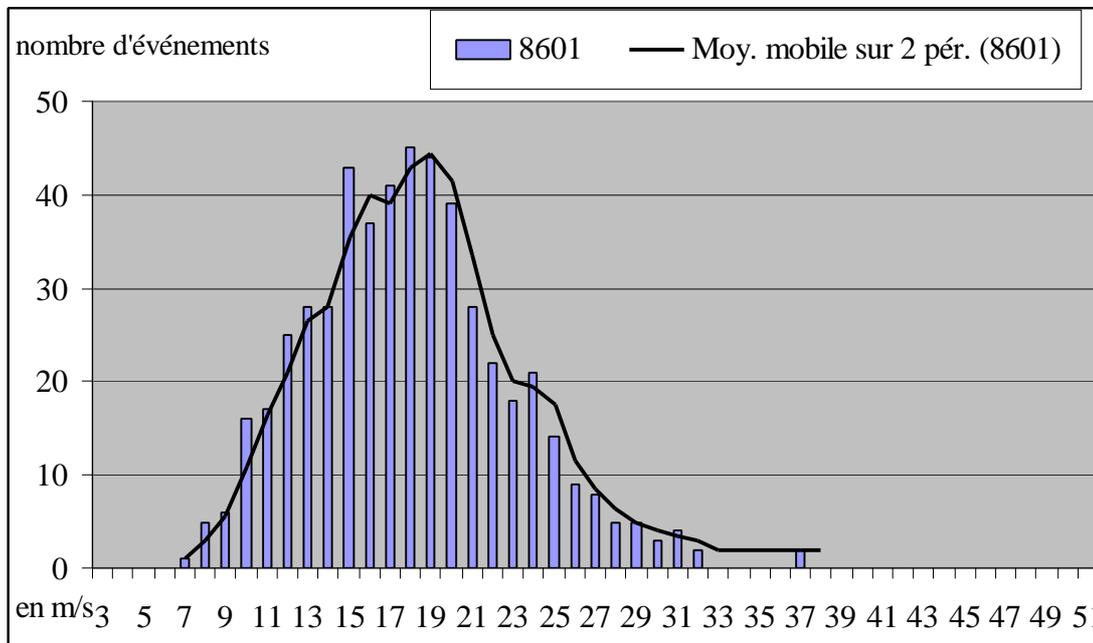


Figure 2.4.6 : Fonction de distribution de l'intensité du vent de la station de Poitiers

Aussi est-il décidé d'ajuster la partie droite de la distribution marginale discrète du seul paramètre " intensité maximale du vent " (V) à une loi exponentielle : en clair, c'est le mode de cette distribution qui fait office de seuil. (Une étude montre par ailleurs que la vitesse " mode " est très voisine de la vitesse d'apparition des dommages).

La loi exponentielle convient bien au type de distribution présent. De plus, nous verrons plus loin que sa fonction de densité se prête bien au calcul de l'espérance de la charge actualisée par tempête et contrat, que nous aurons à mener plus loin. Sa densité est donnée par la formule :

$$f(x) = \lambda \exp(-\lambda x), \text{ où } x > 0.$$

La fonction de répartition de la loi exponentielle s'écrit :

$$F(x) = 1 - \exp(-\lambda x),$$

puisque $F(0) = 0$, ce qui se traduit aussi par la relation suivante :

$$-\lambda x = \ln(1 - F(x)).$$

De la sorte, la distribution empirique (x_1, x_2, \dots, x_n) du vent maximal quotidien de chaque station météorologique est ajustée à une loi exponentielle dont le paramètre λ admet pour estimateur du maximum de vraisemblance :

$$\hat{I} = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\bar{x}}$$

Chaque distribution empirique admet 50 valeurs de vitesses comprises entre 2 et 51 m/s.

La qualité de l'ajustement peut être déterminée grâce au test du χ^2 , pour autant que les intervalles retenus contiennent au moins 5 individus.

Les résultats montrent une assez bonne adéquation entre la distribution empirique et la loi théorique. D'ailleurs, celui-ci peut se faire par graphique en vérifiant que la relation entre les x_i et $\ln(1-F(x_i))$ est linéaire, le coefficient directeur de la droite étant donné par $-I$.

III.1.3 - Détermination d'une prime pure tempête

III.1.3.1 - Formulation mathématique d'une prime pure tempête

Parallèlement, l'utilisation simplifiée du modèle (3) nous permet de déterminer la relation linéaire liant y , logarithme népérien de la prime pure par contrat et par événement et V , l'intensité maximale du vent selon le principe suivant :

$$y = A \cdot \ln V + B + e$$

Les régressions donnent des résultats convenables : sur les 333 modèles, seulement 7 présentent, au seuil de confiance de 95 %, un T de Student supérieur au rapport de l'estimateur de A à son erreur-type. Ainsi, dans la grande majorité des cas, on peut conclure sur l'impact effectif de l'intensité du vent dans le modèle.

On peut linéariser la relation précédente :

$$\exp(y) = \Pi = \exp(B) \cdot V^A \cdot \exp(e)$$

où Π est la prime pure par contrat et par événement.

Par la suite, la loi de distribution à laquelle a été rattachée la répartition empirique des vents est intégrée à cette version simplifiée. En effet, comme V a été « probabilisée », il est ainsi possible de parvenir à la notion d'espérance de la charge actualisée par contrat et par événement :

$$E(V) = \int_0^{\infty} f(v) \cdot v \cdot dv, \text{ et par suite :}$$

$$E(\Pi) = \exp(\hat{B}) \cdot E(V^{\hat{A}}) \cdot E(\exp(e)) = \exp(\hat{B}) \cdot E(V^{\hat{A}})$$

Il reste alors à calculer le terme $E(V^A)$. Son expression est la suivante :

$$E(V^{\hat{A}}) = \int_0^{\infty} v^{\hat{A}} \cdot f(v) \cdot dv = \int_0^{\infty} v^{\hat{A}} \cdot I \cdot \exp(-I \cdot v) \cdot dv$$

Or, si l'on pose :

$$u = I \cdot v$$

alors le terme $E(V^{\hat{A}})$ prend la tournure suivante :

$$E(V^{\hat{A}}) = \int_0^{\infty} \frac{I}{I^{\hat{A}}} u^{\hat{A}} \cdot \exp(-u) \cdot du$$

On reconnaît, à un terme multiplicatif près, l'expression de la fonction eulérienne Gamma (Γ), d'où :

$$E(\Pi) = \exp(\hat{B}) \cdot \frac{I}{I^{\hat{A}}} \cdot \Gamma(\hat{A} + I) = \exp(\hat{B}) \cdot \frac{\hat{A}!}{I^{\hat{A}}}$$

Cette expression n'a de sens que si le coefficient estimé de A est positif : cela correspond dans les faits à une augmentation du montant moyen de la prime pure à mesure que la vitesse du vent croît. C'est en effet systématiquement le cas. De plus, la factorielle n'est normalement définie que pour des nombres entiers. La formule de Stirling, démontrée à partir de la fonction Γ , permet de prolonger son ensemble de définition aux réels positifs :

$$n! \approx n^n \cdot \exp(-n) \cdot \sqrt{2 \cdot \pi \cdot n}$$

L'espérance de la charge actualisée par contrat et événement peut donc être estimée par la formule :

$$E(\Pi) = \exp(\hat{B}) \cdot \left(\frac{\hat{A}}{I}\right)^{\hat{A}} \cdot \exp(-\hat{A}) \cdot \sqrt{2 \cdot \pi \cdot \hat{A}}$$

Il suffit ensuite de multiplier cette prime pure, calculée pour un événement, par la fréquence annuelle d'un événement, tous types confondus, pour obtenir la prime pure annuelle du risque tempête.

III.1.3.2 - Résultats du calcul d'une prime pure tempête

Il est alors possible de passer à l'application numérique de cette formule, puis à la représentation géographique de ces primes pures selon le risque et la zone météorologique. Les figures 2.4.7 à 2.4.9 synthétisent, sur une carte de France, la répartition de la prime pure tempête selon les zones, respectivement pour le risque agricole, d'habitation et professionnel.

La carte de France de *la prime tempête du risque agricole présente un contraste saisissant entre les régions* : si la partie sud-est de la France semble finalement peu impliquée par les conséquences destructrices du risque tempête, en raison peut-être de l'existence de normes strictes de construction, en revanche, la frange centrale du nord

de la France, allant du Périgord à la frontière belge, est très concernée par les dommages aux biens agricoles. Les primes annuelles y dépassent les 450 F, alors que celles de certains départements du sud ne dépassent pas les 40 F annuellement. La moyenne nationale est d'environ 1.127 F, à un niveau supérieur à l'estimation réalisée au chapitre 3.

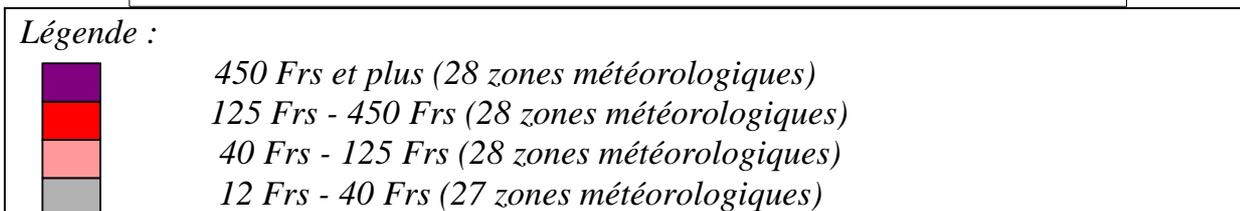
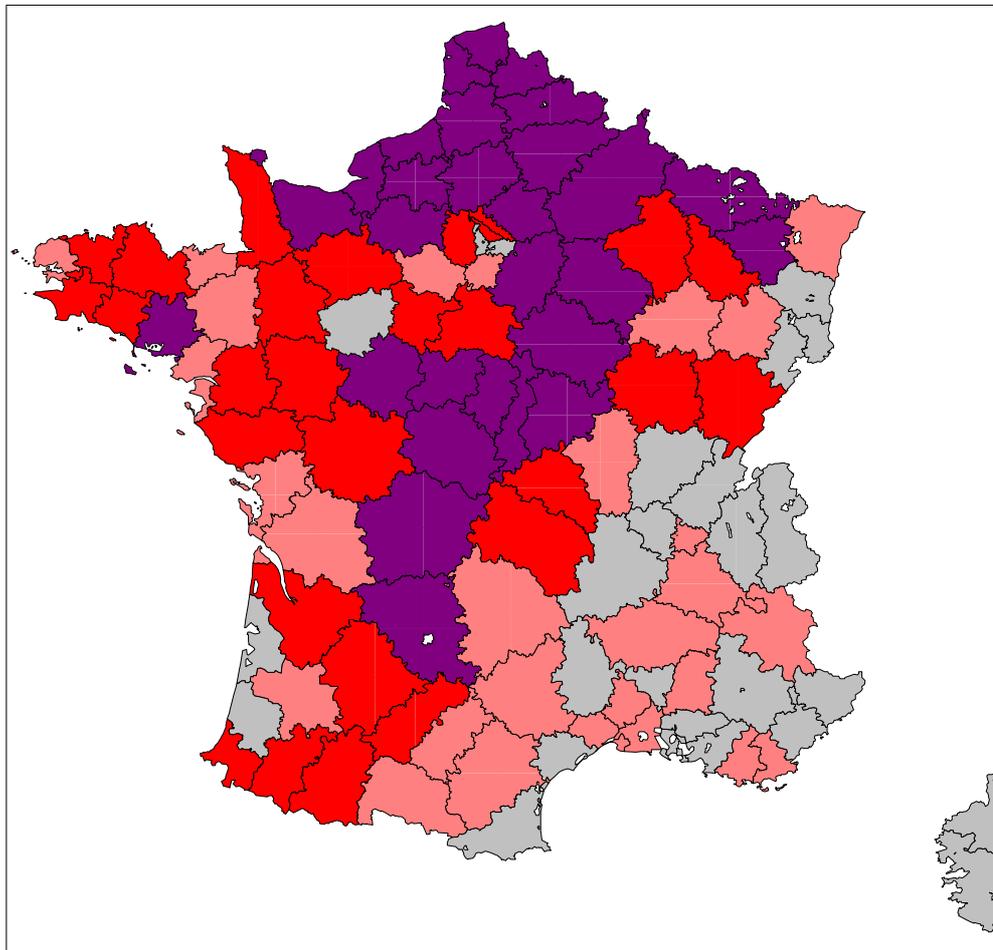


Figure 2.4.7 : Répartition par zone météorologique de la prime pure pour le risque agricole

Cette carte peut être mise en parallèle avec la figure 2.3.6, qui situe, pour chaque département, le montant de la prime pure tempête pour le risque agricole, calculé à partir d'un historique de 18 ans. Les résultats obtenus à partir de l'observation de la vitesse des vents sur 32 ans ne font que confirmer les conclusions du chapitre 3, tout en renforçant les disparités régionales.

Compte tenu des résultats convergents obtenus au chapitre 3 et à ce présent chapitre, il semble nécessaire de réfléchir au renforcement d'une meilleure segmentation tarifaire, ayant pour conséquence le relèvement des primes dans les zones les plus impliquées par la tempête.

La carte de France de *la prime tempête du risque d'habitation présente les mêmes contrastes* que pour le risque agricole : même si l'on peut signaler ici où là quelques exceptions, le quart sud-est de la France, ainsi que la région des Pays de la Loire sont peu atteints par les effets du vent. En revanche, on doit signaler l'existence persistante de cette zone centrale, dont la région parisienne ne fait pas partie, plus sérieusement touchée.

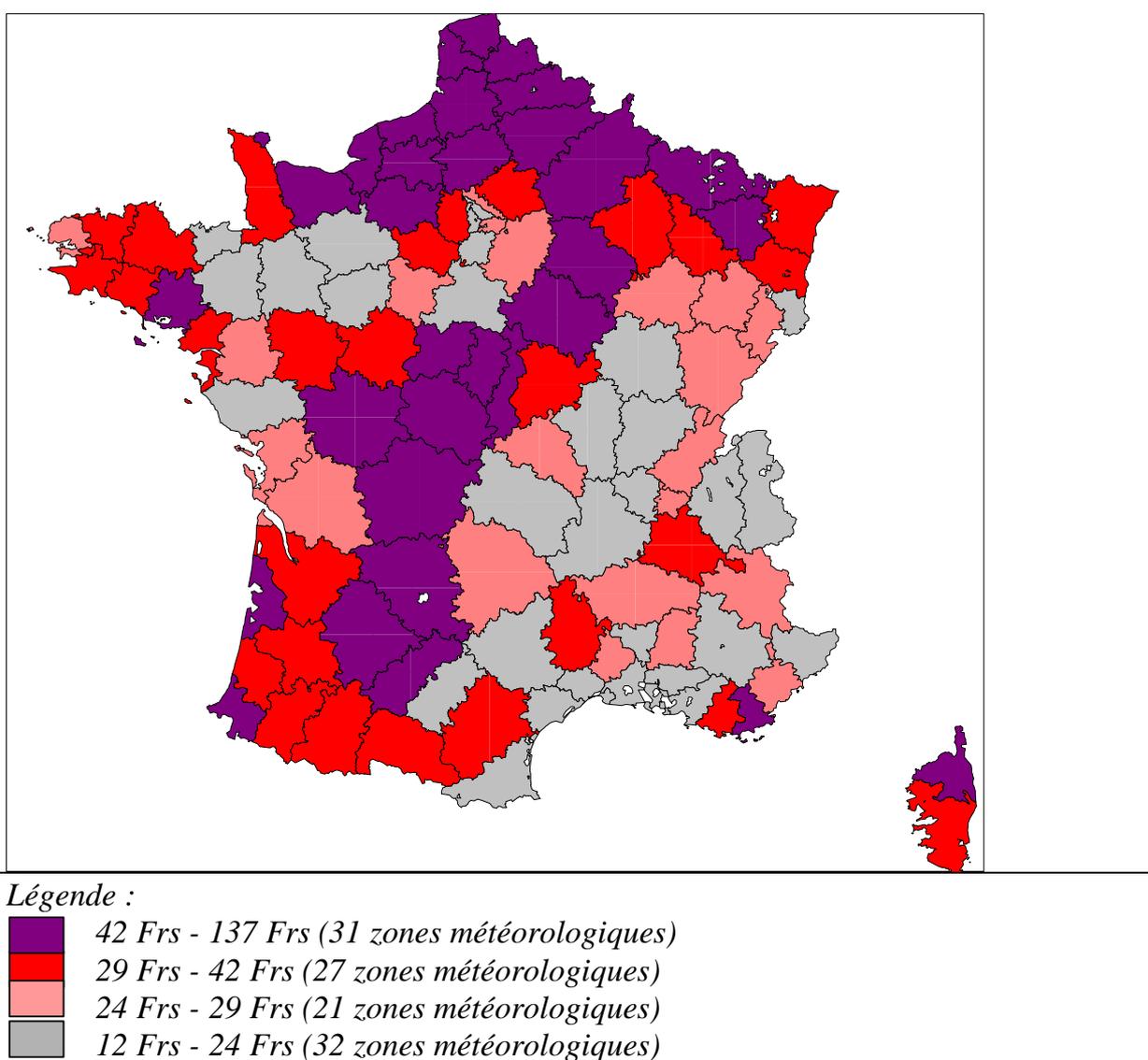
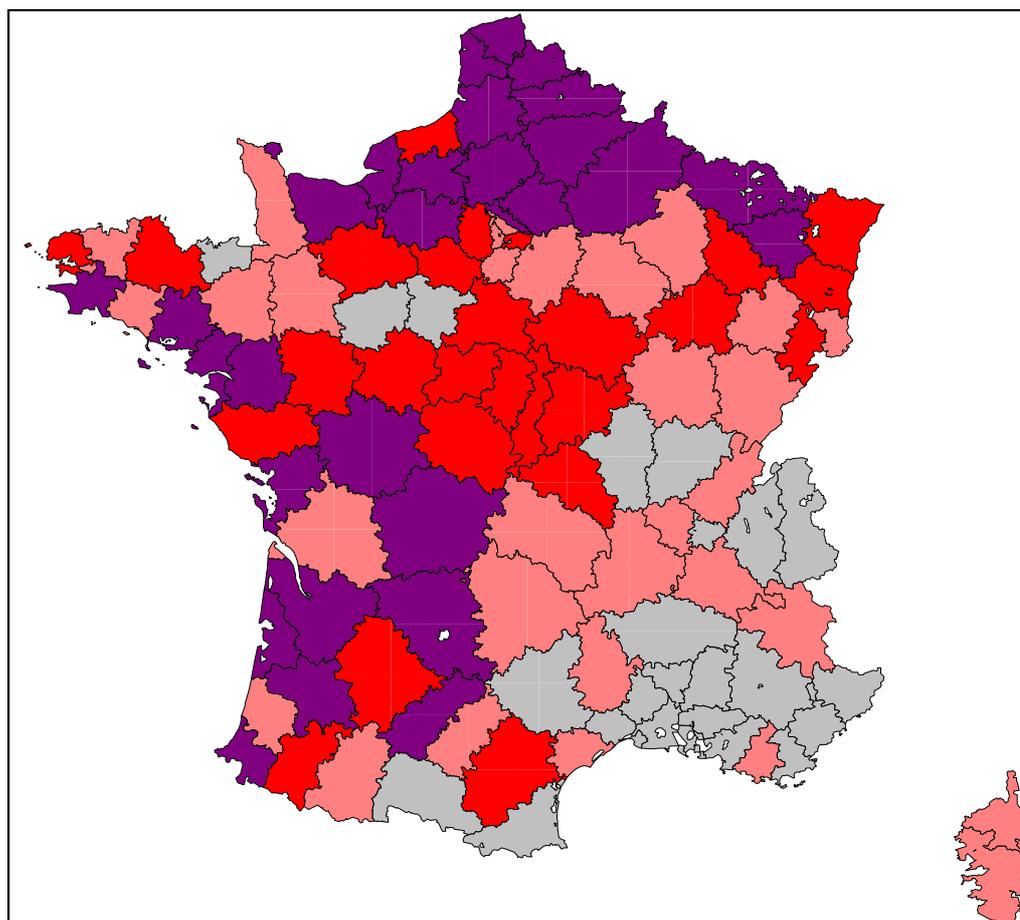


Figure 2.4.8 : Répartition par zone météorologique de la prime pure pour le risque d'habitation

Ces disparités enregistrées pour le risque d'habitation restent plus mesurées que celles du risque agricole : sa plus faible exposition aux effets du vent et sa plus grande homogénéité expliquent en partie cette situation. De plus, le faible montant absolu de la prime moyenne (37 F) en comparaison de la prime moyenne du risque habitation ne justifie pas l'élaboration d'un tarif spécifique à la garantie tempête.

La carte de France de *la prime tempête du risque professionnel présente aussi une opposition entre le nord - et une partie du quart sud-ouest-, très affectés, et le sud-est et le Rhône-Alpes, peu concernés*. La région de la Beauce est une fois encore peu atteinte par cette menace atmosphérique.



Légende :



160 Frs - 2 400 Frs (29 zones météorologiques)

90 Frs - 160 Frs (26 zones météorologiques)

50 Frs - 90 Frs (31 zones météorologiques)

10 Frs - 50 Frs (25 zones météorologiques)

Figure 2.4.9 : Répartition par zone météorologique de la prime pure pour le risque professionnel

Les disparités régionales sont plus marquées que pour le risque d'habitation et moins que pour le risque agricole : la localisation des bâtiments professionnels, souvent à la périphérie des agglomérations, explique peut-être la relative homogénéité en terme d'exposition au vent. De même, même si le montant absolu de la prime moyenne n'est pas à négliger (212 F), il se révèle néanmoins négligeable (de l'ordre de 3 %) en comparaison de la prime moyenne multirisque, ce qui ne constitue pas un montant suffisant à l'élaboration d'un tarif spécifique à la garantie tempête.

Enfin, la prime pure tempête calculée sur la période 1962 - 1993, tous risques confondus, atteint 90 F, ce qui la place à un niveau inférieur à la prime pure calculée sur l'historique compris entre 1980 et 1997 (111 F). Le risque agricole est bien plus affecté par les effets du vent que les deux autres risques.

III.2 - Etude de l'impact de différentes variables sur le montant de la prime pure

III.2.1 - Présentation du modèle de régression

Après avoir vu au paragraphe III.1 dans quelle mesure les variables météorologiques - et plus particulièrement l'intensité maximale - pouvaient contribuer à la modélisation locale du risque tempête, il s'agit maintenant de faire intervenir d'autres variables, ayant trait au type de construction, d'habitat et d'assuré, et de mesurer leur influence sur l'établissement de la prime pure. Ces variables, dites d'habitat, ayant déjà été citées au chapitre 2 de cette partie et au paragraphe III.1.1.3 de ce présent chapitre, on pourra donc s'y référer pour un examen précis.

Par définition, les variables d'habitat ne pouvaient figurer dans la liste des variables retenues dans l'un des 333 modèles décrits précédemment, leur valeur étant considérée comme constante sur l'ensemble du territoire d'une zone météorologique. En revanche, dès que l'on se limite à la constitution de 3 modèles, correspondant à la rupture par risque, ces variables deviennent fluctuantes d'une zone météorologique à une autre et leur intégration devient alors possible.

Les observations de chaque modèle ainsi défini sont au nombre maximal de 13.209, cardinal de l'ensemble des couples croisant le numéro d'événement avec chaque zone météorologique. En revanche, certaines des 111 stations météorologiques, bien que sélectionnées, ne détiennent pas forcément un relevé des 119 événements identifiés, leur historique n'étant pas complet. C'est finalement 12.623 observations qui sont présentes dans chacun des modèles, la carence des relevés étant indépendante du risque considéré.

III.2.2 - Validation du modèle de régression

Le modèle retenu ici reprend l'écriture synthétique (4), utilisée au paragraphe II.3 :

$$Y = Xb + a_i + e \quad (4)$$

où a_i est l'effet principal du niveau i du facteur A " type de tempête ", et X la matrice des autres variables explicatives, non qualitatives.

Comme nous l'avons vu au paragraphe III.1.1.3, un certain nombre de critères doivent être respectés pour que le modèle soit valide. Une des plus importante est la vérification de la normalité des résidus, condition nécessaire à l'utilisation des tests.

La qualité de l'ajustement peut être mesurée par le coefficient de détermination R^2 : celui-ci oscille entre 61 % (pour le risque d'habitation) et 64 % (pour le risque professionnel). Compte tenu du petit nombre de variables explicatives et du grand nombre d'observations présentes, c'est un résultat digne d'intérêt. En outre, il est peu probable que la part d'inertie non expliquée (entre 36 et 39 %) relève d'une seule et même variable cachée.

La présélection des variables susceptibles d'avoir une influence sur les variations de la prime pure peut être maintenant appréciée à la lumière des résultats que l'on va maintenant présenter.

III.2.3 - Principaux résultats

III.2.3.1 - Résultats généraux

L'annexe O présente l'ensemble des résultats fournis par la procédure GLM du logiciel SAS. Si l'on examine en détail la liste des variables, on s'aperçoit ainsi que, pour les trois risques, un certain nombre de considérations identiques peuvent être tirées :

- ***toutes les variables météorologiques et d'habitat***, à l'exception de celle relative au *type de locaux*, pour le *risque agricole*, de celle relative au *type d'assuré*, pour le *risque d'habitation*, et de ces mêmes variables plus la *densité démographique* pour le *risque professionnel*, ***ont un impact significatif sur le montant de la prime pure***;
- ***au premier rang pour la variance expliquée***, on trouve *l'effet local de la tempête*, ce qui montre que bien souvent la tempête prend un caractère local et est loin de frapper toute la zone météorologique, et les relevés météorologiques peuvent ne rien déceler. Au deuxième rang figure *l'intensité du vent maximal*, ce qui peut justifier a posteriori l'emploi de sa courbe de distribution pour retrouver l'espérance de la prime pure, au paragraphe III.1.2. Puis, apparaissent, avec un impact bien moindre que les deux premières variables, le *niveau significatif des précipitations* et la *superficie de la zone météorologique*.

L'invariance de ces résultats selon les risques peut d'ailleurs laisser supposer une relative robustesse du modèle en présence. Cette constance se manifeste aussi dans le signe des coefficients estimés des différentes variables : on constate en effet que, pour une variable donnée, ce dernier est le même quel que soit le risque considéré.

Il faut toutefois signaler que la variable *densité* constitue une exception à cette règle, le risque agricole, avec un coefficient positif, voyant sa prime pure croître avec la densité démographique, alors que le risque d'habitation est dans un contexte inverse, l'augmentation de la densité ayant une influence à la baisse sur le montant de la prime pure. Loin de traduire une instabilité du modèle, cette situation s'explique par le positionnement géographique différent des portefeuilles des deux risques. En effet :

- le portefeuille agricole est établi dans des régions peu peuplées, où l'augmentation de la densité va de pair avec celle du nombre de bâtiments agricoles et du capital sous risque,
- au contraire, le portefeuille d'habitation est majoritairement situé dans des villes où l'augmentation de la densité se traduit par une plus importante part d'assurés locataires à l'exposition au risque tempête moindre.

D'autre part, la traduction du signe des coefficients en influence inflationniste ou réductrice sur le montant de la prime pure n'offre pas de conclusions contraires au bon sens : on remarque par exemple sans surprise que le montant de la prime pure croît à mesure que le caractère local de la tempête s'estompe et que l'effet taille se renforce. Il faut néanmoins citer la *durée de la tempête* dont le coefficient, négatif, tend à diminuer la prime pure à mesure que la durée de la tempête croît.

III.2.3.2 - Mesure de l'impact des variables sur le montant de la prime pure par événement

Muni des résultats de l'annexe O, on peut se livrer au calcul de l'*impact*, sur le montant de la prime pure y , d'une variation de la valeur d'une variable explicative X_i . On définit l'impact de X_i sur y comme la variation algébrique de la prime pure relativement à celle de la variable explicative. Rappelons qu'il s'agit ici du calcul de la *prime pure par événement*, et non de la *prime pure tempête* (cf. à ce sujet le paragraphe III.1.3.1).

