

**Mémoire présenté devant l'Université Paris Dauphine
pour l'obtention du diplôme du Master Actuariat
et l'admission à l'Institut des Actuaire**

le _____

Par : Toyi KOULI

Titre: Etude relative à la construction et au paramétrage du modèle actuariel
nécessaire à la conception et à l'exploitation d'une assurance paramétrique

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

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

Membre présent du jury de l'Institut
des Actuaire :

Signature : Entreprise :

Nom : GROUPE FINACTU

Signature :

Membres présents du jury du Master
Actuariat de Dauphine :

Directeur de mémoire en entreprise :

Nom : Thibault GAUTHIER

Signature :

**Autorisation de publication et de mise en ligne sur un site de diffusion de documents
actuariels** (après expiration de l'éventuel délai de confidentialité)

Signature du responsable entreprise :

Secrétariat :

Bibliothèque :

Signature du candidat :

Mots Clés : Assurance paramétrique, Risque de base, Aléa moral, Aversion au risque, Valeurs extrêmes, Tarification d'assurance paramétrique.

"A mes chers parents, mes sœurs et mon frère,

A toute ma famille, mes amis et amies,

A tous ceux qui me sont chers,

je dédie ce travail "

Avant-propos

Ce document est confidentiel et ne doit être distribué ou communiqué sans le consentement préalable de l'entreprise d'accueil (Groupe FINACTU) et de la MAPM.

Liste des Abréviations

CGMS : Copy Generation Management System

DSS : Direction de la Stratégie et des Statistiques

HARITA : Horn of Africa Risk Transfer for Adaptation

INRA : Institut National de la Recherche Agronomique

MAPM : Ministère de l'Agriculture et de la Pêche Maritime

MRC : Multirisque Climatique

NAIS : National Agricultural Insurance Scheme

WBCIS : Weather Based Crop Insurance Scheme

BASIX : Building Sustainably Index

Résumé

Depuis sa mise en place en 2011, l'assurance multirisque climatique (MRC) au Maroc connaît une incontestable réussite, ce succès de l'assurance MRC doit beaucoup à une forte subvention de l'Etat, et il est important que, progressivement, le mécanisme de couverture des risques puisse fonctionner de façon plus autonome. Cependant l'assurance agricole peine encore à connaître un véritable essor, cela compte tenu de la montée des risques qui pèsent sur l'agriculture marocaine.

Les études réalisées pour comprendre ce faible développement de l'assurance agricole au Maroc, ont permis de mettre en évidence de nombreux facteurs. Certains tiennent au secteur des assurances lui-même, mais beaucoup sont liés au domaine agricole, qui ne forme pas un terrain favorable au développement de l'assurance pour des raisons de dispersion géographique, de réticence culturelle, d'appréhension devant la complexité, etc. Ces entraves rendent le coût de l'assurance traditionnelle très élevé et, de fait, non viable sans un soutien très fort de l'Etat. Ces études ont souligné le fait que, face à ces entraves, l'assurance indiciaire pouvait apparaître comme une solution intéressante.

Ainsi dans un contexte de risque de sécheresse permanent, et malgré la mise en place de la garantie sécheresse et de l'assurance multirisque climatique et surtout dans le but d'améliorer le système de gestion des risques climatiques principalement la sécheresse, des programmes en partenariat avec la banque mondiale, l'agence européenne de développement, DSS, vont permettre à l'INRA de mettre sur pied un modèle de prévision de rendements appelé Modèle CGMS-Maroc. A partir du début d'une saison agricole donnée, ce modèle permet de prévoir la production de la saison à partir d'une série de paramètres, notamment climatiques. Le modèle CGMS-Maroc devrait donc permettre de délimiter un indice grâce auquel un produit d'assurance pourrait être construit.