Bien sûr, cet impact est fluctuant selon le risque. L'écriture du modèle est celle d'une combinaison linéaire de variables parmi lesquelles figure X_i :

$$y = \ln \Pi = b_0 + b_1 X_1 + \dots + b_i X_i + \dots + b_p X_p + e. \quad (5)$$

Ce modèle donne deux estimations \hat{y}_A et \hat{y}_B de la prime pure correspondant respectivement à une valeur de X_i égale à x_i^A et x_i^B . La multiplication des deux membres de l'équation (5) par la fonction exponentielle et le rapport des deux primes estimées conduisent à l'expression suivante :

$$\frac{\hat{\Pi}_B}{\hat{\Pi}_A} = \frac{\exp(b_i x_i^B)}{\exp(b_i x_i^A)} = \exp(b_i (x_i^B - x_i^A)).$$

Intensité du vent

Ainsi, il peut être intéressant d'étudier l'impact financier de deux tempêtes présentant des conditions météorologiques (durée, direction, précipitation) et géographiques (zones postales touchées) identiques, si l'on excepte la vitesse du vent : celui-ci a atteint une intensité maximale journalière de 20 m/s pour le premier événement, contre 25 m/s pour le second. Le tableau 2.4.2 donne l'estimation de l'impact sur le montant de la prime pure d'une augmentation de la vitesse du vent de 20 m/s à 25 m/s.

Tableau 2.4.2 : Impact sur le montant de la prime pure de l'augmentation de l'intensité maximale journalière du vent de 20 à 25 m/s

Risque	agricole	habitation	professionnel
Augmentation de l'intensité maximale journalière du vent de 20 à 25 m/s	+ 27,3 %	+ 11,3 %	+23,6 %

C'est le risque agricole qui enregistre l'augmentation relative du montant de la prime pure la plus forte, un peu devant le risque professionnel et loin devant le risque d'habitation : ceci traduit la plus grande sensibilité au risque tempête du portefeuille agricole, plus exposé que le portefeuille d'habitation.

Il est à noter que l'augmentation relative s'estompe à mesure que la vitesse du vent croît : le passage de l'intensité maximale journalière du vent de 35 à 40 m/s induit un e croissance de la prime pure « habitation » de 6,6 %, contre 11,3 % précédemment.

Direction du vent

La direction du vent peut être une source importante de dommages : dans chaque région il existe en effet un vent dominant, c'est-à-dire un vent soufflant le plus fréquemment dans une direction donnée. L'atypisme de cette direction explique souvent l'absence de dispositions préventives, qui existent pour les vents soufflant dans une direction fréquente, dont la présence aurait pu limiter ou annuler des dommages.

Le tableau 2.4.3 donne une estimation de l'impact du passage d'une direction normale à anormale.

Tableau 2.4.3 : Impact sur le montant de la prime pure du passage d'une direction normale à anormale

Risque	agricole	habitation	professionnel
passage d'une direction normale à anormale	+ 3,4 %	+1,9 %	+ 0,8 %

Les manifestations destructrices de ce paramètre météorologique sont évidemment moins spectaculaires que l'intensité du vent.

Niveau significatif des précipitations

Un niveau significatif des précipitations (plus de 15 ml) accompagnant ou suivant des vents violents est souvent à l'origine de nombreux dégâts au mobilier intérieur, au matériel, ou aux embellissements du bien touché par la tempête. Ce surcoût n'intervient que si l'arrachage de toitures ou l'ouverture de baies ont permis aux précipitations de pénétrer dans les bâtiments.

Le tableau 2.4.4 donne une estimation de l'impact du passage d'un niveau faible à significatif des précipitations accompagnant une tempête.

Tableau 2.4.4 : Impact sur le montant de la prime pure du passage d'un niveau faible à significatif des précipitations

Risque	agricole	habitation	professionnel
passage d'une direction normale à anormale	+ 22,7 %	+11,9 %	+ 34,8 %

Le niveau des précipitations est un facteur d'inflation du montant de la prime pure. Son influence est d'autant plus forte que les précipitations pénètrent à l'intérieur des bâtiments professionnels ou agricoles, dont le contenu est souvent composé de machines d'une grande valeur.

Type de tempête

Les résultats de ce paragraphe peuvent être comparés à ceux du paragraphe II.2, où figuraient les principales caractéristiques de chaque type de tempête défini par M étéo-France. On retrouve dans le tableau 2.4.5 une estimation de l'impact du passage du type de tempête le moins destructeur au plus destructeur.

Tableau 2.4.5 : Impact sur le montant de la prime pure du passage du type de tempête le moins destructeur au plus destructeur

type de tempête	agricole	habitation	professionnel
le moins destructeur	NE	NE	NE
le plus destructeur	OR	OR	OR
impact du passage de l'un à l'autre	+ 45,4 %	+35,4 %	+ 47,1 %

Les résultats montrent une grande similitude par rapport à ceux du paragraphe précédent : toutes choses égales par ailleurs, et quel que soit le risque, le type de tempête NE apparaît comme le moins coûteux. A l'inverse, l'orage est le type de tempête qui provoque les dégâts les plus importants, à vitesse de vent et concentration d'assurance identiques. Si l'on examine plus avant les résultats de l'annexe O, on constate en outre que le classement complet englobant les sept types de tempête est pratiquement le même, quelque risque que l'on considère, signe du caractère discriminant des types ainsi définis par M étéo-France, du point de vue des dommages.

Le tableau précédent donne en outre une estimation du surcoût qui résulterait de la survenance de deux tempêtes, présentant exactement les mêmes caractéristiques en terme de durée, d'intensité, d'étendue touchée et de concentration d'assurance, mais se distinguant par le type, l'une étant de type NE, et l'autre de type OR : le type OR induit ainsi un surcoût compris selon les risques entre 35 % (pour l'habitation) et 47 % (pour le risque professionnel). Une fois de plus, on relève la moindre exposition aux effets du vent du risque d'habitation.

Effet taille de la tempête

On a vu au paragraphe II.2 que les tempêtes pouvaient être très différentes du point de vue de l'étendue de la zone sinistrée : si les tempêtes les plus importantes peuvent couvrir la moitié du territoire métropolitain, en revanche, d'autres se limitent à une portion réduite d'une zone météorologique. C'est notamment le cas des tempêtes de types E et NE. L'effet taille de la tempête reste ainsi un facteur essentiel dans l'explication des variations de la prime pure.

Le tableau 2.4.6 donne une idée plus précise de cet impact financier, consécutif à l'augmentation de l'effet taille de 10 % à 15 % : en d'autres termes, la part du capital sous risque touché augmente de 5 points.

Si l'augmentation de la prime pure est plus forte que celle à laquelle on pourrait s'attendre (+50 %), à l'exception notable du risque d'habitation (+ 39 %), c'est que,

sans doute, à partir d'une certaine taille, les dommages occasionnés par la tempête croissent de manière exponentielle.

Tableau 2.4.6 : Impact sur le montant de la prime pure de l'augmentation de la part du capital sous risque touché de 10 % à 15 %

Risque	agricole	habitation	professionnel
Impact de l'augmentation de 10 % à 15 % de la part du capital sous risque touché	+ 119 %	+ 39 %	+ 80 %

Effet de « richesse »

Il est établi que le niveau du coût de la vie diffère selon les régions. Néanmoins, la prime pure du tarif incendie n'en tient pas compte, puisque établie en premier lieu à partir du nombre ou de la surface de pièces, et non du niveau de vie de l'assuré. Ce dernier se limitera au choix des garanties de base et n'optera pas pour les garanties facultatives (vol, bris de glace, dégâts des eaux) dont la souscription pourrait grever son budget d'assurance : cette tendance traduisant un *effet de richesse* vise à contrebalancer l'établissement d'un tarif ne retenant pas le coût de la vie comme critère tarifaire.

Il peut être intéressant de voir si cette différence de niveau de vie entre les zones météorologiques se traduit effectivement dans le montant de la prime pure. Le tableau 2.4.7, où figure l'impact par risque de l'augmentation de 2.000 F à 3.000 F de la prime moyenne de chaque zone météorologique, nous permet de répondre.

Tableau 2.4.7 : Impact sur le montant de la prime pure de l'augmentation de la prime moyenne de la zone météorologique de 2.000 F à 3.000 F

Risque	agricole	habitation	professionnel
Impact de l'augmentation de la prime moyenne de 2.000 F à 3.000 F	+ 3,0 %	+ 3,1 %	+ 1,2 %

L'augmentation de la prime pure est, on le voit, peu conséquente. L'effet de richesse existe mais son impact sur le montant de la prime pure est très minime.

Type d'habitat

Comme l'a expliqué le paragraphe I.2.3, la variable « type d'habitat » ne peut prendre qu'une valeur par zone météorologique. Cette contrainte explique grandement

le faible poids explicatif de cette variable et le fait qu'elle ne soit interprétable que pour le risque d'habitation. Il est néanmoins possible de procéder à quelques simulations, dont l'une consiste à imaginer la différence entre le montant de la prime pure résultant de la survenance de deux tempêtes identiques, dont l'une soufflerait sur une zone d'habitat totalement constituée de logements en appartements, et l'autre sur une zone pavillonnaire. **Alors le calcul conduit à un surcoût de 43,8 % pour la zone faite de maisons individuelles.** Les zones météorologiques présentant un habitat varié, la réalité se trouve entre ces deux situations extrêmes.

Superficie

Le modèle accepte aussi comme variable explicative la *superficie* des zones atteintes par le vent : celle-ci est un facteur d'augmentation de la prime pure. Le tableau 2.4.8 donne une idée plus précise de cet impact financier, consécutif à l'augmentation de la superficie des zones touchées de 1.000 à 2.000 km².

Tableau 2.4.8 : Impact sur le montant de la prime pure de l'augmentation de la superficie des zones touchées de 1.000 à 2.000 km²

Risque	agricole	habitation	professionnel
Impact de l'augmentation de 1.000 à 2.000 km ² de la zone touchée	+ 5,8 %	+ 1,7 %	+ 2,4 %

L'impact de la superficie de la zone touchée est très mesuré sur le montant de la prime pure.

Normes de construction

Le renforcement des normes de construction requises peut avoir un effet modérateur sur la fréquence et le montant moyen du préjudice subi par chaque assuré : c'est en tout cas ce qui ressort de l'analyse du modèle, puisque la variable « norme de construction » dispose d'un pouvoir explicatif relativement important.

Le tableau 2.4.9 permet de traduire en chiffres cette analyse : il donne l'impact, sur le montant de la prime pure, du résultat de la survenance de deux tempêtes identiques, dont l'une soufflerait sur une zone météorologique fortement exposée, et l'autre sur une zone faiblement exposée. On pourra à ce sujet se reporter à la carte des vents, en annexe A.

Tableau 2.4.9 : Impact sur le montant de la prime pure du passage d'une zone fortement exposée à une zone faiblement exposée

Risque	agricole	habitation	professionnel
Impact du passage d'une zone fortement exposée à une zone faiblement exposée	+ 43,0 %	+ 9,8 %	+ 28,6 %

Comme on le voit, le niveau des normes de construction requises joue un rôle important dans l'établissement de la prime pure tempête par événement. L'impact de cette variable est manifeste pour le risque agricole (43,0 % de différence entre les deux primes pures) et le risque professionnel (28,6 % de différence), mais il est nettement en retrait pour le risque d'habitation (9,8 %), ce qui ne fait que confirmer les précédentes conclusions : le risque agricole est le plus exposé des trois risques face à la tempête, le risque d'habitation y étant nettement moins sensible.

IV - Conclusion

Les résultats précédents ont prouvé, malgré les difficultés de rapprochement des bases et les contraintes liées au modèle utilisé, qu'il était possible de parvenir à une *modélisation satisfaisante du risque tempête*. L'introduction dans le modèle des *variables météorologiques et d'habitat* a montré son bien-fondé, puisque celles-ci *expliquent en grande partie les variations de la mesure du risque tempête*, quelle que soit la composante considérée (fréquence, coût moyen, prime pure).

La segmentation de la branche incendie selon les trois risques (agricole, habitation, professionnel) a montré sa justesse comme en témoignent les résultats qui montrent de nettes différences : en résumé, *le portefeuille agricole est, de par sa répartition géographique, le plus exposé au risque tempête, alors que le portefeuille habitation en est bien moins dépendant.*

Ainsi, il semble intéressant et utile de s'interroger sur la *réalisation d'un tarif adapté au risque tempête*, et en premier lieu pour le portefeuille agricole, la question des portefeuilles professionnel et de l'habitation ne montrant pas le même degré d'urgence.

Les analyses précédentes ont souligné l'utilité de l'*intégration de nouvelles variables lors de l'établissement d'un nouveau tarif*, et particulièrement *la norme de construction requise*.

Enfin, l'étude de l'impact local d'une tempête a prouvé qu'il est possible d'apprécier de manière satisfaisante le montant du préjudice occasionné par un épisode venteux, à partir des relevés de la station météorologique la plus proche.

*Pour le futur, il pourrait être alors parfaitement envisageable de conclure un accord de partenariat avec Météo-France, qui offrirait un accès à la base « Colchique » où figurent les relevés météorologiques de l'ensemble des stations métropolitaines. Un « balayage » du portefeuille incendie de la région touchée par la tempête permettrait de **disposer très rapidement d'une estimation des dommages.***

Enfin, compte tenu du changement de périmètre intervenu suite à la fusion des deux groupes AXA et UAP, une étape préliminaire à l'utilisation de ce modèle serait l'ajout de données supplémentaires permettant l'affinement du modèle.

CONCLUSION

Événement naturel à part entière, la tempête appartient désormais au domaine assurable, comme la foudre ou la grêle. Malgré cela, ce risque n'a jamais fait l'objet d'une tarification spécifique, son intégration systématique à une garantie de base ou le paiement d'une prime proportionnelle à la prime incendie étant les deux situations les plus communes.

Face à cette carence, il nous a semblé utile de prendre la réelle mesure de ce risque en l'analysant selon toutes ses facettes.

Événement naturel, son évolution ne dépend donc pas de l'intervention humaine, encore que certains scientifiques imputent à l'effet de serre, produit par les hommes, le réchauffement de la planète. Ce dernier serait lui-même à l'origine, pour d'autres experts, de l'augmentation des événements naturels particulièrement dévastateurs, parmi lesquels figurent les tempêtes. Mais les modèles climatiques actuels ne permettent pas d'affirmer que l'éventuel dérèglement climatique peut être un facteur d'augmentation du nombre de tempêtes.

Les études les plus récentes en matière de météorologie ont montré que, sur la période 1962-1993, aucune tendance marquée ne permet de conclure quant à une augmentation du nombre de tempêtes en France. De même, elles ont permis de définir une typologie des tempêtes, basée sur des caractéristiques climatiques.

Le chapitre consacré à la vulnérabilité des constructions a montré de quelle manière l'action des forces dynamiques et statiques s'exerce sur tout bâtiment. Il met en outre l'accent sur les dispositifs de prévention que chaque bâtiment doit comporter pour lutter efficacement contre les effets du vent.

Le chapitre relatif aux aspects législatifs et contractuels soulignent la situation floue dans laquelle le risque tempête a été maintenu, entre 1982 et 1990, contribuant à la déresponsabilisation de nombre d'acteurs.

Au nombre des difficultés rencontrées avant la phase d'analyse du risque tempête figure l'édification d'une base de données. Objet du premier chapitre de la seconde partie, ces phases de collecte d'informations et d'amélioration de leur qualité se révèlent particulièrement délicates en raison principalement de leur lourdeur. Mais l'on a essayé de résoudre du mieux possible chaque difficulté en trouvant des solutions adaptées. Celles-ci reposent sur des hypothèses que différentes techniques algorithmiques ou statistiques permettent en général de justifier.

La base de données relative à la tempête comporte au final des données d'assurances appartenant à une seule compagnie, l'UAP, gage d'une plus grande

homogénéité, des relevés météorologiques et une base relative aux types d'habitat et de construction. Un lien géographique (le code postal) ou temporel (la date de survenance) permet d'unir les différentes dimensions de cette base.

Une fois constituée, la base de données « Sinistres » a servi au calcul et à l'actualisation du préjudice occasionné par la tempête. Chaque sinistre de cette base peut alors faire l'objet d'une comparaison dans les conditions prévalant à la fin de l'exercice 1997. En raison de la longueur de l'historique des sinistres, ces opérations requièrent une grande rigueur et un travail initial très important d'inventaire et de collecte de documents contractuels.

Cette question préliminaire étant réglée, ***le calcul d'un tarif du risque tempête, fondé sur la seule base des sinistres***, décliné par type de risque (agricole, habitation, professionnel) et département de l'assuré, a pu être mené à bien. Sur la période 1980-1997, ***le montant de la prime pure liée au risque tempête est évalué à 126 F***, ce qui représente moins de 10 % de la prime globale incendie ou multirisque.

Par ailleurs, le montant de la prime pure par contrat pour les risques d'habitation et professionnel a été apprécié respectivement à 71 F et 429 F : compte tenu de la prime commerciale généralement pratiquée, il est raisonnable de penser qu'une marge substantielle est dégagée. Le poids modique de cette prime pure relativement à la prime incendie ne justifie pas l'établissement d'un tarif tempête. En revanche, le montant de la prime pure du risque agricole (729 F) est plus significatif, ce qui autorise la réflexion sur l'élaboration d'un tarif.

L'appréciation de ces primes pures peut être revue à la lumière des relevés météorologiques fournis par Météo-France pour la période 1962 - 1993. Il en ressort une nette divergence des conclusions obtenues selon les deux approches puisque la prime pure globale s'établit à 90 F. Les disparités entre risques renforcent les conclusions précédentes, proposant une tarification spécifique pour le risque agricole. Elles justifient a posteriori le bien-fondé de la segmentation de la branche incendie selon les trois risques (agricole, habitation, professionnel). En résumé, ***le portefeuille agricole est, de par sa répartition géographique, le plus exposé au risque tempête, alors que le portefeuille habitation en est le moins dépendant.***

Les résultats précédents ont montré, malgré les difficultés de rapprochement des bases et les contraintes liées au modèle utilisé, qu'il était possible de parvenir à une ***modélisation satisfaisante du risque tempête***. L'introduction dans le modèle des ***variables météorologiques et d'habitat*** a montré sa pertinence, puisque celles-ci ***expliquent en grande partie les variations de la mesure du risque tempête***, quelque composante que l'on considère (fréquence, coût moyen, prime pure).

La modélisation a d'autre part conduit à trois conclusions très intéressantes :

- **le renforcement des normes de construction requises** est un des axes majeurs de réduction du risque tempête; son **intégration lors de l'établissement d'un nouveau tarif** peut y contribuer,
- **la typologie de Météo-France**, basée sur des critères météorologiques, **a permis d'identifier les tempêtes les plus destructrices**,
- enfin, **l'étude de l'impact régional d'une tempête**, à partir des relevés de la station météorologique la plus proche du sinistre, **conduit à une appréciation satisfaisante du montant du préjudice** occasionné par un épisode venteux.

Pour le futur, il est prévu la mise en place d'un **accord de partenariat avec Météo-France**, permettant un accès à la base « Colchique » où figurent les relevés météorologiques de l'ensemble des stations métropolitaines. La lecture du portefeuille incendie de la région touchée par la tempête permettra de **disposer très rapidement d'une estimation des dommages**.

Auparavant, il conviendra d'adapter le modèle à la nouvelle réalité de l'entreprise, compte tenu du changement de périmètre intervenu suite à la fusion des deux groupes AXA et UAP, en l'enrichissant de données supplémentaires. Ce modèle pourra alors faire l'objet d'un **ajout régulier de données** permettant l'affinement des résultats du modèle.

Si le présent mémoire ne devait atteindre qu'un des objectifs poursuivis, il faudrait espérer que ce fût celui d'une meilleure connaissance du risque tempête. A défaut d'une improbable responsabilisation des assurés, elle s'inscrirait au moins dans la perspective actuelle d'une **segmentation plus précise** de chaque risque assurable.

BIBLIOGRAPHIE

OUVRAGES :

TEMPÊTES : Nouvelles dimensions de sinistres d'un phénomène naturel - Münchener Ré - mars 1991

ASSURANCES DES CATASTROPHES NATURELLES : Presses Universitaires - Aix Marseille - 1984

ANALYSE DE LA MACROSTRUCTURE DES VENTS DE TEMPÊTE A L'AIDE DE L'OBSERVATION DES DÉGÂTS : Centre scientifique et technique du bâtiment – Jean Paul FLORI - 1992

RÈGLES DÉFINISSANT LES EFFETS DE LA NEIGE ET DU VENT SUR LES CONSTRUCTIONS ET ANNEXES - EYROLLES. 11^{ème} édition -1983

REVUES :

LES TEMPÊTES EN EUROPE - Suisse de Réassurance

CLIMAT ET RISQUES - Suisse de Réassurance

SUR BIEN DES MARCHÉS, LES CATASTROPHES NATURELLES SONT TROP PEU RÉASSURÉES - SIGMA - Suisse de Réassurance

LES REGIMES ÉTRANGERS D'INDEMNISATIONS DES CATASTROPHES NATURELLES - ARGUS

CATASTROPHES NATURELLES - les zones à risques sous haute surveillance. L'Assurance Française n° 696 - mai 1994

L'ASSURANCE DES CATASTROPHES NATURELLES - F.F.S.A. janvier 1994

ASSURANCE DES CATASTROPHES - RISQUES n° 15 - septembre 1993

CATASTROPHES NATURELLES ET SINISTRES MAJEURS EN 1995 - Suisse de Réassurance SIGMA n° 2 - avril 1996

ASSURANCE DES CATASTROPHES NATURELLES- couverture des événements naturels dans les principaux pays développés - SCOR avril 1996

DOCUMENTATION TECHNIQUE :

ÉTUDE DES ÉVÈNEMENTS TEMPÊTES - Rapport final réalisé par **MÉTÉO-FRANCE** pour un groupe d'Assureurs et de Réassureurs - juillet 1995

AJUSTEMENT D'UNE MATRICE SUR SES MARGES - annales INSEE numéros 22 et 23 - avril / septembre 1976

PROBABILITES ANALYSE DES DONNEES ET STATISTIQUE - Gilbert SAPORTA – Editions Technip – février 1990

METHODES STATISTIQUES DES SONDAGES – Jean-Marie GROSBRAS – Editions Economica – 1991

ARTICLES :

LES ASSUREURS ET L'EFFET DU SERRE - Jean-Jacques DUBY Directeur scientifique UAP - mars 1995.

TROPICAL CYCLONES OF THE ATLANTIC OCEAN 1871 - 1986 - US Departement and Commerce – National Oceanic and atmospheric Administration.

SIMPLIFIED ESTIMATORS OF THE NORMAL DISTRIBUTION , WHEN SAMPLES ARE SINGLY CENSORED OR TRUNCATED - COHEN Technometrics volume 1 – 1991

ASPECTS JURIDIQUES DES CATASTROPHES NATURELLES- Bulletin d'information des Régisseurs n° 18 - UAP - février 1986

MEMOIRES ACTUARIAT :

ASSURANCE DES RISQUES D'ORIGINE MÉTÉOROLOGIQUE EN FRANCE - Application aux tempêtes - Christian de la FOETA - Anne LARPIN - Emmanuel TASSIN et Guy VAN HECKE - CEA / IAF 1994

LA RÉASSURANCE DES GRANDES CATASTROPHES NATURELLES AUX ETAT-UNIS - Benoît VALLEAUX - CEA / IAF 1992

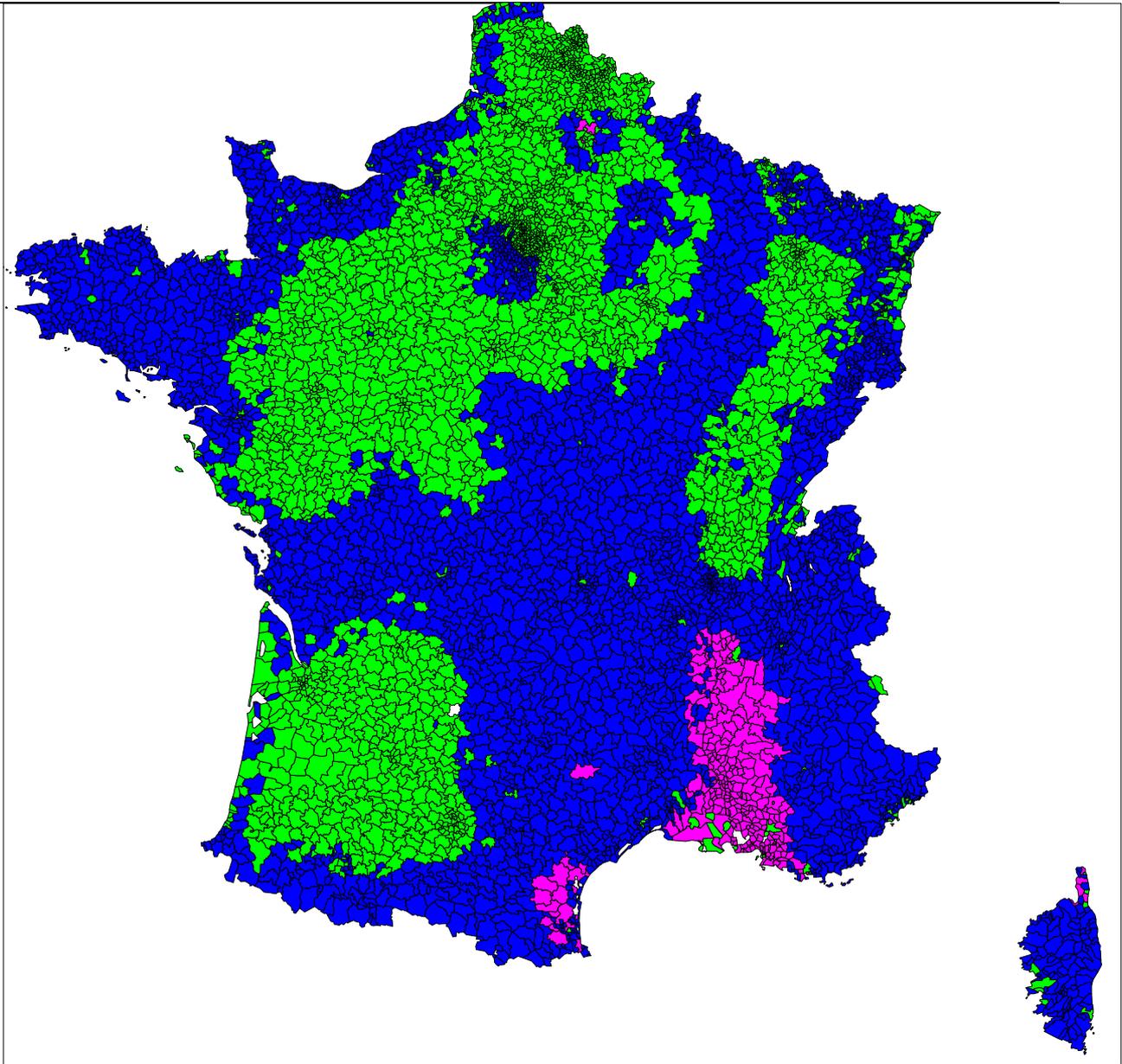
CONTRATS A TERME ET OPTIONS SUR INDICES D'ASSURANCES COUVRANT LES CATASTROPHES NATURELLES AUX ETATS-UNIS - Stéphane CHAPPELIER - Patrick PADEL et Anne STEPHANI - CEA/ IAF 1994

ÉTUDE STATISTIQUE ET ACTUARIELLE DES SINISTRES PROVOQUÉS PAR LES CATASTROPHES NATURELLES AUX ETATS-UNIS EN VUE DE SON APPLICATION A LA RÉASSURANCE- Florence ROTH - ISUP - 1985

ANNEXES

<u>ANNEXE A</u> : Carte des vents	394
<u>ANNEXE B</u> : Constitution de la table code INSEE - code postal et des matrices portefeuille par exercice, entité, risque et zone postale	395
<u>ANNEXE C</u> : Suppression du produit “ mortalité du bétail ” du contenu du risque agricole	397
<u>ANNEXE D</u> : Reconstitution du portefeuille selon le principe des évolutions homogènes	400
<u>ANNEXE E</u> : Détermination du code postal le plus courant pour les contrats plusieurs fois sinistrés et dont le lieu de survenance est au moins une fois connu	402
<u>ANNEXE F</u> : Redressement du code postal du lieu de survenance du sinistre	403
<u>ANNEXE G</u> : Détermination de la station météorologique la plus proche de chaque zone postale ou de chaque station.....	409
<u>ANNEXE H</u> : Caractéristiques démographiques des 130 zones météorologiques	410
<u>ANNEXE I</u> : Extraction et constitution de la segmentation du risque d’habitation.....	414
<u>ANNEXE J</u> : Calcul du taux de présence de la garantie tempête.....	416
<u>ANNEXE K</u> : Calcul et actualisation du préjudice dû à la tempête.....	417
<u>ANNEXE L</u> : Etude de la distribution du montant du préjudice des risques agricole, habitation et professionnel.....	423
<u>ANNEXE M</u> : Etude de l’influence du type de tempête sur la charge actualisée d’une tempête.....	429
<u>ANNEXE N</u> : Modélisation de la charge actualisée par contrat d’une tempête.....	439
<u>ANNEXE O</u> : Etude de l’impact de différentes variables sur le montant de la prime pure	453

ANNEXE A : Carte des vents



LEGENDE :

-  zone de vents faibles (2291 zones postales)
-  zone de vents moyens (3298 zones postales)
-  zone de vents forts (279 zones postales)

ANNEXE B : Constitution de la table code INSEE - code postal et des matrices portefeuille par exercice, entité, risque et zone postale

1 - Programme APL (DEPTTEMP) de constitution de la table de correspondances univoque code INSEE - code postal

```
? RES ? DON DEPTTEMP VEC ;DEPT ;M ;I ;A ;B ;N
[0] RES? 0 2 ?0
[1] D : DEPT? 1?VEC
[2] M? (I ? DEPT=[DON[ ;1]÷1000)≈DON
[3] DON? (~I)≈DON
[4] M? M[×M[ ;1] ;]
[5] M? M[;2 1]
[6] M? M[×M[ ;1] ;]
[7] A? M[ ;1] ,0
[8] B? 0 ,M[ ;1]
[9] N? `1MA∪B
[10] M? N≈M
[11] RES? RES,[1]M
[12] Θ(0< ?VEC? 1 MVEC)/D
```

2 - Procédure HARRY (PORTEMP) de création de matrices portefeuille par exercice, entité, risque et zone postale

```
TAB: $D3&dp+$ENT+$PME
LVARIABLES: PVARIABLE
SELECTION: $DEPAG=&dd
eSELECTION: SPBRAN=14, 15, 16, 50, 51, 52, 62
REPARTITION: SPSIT, SRISQUE: emi c93, emi c94, emi c95, emi c96, ($phi sto1=1), ($phi sto2=1), ($phi sto3=1), ($phi sto4=1)
XLI GNES: DERNIERE
TABLE: INSPOS
LRELATION: TEMP1: C101: C102
CALCUL: C13: COPOS: 0: ((C12=0)*C1)+C12
LRELATION: COPOS: C102: C101
CALCUL: C15: COP01: 0: ((C14=0)*((<<C13/1000)*1000))+C13*(C14>0)
SAUVEGARDE: TEMP
TABLE: TEMP
REPARTITION: COP01, TEMP2: SOM_EMI C93, SOM_EMI C94, SOM_EMI C95, SOM_EMI C96, SOM_TEMP8, SOM_TEMP9, SOM_TEMP10, SOM_TEMP11
XLI GNES: DERNIERE
XCOLONNES: C3
APL: ) COPY VEHVOL PORTEMP REPART
APL: GEN_XCTL 1
APL: TTT? ( < TTT) PORTEMP .. ? 3
APL: TTT? (1 > TTT), [1] (2 > TTT), [1] 3 > TTT
PROC: VARIABLES: TTT
SAUVEGARDE: TEMP&ddb
question: &dp: quelle délégation ?
question: &dd: quel département (RDD)?
```

Programme APL (PORTEMP) intervenant dans la constitution des matrices portefeuille par exercice, entité, risque et zone postale

```
? RES ? M PORTEMP I
[1] MAT? (M[ ;2]=I)≈M
[2] M0? (MAT[ ;1]>0)≈MAT
[3] MAT? (MAT[ ;1]=0)≈MAT
```

```
[4] (C MAT) REPART 2+ ? 8
[5] RES? M0
```

Programme APL (REPART) utilisée pour la répartition par péréquation des nombres ou montants non affectés

```
? MAT REPART K ;PART ;NB ;NBE ;DIF ;CONT ;RESTE ;LISTE ;NUMERO
[1] Θ(~(0=x/ ? M0) ^ 0\0x/ ?MAT)/B0
[2] 'M0 EST VIDE'
[3] Θ0
[4] B0 : PART? ωM0[ ;K] ÷ +/M0[ ;K]
[5] NB ? PARTxCONT? +/MAT[ ;K]
[6] NBE ? [0 .5+NB
[7] DIF ? (+/NBE)-CONT
[8] RESTE ? NB- NBE
[9] LISTE ? ≥RESTE
[10] NUMERO? ( ^1xDIF) ?LISTE
[11] NBE[NUMERO]? NBE[NUMERO]+ ^1xxDIF
[12] M0[ ;K]? M0[ ;K]+NBE
```

ANNEXE C : Suppression du produit “ mortalité du bétail ” du contenu du risque agricole

programme SAS (SUP36134)

```
/*PROGRAMME SUP36134 : SUPPRESSION DES CONTRATS MORTALITE DU BETAIL DU RISQUE AGRICOLE*/

%LET FNB92=467.68;
%LET FNB93=487.60;
%LET T=3;
data p92agr; set tab.p92agr;
rename nombre=nomb92;
run;

data insresag; set tab.insresag; nbmb92=1;
if resag=37504 then resag=20000;
if resag=39918 then resag=10000;
run;

proc sort data=insresag; by insee;
proc means sum data=insresag noprint;
by insee;
var nbmb92;
output out=insmb (drop=_type_ _freq_) sum=nbmbins;
run;

data temp;
merge p92agr insmb;
by insee;
run;

data cgs36134; set tab.cgs36134 (drop=var0);
if som_temp>0 and temp1>-1;
/*supression des lignes avec zéro contrat cgs36134 et ligne total*/
rename som_temp=nomb93 som_emi c=emi c93;
insee=temp1; drop temp1 nombre;
run;

data temp;
merge temp cgs36134;
by insee ;
run;

proc sort data=insresag; by insee resag;
proc means sum data=insresag noprint;
by insee resag;
var nbmb92;
output out=insmb (drop=_type_ _freq_) sum=;
run;

data temp;
merge temp insmb;
by insee;
/*SUPPRESSION DES CONTRATS MORTALITE BETAIL AVEC UN NB CONTRATS SUP OU EGAL A CELUI AGRICOLE*/
if nomb92=. then nomb92=0;
if nbmbins=. then nbmbins=0;
nbagr92p=nomb92/nbmbins;
emi c92p=emi c92;
caressup=0; nbressup=0; emi csur=0; nbsur=0; emi csue=0; nbsue=0;
if nbagr92p<1 then do;
    emi c92p=0;
    emi csur=int(.5+emi c92*nbmb92/nbmbins);
    nbagr92p=0;
    nbsur=nbmb92;
end;

drop goupe;
proc sort data=temp; by resag;

proc sort data=insresag; by resag;
proc means sum data=insresag noprint;
by resag;
var nbmb92;
output out=resmb (drop=_type_ _freq_) sum=nbresmb;
run;

data caresag; set tab.caresag; rename nbmb92=nbres92;
proc sort ; by resag;

data temp; merge temp caresag resmb;
by resag;
if resag>0;
cares92=pnamb92+annmb92;
```

```

drop pnamb92 annmb92;

/*SI LE NOMBRE DE CONTRATS MORTALITE BETAÏL EGALE LE NOMBRE DE CONTRATS DE L'AGENCE */
if (nmb92=nbres92) and (nbsur=0) then do;
    nbressup=nbres92; nbsur=nbres92;
    caressup=cares92; emi csur=cares92;
end;

run;

/*SI LE NOMBRE DE CONTRATS MORTALITE BETAÏL 92 EGALE LE NOMBRE DE CONTRATS MORTALITE BETAÏL
93* /

data temp;set temp;
if nomb93=. then do;
    emi c93=0; nomb93=0;
end;
if ((nbsur=0) and (nbressup=0)) then do;
emi csue=nmb92*((max(nomb93, 1)/nomb92)**&T/nomb92*emi c92
+(emi c93>0)*emi c93*(&FNB92/&FNB93)/max(nomb93, 1))/((max(nomb93, 1)/nomb92)**&T+(emi c93>0));
    emi csue=mi n(emi c92, int(. 5+emi csue));
    emi c92p=emi c92- emi csue;
    nbsue=nmb92;
    nbagr92p=nomb92- nbsue;
end;

run;

proc means sum data=temp noprint;
by resag;
var emi csur nbsur emi csue nbsue;
output out=tempa (drop=_type_ _freq_) sum=emi csu1 nbsu1 emi csu2 nbsu2;

data temp;merge temp tempa;by resag;
cares92b=cares92- emi csu1;
nbres92b=nbres92- nbsu1;
run;

/*REPARTITION PAR NB CONTRATS PAR CODE INSEE DU C. A. 92 AGENCE QUAND EGALITE DU NB CONTRATS
PAR AGENCE ENTRE LE FICHER DE SYNTHESE ET LE FICHER CA PAR AGENCE*/

data temp;set temp;
if (emi csur=0) and ((nbres92=nbresmb) or ((nbresmb- nbres92)*(emi csu1+emi csu2- cares92))<0)
then do;
    emi csur=mi n(emi c92, int(. 5+emi csue*cares92b/emi csu2));
    nbsur=nmb92;
    emi csue=0;
    nbsue=0;
end;

run;

/*SUPPRESSION DU C. A. 92 MORTALITE DU BETAÏL DU C. A. 92 AGRICOLE*/

proc means mean data=temp noprint;
class insee;
var nmb92 emi csue nbsue emi csur nbsur ;
output out=drb sum= ;
run;

data drb;merge p92agr drb;by insee;
if insee=. then delete;
if emi c92=. then delete;
if emi csue=. then emi csue=0;
if emi csur=. then emi csur=0;
if nmb92=. then nmb92=0;
emi c92p=emi c92- emi csue- emi csur;
nbagr92p=nomb92- nmb92;
run;

/*REPARTITION DU C. A. 92 DES CONTRATS AGRICOLES RESILIES EN 1992*/

data tab.p92agrp;set drb;
tot=207134064;
tot1=191895782;
tx=tot/tot1;
emi c92p=int(. 5+emi c92p*tx);
if emi c92p=0 and nbagr92p=0 then delete;
keep insee nomb92 nbagr92p emi c92p emi c92;
run;

```


ANNEXE D : Reconstitution du portefeuille selon le principe des évolutions homogènes

1-programme HARRY (PORTESTAG) de la méthode de reconstitution du portefeuille agent utilisant l'agence selon le principe des évolutions homogènes

```
TAB: TEMP3&dpE
LISTE: C101, C102, C103, c106, c107
APL: GEN_XCTL 1

APL: TTT? ( < TTT) PORTESTAG? ? 3
APL: ECA1? ECA1, [1](1 > TTT) ECARTAG UOAQO 1? [2](1 > TTT)
APL: ECA2? ECA2, [1](2 > TTT) ECARTAG UOAQO 1? [2](2 > TTT)
APL: ECA3? ECA3, [1](3 > TTT) ECARTAG UOAQO 1? [2](3 > TTT)
APL: ECA? ECA, [1]( >, [1]/TTT) ECARTAG UOAQO 1? [2]( >, [1]/TTT)
APL: CP? (2 1)UOAMZ CP, [1]1? [2] >, [1]/TTT
question: &dp: quel département (DD)?
```

Programme APL (PORTESTAG) intervenant dans l'estimation des matrices portefeuille par exercice, agence, risque

```
? RES ? M PORTESTAG I ;AG
[1] M? MODAG956 M ∂TRANSFERT DE PORTEFEUILLE DES ANCIENNES VERS LES NOUVELLES
AGENCES
[2] M? ( ?3) UOAMZ M
[3] M? (M[ ;3]=I)≈M
[4] ∂ SELECTION DES CODES POSTAUX AVEC CONTRATS DE RISQUE I POUR L'EXERCICE
CONNU N+1
[5] AG? UOAQO M[ ;1]
[6] M? ( < M) PORTESTAG1 ? AG
[7] M? > [1] /M
[8] RES? M
```

Programme APL (PORTESTAG1) intervenant dans l'estimation des matrices portefeuille par exercice, agence, risque

```
? RES? M PORTESTAG1 AG ;DIF ; LISTE ;NUMERO
[1] M? (M[ ;1]=AG)≈M
[2] B0 : M? M,( M[ ;5]+0=+/M[ ;5])x(+/M[ ;4]) ÷ +/M[ ;5]+0=+/ M[ ;5]
[3] ∂ ESTIMATION DU NOMBRE DE CONTRATS PAR AGENCE ET CODE POSTAL
[4] M? M[ ; ?5],(L.5+ M[ ;6]) , M[ ;6]
[5] Θ(0=,DIF? -/ +≈ M [ ;4 6])/FIN
[6] M[ ;7]? (M[ ;6]-M[ ;7]) ÷ M[ ;7] O1
[7] LISTE ? ≥ M[ ;7]
[8] NUMERO? ( `1xDIF) ?LISTE
[9] M[NUMERO ;6]? M[NUMERO ;6]+ xDIF
[10] FIN : RES? 6?[2]M
```

Programme APL (UOAMZ) de sommation de matrices

```

? R? V UOAMZ MA ;/O ;I ;W
[1] W? (~WeV)/W? ?1MλMA? (1+O/, MA[ ;V? ∞I? ,V),[I0? 1](⊥2?1,1,λMA)λMA
[2] Θ((0λV? 1MV),0λMA? MA[×MA[ ;V[11]; ])/LC
[3] Θ(>/I? E/Iv1—I? MA[ ;I])/FIN
[4] MA[ ;W] ? +...MA[ ;W]
[5] MA? I=MA
[6] MA[ ;W] ? MA[ ;W] -0,[1] ⊥1 oM MA[ ;W]
[7] FIN : R? ⊥1 oM MA

```

Programme APL (U0AQ0) de détermination des modalités d'une variable

```

? R? U0AQ0 X
[1] R? ⊥1 M (0vX ⊥1∞X)/X? X,1+ ⊥1 λX? X[×X? ,X]

```

2-procédure HARRY (PORTESTDEP) de la méthode de reconstitution du portefeuille de tous les réseaux selon le principe des évolutions homogènes (réseau agent : utilisation des entités départementales)

```

TAB: TEMP3&dpD
LISTE: &dp, C101, C102, c110, c111
APL: GEN_XCTL 1
APL: TTT? (⊂ TTT) PORTESTDEP? ? 3
APL: ECA1? ECA1, [1](1⊃ TTT) ECARTAG U0AQ0 1?[2](1⊃ TTT)
APL: ECA2? ECA2, [1](2⊃ TTT) ECARTAG U0AQ0 1? [2](2⊃ TTT)
APL: ECA3? ECA3, [1](3⊃ TTT) ECARTAG U0AQ0 1? [2](3⊃ TTT)
APL: ECA? ECA, [1](⊃ , [1]/TTT) ECARTAG U0AQ0 1? [2](⊃ , [1]/TTT)
APL: CP? (2 1)UOAMZ CP, [1]1? [2]⊃ , [1]/TTT
question: &dp: quel département (DD)?

```

Programme APL (PORTESTDEP) intervenant dans l'estimation des matrices portefeuille par exercice, entité départementale, risque

```

? RES? M PORTESTDEP I ;DIF ; LISTE ;NUMERO
[1] Θ(0=v/λRES? M? (M[ ;3]=I)≈M)/FIN ∂ SELECTION DES CODES POSTAUX AVEC RISQUE POUR L'EXERCICE N+1
[2] M? M, M[ ;5] v (+/M[ ;4]) ÷ +/M[ ;5]
[3] ∂ ESTIMATION DU NOMBRE DE CONTRATS PAR DEPARTEMENT ET CODE POSTAL
[4] M? M[ ;?5],(⊥.5+ M[ ;6]) , M[ ;6]
[5] Θ(0=,DIF? /+≈ M [ ;4 6])/FIN
[6] M[ ;7]? (M[ ;6]-M[ ;7]) ÷ M[ ;7] O1
[7] LISTE ? ≥ M[ ;7]
[8] NUMERO? ( ⊥1xDIF) ?LISTE
[9] M[NUMERO ;6]? M[NUMERO ;6]+ xDIF
[10] FIN : RES? 6?[2]M

```

ANNEXE E : Détermination du code postal le plus courant pour les contrats plusieurs fois sinistrés et dont le lieu de survenance est au moins une fois connu

programme SAS (TEMP8C)

```
libname htemp 's112.PROJ2.U.SAS.TEMPETE' disp=shr;
run;
data drb;
set htemp.histemp;
copo1=INPUT(copos,5.); DEP=int(copo1/1000);
copn=copo1-(dep*1000);
dept=SUBSTR(NUCON,1,3);
if dept<'325';
run;
proc sort;by nucon;run;
proc means mean noprint data=drb;
where (copn=999);
class nucon;
var paiet;
output out=drb1 mean=;
RUN;
proc means mean noprint data=drb;
where (copn<999);
class nucon;
var paiet;
output out=drb2 mean=;
RUN;
data drb3;merge drb1(in=in1) drb2(in=in2);
by nucon;if in1 and in2;
run;
data drb4;merge drb3(in=in1) drb(in=in2);
by nucon;if in1 and in2 and (copn<999);
run;
proc means mean noprint data=drb4;
class nucon copo1;
var paiet;
output out=drb5 mean=;
run;

data drb6;set drb5;count1=_freq_;where _type_=3;
proc sort;by nucon descending count1 copo1;
proc means max noprint data=drb6;class nucon ;var count1;
output out=drb7 max=count1;
run;
data drb6;merge drb6 drb7;by nucon;
data drb6;set drb6;where _TYPE_=1;
proc print;run;

data _NULL_;
file "B26290.POLCOPNO.CSV";
SET DRB6;
IF _type_=1;
put NUCON ';' copo1 ';' count1 ';';
run;
```

ANNEXE F : Redressement du code postal du lieu de survenance du sinistre

Programme SAS (REDCOPO1) : exploitation des renseignements de la fiche
« contrat »

```
/*TEMPETE : lecture du fichier " sinistres "*/
FILENAME fic 'e:\tempet13\81.csv';
data sin;
infile fic delimiter=';' missover;
input nusin nucon $12. datsi coposin brris vs1 vs2 paiet hono;
NUSINCON=put(nusin, 10.)!!nucon;
run;

/*COPOSRE3 : recherche du code postal d'un code INSEE*/
FILENAME fic 'e:\tempet13\inspos.csv';
data inspos;
infile fic delimiter=';' missover;
input inseeins coposin;
run;
proc sort data=sin; by coposin;
proc sort data=inspos; by coposin;
data sinr; merge sin(in=in1) inspos (in=in2); by coposin; if in1 and not in2;
data sin; merge sin(in=in1) inspos (in=in2); by coposin; if in1 and in2;
run;
proc sort data=sin; by nusincon; run;
data sin; set sin; by nusincon; if first.nusincon; run;

/*PORTEMP : lecture du fichier " portefeuille "*/
FILENAME fic 'e:\tempet13\port1.csv';
data port;
infile fic delimiter=';' missover; length nucon $12. ville $24. adrrisq $22. villrisq $30.
comprisq $30.;
input nucon $ brris copos ville $ insee adrrisq $ villrisq $ comprisq $ coposcou;
run;

/*COPOSRED : rapprochement des fichiers " sinistres " et " portefeuille "*/
proc sort data=port; by nucon;
proc sort data=sinr; by nucon;
data sinr; merge sinr(in=in1) port; by nucon; if in1; run;

proc datasets ;
delete port ;
run ;
/*COPOSRE2 : traduction en code postal des informations présentes sur la fiche écran*/
data sinr; set sinr;
x='abcdefghijklmnopqrstuvwxy-'. /:., &ABCDEFGHIJKLMNPOQRSTUVWXYZ';
y=compress(adrrisq, x);
coposadr=input(substr(y, length(y)-4, 5), 5.);
z=compress(villrisq, x);
coposvil=input(substr(z, length(z)-4, 5), 5.);
w=compress(comprisq, x);
coposcom=input(substr(w, length(w)-4, 5), 5.);

/*récupération du code INSEE présent dans le fichier INSPPOS correspondant au code postal le
plus courant, aux codes postaux de l'adresse et des zones alpha du fichier " contrats "*/

data inspos ; set inspos ; rename coposin=copos ; rename inseeins=inscopos ; run ;
proc sort data=inspos ; by copos ; run ;

proc sort data=sinr ; by copos ; run ;
data sinr ; merge sinr (in=in1) inspos ; by copos ; if in1 ; run ;
proc sort data=sinr ; by nusincon ; run ;
data sinr ; set sinr ; by nusincon ; if first.nusincon ; run ;

data inspos ; set inspos ; rename copos=coposcou ; rename inscopos=inspcou ; run ;

proc sort data=sinr ; by coposcou ; run ;
data sinr ; merge sinr (in=in1) inspos ; by coposcou ; if in1 ;
run ;
proc sort data=sinr ; by nusincon ; run ;
data sinr ; set sinr ; by nusincon ; if first.nusincon ; run ;

data inspos ; set inspos ; rename coposcou=coposadr ; rename inspcou=inscpadr ; run ;

proc sort data=sinr ; by coposadr ; run ;
data sinr ; merge sinr (in=in1) inspos ; by coposadr ; if in1 ;
run ;
proc sort data=sinr ; by nusincon ; run ;
data sinr ; set sinr ; by nusincon ; if first.nusincon ; run ;
```

```

data inspos ;set inspos ;rename coposadr=coposvil ;rename inscpadr=inscpvil ; run ;
proc sort data=sinr ;by coposvil ;run ;
data sinr ; merge sinr (in=in1) inspos ;by coposvil ; if in1 ;
run ;
proc sort data=sinr ;by nusincon ;run ;
data sinr ;set sinr ;by nusincon ;if first.nusincon ;run ;

data inspos ;set inspos ;rename coposvil=coposcom ;rename inscpvil=inscpcom ; run ;
proc sort data=sinr ;by coposcom ;run ;
data sinr ; merge sinr (in=in1) inspos ;by coposcom ; if in1 ;
run ;
proc sort data=sinr ;by nusincon ;run ;
data sinr ;set sinr ;by nusincon ;if first.nusincon ;run ;

data sinr ;set sinr ;
if inscopos=. Then copos=. ;
if inscpou=. Then coposcou=. ;
if inscpadr=. Then coposadr=. ;
if inscpvil=. Then coposvil=. ;
if inscpcom=. Then coposcom=. ;
drop inscopos inscpou inscpadr inscpvil inscpcom ;
run ;

/*RECHERCHE DE NOM DE COMMUNE DANS LES ZONES ALPHA DU FICHIER « CONTRATS »*/
data sinr ;set sinr ;
depsi n=int(coposin/1000);
cna1=coposadr>0 and coposadr<96; cnb1=depsi n=coposadr; cnc1=cna1*10+cnb1;
cna2=coposvil>0 and coposvil<96; cnb2=depsi n=coposvil; cnc2=cna2*10+cnb2;
cna3=coposcom>0 and coposcom<96; cnb3=depsi n=coposcom; cnc3=cna3*10+cnb3;

length zone alpha $ 30;
if cnc1=11 then zone=adrri sq;
if cnc2=11 then zone=villri sq;
if cnc3=11 then zone=compri sq;
alpha=compbl (substr(zone, index(zone, put(depsi n, 2.)), 11));
alpha=put(depsi n, Z2.)!!substr(alpha, 4, 8);
a='0123456789, : &-';
n1=indexc(adrri sq, '1', '2', '3', '4', '5', '6', '7', '8', '9');
zone1=compress(adrri sq, a);
n2=indexc(villri sq, '1', '2', '3', '4', '5', '6', '7', '8', '9');
zone2=compress(villri sq, a);
n3=indexc(compri sq, '1', '2', '3', '4', '5', '6', '7', '8', '9');
zone3=compress(compri sq, a);
zone1=tranwrd(zone1, 'ST', 'ST-');
zone2=tranwrd(zone2, 'ST', 'ST-');
zone3=tranwrd(zone3, 'ST', 'ST-');
zone1=tranwrd(zone1, 'STE', 'STE-');
zone2=tranwrd(zone2, 'STE', 'STE-');
zone3=tranwrd(zone3, 'STE', 'STE-');
zone1=tranwrd(zone1, 'SAINT', 'ST-');
zone2=tranwrd(zone2, 'SAINT', 'ST-');
zone3=tranwrd(zone3, 'SAINT', 'ST-');
zone1=tranwrd(zone1, 'SUR', '-SUR-');
zone2=tranwrd(zone2, 'SUR', '-SUR-');
zone3=tranwrd(zone3, 'SUR', '-SUR-');
zone1=tranwrd(zone1, 'S', '-SUR-');
zone2=tranwrd(zone2, 'S', '-SUR-');
zone3=tranwrd(zone3, 'S', '-SUR-');
zone1=tranwrd(zone1, 'EN', '-EN-');
zone2=tranwrd(zone2, 'EN', '-EN-');
zone3=tranwrd(zone3, 'EN', '-EN-');
zone1=tranwrd(zone1, 'DE', '-DE-');
zone2=tranwrd(zone2, 'DE', '-DE-');
zone3=tranwrd(zone3, 'DE', '-DE-');
zone1=tranwrd(zone1, 'L', 'L');
zone2=tranwrd(zone2, 'L', 'L');
zone3=tranwrd(zone3, 'L', 'L');
zone1=tranwrd(zone1, 'LA', 'LA-');
zone2=tranwrd(zone2, 'LA', 'LA-');
zone3=tranwrd(zone3, 'LA', 'LA-');
zone1=tranwrd(zone1, 'LE', 'LE-');
zone2=tranwrd(zone2, 'LE', 'LE-');
zone3=tranwrd(zone3, 'LE', 'LE-');
zone1=tranwrd(zone1, 'LES', 'LES-');
zone2=tranwrd(zone2, 'LES', 'LES-');
zone3=tranwrd(zone3, 'LES', 'LES-');
if n1>1 then zone1=put(depsi n, Z2.)!!substr(zone1, 1, 8);
if n1<2 then zone1=put(depsi n, Z2.)!!reverse(scan(reverse(compbl (zone1)), 1, ' '));
if n2>1 then zone2=put(depsi n, Z2.)!!substr(zone2, 1, 8);
if n2<2 then zone2=put(depsi n, Z2.)!!reverse(scan(reverse(compbl (zone2)), 1, ' '));
if n3>1 then zone3=put(depsi n, Z2.)!!substr(zone3, 1, 8);
if n3<2 then zone3=put(depsi n, Z2.)!!reverse(scan(reverse(compbl (zone3)), 1, ' '));
alpha=tranwrd(alpha, ' ', '-');
zone1=tranwrd(zone1, ' ', '-');
zone2=tranwrd(zone2, ' ', '-');