L'objectif de ce travail est de proposer un modèle actuariel de tarification d'assurance paramétrique, ce produit devrait venir en complément ou en substitution de l'actuelle garantie : l'assurance MRC. Un tel produit d'assurance paramétrique sur les rendements devrait aussi non seulement couvrir tous types de risques (climatiques, humains, sanitaires, ...), mais aussi permettre d'éliminer l'aléa moral, le risque d'anti-sélection et de réduire considérablement les délais d'indemnisation. La suppression de l'aléa moral et de l'anti-sélection dans ce nouveau produit réside dans le fait que l'agriculteur est indemnisé en fonction du paramètre, qu'il ait ou non subi un sinistre. Il n'a donc non seulement aucun intérêt à augmenter son sinistre pour maximiser l'indemnisation, mais s'il gère correctement sa récolte, il peut même espérer percevoir l'indemnité tout en tirant bénéfice de sa récolte. De ce fait, cette forme d'assurance intéresse et attire aussi bien les bons agriculteurs que les mauvais agriculteurs.

L'assurance paramétrique attendue sera aussi caractérisée par une moindre consommation de données liées aux sinistres : une fois le modèle calibré, l'assurance paramétrique ou indicielle a le mérite de ne plus avoir besoin de données liées aux sinistres pour le déclenchement de l'assurance et la gestion des indemnités. Une fois que l'indice est construit, les indemnités ne sont déclenchées que par le niveau de l'indice.

Abstract

Since it was set up in 2011, the multi-risk climate insurance (MRC) in Morocco knows an undeniable success, this success of the insurance MRC owes a lot to a strong state subsidy, and it is important that, gradually, the mechanism of coverage risks to operate more autonomously. However agricultural insurance penalty still a real boom, given the increasing risks that now exist on Moroccan agriculture. The studies carried out to understand this low development of agricultural insurance in Morocco, have highlighted many factors. Some go to the sector of insurance itself, but many are related to agricultural property, which does not form a favorable to the development of insurance for reasons of geographical dispersion, cultural recreation, apprehension about complexity. These reasons make cost of traditional insurance very high and, in fact, unsustainable without strong support from the state. These studies have highlighted the fact that, in the face of these barriers, index insurance appear as an interesting solution.

So in a context of permanent drought risk in Morocco, the drought warranty and climate-risk insurance have been implemented to ensure the protection against the decline in income of farmers. In order to improve the climate-risk management system and the drought in particular, the European Development Agency, DSS, will allow the INRA to put in place a prevision model of yields called the CGMS-Morocco Model that would delimit an index with which an insurance product could be built.

The goal of this work is to propose an actuarial model of parametric insurance pricing, this product would complement or replace the current warranty : The AMRC. This new product is expected to eliminate moral hazard, the risk of anti-selection and considerably reduce the time for compensation.

Remerciements

Je tiens à exprimer mes vifs remerciements en premier à Mr **Thibault GAUTHIER** et à Mr **Ulrich OUINSOU**, actuaires à FINACTU qui ont accepté d'encadrer ce projet, pour leur proximité, leurs conseils et leur bonne humeur.

Je tiens à remercier Mlle **Stéphanie SOEDJEDE**, actuaire à FINACTU, pour son attention et ses conseils dans la réalisation de ce projet.

Toutes mes gratitudes vont aussi à toute l'équipe dirigeante du groupe FINACTU, l'équipe administrative et tous les autres consultants qui ont facilité mon intégration et le bon déroulement de mon stage.

J'exprime mes profondes gratitudes à tous mes enseignants qui m'ont formé durant mes deux années du Master Actuariat.