```

```

zone3=tranwrd(zone3, ' ','-');
alpha=tranwrd(alpha, 'LA-', 'LA ');
zone1=tranwrd(zone1, 'LA-', 'LA ');
zone2=tranwrd(zone2, 'LA-', 'LA ');
zone3=tranwrd(zone3, 'LA-', 'LA ');
alpha=tranwrd(alpha, 'LE-', 'LE ');
zone1=tranwrd(zone1, 'LE-', 'LE ');
zone2=tranwrd(zone2, 'LE-', 'LE ');
zone3=tranwrd(zone3, 'LE-', 'LE ');
alpha=tranwrd(alpha, 'LES-', 'LES ');
zone1=tranwrd(zone1, 'LES-', 'LES ');
zone2=tranwrd(zone2, 'LES-', 'LES ');
zone3=tranwrd(zone3, 'LES-', 'LES ');
drop adrriq villriq comprisq x y z w cna1 cnb1 cnc1 cna2 cnb2 cnc2 cna3 cnb3 cnc3 zone a n1
n2 n3 ;
run;

/*lecture du fichier COMMUNE pour récupérer le code insee correspondant au libellé de l'écran
portefeuille*/
FILENAME fic 'e:\ign\commune.csv';
data commune;
infile fic delimiter=';' missover;
input a $1. inseeign alpha $10.;
drop a ;
b="-----";
alpha=compbl(alpha);
l=length(alpha);
if l<10 then alpha=substr(alpha, 1, l)!!substr(b, 1, 10-l);
drop b l ;
run;

/*récupération du code insee présent dans le fichier COMMUNE correspondant aux zones alpha du
fichier " contrats "*/
proc sort data=commune ;by alpha ;run ;
proc sort data=sinr ;by alpha ;run ;

data sinr ; length zone1 $ 10 zone2 $ 10 zone3 $ 10 alpha $ 10 ;
merge sinr (in=in1) commune ;by alpha ; if in1 ;/*if first.alpha ;*/
rename inseeign =insalpha ;
run ;

proc sort data=sinr ;by zone1 ;run ;
data commune ;set commune ; rename alpha=zone1 ;
data sinr ; merge sinr (in=in1) commune ;by zone1 ; if in1 ; /*if first.zone1 ;*/
rename inseeign =inszone1;
run ;

proc sort data=sinr ;by zone2 ;run ;
data commune ;set commune ; rename zone1=zone2 ;
data sinr ; merge sinr (in=in1) commune ;by zone2 ; if in1 ;/* if first.zone2 ;*/
rename inseeign =inszone2;
run ;

proc sort data=sinr ;by zone3 ;run ;
data commune ;set commune ; rename zone2=zone3 ;
data sinr ; merge sinr (in=in1) commune ;by zone3 ; if in1 ;
rename inseeign =inszone3;
drop zone1 zone2 zone3 alpha ;
run ;

proc datasets ;
delete commune ;
run ;

/*lecture du fichier AGENCE pour récupérer le code insee correspondant au code agence du
sinistre */
FILENAME fic 'e:\tempet13\agence.csv';
data agenins;
infile fic delimiter=';' missover;
input numag inseeag ;
run;

data sinr ; set sinr ; numag=input(substr(nucon, 2, 4), 4.) ;run ;
proc sort data=sinr ;by numag ;run ;
data sinr ; merge sinr (in=in1) agenins ;by numag ; if in1 ;run ;

/*validité des codes INSEE issus du portefeuille ou du fichier des agences*/
proc sort data=inspos; by inscpcom;run ;
data inspos ;set inspos ; rename inscpcom=insee ;rename coposcom=coposins ;run ;
proc sort data=sinr; by insee;run ;
data sinr ; merge sinr (in=in1) inspos ;by insee ; if in1 ;
run ;

data inspos ;set inspos ; rename insee=inseeag ;rename coposins=coposag ;run ;

proc sort data=sinr; by inseeag;run ;
data sinr ; merge sinr (in=in1) inspos ;by inseeag ; if in1 ;

```

```

run ;
data inspos ;set inspos ; rename inseeag=insalpha ;rename coposag=copalpha ;run ;
proc sort data=sinr; by insalpha;run ;
data sinr ; merge sinr (in=in1) inspos ;by insalpha ; if in1 ;
run ;

data inspos ;set inspos ; rename insalpha =inszone1 ;rename copalpha =copzone1 ;run ;
proc sort data=sinr; by inszone1;run ;
data sinr ; merge sinr (in=in1) inspos ;by inszone1 ; if in1 ;
run ;

data inspos ;set inspos ; rename inszone1 =inszone2 ;rename copzone1 =copzone2 ;run ;
proc sort data=sinr; by inszone2;run ;
data sinr ; merge sinr (in=in1) inspos ;by inszone2 ; if in1 ;
run ;

data inspos ;set inspos ; rename inszone2 =inszone3 ;rename copzone2 =copzone3 ;run ;
proc sort data=sinr; by inszone3;run ;
data sinr ; merge sinr (in=in1) inspos ;by inszone3 ; if in1 ;
run ;

/*SELECTION DES SEULS CODES POSTAUX APPARTENANT AU DEPARTEMENT SINISTRE*/
data sinr ;set sinr ;
drop insalpha inszone1 inszone2 inszone3 numag inseeag inseeins ville brri insee ;
if int(coposins/1000) ne depsi then coposins=. ;
if int(copalpha/1000) ne depsi then copalpha =. ;
if int(copzone1/1000) ne depsi then copzone1 =. ;
if int(copzone2/1000) ne depsi then copzone2 =. ;
if int(copzone3/1000) ne depsi then copzone3 =. ;
if int(coposadr/1000) ne depsi then coposadr =. ;
if int(coposvil /1000) ne depsi then coposvil =. ;
if int(coposcom /1000) ne depsi then coposcom =. ;
if int(coposcou /1000) ne depsi then coposcou =. ;
if int(copos /1000) ne depsi then copos =. ;
if int(coposag /1000) ne depsi then coposag =. ;
run ;

/*EXPORTATION DES DONNEES VERS EXCEL POUR DETERMINER LE MODE DE LA SERIE DES CODES POSTAUX*/
data sin1 ;set sinr ;
keep nusin nusincon coposins copalpha copzone1 copzone2 copzone3 coposadr coposvil coposcom
coposcou copos coposag;
run;

PROC DBLOAD DBMS=EXCEL DATA=WORK.SIN1;
PATH=' E:\TEMPETE9\cpsin81.xls';
PUTNAMES YES;
LIMIT=0;
LOAD;
RUN;

```

Programme REDCOPO2 : redressements ultimes

```

/*TEMPETE : lecture du fichier "code postal sinistres "*/
FILENAME fic 'e:\tempete9\cpsin81a.csv';
data cpsin;
infile fic delimiter=';' missover;
input nusincon $22. coposin1 ;
run;
proc sort data=sinr; by nusincon; run;
proc sort data=cpsin; by nusincon; run;
data sinr;merge sinr (in=in1) cpsin (in=in2);by nusincon;
drop coposins copalpha copzone1 copzone2 copzone3 coposadr coposvil coposcom
coposcou copos coposag;
run;

data sinr;set sinr;datsidp=datsi*100+depsi;run;
proc means mean data=sinr noprint;
where coposin1>0;
class datsidp coposin1;
var paiet;
output out=drb sum=paiet1 ;
run;

data drb;set drb;if _type_=3;run;
proc sort data=drb;by datsidp descending _freq_;run;
data drb;set drb;by datsidp descending _freq_;if first.datsidp;
rename coposin1=coposin2;run;

proc sort data=sinr;by datsidp;run;
data sinr;merge sinr (in=in1) drb;by datsidp;if in1;run;

```

```

data sinr;set sinr; if coposin2=. then coposin2=0;
coposin=(coposin1 ne 0)*(coposin1)+(coposin1 =0)*(coposin2 ne 0)*
(coposin2)+(coposin1=0)*(coposin2=0)*(depsin*1000);
if coposin=75000 then coposin=75016;
if coposin=69000 then coposin=69001;
if coposin=13000 then coposin=13001;

run;

proc sort data=sinr; by nusincon; run;
data sinr;set sinr; by nusincon;if first.nusincon;

drop  depsin coposin1 datsidp coposin2 _type_ _freq_ nusincon paiet1;

run;
data sin;set sin;
drop nusincon inseei ns; run;

FILENAME fic 'e:\tempet14\81a.csv';

data _null_;set sin sinr;
file fic ;
put nusin ';' nucon ';' datsi ';' coposin ';' brris ';' vs1 ';' vs2 ';' paiet ';' hono ;
run;

```

ANNEXE G : Détermination de la station météorologique la plus proche de chaque zone postale ou de chaque station

programme APL (MINDISTZPSM) de détermination de la station météorologique la plus proche de chaque zone postale

```
? MINDISTZPSM [ ]?
[0] R? MINDISTZPSM;ZPLAMB ;SMLAMB
[1] ZPLAMB? LECIGN 'S112.PROJ2.U.IGN'
[2] ZPLAMB? ZPLAMB[ZPLAMB[ ;1] ?U0AQ0 ZPLAMB[ ;1] ;]
[3] SMLAMB? LECSTATION 'S112.PROJ2.U.STATLAMB'
[4] SMNP? 603 1301 1306 1307 1402 1702 2001 2003 2005 2006 2007 2008 2009 2010 2011
[5] SMNP? SMNP, 2201 2901 2902 2904 2906 2907 2908 3303 3503 4003 5001 5402 5602
[6] SMNP? SMNP, 5603 6403 6602 7803 8304 8305 8307 8308 8403 8501
[7] SMLAMB? (~ SMLAMB[;1] e SMNP)≈SMLAMB
[8] R? (c [2]ZPLAMB)MINDISTSMSM1.. c SMLAMB
[9] R? ZPLAMB[; ,1], ⊃ R

? MINDISTZPSM1 [ ]?
[0] R? LAMBZP MINDISTZPSM1 LAMBSM;ZP
[1] ZP? 2 ⊃ LAMBZP
[2] R? LAMBSM[;2] DISTLAMB.. ZP
[3] R? (L.5+L/R),(R=L/R)/LAMBSM[;1]

? DISTLAMB [ ]?
[0] R? A DISTLAMB B
[1] R? (((L/A÷1E5)-L/B÷1E5)÷10)*2)+((((1E5|A)-1E5|B)÷10)*2))*0.5
```

programme APL (MINDISTSMSM) de détermination de la station météorologique la plus proche de chaque station météorologique

```
? MINDISTSMSM [ ]?
[0] R? MINDISTSMSM;SMLAMB
[1] SMLAMB? LECSTATION 'S112.PROJ2.U.STATLAMB'
[2] SMLAMB? (~ SMLAMB[;1]=8306)≈SMLAMB
[3] R? (c [2]SMLAMB)MINDISTSMSM1.. c SMLAMB
[4] R? SMLAMB[; ,1], ⊃ R

? MINDISTSMSM1 [ ]?
[0] R? UNESM MINDISTSMSM1 LISTSM;SM
[1] SM? 2 ⊃ UNESM
[2] LISTSM? (~LISTSM[;1] ⊃ UNESM)≈LISTSM
[3] R? LISTSM[;2] DISTLAMB.. SM
[4] R? (L.5+L/R),(R=L/R)/LISTSM[;1]
```

ANNEXE H : Caractéristiques démographiques des 130 zones météorologiques

NUSTA	NOMSTA	SUPERF	NBHAB90	NBZP	TDISSMZP	MOYDIS	MOYDISP	NUSTAP	DIST
101	AMBERIEU	320 325	243 756	31	801	26	25	6902	34
201	ST-QUENTIN	605 256	549 687	72	2 242	31	30	6202	37
301	VICHY	549 737	287 274	40	1 245	31	28	6301	43
401	ST-AUBAN	640 594	131 871	38	1 031	27	26	501	65
501	EMBRUN	578 423	107 064	33	950	29	29	401	65
601	CANNES	96 553	343 786	26	345	13	7	8303	25
602	NICE	342 343	628 879	39	653	17	6	601	26
701	AUBENAS	394 264	145 696	33	860	26	22	2601	29
801	CHARLEVILLE- MEZIERES	591 790	338 765	45	1 365	30	24	5101	75
901	ST-GIRONS	632 421	190 143	38	1 228	32	33	3102	73
1001	TROYES- BARBEREY	761 516	321 061	47	1 249	27	21	8901	67
1101	CARCASSONNE	777 557	367 136	53	1 650	31	30	6601	73
1201	MILLAU- SOULOBRES	520 993	153 438	33	1 074	33	34	3001	45
1302	SALON-DE- PROVENCE	160 053	233 607	37	578	16	14	1303	27
1303	MARIGNANE	56 536	247 142	18	190	11	10	1305	18
1304	AIX-LES-MILLES	239 620	328 031	43	726	17	11	1303	22
1305	CAP-POMEGUES	52 530	873 034	25	242	10	7	1303	18
1401	CAEN	513 918	563 944	83	1 832	22	19	7601	49
1501	AURILLAC	854 396	257 271	53	1 914	36	32	1901	76
1601	COGNAC	824 900	513 471	67	2 488	37	34	7901	71
1701	LA-ROCHELLE	289 660	316 392	56	1 464	26	19	7901	56
1801	BOURGES	360 136	193 053	20	451	23	13	1802	20
1802	AVORD	337 223	105 717	19	751	40	38	1801	20
1901	BRIVE-LA- GAILLARDE	635 544	268 929	50	1 454	29	22	4601	48
2002	AJACCIO	535 089	125 423	64	2 041	32	24	2004	102
2004	BASTIA	307 893	113 186	40	1 241	31	19	2002	102
2101	DIJON	757 031	525 187	59	1 867	32	20	5202	64
2202	ST-BRIEUC	519 624	382 597	55	1 575	29	24	3501	57
2301	GUERET	728 634	225 018	48	1 675	35	37	8701	64
2401	BERGERAC	752 719	379 639	54	1 900	35	33	4601	72
2501	BESANCON	605 206	372 272	57	1 603	28	20	7001	69
2601	MONTELMAR	434 767	341 836	36	1 002	28	29	701	29
2701	EVREUX	479 134	393 083	58	1 696	29	24	7602	41
2801	CHATEAUDUN	305 674	104 842	14	348	25	22	2802	43
2802	CHARTRES	328 120	283 922	23	496	22	17	2801	43
2903	BREST-GUIPAVAS	153 651	352 593	23	425	18	12	2905	26
2905	LANDIVISIAU	271 398	218 959	33	943	29	29	2903	26
2909	QUIMPER	259 994	285 266	34	750	22	18	2903	55
3001	MONT-AIGOUAL	272 534	124 107	29	663	23	29	1201	45
3002	NIMES- COURBESSAC	193 814	262 939	29	513	18	10	3003	9
3003	NIMES-GARONS	215 406	167 170	22	383	17	17	3002	9
3101	TOULOUSE- BLAGNAC	127 812	196 054	19	264	14	10	3102	6

NUSTA	NOMSTA	SUPERF	NBHAB90	NBZP	TDISSMZP	MOYDIS	MOYDISP	NUSTAP	DIST
3102	TOULOUSE-FRANCAZAL	303 150	647 372	40	729	18	8	3101	6
3201	AUCH	506 173	135 524	34	906	27	23	4701	63
3301	CAZAUX	312 775	125 895	18	389	22	17	3302	51
3302	BORDEAUX	681 565	1 058 955	74	1 785	24	14	3301	51
3401	MONTPELLIER	173 150	430 930	34	376	11	7	3402	27
3402	SETE	241 574	339 670	38	1 168	31	28	3401	27
3501	DINARD	202 797	224 136	31	553	18	16	5003	42
3502	RENNES	566 138	610 099	63	1 437	23	16	3501	65
3601	CHATEAUX	610 258	214 515	28	821	29	22	1801	61
3701	TOURS	548 975	523 104	41	916	22	16	4101	53
3801	ST-MARTIN-HERES	459 643	546 247	57	1 210	21	10	3802	41
3802	GRENOBLE-ST-GEOIRS	359 434	332 381	38	998	26	28	3801	41
3901	LONS-LE-SAUNIER	683 645	396 665	54	2 061	38	42	7102	68
4001	DAX	345 443	146 432	24	520	22	16	6401	42
4002	MONT-DE-MARSAN	539 755	132 319	24	698	29	20	4001	49
4101	BLOIS	307 216	201 662	17	319	19	16	4102	41
4102	ROMORANTIN	343 097	139 761	15	354	24	22	4101	41
4201	ST-ETIENNE	546 924	791 296	89	2 287	26	17	6901	54
4301	LE-PUY-LOUDES	566 567	189 768	40	1 149	29	27	4201	64
4401	ST-NAZAIRE	284 401	331 068	41	1 023	25	20	4402	51
4402	NANTES	483 937	827 826	61	1 333	22	13	4401	51
4501	ORLEANS	521 047	477 590	46	1 058	23	17	2801	47
4601	GOURDON	455 995	152 308	28	744	27	28	1901	48
4701	AGEN	544 335	276 338	46	1 129	25	21	3201	63
4801	MENDE	484 081	70 918	25	691	28	22	3001	50
4901	ANGERS	687 793	676 208	61	1 751	29	25	5301	69
5002	CHERBOURG-VALOGNES	246 795	234 291	31	727	23	18	5003	79
5003	PTE-DU-ROC	360 823	262 383	44	1 345	31	32	3501	42
5101	REIMS	754 005	589 970	54	1 653	31	22	801	75
5201	ST-DIZIER	709 966	239 839	32	929	29	23	5403	73
5202	LANGRES	621 405	145 853	28	824	29	29	2101	64
5301	LAVAL	641 119	392 451	50	1 401	28	27	3502	68
5401	NANCY-ESSEY	351 548	466 515	56	1 064	19	13	5403	25
5403	NANCY-OCHEY	342 454	145 101	27	537	20	22	5401	25
5601	LORIENT-LANN-BIHOUE	254 624	318 803	37	738	20	15	5604	46
5604	VANNES	379 498	296 102	39	902	23	21	5601	46
5701	METZ	644 413	1 037 670	133	4 396	33	31	5401	45
5801	NEVERS	659 834	219 069	40	1 238	31	21	1802	38
5901	DUNKERQUE	181 942	442 411	39	805	21	18	6203	64
5902	LILLE	274 797	1 971 787	160	3 318	21	17	6202	52
6001	BEAUVAIS	498 712	421 810	62	1 661	27	23	9502	58
6101	ALENCON	671 253	279 171	46	1 498	33	28	7201	49
6201	LE-TOUQUET	134 887	105 151	18	349	19	17	6203	23
6202	CAMBRAI-EPINOY	463 944	1 104 185	151	4 096	27	29	201	37
6203	BOULOGNE-SUR-MER	120 414	275 652	24	400	17	17	6201	23
6301	CLERMONT-FERRAND	713 126	564 137	89	2 441	27	18	301	43

NUSTA	NOMSTA	SUPERF	NBHAB90	NBZP	TDISSMZP	MOYDIS	MOYDISP	NUSTAP	DIST
6401	BIARRITZ	240 480	263 555	25	442	18	12	4001	42
6402	PAU-UZEIN	467 801	303 783	30	732	24	19	6501	41
6501	TARBES-OSSUN	518 091	261 366	49	1 223	25	18	6402	41
6601	PERPIGNAN	487 457	384 945	75	1 816	24	15	1101	73
6701	STRASBOURG	531 082	972 427	105	3 038	29	20	6801	64
6801	COLMAR	276 965	323 842	50	926	19	16	6802	37
6802	BALE-MULHOUSE	160 278	435 675	48	665	14	10	6801	37
6901	LYON-BRON	246 956	1 528 690	79	1 294	16	13	6902	11
6902	LYON-SATOLAS	102 594	166 885	16	200	13	14	6901	11
7001	LUXEUIL	368 910	170 239	24	516	22	21	8801	40
7101	ST-YAN	660 647	416 537	54	1 938	36	38	301	56
7102	MACON	442 884	356 993	48	1 133	24	27	101	56
7201	LE MANS	472 162	438 948	39	869	22	15	6101	49
7301	BOURG-ST-MAURICE	541 129	248 879	52	2 057	40	40	7302	69
7302	CHAMBERY-AIX	477 415	611 508	63	1 882	30	31	3802	53
7501	PARIS-MONTSOURIS	36 487	3 935 487	70	578	8	6	9501	9
7601	CAP-DE-LA-HEVE	263 573	528 533	35	846	24	18	1401	49
7602	ROUEN-BOOS	353 820	717 530	66	1 232	19	16	2701	41
7603	DIEPPE	244 660	222 472	30	620	21	21	8001	58
7701	MELUN	561 390	650 798	83	2 427	29	26	9101	27
7801	TRAPPES	201 564	832 093	79	1 371	17	17	7802	13
7802	VILLACOUBLAY	37 807	1 026 991	43	338	8	8	7801	13
7901	NIORT	554 749	339 393	42	1 218	29	26	1701	56
8001	ABBEVILLE	496 107	467 049	71	1 955	28	31	6201	50
8101	ALBI	636 832	324 119	49	1 411	29	25	8201	64
8201	MONTAUBAN	413 609	212 780	32	710	22	16	3101	43
8301	TOULON	83 803	389 755	19	217	11	5	8306	16
8302	LE-LUC	244 376	140 574	25	475	19	17	8306	34
8303	ST-RAPHAEL	144 143	140 471	17	287	17	10	601	25
8306	ILE-PORQUEROLLES	62 715	121 185	9	111	12	7	8301	16
8401	ORANGE	154 855	269 244	29	435	15	16	8402	21
8402	CARPENTRAS	326 077	257 577	40	743	19	17	8401	21
8502	LA-ROCHE-SUR-YON	517 890	401 289	43	1 305	30	28	1701	61
8601	POITIERS	848 664	419 792	57	1 739	31	24	7901	68
8701	LIMOGES-BELLEGARDE	807 894	426 768	64	1 899	30	22	2301	64
8801	EPINAL	369 006	267 242	34	829	24	20	7001	40
8901	AUXERRE	903 107	361 978	65	2 146	33	32	1001	67
9001	BELFORT	273 820	389 217	60	1 162	19	16	6802	38
9101	BRETIGNY-S/-ORGE	134 260	535 289	52	656	13	10	9102	16
9102	PARIS-ORLY	49 035	1 393 493	60	545	9	8	7501	14
9501	LE-BOURGET	57 997	1 740 143	79	1 027	13	9	7501	9
9502	ROISSY-EN-FRANCE	364 666	923 864	100	2 673	27	23	9501	10
	TOTAL	54 307 972	56 393 976	6033	150 581	25	18	-	45

Légende :

<i>NUSTA</i>	<i>Code numérique de la station météorologique (structure : code département + numéro d'ordre dans le département)</i>
<i>NOMSTA</i>	<i>Nom de la station météorologique</i>
<i>SUPERF</i>	<i>Superficie, en hectare, de la zone météorologique</i>
<i>NBHAB90</i>	<i>Population, en nombre d'habitants au recensement de 1990, de la zone météorologique</i>
<i>NBZP</i>	<i>Nombre de zones postales rattachée à la zone météorologique</i>
<i>TDISSMZP</i>	<i>Total des distances, en kilomètre, entre la station météorologique et chaque zone postale</i>
<i>MOYDIS</i>	<i>Distance moyenne, en kilomètre, entre la station météorologique et chaque zone postale</i>
<i>MOYDISP</i>	<i>Distance moyenne, en kilomètre, pondérée par le nombre d'habitants de la zone postale, entre la station météorologique et chaque zone postale</i>
<i>NUSTAP</i>	<i>Code numérique de la station météorologique la plus proche</i>
<i>DIST</i>	<i>Distance, en kilomètre, entre la station météorologique et la station la plus proche</i>

ANNEXE I : Extraction et constitution de la segmentation du risque d'habitation

programme Harry (REPZPHAB)

```
TABLE: SD301+SD302+SD303+SD305+SD307+SD308+SD309+SD310+SD311+SD312+SD313
+SD315+SD350+SD351+SD352+$ent+$pme+$bnp
lvar: pvari ard
SELECTION: $DIZDEP=$dz
eselecti on: $Pcla=3
eSELECTION: $pgro=1
eSELECTION: $pfi c>3
eSELECTION: $phisto1=1
rep: $psit, typo(pyramhab): 1
XLI GNES: DERNIERE
XC01 onnes: c3
TABLE: INSPOS
LRELATION: TEMP1: C101: C102
CALCUL: C5: COPOS: 0: ((C4=0)*C1)+C4
LRELATION: COPOS: C102: C101
CALCUL: C7: COPO1: 0: ((C6=0)*((<<C5/1000)*1000))+C5*(C6>0)
SAUVEGARDE: TEMP
TABLE: TEMP
```

Table d'encodage (PYRAMHAB) des différentes modalités de la segmentation précédente

```
LOA: ((ptyp='MMD', 'MHO', 'MHA') et (((Spn12a=1, 2) et
(Spn31a=10, 11, 20, 21, 30, 31, 40, 41, 50, 51, 60, 61, 70, 80, 200, 201, 202, 203, 204)) ou
(Spn31a=120, 121, 122, 123)))

POA: ((ptyp='MMD', 'MHO', 'MHA') et (((Spn12a=1, 2) et
(Spn31a=111, 112, 113, 114, 215, 216, 219, 225, 226, 229, 235, 245, 246, 250, 255, 256, 100, 101, 300, 301, 302, 30
3, 102)) ou (Spn31a=221, 222, 321, 322)))

LOM: ((ptyp='MMD', 'MHO', 'MHA') et (((Spn12a=3, 4) et
(Spn31a=10, 11, 20, 21, 3 0, 31, 40, 41, 50, 51, 60, 61, 70, 80, 200, 201, 202, 203, 204)) ou
(Spn31a=131, 132, 141, 142)))

POM: ((ptyp='MMD', 'MHO', 'MHA') et (((Spn12a=3, 4) et
(Spn31a=111, 112, 113, 114, 215, 216, 219, 225, 226, 229, 235, 245, 246, 250, 255, 256, 100, 101, 300, 301, 302, 30
3, 102)) ou (Spn31a=231, 232, 241, 242, 331, 332, 341, 342)))

PNOM: (((ptyp='IMD') et (Spn12a=1)) ou ((ptyp='MNO') et (Spn11a=1)))

PNOX: (((ptyp='IMD') et (Spn12a>3)) ou ((ptyp='MNO') et (Spn31a=0, 999))) ou
((ptyp<>'MNO', 'IMD') et
($pbraris=1403, 1404, 1413, 1414, 1423, 1424, 1507, 1515, 1517, 1524, 1575, 5015, 5025, 5029, 5065, 5067, 5075
, 6201, 6202, 6226)))

PNOI: (((ptyp='IMD') et (Spn12a=3)) ou ((ptyp='MNO') et (Spn11a=2) et (Spn31a=410, 420)))

PNOA: (((ptyp='IMD') et (Spn12a=2)) ou ((ptyp='MNO') et (Spn11a=2) et
(Spn31a=437, 438, 445, 447, 448, 450, 455)))

LOX: (((ptyp='MHO') et (Spn12a>4) et (Spn31a<100)) ou (((ptyp='MHO') et (Spn31a<3)) ou
(ptyp<>'MHO', 'MHA', 'MMD')) et
($pbraris=1401, 1411, 1421, 1431, 1441, 1442, 1451, 1452, 1459, 1501, 1502, 1511, 1512, 5014, 5024)))

POX: (((ptyp='MHO') et (Spn12a>4) et (Spn31a>100)) ou (((ptyp='MHO') et (Spn31a<3)) ou
(ptyp<>'MHO', 'MHA', 'MMD')) et
($pbraris=1405, 1415, 1425, 1435, 1443, 1445, 1453, 1455, 1458, 1503, 1504, 1506, 1508, 1513, 1514, 1516, 1523
, 5016, 5026)))

OCX: (((ptyp='MHO') et (Spn31a<3)) ou (ptyp<>'MHO', 'MHA', 'MMD')) et
($pbraris=1434, 1436, 1444, 1447, 1454, 1490, 1505, 1526, 1527, 1540, 1542, 1543, 1544, 1546, 1547, 1560, 1561
, 1562, 1563, 1564, 1592, 1608, 1609, 1610, 1612, 5001, 5004, 5007, 5008, 5011, 5018, 5021, 5027, 5060, 5063, 507
1, 5072, 5073, 5087, 5092, 5218, 6227))

OCA: ((ptyp='MHO') et (Spn12a=1, 2) et (Spn31a>300))

OCM: ((ptyp='MHO') et (Spn12a=3, 4) et (Spn31a>300))
```

ANNEXE J : Calcul du taux de présence de la garantie tempête

programme SAS (TPGTEMP)

```
FILENAME fic 'e:\tempet11\histprtbr.csv';
data temp;
infile fic delimiter=';' missover;
input brris nomb ex; if brris ne .;
run;
data tpgt; set tab.tpgt; run;
proc sort data=temp; by brris; run;

data brri; set tab.brri; run;
proc sort data=brri; by brris; run;
data temp; merge temp (in=in1) tpgt brri; by brris ; if in1; run;

data temp; set temp; coef=1;
if pres='abs' then delete;
if groupe=1 and pres='n/o' then coef=.8;
if groupe=1 and pres='n' then coef=.7;

if groupe=61 and pres in ('n/o','n') then coef=.8;
if groupe in (31,32) and pres in ('n/o','n') then coef=.67;

nombtemp=coef*nomb;
run;

proc means sum data=temp noprint;
class ex groupe;
var nombtemp nomb;
output out=drb sum=;
run;

data tab.tpgtgr; set drb; if _type_=3; tpgt=int(.5+100*nombtemp/nomb);
keep ex groupe tpgt ;
run;
```

ANNEXE K : Calcul et actualisation du préjudice dû à la tempête

programme SAS (PREJUDICE)

```
/*DETERMINATION DU MONTANT DE LA FRANCHISE A APPLIQUER AU SINISTRE */
/*lecture du fichier des sinistres */
  FILENAME fic 'e:\tempet16\80b.csv'; /*à changer */
data sin1;
infile fic delimiter=';' missover;
input nusin nucon $12. D datsi coposin brris vs1 vs2 paiet hono;
drop d ;if hono=. then hono=0; if paiet=. then paiet=0;
tcout=int((paiet-hono)/1000);
prncip=paiet-hono;
numag=input(substr(nusin, 5, 4), 4.);
if prncip>0;
exsi=1900+int(datsi/10000);
br=int(brris/100);
if br in (14, 15, 16, 50, 51, 52, 62);
run;

/*récupération du code réseau de l_agence pour la distinction entre Urbaine et Union*/
data resag;set tab.agence;run;
proc sort data=sin1;by numag;run;
data sin1;merge sin1(in=in1) resag;by numag;if in1;run;

/*lecture du fichier des indices FNB et RI */
FILENAME fic 'e:\tempete7\indval.csv';
data indval;
infile fic delimiter=';' missover;
input exsi mtindfndb mtindri;
run;

/*récupération de données supplémentaires pour caractériser les types de franchise */

data sin1;merge sin1(in=in1) indval;by exsi;if in1;
data brris;set tab.brris;run;
data tpgt;set tab.tpgt;run;
proc sort data=sin1;by brris;run;
proc sort data=tpgt;by brris;run;
proc sort data=brris;by brris;run;

data sin1;merge sin1(in=in1) tpgt brris;by brris;if in1;
mtfr=0; mtfrmi n=5*mtindfndb;
if (datsi<830400)
then mtfr=min(max(5*mtindfndb, .20*prncip/.80), 30*mtindfndb); /*masse des contrats*/

if ( reseau='A' and (datcrea='av 1970' or datcrea<'1971') and datsi<830400) then do;
  mtfr=min(max(500, .20*prncip/.8), 2000);
  mtfrmi n=500; end; /*annexe Union 7103*/

if ( reseau='A' and datsi<830400
and (brris in (1405, 1415, 1425, 5015, 1404, 1414, 1424, 1403, 1413, 1423, 1401, 1411, 1421)) then do;
mtfr=int(.5+mtindfndb/63)*100;
mtfrmi n=int(.5+mtindfndb/63)*100; end; /*garantie intégrée Combinée Union*/

if ( reseau='B' and datsi<830400) and (brris in (1423, 1424, 1425)) then do;
mtfr=min(max(3*mtindfndb, .05*prncip/.95), 15*mtindfndb);
mtfrmi n=3*mtindfndb; end; /*garantie intégrée Globale Habitation*/

if (groupe=61 and datsi<830400) then do;
  mtfr=min(max(10*mtindfndb, .20*prncip/.8), 40*mtindfndb);
  mtfrmi n=10*mtindfndb; end; /*risque agricole*/

if (groupe =01 and (datsi>830400)) then do;
  mtfr=5*mtindfndb;
  mtfrmi n=5*mtindfndb; end; /*risque habitation après 1983*/

if (groupe in(31, 32, 61) and (datsi>830400)) then do;
  mtfr=max(15*mtindfndb, .10*prncip/.9);
  mtfrmi n=15*mtindfndb; end; /*risques agricole et professionnel après 1983*/

if (groupe=61 and (datsi>840600)) then do;
  mtfr=max(7.5*mtindfndb, .20*prncip/.8);
  mtfrmi n=7.5*mtindfndb; end; /*risques agricole après 1984*/

if (groupe=61 and (datsi>880300)) then do;
  mtfr=min(max(7.5*mtindfndb, .20*prncip/.8), 40*mtindfndb);
  mtfrmi n=7.5*mtindfndb; end; /*risques agricole après 1988*/

if (pyram='RI') then do;
  mtfr=max(2*mtindri, .10*prncip/.9);
  mtfrmi n=2*mtindri; end; /*risque industriel*/
```

```

if (datsi<830400) and (brris=1508) then do;
mtfr=min(max(650, .20*princip/.8), 3750);
mtfrmin=650; end; /*Logipack*/

if (datsi>910000) and (brris in (1545, 1546, 1547, 1560, 1561, 1562)) then do;
mtfr=750;
mtfrmin=750; end; /*contrat CH*/

if (datsi>920000) and (brris in (1545, 1546, 1547, 1560, 1561, 1562)) then do;
mtfr=800;
mtfrmin=800; end; /*contrat CH*/

if (datsi>930000) and (brris in (1545, 1546, 1547, 1560, 1561, 1562)) then do;
mtfr=900;
mtfrmin=900; end; /*contrat CH*/

if brris=1542 then do;
mtfr=2000;
mtfrmin=2000; end; /*habitation N° 1*/

if (datsi>910000) and (brris in (1542, 1543)) then do;
mtfr=500;
mtfrmin=500; end; /*contrat CH N° 1*/

if (datsi>930000) and (brris in (1542, 1543)) then do;
mtfr=600;
mtfrmin=600; end; /*contrat CH N° 1*/

if (datsi>850000) and (brris in (1572, 1593, 1594, 5058, 5059, 5078, 5079, 5089)) then do;
mtfr=max(4500, .10*princip/.9);
mtfrmin=4500; end; /*contrat MP85*/

if (datsi>900000) and (brris in (1572, 1593, 1594, 5058, 5059, 5078, 5079, 5089)) then do;
mtfr=max(1300, .10*princip/.9);
mtfrmin=1300; end; /*contrat MP85*/

if (vs1=30 or vs2 in (30, 35, 99, 1230, 1930, 3099)) and marche='PRO' then do;
mtfr=max(mtfr, .10*princip/.9, 4000);
mtfrmin=max(mtfrmin, 4000); end;

if (vs1=30 or vs2 in (30, 35, 99, 1230, 1930, 3099)) and marche='PRO' and datsi>831000 then do;
mtfr=max(mtfr, .10*princip/.9, 4500);
mtfrmin=max(mtfrmin, 4500); end;

if (vs1=30 or vs2 in (30, 99, 1230, 1930, 3099)) and marche='PAR' then do;
mtfr=max(mtfr, 800);
mtfrmin=max(mtfrmin, 800); end;

if (vs1=30 or vs2 in (30, 99, 1230, 1930, 3099)) and marche='PAR' and datsi>831000 then do;
mtfr=max(mtfr, 800);
mtfrmin=max(mtfrmin, 800); end;
drop nusin nucon paiet hono tcout ct moy nomb mtindfnb mtindri pres cgs
nb br marche typ ;
run;

proc means mean data=sin1 noprint;
class mtfrmin ;
var mtfr princip;
output out=drb sum= ;
run;
/*proc print data=drb; run; */

/*DETERMINATION DE LA SURCHARGE DUE A L'ABAISSMENT DE LA FRANCHISE A ZERO */
/*regroupement des sinistres selon le niveau minimal de franchise*/

data sin1;set sin1;
/*lignes à modifier en fonction des regroupements de niveaux minima de franchise */
if mtfrmin=650 then mtfrmin=651; /*à changer*/
y=log(princip + mtfrmin);
risque=3;
if groupe=1 then risque=2;
if groupe=61 then risque=1;
ninf=princip<mtfrmin ;
num=1 ;
run;

proc means mean data=sin1 noprint;
class risque mtfrmin ;
var y princip ninf num;
output out=drb mean=ym var=yv sum=ys paiets ninf num;
run;

data drb;set drb;
omega=yv/(ym-log(mtfrmin))**2;
p4=-.00374615+.17462558*omega-.2.87168509*omega**2+17.4893265*omega**3-11.9171654*omega**4;
p3=1+.5.74050101*omega-.13.53427037*omega**2+.6.88665552*omega**3;
q=p4/p3;
mu=ym-q*(ym-log(mtfrmin));

```

```

sigma2=yv+q*(ym-log(mtfm n))**2;
espx=exp(mi+sigma2/2);
na=(paiets+_freq_*(mtf m n- espx))/(espx- mtf m n/2);
na=int(.5+na);
if (na<0 or omega>.60) then na=ninfs ;
ca=int(.5+na*mtf m n/2);
run;

proc sort data=sin1;by risque mtf m n;run;
proc sort data=drb;by risque mtf m n;run;
data sin1;merge sin1(in=in1) drb;by risque mtf m n;if in1;
indfra=1+na/nums ;
cafra=ca/nums ;
drop y ninf ym yv ys paiets ninfs omega p4 p3 q mu sigma2 espx _freq_ num ca;
run;

/*PRISE EN COMPTE DE LA SURCHARGE DUE A LA SUPPRESSION DE LA LIMITE DE GARANTIE */
proc sort data = sin1 ;by groupe ;run ;

data lingar ;set tab.lingar ;run ;

data sin1 ;merge sin1(in=in1) lingar ;by groupe ; if in1 ;
indling=1 ;
if (datsi<830400) and (groupe in (1,31,61)) then indling =1+tx900fnb; /*masse des contrats */

if (reseau='A') and (datcrea=' av 1970' or datcrea<' 1971') and (datsi<830400) and (groupe =1)
and (brris not in (1405, 1415, 1425, 5015, 1404, 1414, 1424, 1403, 1413, 1423, 1401, 1411, 1421)) then do;
indling =1+tx40kf80*(exsi=1980)+ tx40kf81*(exsi=1981)+ tx40kf82*(exsi=1982); end;
/*annexe Uni on 7103*/

if ( reseau='B') and (datcrea=' av 1970' or datcrea<' 1971') and (datsi<830400) and (groupe =1)
and (brris not in (1423, 1424, 1425)) then do;
indling =1+tx600fnb; end; /*garantie T. G. A. Urbaine*/

if ( numag=10000) and (datcrea=' av 1970' or datcrea<' 1971') and (datsi<830400) then do;
indling =1+txmilfnb; end; /*garantie T Séquanaise*/
if (groupe =32) or (vs1=30 or vs2 in (30, 99, 1230, 1930, 3099)) then indling =1 ;

run ;

/* PRISE EN COMPTE DES EXCLUSIONS DANS LE MONTANT DU PREJUDICE*/
%LET NIVA=0.011;
%LET NIVC=0.028 ;
%LET NIVE=0.026 ;
%LET NIVG=0.013 ;
%LET NIVM=0.017 ;
%LET NIVI=0.001 ;
%LET NIVO=0.038 ;

data sin1 ;set sin1 ;
indexc=1 ; /*suppression des exclusions pour la masse des contrats après 1990*/
if (datsi<901000 and groupe=1) then indexc= sum(1, &NIVA, &NIVC, &NIVE, &NIVG, &NIVM, &NIVI, &NIVO) ;
/* contrats habitation avant 1990*/

if (datsi<901000 and brris in(1542, 1543, 1544, 1563)) then indexc=
sum(1, &NIVE, &NIVG, &NIVM, &NIVI) ;
/* contrats CHLA- CHPM et CH N°1 avant 1990*/

if (datsi<901000 and (groupe in (31, 32))) then indexc= sum(1, &NIVC, &NIVG, &NIVM) ;
/* contrats professionnels avant 1990*/

if (datsi<930600 and groupe=61) then indexc= sum(1, &NIVC, &NIVG, &NIVM) ;
/* contrats agricoles avant 1993*/

if (vs1=30 or vs2 in (30, 1230, 1930, 3099)) then indexc=1 ;
/* sinistres réglés en cat.nat. */

/* PRISE EN COMPTE DE LA VETUSTE DANS LE MONTANT DU PREJUDICE*/
indvet=1 ;
if (datsi<901000 and groupe=1) then indvet= 1.069 ;
/* contrats habitation avant 1990*/

if (datsi<930600 and groupe=61) then indvet= 0.755 ;
/* contrats agricole avant 1990*/

/* PRISE EN COMPTE DE LA DOUBLE VETUSTE DANS LE MONTANT DU PREJUDICE*/
inddvet=1 ;
if (datsi<830300) then inddvet= 1.01 ;

/* PRISE EN COMPTE DE L'EVOLUTION LIEE A LA T. V. A. */

indtva=1 ;
if (datsi<950700 and groupe=1) then indtva= 1.017 ;

```

```

/* contrats habitation avant 1995*/
if (datsi<820700 and groupe=1) then indtva= 1.026 ;
/* contrats habitation avant 1982*/

if (datsi<950700 and groupe =61) then indtva= 1.001 ;
/* contrats habitation avant 1995*/

if (datsi<820700 and groupe=61) then indtva= 1.002 ;
/* contrats habitation avant 1982*/

run ;

/* PRISE EN COMPTE DE L' EVOLUTION LIEE A L_INFLATION. */

FILENAME fic 'e:\tempete7\indfnbri.csv';
data ind;
infile fic delimiter=';' missover;
input aatt indfnb indri;
run;

/*choix de l_indice de référence (F.N.B. ou R.I.)*/
data sin1 ;set sin1 ;
mmsi=mod(int(datsi/100), 100) ;
if mmsi in (1,2,3) then aatt=(exsi-1900)*100+3 ;
if mmsi in (4,5,6) then aatt=(exsi-1900)*100+6 ;
if mmsi in (7,8,9) then aatt=(exsi-1900)*100+9 ;
if mmsi in (10,11,12) then aatt=min((exsi-1900+1)*100+3, 9712) ;
run ;

proc sort data=sin1 ;by aatt ;run ;

data sin1 ;merge sin1(in=in1) ind ;by aatt ; if in1 ;
indinf= indfnb ;
if pyram='RI' then indinf= indri ;
run ;

/*PRISE EN COMPTE DU TAUX DE SOUSCRIPTION DE LA GARANTIE TEMPETE */
proc sort data = sin1 ;by exsi groupe ;run ;

data tpgtgr ;set tab.tpgtgr ;rename ex=exsi ;run ;
proc sort data = tpgtgr;by exsi groupe ;run ;

data sin1 ;merge sin1(in=in1) tpgtgr ;by exsi groupe ; if in1 ;
indtpgt=1 ;
if (datsi<830300) then indtpgt=100/tpgt ;
if (datsi<900800 and pyram='RI') then indtpgt=100/tpgt ;
run ;

/*SOMMATION PAR RISQUE, ZONE POSTALE ET DATE DE SURVENANCE */

data sin1 ;set sin1 ;
surnb= indtpgt*indfra ;
surcout=(princip+mfr+cafra) * indexc*indvet*inddvet*indtva*indinf* indtpgt* indling ;
run ;

proc means mean data=sin1 noprint;
class risque datsi coposin;
var surnb surcout ;
output out=sin2 sum=;
run;

data sin2 ;set sin2 ; if _type_=7 ;
surnb=(int(.5+100*surnb))/100 ;
surcout=(int(.5+100*surcout))/100 ;
run ;

proc means data=sin2 noprint;
var surnb surcout ;
output out=sin3 sum=;
run ;
proc print data=sin3 ;run ;

FILENAME fic 'e:\tempet17\A80c.csv'; /*à changer*/
data _null_;set sin2;
file fic ;
put risque ';' datsi ';' coposin ';' surnb ';' surcout ;
run;

```

ANNEXE L : Etude de la distribution du montant du préjudice des risques agricole, habitation et professionnel

Risque agricole

```
data trca;set trc1;
if risque=1;
y=log(trc);
run;
proc univariate data=trca normal plot;
var y;
freq nb;
run;
```

```
The SAS System                    5
                                08:16 Friday, October 15, 1999

                                Univariate Procedure

Variable= Y

                                Moments

      N          54696      Sum Wgts      54696
      Mean       4.400441   Sum          240686.5
      Std Dev    1.233001   Variance     1.520291
      Skewness   -0.48915   Kurtosis     1.170652
      USS        1142279   CSS          83152.34
      CV         28.01994   Std Mean     0.005272
      T: Mean=0   834.6618     Pr>|T|        0.0001
      Num ^= 0   54362       Num > 0       54362
      M(Sign)    27181     Pr>=|M|       0.0001
      Sgn Rank   7.3882E8   Pr>=|S|       0.0001
      D: Normal  0.071655   Pr>D          <.01

                                Quantiles(Def=5)

      100% Max   10.39757      99%  7.215975
      75%  Q3    5.117994      95%  6.304449
      50%  Med   4.49981       90%  5.817111
      25%  Q1    3.78419       10%  2.833213
      0%  Min    0              5%   2.197225
                                1%   0.693147

      Range      10.39757
      Q3- Q1     1.333804
      Mode       4.143135

                                Extremes

      Lowest      Obs      Highest      Obs
      0(          2) 9.544954( 1685)
      0.693147(   3) 9.652973( 1686)
      1.098612(   4) 9.718001( 1687)
      1.386294(   5) 9.867394( 1688)
      1.609438(   6) 10.39757( 1689)

      Missing Value
      Count          55099
      % Count/Nobs   50.18
```

```
The SAS System                    6
                                08:16 Friday, October 15, 1999

                                Univariate Procedure

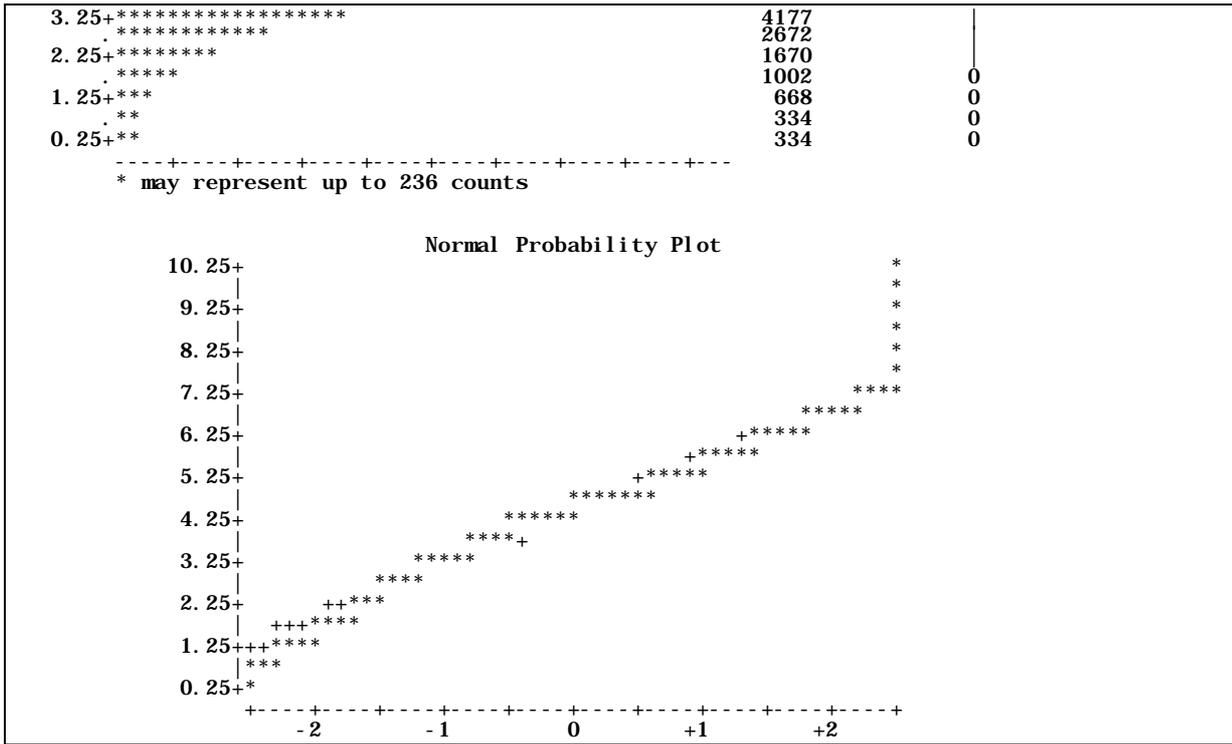
Variable= Y

                                Histogram

      10.25+*
      . *
      . *
      9.25+*
      . *
      . *
      8.25+*
      . *
      . *
      7.25+***
      . *****
      . *****
      6.25+*****
      . *****
      . *****
      5.25+*****
      . *****
      . *****
      4.25+*****
      . *****
      . *****

                                #
      1
      4
      5
      17
      46
      210
      561
      1171
      2280
      4114
      7458
      11315
      10910
      5747

                                Boxplot
      *
      *
      0
      0
      0
      0
      0
      |
      +-----+
      |-----|
      *-----*
      +-----+
```



Risque d'habitation

```

data trch;set trc1;
if risque=2;
y=log(trc);
run;
proc univariate data=trch normal plot;
var y;
freq nb;
run;

```

The SAS System 7

08:16 Friday, October 15, 1999

Uni vari ate Procedure

Vari abl e=Y

Moments			
N	191953	Sum Wgts	191953
Mean	4.135276	Sum	793778.7
Std Dev	1.20568	Variance	1.453665
Skewness	-0.49888	Kurtosis	1.103887
USS	3561528	CSS	279033.8
CV	29.15598	Std Mean	0.002752
T: Mean=0	1502.692	Pr> T	0.0001
Num ^= 0	190447	Num > 0	190447
M(Si gn)	95223.5	Pr>= M	0.0001
Sgn Rank	9.0676E9	Pr>= S	0.0001
D: Normal	0.070432	Pr>D	<.01

Quantiles(Def=5)			
100% Max	10.79581	99%	6.883463
75% Q3	4.867534	95%	5.934894
50% Med	4.234107	90%	5.509388
25% Q1	3.496508	10%	2.564949
0% Mi n	0	5%	1.94591
		1%	0.693147
Range	10.79581		

Q3-Q1 1.371027
 Mode 3.89182

Extremes

Lowest	Obs	Highest	Obs
0.693147(0(9.605216(2047)
1.098612(3)	9.923682(2048)
1.386294(4)	10.12206(2049)
1.609438(5)	10.66359(2050)
	6)	10.79581(2051)

Missing Value
 Count 192294
 % Count/Nobs 50.04

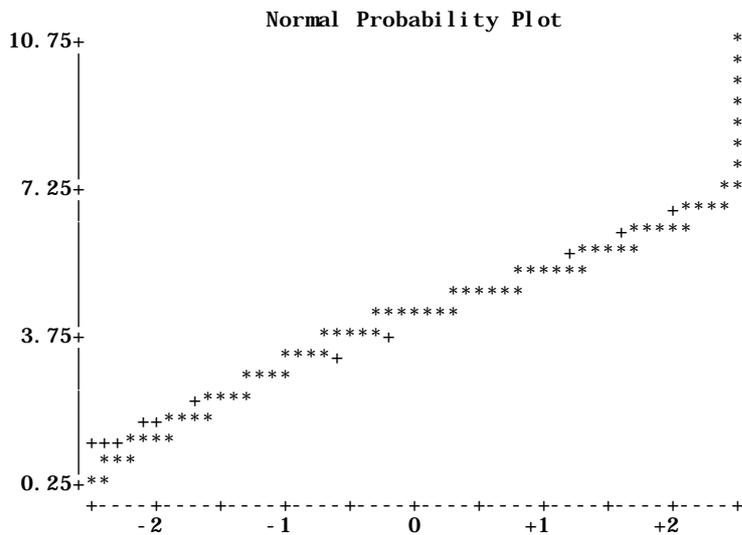
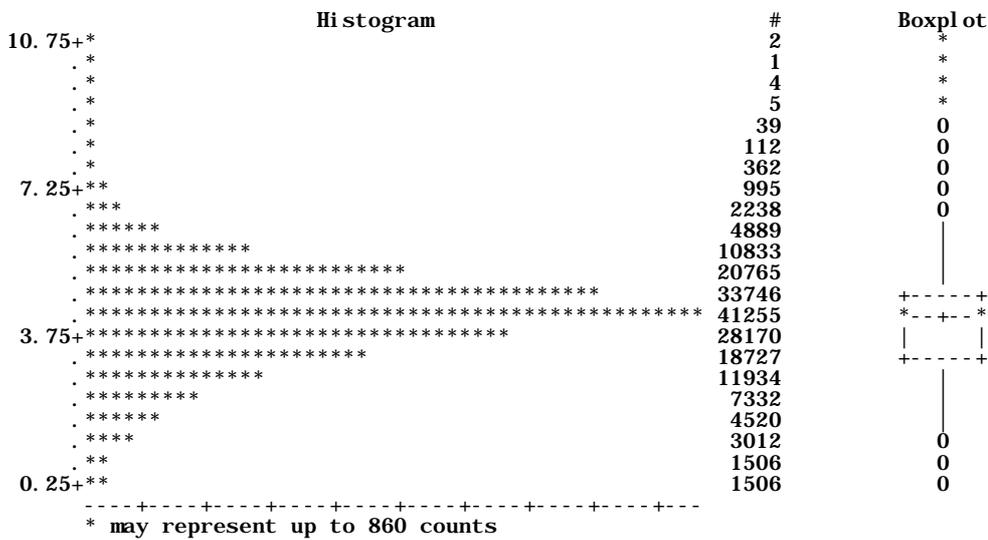
The SAS System

8

08:16 Friday, October 15, 1999

Univariate Procedure

Variable=Y



Risque professionnel

```
data trcp;set trc1;
if risque=3;
```

```

y=log(trc);
run;
proc univariate data=trcp normal plot;
var y;
freq nb;
run;

```

The SAS System		9		08:16 Friday, October 15, 1999	
Uni vari ate Procedure					
Variable=Y					
Moments					
N	66739	Sum Wgts	66739		
Mean	4.549056	Sum	303599.4		
Std Dev	1.282655	Variance	1.645203		
Skewness	-0.32074	Kurtosis	1.517942		
USS	1490888	CSS	109797.6		
CV	28.19607	Std Mean	0.004965		
T: Mean=0	916.2233	Pr> T	0.0001		
Num ^= 0	66372	Num > 0	66372		
M(Si gn)	33186	Pr>= M	0.0001		
Sgn Rank	1.1013E9	Pr>= S	0.0001		
D: Normal	0.05814	Pr>D	<.01		
Quantiles(Def=5)					
100% Max	12.12156	99%	7.682022		
75% Q3	5.273	95%	6.486161		
50% Med	4.691348	90%	5.978886		
25% Q1	3.89182	10%	2.944439		
0% Mi n	0	5%	2.302585		
		1%	0.693147		
Range	12.12156				
Q3-Q1	1.381179				
Mode	2.564949				
Extremes					
Lowest	Obs	Highest	Obs		
0(2)	11.43671(2307)		
0.693147(3)	11.46284(2308)		
1.098612(4)	11.49081(2309)		
1.386294(5)	11.52975(2310)		
1.609438(6)	12.12156(2311)		
Missing Value					
Count			67630		
% Count/Nobs			50.33		