Table des matières

1	Assurance agricole et problématique de l'étude	16
1.1	Contexte d'étude	16
1.2	Point sur l'assurance paramétrique en Afrique	17
1.3	Les outils de gestion de risques agricoles	19
1.3.1	L'acceptation du risque	20
1.3.2	Le transfert du risque	20
1.3.3	Stratégies publiques et filets de sécurité	20
1.3.4	L'assurance Agricole Classique	21
1.3.5	Assurance Agricole paramétrique	22
1.3.5.1	Les Indices	22
1.3.5.2	Le risque de base	24
1.4	Référence d'assurance indicielle	25
1.4.1	HARITA en Éthiopie, basée sur des données satellite	25
1.4.2	L'assurance Alliance One au Malawi	26
1.4.3	NAIS, Assurance indicielle basée sur le rendement en Inde	27
1.4.4	WBCIS, une assurance indicielle basée sur le climat en Inde	27
1.4.5	BASIX, une assurance indicielle basée sur la pluviométrie	27
1.4.6	Assurance récolte Sahel, une assurance indicielle basée sur l'évapotranspiration	27
1.5	Assurance paramétrique agricole en France	28
1.5.1	L'expérience de PACIFICA	28
1.5.2	L'expérience de METEO PROTECT	29
2	Étape Data	30
2.1	Système CGMS-Maroc et historique de l'assurance indicielle au Maroc	30
2.1.1	Les inputs	31
2.1.2	Traitement	32
2.1.3	Outputs :	34
2.2	Les drivers de risques dans les données	34

2.2.1	Risque de choc	34
2.2.2	Risque de volatilité et de tendance	35
2.3	Data Mining	35
2.3.1	Démarche de modélisation :	35
2.3.2	Point sur l'existant et choix de la modélisation	36
2.3.3	Mesure d'erreur d'estimation du modèle	36
2.3.3.1	Erreurs moyennes par céréale et taux de commune par seuil d'erreur relative	36
3	Évaluation et gestion du risque de base	41
3.1	Assurance indicielle et approche microéconomique	41
3.2	Approche « macro » du risque de base	43
3.2.1	Évaluation du risque de base à partir de la fonction d'indemnisation 1	47
3.2.2	Évaluation du risque de base à partir de la fonction d'indemnisation 2	48
4	Modèle actuariel d'assurance indicielle	50
4.1	Rappel des modèles d'assurance dommage et de catastrophe naturelle	50
4.1.1	Modèle fréquence-coût	52
4.1.2	Titrisation des catastrophes naturelles par les Cat bonds	53
4.1.3	Modèle de tarification de catastrophes naturelles	54
4.2	Modèle AMRC	55
4.3	Modèle d'assurance indicielle basée sur le rendement	56
4.3.1	Approche Burning Cost (déterministe)	56
4.3.2	Approche actuarielle et Monte Carlo	57
4.3.3	Area-yield insurance	58
4.4	Cadre de la Théorie sur les valeurs extrêmes	58
4.4.1	Convergence en loi du maximum d'un échantillon	59
4.4.2	Détermination de la loi de distribution	61
4.4.3	Test du « Q-Q Plot »	61
4.4.4	Tests d'adéquation	61
5	De la calibration à la tarification	62
5.1	Techniques de clustering sur la variable cible	62
5.1.1	Technique de classification : classification hiérarchique ascendante . .	62
5.1.2	Résultats et validation de la classification	63
5.2	Paramétrisation du modèle	67
5.2.1	Algorithme de la simulation	68

<i>TABLE DES MATIÈRES</i>	12
5.2.2 Prime d'assurance indicielle	68
5.2.3 Analyse du tarif sur les trois scénarios	73
5.3 Mesure du risque de tarification : scénario de baisse de rendements	73
5.4 Comparaison des deux garanties : l'assurance paramétrique et la MRC . . .	74
Annexes	83