Univariate Procedure

Variable=Y

Histogram

12.25+*

Boxplot



* may represent up to 283 counts

Univariate Procedure

Variable=Y

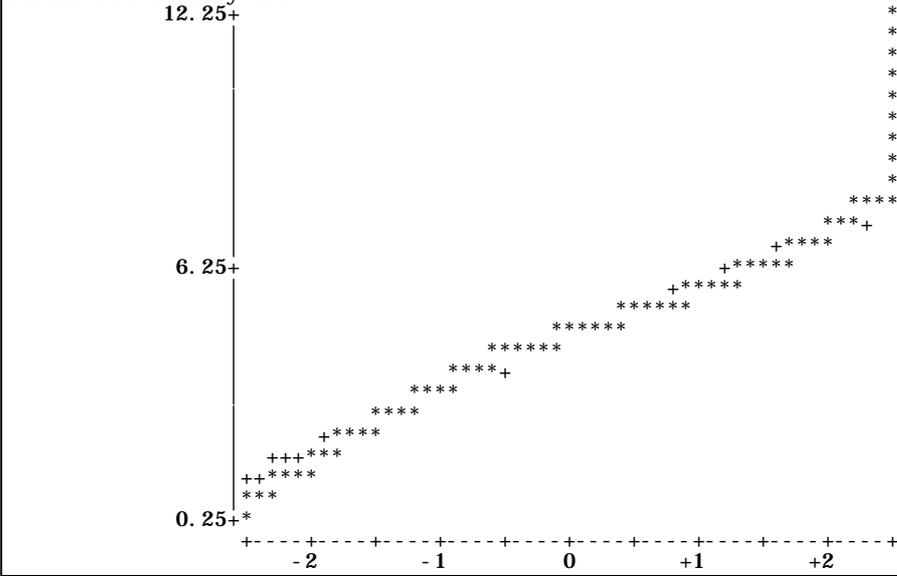
Normal Probability Plot

12.25+*

6.25+*

0.25+*

-2 -1 0 +1 +2



ANNEXE M : Etude de l'influence du type de tempête sur la charge actualisée d'une tempête

Programme CUMNBSIN assurant le rapprochement des événements définis par Météo-France (72 heures de 1980 à 1996) et des tempêtes définies par nos propres soins

```
/*RAPPROCHEMENT DES DONNEES SINISTRES ET METEO (FICHER ARCHIVES) A PARTIR DES DATES
EVENEMENT*/
FILENAME fic 'e:\tempet17\sinzpr.csv';
data sin;
infile fic delimiter=';' missover;
input risque datsi copos nbsin mtsin ;
run ;

proc means data=sin noprint;
class datsi ;
var nbsin mtsin;
output out=sin1 sum=nbsin1 mtsin1 ;
run ;
data sin1 ;set sin1 ;if _type_=1 ;nbzpsi=_freq_ ;run ;

proc means data=sin noprint;
class risque datsi ;
var nbsin mtsin;
output out=sin sum= ;
run ;

data sin;set sin;
if datsi ne . and risque=.;
run;
options center;
/*libname don 'd:\acn\exemple' ;*/ /* Allocation de la librairie */

/* on trie la table SAS par risque et date */
proc sort data=sin out=temp ;
/* AJOUTER LES AUTRES VARIABLES DE TRI */
by risque datsi;
run;

data a(drop= jour mois an date); /* modification des dates */
set temp;
%let val=15; /*SEUIL DU NB DE SINISTRES A PARTIR DUQUEL ON CONSIDERE QU' UNE JOURNEE EST
VENTEUSE */
date=put(datsi, $6.);
jour=substr(date, 5, 2);
mois=substr(date, 3, 2);
an=substr(date, 1, 2);
date2=mdy(mois, jour, an);
format date2 date8.;
/*regroupement de 2 tempêtes, Météo-France ne donnant qu'un relevé*/
run;

data b;
set a;
by risque ; /* METTRE LES MEMES VARIABLES QUE POUR LE TRI LA DATE EN MOINS */

retain testdt ; /* variable permettant de tester la date*/
/* sa valeur sera gardee d'un enregistrement à l'autre */

retain cumul; /* variable cumul du nombre de sinistres */
/* sa valeur sera gardee d'un enregistrement à l'autre */

retain cumul1; /* variable cumul du montant des sinistres */
/* sa valeur sera gardee d'un enregistrement à l'autre */

retain numtemp; /*comptage nombre de tempetes */

retain nbjour ; /*comptage nombre de jours de la tempete */

/* 1er enregistrement pour un risque donné */
/* on initialise cumul et testdt */
if first.risque=1 then do;
cumul=(nbsin)*(nbsin>&val);
cumul1=(mtsin)*(nbsin>&val);
testdt=date2;
numtemp=0;
nbjour=0 ;
end;

if nbjour=3 then do ;
```

```

        cumul=0 ;
        cumul1=0 ;
        nbjour=0 ;
end ;

/* date = date précédente +1 */
/* on fait le cumul */
    if (date2 = testdt+1) then do; /* AJOUTER LA CONDITION NBSIN > VALEUR */
        numtemp=numtemp+(cumul=0)*(nbsin>&val);
        nbjour=(nbjour+1)*(nbsin>&val);
        cumul=(cumul+nbsin)*(nbsin>&val);
        cumul1=(cumul1+mtsin)*(nbsin>&val);
        testdt=date2;

    end;

/* date ne date précédente +1 */
/* on reinitialise cumul et testdt */
    else do;
        nbjour=(nbsin>&val);
        numtemp=numtemp+(nbsin>&val);
        cumul=nbsin*(nbsin>&val);
        cumul1=mtsin*(nbsin>&val);
        testdt=date2;
end;
format testdt date8.;
run;

data c;set b;if cumul>0;
drop _type_ _freq_ nbsin mtsin date2 ;
run;

proc sort data=c out= d ;by numtemp descending nbjour ;run ;
data d ;set d ;by numtemp descending nbjour ;if first.numtemp ;
rename cumul = cumuln testdt =dernjour nbjour=nbjourt cumul1 = cumulm ;run ;

data e ;set c ;by numtemp nbjour ;if first.numtemp ;
rename testdt=premjour ;run ;

data f ;merge e(keep=numtemp premjour) d(keep=numtemp dernjour cumuln nbjourt) ;
by numtemp ;
if numtemp in (301, 142, 622, 372, 462, 106, 532) then numtemp=numtemp-1 ;
run ;

/*regroupement de 2 tempêtes sinistres pour lesquelles METEO-France ne donne qu'un relevé ;*/
proc means data=f noprint;
class numtemp ;
var cumuln cumulm nbjourt;
output out=g sum= ;
run ;
data g ;set g ;if _type_ =1 ;
data f ;merge f g;by numtemp ; if first.numtemp ;drop _type_ _freq_;
run ;
data d ;set d ;
if numtemp in (301, 142, 622, 372, 462, 106, 532) then numtemp=numtemp-1 ;
run ;

data f ;merge f d(keep=numtemp dernjour) ;by numtemp ;run ;
data f ;set f ;by numtemp ;if last.numtemp ;run ;

data e ;set e ;
if numtemp in (301, 142, 622, 372, 462, 106, 532) then numtemp=numtemp-1 ;
run ;

data f ;merge f e(keep=numtemp premjour) ;by numtemp ;run ;
data f ;set f ;by numtemp ;if first.numtemp ;run ;

/*lecture du fichier des variables explicatives météorologiques (1980-1993) */
FILENAME fic 'e:\tempete\varexp.csv';
data varexp;
infile fic delimiter=';' missover;
input d $ 1. datsi tytemp $ ora nei gre lmaxprec lminpres presmoy difpres vmoyno vmoyne
vmoyso vmoyse dirveo dirvmoy dirvmaxi vmintema vmaxtemi;drop d ;
run ;

/*lecture du fichier du type de tempête (1994-1996) */
FILENAME fic 'e:\tempete3\type.csv';
data type1;
infile fic delimiter=';' missover;
input d $ 1. datsi tytemp $ e $ ; drop d e ;
run ;

/*decalage de date entre le fichier de METEO-France et le fichier des sinistres*/
data type ;set varexp type1 ;
if datsi in (871008, 951213) then datsi=datsi+2 ;
if datsi in (810730, 811127, 830725, 841122, 850813, 860118, 880208, 880211, 900210, 900213, 900414,
950117, 950302) then datsi=datsi+1 ;

```

```

if datsi in (830902, 840402, 920314, 950308, 951104) then datsi=datsi-1 ;
if datsi =811003 then datsi=datsi-2 ;
if datsi =880229 then datsi=880301 ;

run ;

/*rapprochement des fichiers type de tempête et sinistres par la date de sinistres : on génère
le n° de tempête*/
data dattemp;merge type(in=in1) c(in=in2) /*sin1(keep=datsi nbsin1 mtsin1 nbzpsi)*/; by datsi;
/*if in1 and in2 ;*/
drop testdt cumul cumul1 nbjour ;run;

proc sort data=dattemp ;by numtemp ;run ;
data dattemp ;merge dattemp(in=in1) f(in=in2) ;by numtemp ;
run ;

/* rapprochement avec le fichier base avec la datsi et le risque : génération du n° tempête*/

/*CREATION DE LA BASE POUR LA MODELISATION DU RISQUE TEMPETE*/

/*lecture des fichiers des contrats habitation*/
FILENAME fic 'e:\tempet15\porthab.csv';
data cont;
infile fic delimiter=';' missover;
input copos nbentite ca93-ca96 nb92 -nb96;
if ca93 ne . ;
run ;

FILENAME fic 'e:\tempet15\zphab.csv';
data cont1;
infile fic delimiter=';' missover;
input copos nb79 - nb91;
if nb79 ne . ;
run ;

FILENAME fic 'e:\tempet15\cahab.csv';
data cont2;
infile fic delimiter=';' missover;
input copos ca79 - ca91;
if ca79 ne . ;
run ;

FILENAME fic 'e:\tempet15\ca92hab.csv';
data cont3;
infile fic delimiter=';' missover;
input copos ca92;
if ca92 ne . ;
run ;

proc sort data=cont ;by copos ;run ;
proc sort data=cont1 ;by copos ;run ;
proc sort data=cont2 ;by copos ;run ;
proc sort data=cont3 ;by copos ;run ;

data conthab ;merge cont cont1 cont2 cont3; by copos ;
risque=2 ;
run ;

/*lecture des fichiers des contrats agricoles*/
FILENAME fic 'e:\tempet15\portagr.csv';
data cont;
infile fic delimiter=';' missover;
input copos nbentite ca93-ca96 nb92 -nb96;
if ca93 ne . ;
run ;

FILENAME fic 'e:\tempet15\zpagr.csv';
data cont1;
infile fic delimiter=';' missover;
input copos nb79 - nb91;
if nb79 ne . ;
run ;

FILENAME fic 'e:\tempet15\caagr.csv';
data cont2;
infile fic delimiter=';' missover;
input copos ca79 - ca91;
if ca79 ne . ;
run ;

FILENAME fic 'e:\tempet15\ca92agr.csv';
data cont3;
infile fic delimiter=';' missover;
input copos ca92;
if ca92 ne . ;
run ;

proc sort data=cont ;by copos ;run ;

```

```

proc sort data=cont1 ;by copos ;run ;
proc sort data=cont2 ;by copos ;run ;
proc sort data=cont3 ;by copos ;run ;

data contagr ;merge cont cont1 cont2 cont3; by copos ;
risque=1 ;
run ;

/*lecture des fichiers des contrats professionnels*/
FILENAME fic 'e:\tempet15\portent.csv';
data cont;
infile fic delimiter=';' missover;
input copos nbenite ca93-ca96 nb92 -nb96;
if ca93 ne . ;
run ;

FILENAME fic 'e:\tempet15\zpent.csv';
data cont1;
infile fic delimiter=';' missover;
input copos nb79 - nb91;
if nb79 ne . ;
run ;

FILENAME fic 'e:\tempet15\caent.csv';
data cont2;
infile fic delimiter=';' missover;
input copos ca79-ca91;
if ca79 ne . ;
run ;

FILENAME fic 'e:\tempet15\ca92ent.csv';
data cont3;
infile fic delimiter=';' missover;
input copos ca92;
if ca92 ne . ;
run ;

proc sort data=cont ;by copos ;run ;
proc sort data=cont1 ;by copos ;run ;
proc sort data=cont2 ;by copos ;run ;
proc sort data=cont3 ;by copos ;run ;

data content ;merge cont cont1 cont2 cont3; by copos ;
risque=3 ;
run ;

data cont ;set conthab contagr content ;
run ;

proc datasets ;delete conthab contagr content cont1 cont2 cont3;
run ;

FILENAME fic 'e:\tempet17\sinzpr.csv';
data sin;
infile fic delimiter=';' missover;
input risque datsi copos nbsin mtsin ;
exsi=int(datsi/10000) ;
run ;

/*regroupement des bases contrats et sinistres selon le code postal pour déterminer le bon
nombre de contrats */
proc sort data=sin ;by copos ;run ;
proc sort data=cont ;by copos ;run ;

data base ;merge sin cont ;by copos ;
if nbsin ne . ;
nbcont=0 ;
ca=0 ;
if exsi=80 then do ;nbcont=nb79 ;ca=ca79 ;end ;
if exsi=81 then do ;nbcont=nb80 ;ca=ca80 ;end ;
if exsi=82 then do ;nbcont=nb81 ;ca=ca81 ;end ;
if exsi=83 then do ;nbcont=nb82 ;ca=ca82 ;end ;
if exsi=84 then do ;nbcont=nb83 ;ca=ca83 ;end ;
if exsi=85 then do ;nbcont=nb84 ;ca=ca84 ;end ;
if exsi=86 then do ;nbcont=nb85 ;ca=ca85 ;end ;
if exsi=87 then do ;nbcont=nb86 ;ca=ca86 ;end ;
if exsi=88 then do ;nbcont=nb87 ;ca=ca87 ;end ;
if exsi=89 then do ;nbcont=nb88 ;ca=ca88 ;end ;
if exsi=90 then do ;nbcont=nb89 ;ca=ca89 ;end ;
if exsi=91 then do ;nbcont=nb90 ;ca=ca90 ;end ;
if exsi=92 then do ;nbcont=nb91 ;ca=ca91 ;end ;
if exsi=93 then do ;nbcont=nb92 ;ca=ca92 ;end ;
if exsi=94 then do ;nbcont=nb93 ;ca=ca93 ;end ;
if exsi=95 then do ;nbcont=nb94 ;ca=ca94 ;end ;
if exsi=96 then do ;nbcont=nb95 ;ca=ca95 ;end ;
if exsi=97 then do ;nbcont=nb96 ;ca=ca96 ;end ;

```

```

drop nb79-nb96 ca79-ca96 ;
run ;

/*lecture du fichier des indices FNB et RI */
FILENAME fic 'e:\tempete7\indval.csv';
data indval;
infile fic delimiter=';' missover;
input exsi mtindfnb mtindri;
exsi=exsi-1900 ;
run;

proc sort data=base ;by exsi ;run ;

data base ;merge base (in=in1) indval ;by exsi ;if in1 ;
ca=int(.5+ca* 543/mtindfnb) ;
drop mtindri mtindfnb;
run ;

proc datasets ;delete sin cont ;run ;

/*lecture de la base type de local */
data typloc ;set tab.typloc ;
poitloc=(typloc=1)+2*(typloc=2)+1.5*(typloc=0) ;
run ;

proc means data=typloc noprint;
class copos ;
var poitloc ;
freq nb ;
output out=typloc mean= ;
run ;
data typloc ;set typloc ;if _type_ =1 ;
run ;

/*lecture de la base type d'assuré */
data typass ;set tab.typass ;
poitass= (typass=1)+3*(typass=2)+2*(typass=3)+2.5*(typass=4) ;
run ;

proc means data=typass noprint;
class copos ;
var poitass ;
freq nb ;
output out=typass mean= ;
run ;

data typass ;set typass ;if _type_ =1 ;
run ;

/*lecture du fichier des normes de construction */
FILENAME fic 'e:\ign\normcons.csv';
data normcons;
infile fic delimiter=';' missover;
input d $1. copos normcons e $;
drop d e ;
run ;

/*lecture de la base démographique */
data hab82 ;set tab.hab82 ;run ;
proc sort data=hab82 ;by copos ;run ;

data hab90 ;set tab.hab90 ;run ;
proc sort data=hab90 ;by copos ;run ;

data demo ;merge hab82 hab90 ;by copos ;
run ;

/*regroupement des bases selon le code postal */
data base1 ;merge normcons demo typloc typass;by copos ;
if copos >0 ;
run ;

proc datasets ;delete normcons demo typloc typass hab82 hab90 ;
run ;

/*somme des données contrats + base en laissant le code risque*/
proc means data=base noprint;
class datsi copos ;
var nbcont ca nbsin;
output out=base sum= ;
run ;
data base ;set base ;if _type_=3 ;run ;

/*rapprochement de la base tempête avec le fichier sinistres + contrats triés par datsi */
proc sort data=base ;by datsi ;run ;
proc sort data=dattemp ;by datsi ;run ;

```

```

data tot ;merge dattemp(in=in1) base(keep=datsi copos nbcont ca nbsin) ;by datsi ;if
in1 ;run ;

/*rapprochement de la base tempête avec les autres bases triées par copos */
proc sort data=tot ;by copos ;run ;
proc sort data=base1 ;by copos ;run ;

data tot ;merge tot (in=in1) base1 ;by copos ;if in1 and copos<96000 and copos ne . ;
exsi=int(datsi/10000) ;
if exsi<87 then do ;
    popsdc=popsdc82 ;
    popresp=poprep82 ;
    nbresp=nbresp82 ;
    nbress=nbress82 ;
end ;
drop popsdc82 poprep82 nbresp82 nbress82 ;
freqe=(nbcont>0)*nbsin/max(1,nbcont) ;
/*contrôle des données aberrantes*/
if freqe>1 then freqe=0 ;
densite=popsdc/superf ;
denfreq=mi n(1,densite*freqe) ;
assfreq= mi n(1,freqe*poitass) ;
locfreq= mi n(1,freqe*poitloc) ;
normfreq= mi n(1,freqe*normcons) ;
run ;

proc means data=tot noprint ;
class numtemp datsi ;
var nbcont ca popsdc popresp nbresp nbress superf ;
output out=tot1 sum= ;
run ;

data tot1 ;set tot1 ; if _type_=3 ; rename _freq_ = nbzpsi ;run ;

proc means data=tot1 noprint ;
class numtemp ;
var nbcont ca popresp popsdc nbresp nbress superf nbzpsi ;
output out=tot2 mean= ;
run ;

data tot2 ;set tot2 ; if _type_=1 ; run ;

proc means data=tot noprint ;
class numtemp ;
var freqe denfreq assfreq locfreq normfreq ;
output out=tot3 sum= ;
run ;

data tot3 ;set tot3 ; if _type_=1 ;
densite=denfreq/freqe ;
normcons= normfreq/freqe ;
poitass= assfreq/freqe ;
poitloc= locfreq/freqe ;
drop denfreq normfreq assfreq locfreq freqe ;run ;

proc sort data=dattemp ;by numtemp ;run ;

data typtemp ;merge dattemp (in=in1) tot2 tot3 ;by numtemp ;
if in1 and vmoyne>0 and cumuln>0 ;run ;

%LET fac=typtemp ;
%LET varcon= ora nei gre presmoy difpres vmoyno vmoyne vmoyso vmoyse nbcont ca densite
superf poitloc poitass normcons ;
%LET var=&fac &varcon ;
proc glm data=typtemp ;
class &fac ;
model cumuln = &fac &varcon/solution ;

TITLE ANALYSE DE LA VARIANCE : TYPE DE TEMPÊTE ;
run ;

```

Résultats sous SAS de l'impact de la variable « type de tempête » sur le coût actualisé par tempête

ANALYSE DE LA VARIANCE : TYPE DE TEMPÊTE		12
		09:02 Monday, November 15, 1999
General Linear Models Procedure Class Level Information		
Class	Levels	Values
TYPTEMP	7	E N NE OR S SW W

Number of observations in data set = 193

ANALYSE DE LA VARIANCE : TYPE DE TEMPÊTE 13
09:02 Monday, November 15, 1999

General Linear Models Procedure

Dependent Variable: CUMULM

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	7.54941E+17	6.29117E+16	43.62	0.0001
Error	180	2.59605E+17	1.44225E+15		
Corrected Total	192	1.01455E+18			
	R-Square	C. V.	Root MSE	CUMULM Mean	
	0.744117	211.5354	37976954	17953002	

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TYPTEMP	6	3.12912E+16	5.21520E+15	3.62	0.0021
ORA	1	4.07945E+15	4.07945E+15	2.83	0.0943
PRESMOY	1	5.39083E+15	5.39083E+15	3.74	0.0548
DI FPRES	1	9.94201E+15	9.94201E+15	6.89	0.0094
VVMOYSE	1	1.89923E+16	1.89923E+16	13.17	0.0004
NBCONT	1	6.80267E+17	6.80267E+17	471.67	0.0001
SUPERF	1	4.97837E+15	4.97837E+15	3.45	0.0648

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TYPTEMP	6	1.72568E+16	2.87613E+15	1.99	0.0687
ORA	1	2.77016E+15	2.77016E+15	1.92	0.1675
PRESMOY	1	5.84027E+14	5.84027E+14	0.40	0.5254
DI FPRES	1	1.04188E+16	1.04188E+16	7.22	0.0079
VVMOYSE	1	5.85710E+14	5.85710E+14	0.41	0.5248
NBCONT	1	2.00474E+16	2.00474E+16	13.90	0.0003
SUPERF	1	4.97837E+15	4.97837E+15	3.45	0.0648

ANNEXE N : Modélisation de la charge actualisée par contrat d'une tempête

Programme MODTATEMP assurant le rapprochement des événements du fichier des assureurs (1980-1993) et définis par Météo-France (72 heures de 1994 à 1996) et des tempêtes définies par nos propres soins.

```
/*CREATION DE LA BASE POUR LA MODELISATION DU RISQUE TEMPETE*/
```

```
/*lecture des fichiers des contrats habitation*/
```

```
FILENAME fic 'e:\tempet15\porthab.csv';  
data cont;  
infile fic delimiter=';' missover;  
input copos nbentite ca93-ca96 nb92 -nb96;  
if ca93 ne . ;  
run ;
```

```
FILENAME fic 'e:\tempet15\zphab.csv';  
data cont1;  
infile fic delimiter=';' missover;  
input copos nb79 - nb91;  
if nb79 ne . ;  
run ;
```

```
FILENAME fic 'e:\tempet15\cahab.csv';  
data cont2;  
infile fic delimiter=';' missover;  
input copos ca79 - ca91;  
if ca79 ne . ;  
run ;
```

```
FILENAME fic 'e:\tempet15\ca92hab.csv';  
data cont3;  
infile fic delimiter=';' missover;  
input copos ca92;  
if ca92 ne . ;  
run ;
```

```
proc sort data=cont ;by copos ;run ;  
proc sort data=cont1 ;by copos ;run ;  
proc sort data=cont2 ;by copos ;run ;  
proc sort data=cont3 ;by copos ;run ;  
  
data conthab ;merge cont cont1 cont2 cont3; by copos ;  
risque=2 ;  
run ;
```

```
/*lecture des fichiers des contrats agricoles*/
```

```
FILENAME fic 'e:\tempet15\portagr.csv';  
data cont;  
infile fic delimiter=';' missover;  
input copos nbentite ca93-ca96 nb92 -nb96;  
if ca93 ne . ;  
run ;
```

```
FILENAME fic 'e:\tempet15\zpagr.csv';  
data cont1;  
infile fic delimiter=';' missover;  
input copos nb79 - nb91;  
if nb79 ne . ;  
run ;
```

```
FILENAME fic 'e:\tempet15\caagr.csv';  
data cont2;  
infile fic delimiter=';' missover;  
input copos ca79 - ca91;  
if ca79 ne . ;  
run ;
```

```
FILENAME fic 'e:\tempet15\ca92agr.csv';  
data cont3;  
infile fic delimiter=';' missover;  
input copos ca92;  
if ca92 ne . ;  
run ;
```

```
proc sort data=cont ;by copos ;run ;  
proc sort data=cont1 ;by copos ;run ;  
proc sort data=cont2 ;by copos ;run ;  
proc sort data=cont3 ;by copos ;run ;  
  
data contagr ;merge cont cont1 cont2 cont3; by copos ;  
risque=1 ;
```

```

run ;
/*lecture des fichiers des contrats professionnels*/
FILENAME fic 'e:\tempet15\portent.csv';
data cont;
infile fic delimiter=';' missover;
input copos nbentite ca93-ca96 nb92 -nb96;
if ca93 ne . ;
run ;

FILENAME fic 'e:\tempet15\zpent.csv';
data cont1;
infile fic delimiter=';' missover;
input copos nb79 - nb91;
if nb79 ne . ;
run ;

FILENAME fic 'e:\tempet15\caent.csv';
data cont2;
infile fic delimiter=';' missover;
input copos ca79-ca91;
if ca79 ne . ;
run ;

FILENAME fic 'e:\tempet15\ca92ent.csv';
data cont3;
infile fic delimiter=';' missover;
input copos ca92;
if ca92 ne . ;
run ;

proc sort data=cont ;by copos ;run ;
proc sort data=cont1 ;by copos ;run ;
proc sort data=cont2 ;by copos ;run ;
proc sort data=cont3 ;by copos ;run ;

data content ;merge cont cont1 cont2 cont3; by copos ;
risque=3 ;
run ;

data cont ;set conthab contagr content ;
run ;

proc datasets ;delete conthab contagr content cont1 cont2 cont3;
run ;

/*****/
/*lecture du fichier des zones postales sinistrées */
FILENAME fic 'e:\tempet17\sinzpr.csv';
data sin;
infile fic delimiter=';' missover;
input risque datsi copos nbsin mtsin ;
ex=int(datsi/10000) ;if ex<97 ;
run ;

/*regroupement des bases contrats et sinistres selon le code postal pour déterminer la part de
contrats sinistrés de la zone météo */
proc sort data=sin ;by risque copos ;run ;
proc sort data=cont ;by risque copos ;run ;

data sin ;merge sin cont ;by risque copos ;
if nbsin ne . ;
nbcontsi=0 ;
casi=0 ;
if ex=80 then do ;nbcontsi=nb79 ;casi=ca79 ;end ;
if ex=81 then do ;nbcontsi=nb80 ;casi=ca80 ;end ;
if ex=82 then do ;nbcontsi=nb81 ;casi=ca81 ;end ;
if ex=83 then do ;nbcontsi=nb82 ;casi=ca82 ;end ;
if ex=84 then do ;nbcontsi=nb83 ;casi=ca83 ;end ;
if ex=85 then do ;nbcontsi=nb84 ;casi=ca84 ;end ;
if ex=86 then do ;nbcontsi=nb85 ;casi=ca85 ;end ;
if ex=87 then do ;nbcontsi=nb86 ;casi=ca86 ;end ;
if ex=88 then do ;nbcontsi=nb87 ;casi=ca87 ;end ;
if ex=89 then do ;nbcontsi=nb88 ;casi=ca88 ;end ;
if ex=90 then do ;nbcontsi=nb89 ;casi=ca89 ;end ;
if ex=91 then do ;nbcontsi=nb90 ;casi=ca90 ;end ;
if ex=92 then do ;nbcontsi=nb91 ;casi=ca91 ;end ;
if ex=93 then do ;nbcontsi=nb92 ;casi=ca92 ;end ;
if ex=94 then do ;nbcontsi=nb93 ;casi=ca93 ;end ;
if ex=95 then do ;nbcontsi=nb94 ;casi=ca94 ;end ;
if ex=96 then do ;nbcontsi=nb95 ;casi=ca95 ;end ;
nbzpsi=(nbsin>0) ;
drop nb79-nb96 ca79-ca96 ;
run ;

```

```

/*proc datasets ;delete sin cont ;
run ;*/

/*regroupement au niveau de la zone météo */

data dismi nzp ;set tab. dismi nzp ;run ;
proc sort data=sin ;by copos ;run ;
data sin ;merge sin (in=in1) dismi nzp ;by copos ;if in1 ;run ;

proc means data=sin noprint;
class risque datsi sm ;
var nbzpsi nbsin mtsin nbcontsi casi;
output out=sin sum= ;
run ;

data sin ;set sin ;if _type_ =7 ; ex=int(datsi/10000) ;drop _freq_ _type_ ;
run ;

proc sort data=cont ;by copos ;run ;
data cont ;merge cont (in=in1) dismi nzp ;by copos ;if in1 ;run ;

proc means data=cont noprint;
class risque sm ;
var nb79-nb96 ca79-ca96;
output out=cont sum= ;
run ;

data cont ;set cont ;if _type_ =3 ; rename _freq_ = nbzp ;drop _type_ ;run ;

/*CONSTITUTION DE LA MATRICE PAR RISQUE SM ET EXERCICE */
data sme80 ;set tab. smf(keep=sm) ;ex=80 ;run ;
data sme81 ;set tab. smf(keep=sm) ;ex=81 ;run ;
data sme82 ;set tab. smf(keep=sm) ;ex=82 ;run ;
data sme83 ;set tab. smf(keep=sm) ;ex=83 ;run ;
data sme84 ;set tab. smf(keep=sm) ;ex=84 ;run ;
data sme85 ;set tab. smf(keep=sm) ;ex=85 ;run ;
data sme86 ;set tab. smf(keep=sm) ;ex=86 ;run ;
data sme87 ;set tab. smf(keep=sm) ;ex=87 ;run ;
data sme88 ;set tab. smf(keep=sm) ;ex=88 ;run ;
data sme89 ;set tab. smf(keep=sm) ;ex=89 ;run ;
data sme90 ;set tab. smf(keep=sm) ;ex=90 ;run ;
data sme91 ;set tab. smf(keep=sm) ;ex=91 ;run ;
data sme92 ;set tab. smf(keep=sm) ;ex=92 ;run ;
data sme93 ;set tab. smf(keep=sm) ;ex=93 ;run ;
data sme94 ;set tab. smf(keep=sm) ;ex=94 ;run ;
data sme95 ;set tab. smf(keep=sm) ;ex=95 ;run ;
data sme96 ;set tab. smf(keep=sm) ;ex=96 ;run ;

data sme ;set sme80 sme81 sme82 sme83 sme84 sme85 sme86 sme87 sme88 sme89 sme90 sme91 sme92
sme93 sme94 sme95 sme96 ;run ;

data smer1 ;set sme ;risque=1 ;run ;
data smer2 ;set sme ;risque=2 ;run ;
data smer3 ;set sme ;risque=3 ;run ;

data smer ;set smer1 smer2 smer3 ;run ;

proc datasets ; delete sme80 sme81 sme82 sme83 sme84 sme85 sme86 sme87 sme88 sme89 sme90 sme91
sme92 sme93 sme94 sme95 sme96 smer1 smer2 smer3 sme ;run ;

/*regroupement des bases portefeuille et table des modalités*/
proc sort data=smer ;by risque sm ;run ;
data cont ;merge cont smer ;by risque sm ;
nbcont=0 ;
ca=0 ;
if ex=80 then do ;nbcont=nb79 ;ca=ca79 ;end ;
if ex=81 then do ;nbcont=nb80 ;ca=ca80 ;end ;
if ex=82 then do ;nbcont=nb81 ;ca=ca81 ;end ;
if ex=83 then do ;nbcont=nb82 ;ca=ca82 ;end ;
if ex=84 then do ;nbcont=nb83 ;ca=ca83 ;end ;
if ex=85 then do ;nbcont=nb84 ;ca=ca84 ;end ;
if ex=86 then do ;nbcont=nb85 ;ca=ca85 ;end ;
if ex=87 then do ;nbcont=nb86 ;ca=ca86 ;end ;
if ex=88 then do ;nbcont=nb87 ;ca=ca87 ;end ;
if ex=89 then do ;nbcont=nb88 ;ca=ca88 ;end ;
if ex=90 then do ;nbcont=nb89 ;ca=ca89 ;end ;
if ex=91 then do ;nbcont=nb90 ;ca=ca90 ;end ;
if ex=92 then do ;nbcont=nb91 ;ca=ca91 ;end ;
if ex=93 then do ;nbcont=nb92 ;ca=ca92 ;end ;
if ex=94 then do ;nbcont=nb93 ;ca=ca93 ;end ;
if ex=95 then do ;nbcont=nb94 ;ca=ca94 ;end ;
if ex=96 then do ;nbcont=nb95 ;ca=ca95 ;end ;
drop nb79-nb96 ca79-ca96 ;
run ;

```

```

/*regroupement des bases contrats et sinistres selon le code postal pour évaluer la part de
contrats sinistrés de la zone météo */
proc sort data=sin ;by risque sm ex;run ;
proc sort data=cont ;by risque sm ex;run ;

data base ;merge sin cont ;by risque sm ex;run ;

/*lecture du fichier des indices FNB et RI */
FILENAME fic 'e:\tempete7\indval.csv';
data indval;
infile fic delimiter=';' missover;
input ex mtindfnb mtindri;
ex=ex-1900 ;
run;

proc sort data=base ;by ex ;run ;

data base ;merge base (in=in1) indval ;by ex ;if in1 ;
ca=int(.5+ca* 543/mtindfnb) ;
casi=int(.5+casi*543/ mtindfnb) ;
drop mtindri mtindfnb;
run ;

/*lecture de la base type de local */
data typloc ;set tab.typloc ;
poitloc=(typloc=1)+2*(typloc=2)+1.5*(typloc=0) ;
run ;

proc means data=typloc noprint;
class copos ;
var poitloc ;
freq nb ;
output out=typloc mean= ;
run ;
data typloc ;set typloc ;if _type_ =1 ;
run ;

/*lecture de la base type d'assuré */
data typass ;set tab.typass ;
poitass= (typass=1)+3*(typass=2)+2*(typass=3)+2.5*(typass=4) ;
run ;

proc means data=typass noprint;
class copos ;
var poitass ;
freq nb ;
output out=typass mean= ;
run ;

data typass ;set typass ;if _type_ =1 ;
run ;

/*lecture du fichier des normes de construction */
FILENAME fic 'e:\ign\normcons.csv';
data normcons;
infile fic delimiter=';' missover;
input d $1. copos normcons e $;
drop d e ;
run ;

/*lecture de la base démographique */
data hab82 ;set tab.hab82 ;run ;
proc sort data=hab82 ;by copos ;run ;

data hab90 ;set tab.hab90 ;run ;
proc sort data=hab90 ;by copos ;run ;

data demo ;merge hab82 hab90 ;by copos ;
run ;

/*regroupement des bases selon le code postal */
data base1 ;merge normcons demo typloc typass;by copos ;
run ;

proc datasets ;delete normcons demo typloc typass hab82 hab90 ;
run ;

/*regroupement au niveau de la zone météo */
data base1 ;merge base1 (in=in1) dism1nzp ;by copos ;if in1 ;run ;

proc means data=base1 noprint;
class sm ;
var normcons poitloc poitass;
freq popsdc82 ;
output out=base182 mean= normco82 poitlo82 poitas82;
run ;

```

```

data base182 ;set base182 ;if _type_ =1 ;
proc means data=base1 noprint;
class sm ;
var normcons poitloc poitass;
freq popsc ;
output out=base190 mean= normcons poitloc poitass;
run ;

data base190 ;set base190 ;if _type_ =1 ;
run ;

proc means data=base1 noprint;
class sm ;
var popsc82 poprep82 nresp82 nbress82 nblvac82 superf popsc popresp nresp nbress;
output out=base1 sum= ;
run ;

data base1 ;set base1 ;if _type_ =1 ;
run ;

/*rapprochement de la base contrats + sinistres plus des autres bases selon le code station*/
proc sort data=base ;by sm ;run ;

data base2 ;merge base base1 base182 base190 ;by sm ;
nblgvac=nblvac82 ;
if ex<87 then do ;
    popsc=popsc82 ;
    popresp=poprep82 ;
    nresp=nresp82 ;
    nbress=nbress82 ;
    normcons=normco82 ;
    poitloc=poitlo82 ;
    poitass=poitas82 ;
    nblgvac=nblvac82 ;
end ;
densite=popsc/superf ;
drop popsc82 poprep82 nresp82 nbress82 normco82 poitlo82 poitas82 nblvac82;
run ;

proc datasets ; delete base base1 base182 base190 ;
run ;
/*****
/*DETERMINATION DES JOURS DE TEMPÊTE DANS LE FICHIER SINISTRES*/
proc means data=base2 noprint;
class risque datsi ;
var nbsin mtsin;
output out=sin sum= ;
run ;

data sin;set sin;
if risque =. And datsi ne . ;
run;
options center;
/*libname don 'd:\acn\exemple' ;*/ /* Allocation de la librairie */

data a(drop= jour mois an date); /* modification des dates */
set sin;
%let val=15; /*SEUIL DU NB DE SINISTRES A PARTIR DUQUEL ON CONSIDERE QU' UNE JOURNEE EST
VENTEUSE */
date=put (datsi, $6.);
jour=substr(date, 5, 2);
mois=substr(date, 3, 2);
an=substr(date, 1, 2);
date2=mdy(mois, jour, an);
format date2 date8.;
/*regroupement de 2 tempêtes, Météo-France ne donnant qu'un relevé*/
run;

data b;
set a;
by risque ; /* METTRE LES MEMES VARIABLES QUE POUR LE TRI LA DATE EN MOINS */

retain testdt ; /* variable permettant de tester la date*/
/* sa valeur sera gardee d'un enregistrement à l'autre */

retain cumul; /* variable cumul du nombre de sinistres */
/* sa valeur sera gardee d'un enregistrement à l'autre */

retain cumul1; /* variable cumul du montant des sinistres */
/* sa valeur sera gardee d'un enregistrement à l'autre */

retain numtemp; /*comptage nombre de tempetes */

retain nbjour ; /*comptage nombre de jours de la tempete */

/* 1er enregistrement pour un risque donné */

```

```

/* on initialise cumul et testdt */
  if first.risque=1 then do;
    cumul=(nbsin)*(nbsin>&val);
    cumul1=(mtsin)*(nbsin>&val);
    testdt=date2;
    numtemp=0;
    nbjour=0 ;
  end;

if nbjour=3 then do ;
  cumul=0 ;
  cumul1=0 ;
  nbjour=0 ;
end ;

/* date = date précédente +1 */
/* on fait le cumul */
  if (date2 = testdt+1) then do; /* AJOUTER LA CONDITION NBSIN > VALEUR */
    numtemp=numtemp+(cumul=0)*(nbsin>&val);
    nbjour=(nbjour+1)*(nbsin>&val);
    cumul=(cumul+nbsin)*(nbsin>&val);
    cumul1=(cumul1+mtsin)*(nbsin>&val);
    testdt=date2;

  end;

/* date ne date précédente +1 */
/* on reinitialise cumul et testdt */
  else do;
    nbjour=(nbsin>&val);
    numtemp=numtemp+(nbsin>&val);
    cumul=nbsin*(nbsin>&val);
    cumul1=mtsin*(nbsin>&val);
    testdt=date2;
end;
format testdt date8.;
run;

data c;set b;if cumul>0;
if numtemp in (140, 175, 204, 298, 368, 457, 466, 526, 573, 615) then numtemp=numtemp- 1 ;
drop _type_ _freq_ nbsin mtsin date2 ;
run;

/*****
/*rapprochement du fichier des dates de sinistres sélectionnées et du fichier base2 */

proc sort data=c ;by datsi ;run ;
proc sort data=base2 ;by datsi ;run ;

data g ;merge base2(keep=datsi sm risque nbsin mtsin) c(drop=risque) ;by datsi ;if numtemp ne
. ;
run ;

/*****
/*lecture du fichier des paramètres descriptifs météorologiques */
FILENAME fic 'e:\tempete3\param.csv' ;
data meteo;
infile fic delimiter=';' missover;
input d $ 1. datsi typtemp $ sm vit dir dur prec e $; drop d e;
run ;

/*decalage de date entre le fichier de METEO-France et le fichier des sinistres*/
data meteo ;set meteo ;if datsi>800000 ;
if datsi in (830817,871008) then datsi=datsi+2 ;
if datsi in (810730,821209,841122,850813,860118,931207,940126) then datsi=datsi+1 ;

run ;

/*rapprochement avec la base météo selon le code station et la date du sinistre*/
proc sort data=meteo ;by datsi sm;run ;

/*rapprochement des fichiers type de tempête et sinistres par la date de sinistres : on génère
le n° de tempête*/
data z;merge meteo(in=in1) g(in=in2); by datsi sm;
if in1 and in2 ;run;
/*drop testdt cumul cumul1 nbjour ;run;*/

/*détermination des tempêtes communes aux deux fichiers MF et sinistres*/
proc means data =z noprint ;
class numtemp ;
var nbsin ;
output out=y sum=nbsin2 ;
run ;

data y ;set y ;if _type_=1 ;run ;

```

```

proc sort data=g ; by numtemp ;run ;
data g1 ;merge g y(in=in2) ;by numtemp ;if in2 ;run ;

/*tables de correspondances datsi-numtemp*/
data z1;set z(keep=datsi numtemp);by datsi;if first.datsi;run;
data z2;set g1(keep=datsi numtemp);by datsi;if first.datsi;run;
data meteo ;merge meteo(in=in1) z1(in=in2) ;by datsi;rename datsi=datsi1 ;if in2 ;
run ;

/*puis il faut extraire de base2 les jours présents dans z2 et attacher à dattemp*/

data base3 ;merge base2 z2(in=in2);by datsi ;
nbsin=max(0, nbsin) ;
mtsin=max(0, mtsin) ;
casi=max(0, casi) ;
nbcontsi=max(0, nbcontsi) ;
nbzpsi= max(0, nbzpsi) ;
numtemp= max(0, numtemp) ;
run ;

/*rapprochement des bases MF et sinistres*/
proc means data =base3 noprint ;
class sm numtemp risque;
var nbsin mtsin ;
output out=base4 sum=;
run ;

data base4 ;set base4 ;if _type_=7 ;run ;

proc means data =base3 noprint ;
class sm numtemp risque;
var nbzpsi nbcontsi casi;
output out=base5 mean=;
run ;

data base5 ;set base5 ;if _type_=7 ;run ;

data meteo ;set meteo ;ex=int(datsi1/10000) ;run ;
proc sort data=meteo ;by sm ex ;run ;

proc means data=base3 noprint;
class sm ex risque;
var nbcont ca nbzp superf popsd pc popresp nbresp nbress normcons poitloc poitass densite ;
output out=base3b mean= ;
run ;

data base3b ;set base3b ; if _type_=7 ; smexri=risque+10*ex+1000*sm ;run ;

data meteo1;set meteo;risque=1; run;
data meteo2;set meteo;risque=2; run;
data meteo3;set meteo;risque=3; run;
data meteo ;set meteo1 meteo2 meteo3 ;smexri=risque+10*ex+1000*sm ;run ;
proc sort data=base3b ;by smexri ;run ;
proc sort data=meteo ;by smexri ;run ;

data base6 ;merge meteo base3b ;by smexri ;run ;

proc sort data=base6 ;by sm numtemp risque;run ;

data temploc ;merge base4 base5 base6;by sm numtemp risque;if numtemp>0 and nbcont ne . ;
nbsin=max(0, nbsin) ;
mtsin=max(0, mtsin) ;
casi=max(0, casi) ;
nbcontsi=max(0, nbcontsi) ;
nbzpsi= max(0, nbzpsi) ;
drop _type_ _freq_ smexri;
run ;

proc datasets ;delete base2 base3 base3b base4 base5 base6 meteo a b c disminzp g g1 indval
meteo1 meteo2 meteo3 smer y z z1 z2;
run ;

proc sort data=temploc ;by sm risque ;run ;

data temploc ;set temploc ;
if vit>0 ;
primpur=log(max(1, mtsin/nbcont)) ;
freq=log(max(1, nbsin/nbcont)) ;
lvit=log(max(1, vit)) ;
ldur=log(max(1, 100*dur)) ;
richesse=log(max(1, ca/nbcont)) ;
efflocm=casi/ca ;
efflocn=nbcontsi/nbcont ;
run ;

```

```

%LET fac=typtemp prec dir ;
%LET varcon= lvit ldur richesse efflocm efflocn densite superf poitloc poitass normcons;
%LET var=&fac &varcon;
proc glm data=temploc ;
class &fac;
model primpur = &fac &varcon/solution;
by risque ;
/*output out=cmr p=coutmr; */
run ;
TITLE ANALYSE DE LA VARIANCE : PRIME PURE PAR RISQUE ET TEMPÊTE;
run;

```

Résultats sous SAS de la modélisation du coût actualisé de la tempête par contrat (station de Dinard)

ANALYSE DE LA VARIANCE : PRIME PURE PAR RISQUE ET TEMPÊTE						52
						09:02 Monday, November 15, 1999
----- SM=3501 RISQUE=1 -----						
General Linear Models Procedure						
Class Level Information						
Class	Levels	Values				
TYPTEMP	7	E N NE OR S SW W				
Number of observations in data set = 119						
ANALYSE DE LA VARIANCE : PRIME PURE PAR RISQUE ET TEMPÊTE						53
						09:02 Monday, November 15, 1999
General Linear Models Procedure						
Dependent Variable: PRIMPUR						
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	
Model	10	213.9328245	21.3932825	28.78	0.0001	
Error	108	80.2773108	0.7433084			
Corrected Total	118	294.2101353				
	R-Square	C. V.	Root MSE	PRIMPUR Mean		
	0.727143	127.8887	0.862153	0.674143		
Source	DF	Type I SS	Mean Square	F Value	Pr > F	
TYPTEMP	6	46.9265775	7.8210962	10.52	0.0001	
LVIT	1	27.7659236	27.7659236	37.35	0.0001	
LDUR	1	3.6333833	3.6333833	4.89	0.0292	
RICHESS	1	5.8306200	5.8306200	7.84	0.0060	
EFFLOCM	1	129.7763202	129.7763202	174.59	0.0001	
Source	DF	Type III SS	Mean Square	F Value	Pr > F	
TYPTEMP	6	12.5277752	2.0879625	2.81	0.0140	
LVIT	1	4.1042967	4.1042967	5.52	0.0206	
LDUR	1	0.7424651	0.7424651	1.00	0.3198	
RICHESS	1	0.4255889	0.4255889	0.57	0.4509	
EFFLOCM	1	129.7763202	129.7763202	174.59	0.0001	
Parameter	Estimate	T for H0: Parameter=0	Pr > T	Std Error of Estimate		

INTERCEPT		-1.09147100 B	-0.45	0.6512	2.40784761
TYPTEMP	E	0.31364600 B	0.46	0.6475	0.68413120
	N	0.30801253 B	1.32	0.1901	0.23358433
	NE	1.58842756 B	3.39	0.0010	0.46916317
	OR	0.51443198 B	1.45	0.1487	0.35371816
	S	0.43998771 B	1.70	0.0924	0.25909804
	SW	0.88137375 B	2.82	0.0058	0.31307636
	W	0.00000000 B	.	.	.
LVIT		1.00158165	2.35	0.0206	0.42623736
LDUR		-0.05754322	-1.00	0.3198	0.05757590
RICHESS		-0.22557211	-0.76	0.4509	0.29810880
EFFLOCM		13.31128095	13.21	0.0001	1.00741092
NOTE: The X'X matrix has been found to be singular and a generalized inverse was used to solve the normal equations. Estimates followed by the letter 'B' are biased, and are not unique estimators of the parameters.					
ANALYSE DE LA VARIANCE : PRIME PURE PAR RISQUE ET TEMPÊTE 54					
09:02 Monday, November 15, 1999					
----- SM=3501 RISQUE=2 -----					
General Linear Models Procedure					
Class Level Information					
Class	Levels	Values			
TYPTEMP	7	E N NE OR S SW W			
Number of observations in data set = 119					
ANALYSE DE LA VARIANCE : PRIME PURE PAR RISQUE ET TEMPÊTE 55					
09:02 Monday, November 15, 1999					
General Linear Models Procedure					
Dependent Variable: PRIMPUR					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	64.37628672	7.15292075	28.04	0.0001
Error	109	27.81006305	0.25513819		
Corrected Total	118	92.18634978			
	R-Square	C.V.	Root MSE	PRIMPUR Mean	
	0.698328	158.6406	0.505112	0.318400	
Source	DF	Type I SS	Mean Square	F Value	Pr > F
TYPTEMP	6	12.59718491	2.09953082	8.23	0.0001
LVIT	1	9.71792988	9.71792988	38.09	0.0001
LDUR	1	1.57683723	1.57683723	6.18	0.0144
EFFLOCM	1	40.48433470	40.48433470	158.68	0.0001
Source	DF	Type III SS	Mean Square	F Value	Pr > F
TYPTEMP	6	3.13840798	0.52306800	2.05	0.0650
LVIT	1	0.50863039	0.50863039	1.99	0.1608
LDUR	1	0.20435110	0.20435110	0.80	0.3728
EFFLOCM	1	40.48433470	40.48433470	158.68	0.0001
Parameter	Estimate	T for H0: Parameter=0	Pr > T	Std Error of Estimate	
INTERCEPT	-1.018261292 B	-1.37	0.1732	0.74265074	
TYPTEMP	E	0.170757031 B	0.43	0.40030244	
	N	0.144664853 B	1.07	0.13499096	
	NE	0.029008077 B	0.11	0.27404310	
	OR	0.256225359 B	1.24	0.20633436	
	S	-0.004433773 B	-0.03	0.15177854	
	SW	0.540419809 B	2.99	0.18061462	
	W	0.000000000 B	.	.	
LVIT	0.360859885	1.41	0.1608	0.25557902	
LDUR	-0.030230473	-0.89	0.3728	0.03377880	
EFFLOCM	7.688035555	12.60	0.0001	0.61032283	
NOTE: The X'X matrix has been found to be singular and a generalized inverse was used to solve the normal equations. Estimates followed by the					

letter 'B' are biased, and are not unique estimators of the parameters.

ANALYSE DE LA VARIANCE : PRIME PURE PAR RISQUE ET TEMPÊTE 56
09:02 Monday, November 15, 1999

----- SM=3501 RISQUE=3 -----

General Linear Models Procedure
Class Level Information

Class	Levels	Values
TYPTEMP	7	E N NE OR S SW W
PREC	2	0 1

Number of observations in data set = 119

ANALYSE DE LA VARIANCE : PRIME PURE PAR RISQUE ET TEMPÊTE 57
09:02 Monday, November 15, 1999

General Linear Models Procedure

Dependent Variable: PRIMPUR

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	166.9625104	15.1784100	28.74	0.0001
Error	107	56.5001721	0.5280390		
Corrected Total	118	223.4626825			
	R-Square	C. V.	Root MSE	PRIMPUR Mean	
	0.747161	143.3073	0.726663	0.507066	

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TYPTEMP	6	31.17144308	5.19524051	9.84	0.0001
PREC	1	2.61094507	2.61094507	4.94	0.0283
LVIT	1	23.55968041	23.55968041	44.62	0.0001
LDUR	1	4.00559709	4.00559709	7.59	0.0069
EFFLOCM	1	86.28306807	86.28306807	163.40	0.0001
EFFLOCN	1	19.33177664	19.33177664	36.61	0.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TYPTEMP	6	1.43345444	0.23890907	0.45	0.8418
PREC	1	2.76438354	2.76438354	5.24	0.0241
LVIT	1	3.15964731	3.15964731	5.98	0.0161
LDUR	1	0.35441327	0.35441327	0.67	0.4145
EFFLOCM	1	9.51627249	9.51627249	18.02	0.0001
EFFLOCN	1	19.33177664	19.33177664	36.61	0.0001

Parameter	Estimate	T for H0: Parameter=0	Pr > T	Std Error of Estimate
INTERCEPT	-1.77010278 B	-1.64	0.1037	1.07843426
TYPTEMP	0.23241747 B	0.40	0.6880	0.57723260
E	0.24478822 B	1.23	0.2209	0.19880215
N	0.15649778 B	0.39	0.6953	0.39846089
NE	0.39942476 B	1.32	0.1886	0.30183782
OR	0.26211734 B	1.19	0.2351	0.21952144
S	0.30122889 B	1.11	0.2715	0.27253485
SW	0.0000000 B	.	.	.
W	0.0000000 B	.	.	.
PREC	-0.67269872 B	-2.29	0.0241	0.29400482
0	0.0000000 B	.	.	.
1	0.87850465	2.45	0.0161	0.35913502
LVIT	-0.03992264	-0.82	0.4145	0.04873010
LDUR	-31.50220392	-4.25	0.0001	7.42062526
EFFLOCM	47.82956576	6.05	0.0001	7.90484849
EFFLOCN				

NOTE: The X'X matrix has been found to be singular and a generalized inverse was used to solve the normal equations. Estimates followed by the letter 'B' are biased, and are not unique estimators of the parameters.

ANNEXE O : Etude de l'impact de différentes variables sur le montant de la prime pure

----- RISQUE=1 -----					
ANALYSE DE LA VARIANCE : PRIME PURE PAR RISQUE ET TEMPÊTE 47					
18: 41 Monday, November 22, 1999					
General Linear Models Procedure					
Class Level Information					
Class	Levels	Values			
TYPTEMP	7	E N NE OR S SW W			
PREC	2	0 1			
DIR	2	0 1			
Number of observations in data set = 12623					
ANALYSE DE LA VARIANCE : PRIME PURE PAR RISQUE ET TEMPÊTE 48					
18: 41 Monday, November 22, 1999					
General Linear Models Procedure					
Dependent Variable: PRIMPUR					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	16	27699.15934	1731.19746	1260.09	0.0001
Error	12606	17318.96796	1.37387		
Corrected Total	12622	45018.12730			
	R-Square	C. V.	Root MSE	PRIMPUR Mean	
	0.615289	132.8618	1.172121	0.882211	
Source	DF	Type I SS	Mean Square	F Value	Pr > F
TYPTEMP	6	466.89486	77.81581	56.64	0.0001
PREC	1	477.15179	477.15179	347.31	0.0001
DIR	1	115.42935	115.42935	84.02	0.0001
LVIT	1	6429.82953	6429.82953	4680.10	0.0001
LDUR	1	52.37070	52.37070	38.12	0.0001
RICHESS	1	58.33726	58.33726	42.46	0.0001
EFFLOCN	1	19622.34181	19622.34181	14282.56	0.0001
DENSITE	1	2.70419	2.70419	1.97	0.1607
SUPERF	1	358.68948	358.68948	261.08	0.0001
POITASS	1	28.05741	28.05741	20.42	0.0001
NORMCONS	1	87.35296	87.35296	63.58	0.0001
Source	DF	Type III SS	Mean Square	F Value	Pr > F
TYPTEMP	6	144.02346	24.00391	17.47	0.0001
PREC	1	39.97351	39.97351	29.10	0.0001
DIR	1	2.11796	2.11796	1.54	0.2144
LVIT	1	992.28482	992.28482	722.26	0.0001
LDUR	1	8.43988	8.43988	6.14	0.0132
RICHESS	1	11.89449	11.89449	8.66	0.0033
EFFLOCN	1	18893.75033	18893.75033	13752.24	0.0001
DENSITE	1	6.37677	6.37677	4.64	0.0312
SUPERF	1	266.91312	266.91312	194.28	0.0001
POITASS	1	24.98392	24.98392	18.19	0.0001
NORMCONS	1	87.35296	87.35296	63.58	0.0001
Parameter	Estimate	T for H0: Parameter=0	Pr > T	Std Error of Estimate	
INTERCEPT	-1.82034941 B	-4.88	0.0001	0.37305802	
TYPTEMP E	0.11896330 B	1.42	0.1561	0.08386190	
N	-0.15689451 B	-5.16	0.0001	0.03039198	
NE	-0.23471635 B	-3.76	0.0002	0.06241573	
OR	0.13917312 B	3.74	0.0002	0.03725507	
S	-0.10761031 B	-3.24	0.0012	0.03316791	
SW	-0.15701228 B	-3.90	0.0001	0.04024510	

PREC	W	0.0000000 B			
	0	-0.20469125 B	-5.39	0.0001	0.03794770
	1	0.0000000 B			
DIR	0	-0.03339209 B	-1.24	0.2144	0.02689413

ANALYSE DE LA VARIANCE : PRIME PURE PAR RISQUE ET TEMPÊTE 49
18:41 Monday, November 22, 1999

General Linear Models Procedure

Dependent Variable: PRIMPUR

Parameter	Estimate	T for H0: Parameter=0	Pr > T	Std Error of Estimate
DIR 1	0.0000000 B			
LVIT	1.08313706	26.87	0.0001	0.04030302
LDUR	-0.01587361	-2.48	0.0132	0.00640442
RICHESS	0.07184056	2.94	0.0033	0.02441570
EFFLOCN	15.70646722	117.27	0.0001	0.13393429
DENSITE	0.00235620	2.15	0.0312	0.00109366
SUPERF	0.00000056	13.94	0.0001	0.00000004
POITASS	-0.47211531	-4.26	0.0001	0.11071084
NORMCONS	-0.17886823	-7.97	0.0001	0.02243194

NOTE: The X'X matrix has been found to be singular and a generalized inverse was used to solve the normal equations. Estimates followed by the letter 'B' are biased, and are not unique estimators of the parameters.