Table des figures

1.1	Expositions du secteur agricole	17
1.2	Assurance paramétrique en Afrique	18
1.3	Schéma classique d'assurance paramétrique	22
1.4	Dispositif HARITA	26
2.1	Mesure de l'indice NDVI au Maroc	32
2.2	Dispositif CGMS-Maroc	34
2.3	Évolution des erreurs moyennes	37
2.4	Ecart entre les rendements estimés et observés en 2016	38
2.5	Superficies exploitées pour les trois principales céréales	39
3.1	fonction d'indemnisation 2	46
3.2	Taux de sinistres indemnisés	49
4.1	Fonctionnement des Cat bonds	53
4.2	Convergence d'après la loi forte des grands nombres	58
4.3	Distribution des lois de Gumbel, Frechet et de Weibull	60
5.1	Classification du blé tendre	63
5.2	Fonctions de répartition	65
5.3	Q-Q plot orge classe 1	66
5.4	Q-Q plot orge classe 2	66
5.5	Q-Q plot orge classe 3	67
5.6	LEC à partir du scénario 1 avec la fonction d'indemnisation 1	70
5.7	Tarification à partir de la fonction d'indemnisation 1	71
5.8	Primes sur les trois scénarios à partir de la fonction d'indemnisation 1	73
5.9	Impact d'une baisse de rendement sur le niveau des primes	74
10	Carte factorielle du blé dur	92
11	Carte factorielle du blé tendre	93
12	Carte factorielle orge	94

13	Q-Q plot blé dur : classe 1	95
14	Q-Q plot blé dur : classe 2	96
15	Q-Q plot blé dur : classe 3	97
16	Q-Q plot blé tendre : classe 1	98
17	Q-Q plot blé tendre : classe 2	99
18	Q-Q plot blé tendre : classe 3	100
19	Q-Q plot blé tendre : classe 4	101

Introduction générale

En Afrique, le secteur agricole est connu pour être un des piliers du développement économique et de la sécurité alimentaire. En l'occurrence, dans le pays objet de l'étude, l'agriculture occupe une place importante dans la politique de développement. Toutefois, elle est confrontée à un certain nombre de risques dont les risques climatiques car étant fortement dépendante des pluies. Un produit d'assurance agricole pour couvrir les récoltes contre les risques climatiques, principalement la sécheresse, a déjà été mis en place depuis 5 ans. Le besoin d'innovation à travers la faisabilité d'une assurance indicielle est désormais analysé.

Cette assurance indicielle permettra de réduire les risques d'aléa moral et d'anti-sélection, et présentera des coûts de gestion moins élevés que ceux d'une assurance agricole classique, étant donné l'absence d'expertise sur le terrain pour évaluer les pertes. Des travaux de recherche ont été menés afin d'élaborer un modèle de prévision des rendements agricoles qui permettra d'avoir un indice servant de base pour l'assurance qui sera élaborée. Le travail consiste donc à analyser les données disponibles pour une assurance indicielle, de tester leur fiabilité et pertinence afin de proposer un schéma d'assurance viable. Ce schéma a pour but d'élaborer un modèle de tarification qui devrait permettre d'obtenir les niveaux de primes permettant d'assurer la viabilité d'une telle assurance sur un risque considéré comme catastrophique.

Cette étude se déclinera donc en trois étapes : elle vise d'abord à mettre à disposition un schéma d'assurance indicielle adapté au contexte du pays, ensuite à mettre à disposition des tarifs d'une assurance paramétrique permettant d'élaborer l'offre commerciale à adresser aux agriculteurs et enfin à fournir des recommandations et propositions sur les stratégies/points d'attention pour la mise en œuvre d'un projet d'assurance paramétrique.

Chapitre 1

Assurance agricole et problématique de l'étude

1.1 Contexte d'étude

L'agriculture est un secteur clé de développement économique et du bien être de la population, contribuant en moyenne à plus de 15% du PIB du royaume du Maroc. Elle est cependant entravée par une multitude de risques susceptibles de dégrader continuellement et considérablement non seulement le revenu des producteurs mais aussi l'autosuffisance alimentaire de tout le royaume. Parmi les risques auxquels fait face l'activité agricole en général, le Maroc est particulièrement menacé par la sécheresse, qui représente à elle seule 44% des expositions des chiffres d'affaires toutes filiales agricoles confondues.

Les maladies et criquets ravageurs représentant 10% des expositions des chiffres d'affaires toutes filières confondues constituent la seconde entrave de l'activité agricole, mais la sécheresse prise ensemble avec les hautes températures, le chergui¹ et le stress hydrique constituent 64% des expositions des chiffres d'affaires toutes filiales agricoles confondues.