----- RISQUE=2 -----
ANALYSE DE LA VARIANCE : PRIME PURE PAR RISQUE ET TEMPÊTE 26
18:41 Monday, November 22, 1999

General Linear Models Procedure
Class Level Information

Class	Levels	Values
TYPTEMP	7	E N NE OR S SW W
PREC	2	0 1
DIR	2	0 1

Number of observations in data set = 12623

ANALYSE DE LA VARIANCE : PRIME PURE PAR RISQUE ET TEMPÊTE 27
18:41 Monday, November 22, 1999

General Linear Models Procedure

Dependent Variable: PRIMPUR

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	16	8761.869484	547.616843	1218.26	0.0001
Error	12606	5666.476488	0.449506		
Corrected Total	12622	14428.345972			
	R-Square	C. V.	Root MSE	PRIMPUR Mean	
	0.607268	147.1213	0.670452	0.455714	

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TYPTEMP	6	92.111386	15.351898	34.15	0.0001
PREC	1	261.459797	261.459797	581.66	0.0001
DIR	1	11.527130	11.527130	25.64	0.0001
LVIT	1	2258.823897	2258.823897	5025.12	0.0001
LDUR	1	41.546690	41.546690	92.43	0.0001
RICHESS	1	1.502872	1.502872	3.34	0.0675
EFFLOCN	1	6010.266728	6010.266728	13370.82	0.0001
DENSITE	1	15.813911	15.813911	35.18	0.0001
SUPERF	1	42.471861	42.471861	94.49	0.0001
POITLOC	1	20.261149	20.261149	45.07	0.0001
NORMCONS	1	6.084061	6.084061	13.53	0.0002

Source	DF	Type III SS	Mean Square	F Value	Pr > F
--------	----	-------------	-------------	---------	--------

TYPTEMP	6	109.766452	18.294409	40.70	0.0001
PREC	1	12.104286	12.104286	26.93	0.0001
DIR	1	0.657497	0.657497	1.46	0.2265
LVIT	1	180.929074	180.929074	402.51	0.0001
LDUR	1	9.731476	9.731476	21.65	0.0001
RICHESS	1	2.553453	2.553453	5.68	0.0172
EFFLOCN	1	5930.075599	5930.075599	13192.42	0.0001
DENSITE	1	1.428637	1.428637	3.18	0.0746
SUPERF	1	23.372643	23.372643	52.00	0.0001
POITLOC	1	13.853894	13.853894	30.82	0.0001
NORMCONS	1	6.084061	6.084061	13.53	0.0002

Parameter	Estimate	T for H0: Parameter=0	Pr > T	Std Error of Estimate
INTERCEPT	-2.293365427 B	-7.80	0.0001	0.29394037
TYPTEMP	0.147569513 B	3.07	0.0021	0.04806500
E	0.005738190 B	0.33	0.7413	0.01738069
N	-0.023138456 B	-0.64	0.5190	0.03588056
NE	0.279642522 B	13.12	0.0001	0.02131649
OR	0.056130277 B	2.96	0.0031	0.01896860
S	0.148823548 B	6.45	0.0001	0.02306894
SW	0.000000000 B	.	.	.
W	-0.112519817 B	-5.19	0.0001	0.02168340
PREC	0.000000000 B	.	.	.
0	-0.018629297 B	-1.21	0.2265	0.01540343
DIR	0	.	.	.

ANALYSE DE LA VARIANCE : PRIME PURE PAR RISQUE ET TEMPÊTE 28
18:41 Monday, November 22, 1999

General Linear Models Procedure

Dependent Variable: PRIMPUR

Parameter	Estimate	T for H0: Parameter=0	Pr > T	Std Error of Estimate
DIR	1	0.000000000 B	.	.
LVIT	0.480198010	20.06	0.0001	0.02393504
LDUR	-0.017138434	-4.65	0.0001	0.00368341
RICHESS	0.075691730	2.38	0.0172	0.03175795
EFFLOCN	6.648218385	114.86	0.0001	0.05788194
DENSITE	-0.001195990	-1.78	0.0746	0.00067086
SUPERF	0.000000171	7.21	0.0001	0.00000002
POITLOC	0.363204210	5.55	0.0001	0.06542336
NORMCONS	-0.046630038	-3.68	0.0002	0.01267468

NOTE: The X'X matrix has been found to be singular and a generalized inverse was used to solve the normal equations. Estimates followed by the letter 'B' are biased, and are not unique estimators of the parameters.

----- RISQUE=3 -----

ANALYSE DE LA VARIANCE : PRIME PURE PAR RISQUE ET TEMPÊTE 44
18:41 Monday, November 22, 1999

General Linear Models Procedure
Class Level Information

Class	Levels	Values
TYPTEMP	7	E N NE OR S SW W
PREC	2	0 1
DIR	2	0 1

Number of observations in data set = 12623

ANALYSE DE LA VARIANCE : PRIME PURE PAR RISQUE ET TEMPÊTE 45
18:41 Monday, November 22, 1999

General Linear Models Procedure

Dependent Variable: PRIMPUR

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	14	21931.27002	1566.51929	1605.36	0.0001

Error	12608	12302.92836	0.97580		
Corrected Total	12622	34234.19838			
	R-Square	C. V.	Root MSE	PRIMPUR Mean	
	0.640625	124.3756	0.987828	0.794229	
Source	DF	Type I SS	Mean Square	F Value	Pr > F
TYPTEMP	6	303.62165	50.60361	51.86	0.0001
PREC	1	649.39654	649.39654	665.50	0.0001
DIR	1	44.58622	44.58622	45.69	0.0001
LVIT	1	5444.17021	5444.17021	5579.17	0.0001
LDUR	1	65.19036	65.19036	66.81	0.0001
RICHESSSE	1	5.78383	5.78383	5.93	0.0149
EFFLOCN	1	15289.40144	15289.40144	15668.53	0.0001
SUPERF	1	82.19393	82.19393	84.23	0.0001
NORMCONS	1	46.92582	46.92582	48.09	0.0001
Source	DF	Type III SS	Mean Square	F Value	Pr > F
TYPTEMP	6	137.78507	22.96418	23.53	0.0001
PREC	1	85.49602	85.49602	87.62	0.0001
DIR	1	0.13224	0.13224	0.14	0.7128
LVIT	1	758.45732	758.45732	777.26	0.0001
LDUR	1	15.72362	15.72362	16.11	0.0001
RICHESSSE	1	1.47506	1.47506	1.51	0.2189
EFFLOCN	1	15015.03286	15015.03286	15387.36	0.0001
SUPERF	1	51.84111	51.84111	53.13	0.0001
NORMCONS	1	46.92582	46.92582	48.09	0.0001
Parameter		Estimate	T for H0: Parameter=0	Pr > T	Std Error of Estimate
INTERCEPT		-2.21057307 B	-8.74	0.0001	0.25285152
TYPTEMP	E	0.10875147 B	1.54	0.1239	0.07068228
	N	-0.08244258 B	-3.22	0.0013	0.02562792
	NE	-0.13865795 B	-2.64	0.0084	0.05259464
	OR	0.24661218 B	7.85	0.0001	0.03141841
	S	0.01103678 B	0.39	0.6930	0.02795753
	SW	0.08489659 B	2.50	0.0125	0.03396894
	W	0.00000000 B	.	.	.
PREC	0	-0.29845300 B	-9.36	0.0001	0.03188483
	1	0.00000000 B	.	.	.
DIR	0	-0.00834327 B	-0.37	0.7128	0.02266402
	1	0.00000000 B	.	.	.
LVIT		0.94989655	27.88	0.0001	0.03407155
LDUR		-0.02170663	-4.01	0.0001	0.00540751
RICHESSSE		0.02974837	1.23	0.2189	0.02419576
ANALYSE DE LA VARIANCE : PRIME PURE PAR RISQUE ET TEMPÊTE					46
					18:41 Monday, November 22, 1999
General Linear Models Procedure					
Dependent Variable: PRIMPUR					
Parameter		Estimate	T for H0: Parameter=0	Pr > T	Std Error of Estimate
EFFLOCN		11.81963352	124.05	0.0001	0.09528445
SUPERF		0.00000024	7.29	0.0001	0.00000003
NORMCONS		-0.12589900	-6.93	0.0001	0.01815505
NOTE: The X'X matrix has been found to be singular and a generalized inverse was used to solve the normal equations. Estimates followed by the letter 'B' are biased, and are not unique estimators of the parameters.					

LISTE DES TABLEAUX

PREMIERE PARTIE : LES ASPECTS THEORIQUES DU RISQUE TEMPETE

INTRODUCTION	15
CHAPITRE 1 : LES CONNAISSANCES SCIENTIFIQUES DU RISQUE TEMPETE.....	17
<u>Tableau 1.1.1</u> : Les 40 sinistres les plus coûteux de la période 1970-1997	34
<u>Tableau 1.1.2</u> : Historique des catastrophes naturelles survenues en France depuis 1982	37
CHAPITRE 2 : LES CONNAISSANCES METEOROLOGIQUES DU RISQUE TEMPETE.....	39
<u>Tableau 1.2.1</u> : Liste des situations du fichier d'expertise	42
<u>Tableau 1.2.2</u> : Liste des caractéristiques de la tempête de type Wd et variables explicatives associées	45
<u>Tableau 1.2.3</u> : Nombre de situations du fichier élargi pour chaque type de tempête (résultats provisoires)	48
<u>Tableau 1.2.4</u> : Nombre de situations du fichier élargi pour chaque type de tempête (résultats définitifs)	50
<u>Tableau 1.2.5</u> : Nombre de situations et de stations retenues après la première sélection pour chaque type de tempête	54
<u>Tableau 1.2.6</u> : Nombre de stations sélectionnées pour chaque type de tempête	55
<u>Tableau 1.2.7</u> : Valeur minimale du coefficient de corrélation multiple lors de la reconstitution de l'ensemble des stations touchées par des vents forts à partir des stations du panel	56
<u>Tableau 1.2.8</u> : Table de contingence en fréquence du type N pour la station de Dinard	58
<u>Tableau 1.2.9</u> : Nombre moyen de tempêtes par an pour les différents types et valeurs extrêmes	62
<u>Tableau 1.2.10</u> : Valeurs du χ^2 calculées pour les différents types	66
<u>Tableau 1.2.11</u> : Rapport entre le nombre observé et le nombre théorique de tempêtes après un type donné pour différentes fenêtres pour les mois d'octobre à avril	67
<u>Tableau 1.2.12</u> : Valeurs de C^2 pour les différents types de tempête et pour les différentes fenêtres étudiées (les valeurs significatives à 95 % sont indiquées en gras)	69
<u>Tableau 1.2.13</u> : Fréquence d'observation (en %) d'une tempête de même type, d'un autre type ou d'absence de tempête dans les deux jours qui suivent	72
<u>Tableau 1.2.14</u> : Nombre d'années pour lesquelles on observe au moins une tempête pour chaque type de tempête	74
<u>Tableau 1.2.15</u> : Nombre de stations retenues pour le calcul de l'intensité (Card I) par type	76
CHAPITRE 3 : ETUDE DE LA VULNERABILITE DES CONSTRUCTIONS.....	81
<u>Tableau 1.3.1</u> : Pressions dynamiques selon les régions de France	91
<u>Tableau 1.3.2</u> : Coefficient multiplicateur relatif à l'effet de site	92
CHAPITRE 4 : L'EVOLUTION DES COUVERTURES D'ASSURANCE	97
<u>Tableau 1.4.1</u> : Territorialité des lois sur les catastrophes naturelles et les effets du vent	105
<u>Tableau 1.4.2</u> : Récapitulatif des mesures de réforme du régime des catastrophes naturelles (9 août 1999)	110
<u>Tableau 1.4.3</u> : Evolution du montant des franchises pour la gamme de produits UAP	117
<u>Tableau 1.4.4</u> : Comparaison des risques garantis selon les différents régimes d'indemnisation pour les contrats UAP.....	127
CHAPITRE 5 - LA REASSURANCE.....	131
<u>Tableau 1.5.1</u> : Estimation du montant des primes et des sinistres de la compagnie et de l'indice ISO (exemple n°1).....	143
<u>Tableau 1.5.2</u> : Bilan financier de l'opération de couverture par l'indice du C.B.O.T. (exemple n°1)	145
<u>Tableau 1.5.3</u> : Estimation du montant des primes et des sinistres de la compagnie et de l'indice ISO (exemple n°2).....	146
<u>Tableau 1.5.4</u> : Bilan financier de l'opération de couverture par l'indice du C.B.O.T. (exemple n°2)	147
<u>Tableau 1.5.5</u> : Estimation du montant des primes et des sinistres de la compagnie (exemple n°2, stratégie alternative)	147
<u>Tableau 1.5.6</u> : Risques couverts par branche entrant dans la composition de l'indice CBOT	149

DEUXIÈME PARTIE : MODELISATION DU RISQUE TEMPETE

CHAPITRE 1 : CONSTITUTION DE LA BASE DE DONNÉES SUR LES TEMPÊTES.....157

Tableau 2.1.1 : Illustration de la méthode de redressement des matrices avec code postal absent	164
Tableau 2.1.2 : Illustration de la méthode de calcul de la prime 1992 du contrat “ mortalité du bétail ”	167
Tableau 2.1.3 : Tableau de contingence de deux variables X et Y qualitatives à r et s modalités	170
Tableau 2.1.4 : Principe d'utilisation de l'entité régionale pour l'évaluation de la qualité de l'estimation du chiffre d'affaires du contrat “ mortalité du bétail ” réparti selon la région du risque	174
Tableau 2.1.5 : Répartition du chiffre d'affaires du contrat “ mortalité du bétail ” selon l'entité régionale et la région du risque (en kF)	174
Tableau 2.1.6 : Comparaison des χ^2 des méthodes d'estimation par département ou agence	190
Tableau 2.1.7 : Répartition du nombre de départements selon le risque et le signe de la quantité : χ^2 (département) - χ^2 (agence)	191
Tableau 2.1.8 : Comparaison des distances de l'écart absolu des méthodes d'estimation par département ou agence	192
Tableau 2.1.9 : Comparaison des χ^2 selon le risque et le réseau (méthode à évolution homogène)	194
Tableau 2.1.10 : Comparaison selon la région des χ^2 relatifs au calcul du chiffre d'affaires par département et risque (1992)	201
Tableau 2.1.11 : Exemple de reconstitution par risque et département du portefeuille régional du réseau S	204
Tableau 2.1.12 : Architecture du portefeuille du réseau S selon la nature et le risque	205
Tableau 2.1.13 : Comparaison des distributions départementales des réseaux S et C fournies par deux méthodes grâce à la distance du χ^2 , selon le risque (portefeuille 1984)	207
Tableau 2.1.14 : Taux de sinistres avec V.S. 30 selon l'ordre de génération de numéros suivant les sinistres tempête de 1982	217
Tableau 2.1.15 : Taux de découverte des sinistres avec V.S. 30 selon les caractéristiques des sinistres tempête à l'origine de la génération de leur numéro et selon l'entité	219
Tableau 2.1.16 : Répartition du nombre de sinistres “ cat. nat. ” d'origine tempête par exercice de survenance	226
Tableau 2.1.17 : Répartition du nombre de contrats sinistrés “ délocalisés ” par réseau	229
Tableau 2.1.18 : Contribution de chaque type de renseignement au redressement du code postal du lieu du sinistre (en nombre de sinistres)	235
Tableau 2.1.19 : Segmentation des contrats d'habitation selon l'exposition au risque tempête	250

CHAPITRE 2 : DÉTERMINATION ET ACTUALISATION DU PRÉJUDICE DÛ À LA TEMPÊTE

.....253

Tableau 2.2.1 : Comparaison de 3 calculs d'indemnités à partir d'un même préjudice dû à la tempête	257
Tableau 2.2.2 : Répartition des branches-risques selon la nature de la présence de la garantie tempête	266
Tableau 2.2.3 : Relation entre les taux de présence de la garantie tempête avant et après la pollicitation	269
Tableau 2.2.4 : Taux de présence de la garantie tempête par risque entre 1980 et 1989	270
Tableau 2.2.5 : Intervalle de confiance lié au taux de présence de la garantie tempête par risque entre 1980 et 1989	271
Tableau 2.2.6 : Taux de présence de la garantie tempête (TPGT) pour les tempêtes déclarées catastrophes naturelles	274
Tableau 2.2.7 : Taux de présence de la garantie tempête (TPGT): niveaux de rupture et hypothèse retenus dans les calculs	275
Tableau 2.2.8 : Etude des variations de la fonction $h : E(X) ? n_g$	280
Tableau 2.2.9 : Taux de franchise, par risque, relativement au montant de l'indemnisation avant l'application de la franchise	286
Tableau 2.2.10 : Exemples des dispositions prises en matière d'exclusions pour quelques produits UAP	289
Tableau 2.2.11 : Estimation du poids de chaque élément d'exclusion extrapolé d'un sondage	290
Tableau 2.2.12 : Présentation des situations de règlements en tempête après 1990 selon le niveau de la vétusté	292
Tableau 2.2.13 : Exemple chiffré des situations de règlements en tempête après 1990 selon le niveau de la vétusté	292
Tableau 2.2.14 : Estimation du taux de vétusté par risque extrapolé d'un sondage	295
Tableau 2.2.15 : Estimation du taux de règlement par risque extrapolé d'un sondage	295
Tableau 2.2.16 : Règles d'indemnisation de la tempête en risque agricole avant et après 1993	297
Tableau 2.2.17 : Estimation de la part de préjudice à la charge de l'assuré compte tenu de la limite de garantie	301
Tableau 2.2.18 : Estimation de l'indice correctif lié à l'évolution de la T.V.A.	302

CHAPITRE 3 : LA TARIFICATION DU RISQUE TEMPETE.....307

Tableau 2.3.1 : Tarification du risque tempête pour le contrat Hôtel- Multirisque (1975)	310
---	-----

Tableau 2.3.2 : Primes d'équilibre de l'incendie et de la tempête des contrats Multirisque Habitation et artisan - commerçant.....	311
Tableau 2.3.3 : Tarification de la garantie tempête HT en proportion de la prime Incendie ou des capitaux garantie	313
Tableau 2.3.4 : Impact de l'application d'une franchise sur le montant de la prime pure tempête	337
Tableau 2.3.5 : Test du χ^2 d'ajustement du nombre de sinistres par tempête à une loi de Pareto	342
CHAPITRE 4 : MODELISATION DU RISQUE TEMPÊTE.....	345
Tableau 2.4.1 : Caractéristiques de la sinistralité des différents types de tempêtes définis par Météo-France	354
Tableau 2.4.2 : Impact sur le montant de la prime pure de l'augmentation de l'intensité maximale journalière du vent de 20 à 25 m/s	380
Tableau 2.4.3 : Impact sur le montant de la prime pure du passage d'une direction normale à anormale	381
Tableau 2.4.4 : Impact sur le montant de la prime pure du passage d'un niveau faible à significatif des précipitations	381
Tableau 2.4.5 : Impact sur le montant de la prime pure du passage du type de tempête le moins destructeur au plus destructeur	382
Tableau 2.4.6 : Impact sur le montant de la prime pure de l'augmentation de la part du capital sous risque touché de 10 % à 15 %	383
Tableau 2.4.7 : Impact sur le montant de la prime pure de l'augmentation de la prime moyenne de la zone météorologique de 2.000 F à 3.000 F.....	383
Tableau 2.4.8 : Impact sur le montant de la prime pure de l'augmentation de la superficie des zones touchées de 1.000 à 2.000 km ²	384
Tableau 2.4.9 : Impact sur le montant de la prime pure du passage d'une zone fortement exposée à une zone faiblement exposée	385

LISTE DES FIGURES

PREMIERE PARTIE : LES ASPECTS THEORIQUES DU RISQUE TEMPETE

INTRODUCTION	15
CHAPITRE 1 : LES CONNAISSANCES SCIENTIFIQUES DU RISQUE TEMPETE.....	17
Figure 1.1.1 : Les principales conséquences du réchauffement climatique	30
Figure 1.1.2 : Evolution du nombre de catastrophes naturelles recensées dans le monde de 1970 à 1997	32
Figure 1.1.3 : Evolution du montant des dommages (en millions USD 1997) suite à des tempêtes dans le monde de 1970 à 1997	33
Figure 1.1.4 : Dommages assurés dus aux catastrophes naturelles en % du PIB des pays industrialisés.	33
Figure 1.1.5 : Evolution du montant des règlements des événements naturels selon le régime d'indemnisation (Cat. Nat., TGN) existant en France	36
CHAPITRE 2 : LES CONNAISSANCES METEOROLOGIQUES DU RISQUE TEMPETE.....	39
Figure 1.2.1 : Arborescence de la typologie des tempêtes	43
Figure 1.2.2 : Répartition des différents types de tempête sur le fichier d'expertise	44
Figure 1.2.3 : Algorithme de la classification	48
Figure 1.2.4 : Répartition (en pourcentage) des différents types de tempête sur le fichier d'expertise (en foncé) et sur le fichier élargi (en clair).	49
Figure 1.2.5 : Répartition des différents types de tempête sur le fichier élargi (résultats définitifs)	50
Figure 1.2.6 : Localisation des 91 stations initiales	53
Figure 1.2.7 : Histogramme intensité / durée pour la station de Dinard pour le type N.	59
Figure 1.2.8 : Histogramme intensité / durée pour la station de Dinard pour le type W.	60
Figure 1.2.9 : Nombre de tempêtes observées chaque année de 1962 à 1993 (nombre moyen en pointillé). ...	62
Figure 1.2.10 : Répartition mensuelle des tempêtes	64
Figure 1.2.11 : Nombre de situations pour les différents types de tempête au pas de temps quotidien (en gris) et en considérant une durée de deux (en blanc) ou trois jours (en noir).	70
CHAPITRE 3 : ETUDE DE LA VULNERABILITE DES CONSTRUCTIONS	81
Figure 1.3.1 : Lignes d'action du vent sur une construction.....	84
Figure 1.3.2 : Action des vents sur les constructions et effets des ouvertures	90
Figure 1.3.3 : Valeur des pressions dynamiques normales et extrêmes en fonction de la hauteur par rapport au sol pour chaque site de la région I.....	93
Figure 1.3.4 : Principes de construction permettant de limiter les dommages dus aux tempêtes	95
CHAPITRE 4 : L'EVOLUTION DES COUVERTURES D'ASSURANCE	97
CHAPITRE 5 - LA REASSURANCE.....	131

DEUXIÈME PARTIE : MODELISATION DU RISQUE TEMPETE

CHAPITRE 1 : CONSTITUTION DE LA BASE DE DONNÉES SUR LES TEMPÊTES.....157

Figure 2.1.1 : Recensement des différents cas de transferts de portefeuille entre agences	184
Figure 2.1.2 : Interprétation géographique de l'indicateur de proximité de deux agences	187
Figure 2.1.3 : Exemple du passage de l'estimateur à évolution homogène à celui issu du descriptif départemental pour le portefeuille des réseaux S et C	209
Figure 2.1.4 : Image de la transaction 2STEMP.....	211
Figure 2.1.5 : Fiche signalétique d'un sinistre de l'UAP	214
Figure 2.1.6 : Processus de génération automatique de numéros de sinistres " cat. nat." potentiels	216
Figure 2.1.7 : Présentation d'axes aux pouvoirs discriminants différents	222
Figure 2.1.8 : Fiche signalétique « contrats ».....	228
Figure 2.1.9 : Carte de France des 168 zones météorologiques.....	242
Figure 2.1.10 : Carte de France des 130 zones météorologiques.....	244
Figure 2.1.11 : Répartition des ménages selon la date d'achèvement de l'immeuble occupé	248

CHAPITRE 2 : DÉTERMINATION ET ACTUALISATION DU PRÉJUDICE DÛ À LA TEMPÊTE

.....253

Figure 2.2.1 : Représentation d'une loi log-normale et d'une loi log-normale tronquée	277
Figure 2.2.2 : Histogramme de distribution du taux de vétusté appliquée	294

CHAPITRE 3 : LA TARIFICATION DU RISQUE TEMPETE.....307

Figure 2.3.1 : Evolution annuelle du nombre de sinistres tempête (1980 - 1997)	324
Figure 2.3.2 : Répartition mensuelle des sinistres tempête	326
Figure 2.3.3 : Répartition départementale de la fréquence de sinistres pour le risque agricole	327
Figure 2.3.4 : Répartition départementale de la fréquence de sinistres pour le risque d'habitation	328
Figure 2.3.5 : Répartition départementale de la fréquence de sinistres pour le risque professionnel	329
Figure 2.3.6 : Répartition départementale de la prime pure pour le risque agricole	330
Figure 2.3.7 : Répartition départementale de la prime pure pour le risque habitation	331
Figure 2.3.8 : Répartition départementale de la prime pure pour le risque professionnel	332
Figure 2.3.9 : Distribution du montant du préjudice dû à la tempête pour le risque agricole	334
Figure 2.3.10 : Distribution du montant du préjudice dû à la tempête pour le risque d'habitation	335
Figure 2.3.11 : Distribution du montant du préjudice dû à la tempête pour le risque professionnel	336
Figure 2.3.12 : Evolution annuelle du nombre de tempêtes (1980 - 1997)	338
Figure 2.3.13 : Répartition mensuelle du nombre de tempêtes	340
Figure 2.3.14 : Distribution du nombre de sinistres par tempête (1980 -1997)	341

CHAPITRE 4 : MODELISATION DU RISQUE TEMPÊTE.....345

Figure 2.4.1 : Comparaison de deux tempêtes au coût identique mais touchant partiellement ou totalement une zone météorologique.....	351
Figure 2.4.2 : Contour des 111 zones météorologiques finalement retenues	361
Figure 2.4.3 : Carte des vitesses de vents déclencheurs de dommages pour le risque agricole	367
Figure 2.4.4 : Carte des vitesses de vents déclencheurs de dommages pour le risque d'habitation	368
Figure 2.4.5 : Carte des vitesses de vents déclencheurs de dommages pour le risque professionnel	369
Figure 2.4.6 : Fonction de distribution de l'intensité du vent de la station de Poitiers	371
Figure 2.4.7 : Répartition par zone météorologique de l'espérance de la prime pure pour le risque agricole	374
Figure 2.4.8 : Répartition par zone météorologique de l'espérance de la prime pure pour le risque	375
Figure 2.4.9 : Répartition par zone météorologique de l'espérance de la prime pure pour le risque professionnel	376