Ces quatre risques (sécheresse, chergui, hautes températures et le stress hydrique) que l'on peut dénommer « Maroc pays chaud » constituent le véritable « iceberg » de l'agriculture. Cette étude vise donc à proposer pour le secteur céréalier qui se trouve être le plus touché par la sécheresse un schéma de tarification destiné à servir de produit de couverture ou d'assurance pour les producteurs.

1. Le chergui vient de l'est ou du sud-est passe au-dessus de l'Atlas et redescend complètement asséché sur les plaines côtières

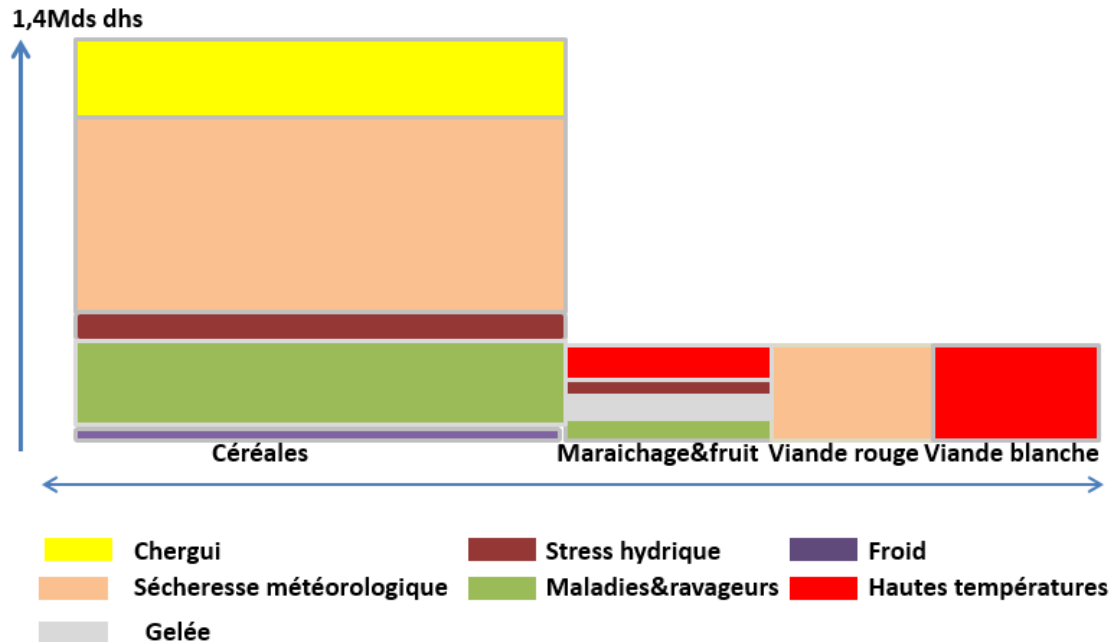


FIGURE 1.1 – Expositions du secteur agricole

1.2 Point sur l'assurance paramétrique en Afrique

L'assurance paramétrique (indicielle) sur le continent africain suscite déjà depuis plus d'une décennie un grand intérêt, les premiers acteurs à s'être positionnés sur le marché de l'indicielle ont été les réassureurs, ce qui leur confère de facto un avantage concurrentiel. Swiss Re par exemple bénéficie d'un partenariat historique avec la Banque mondiale qui a été à l'initiative de la mise en place de l'assurance indicielle à ses débuts, dans plusieurs pays en développement. Depuis peu, l'assurance paramétrique a séduit aussi les compagnies d'assurances et n'est donc plus exclusivement distribuée par les réassureurs.

Ainsi, Allianz développe depuis 2011 une assurance indicielle au Mali et au Burkina Faso. Axa Corporate Solutions s'est lancée dans l'assurance indicielle depuis juin 2013 et a signé, début 2015, un protocole d'accord avec le programme global pour l'assurance indicielle (Giif) géré par la Banque mondiale pour promouvoir cette assurance dans les pays en développement.

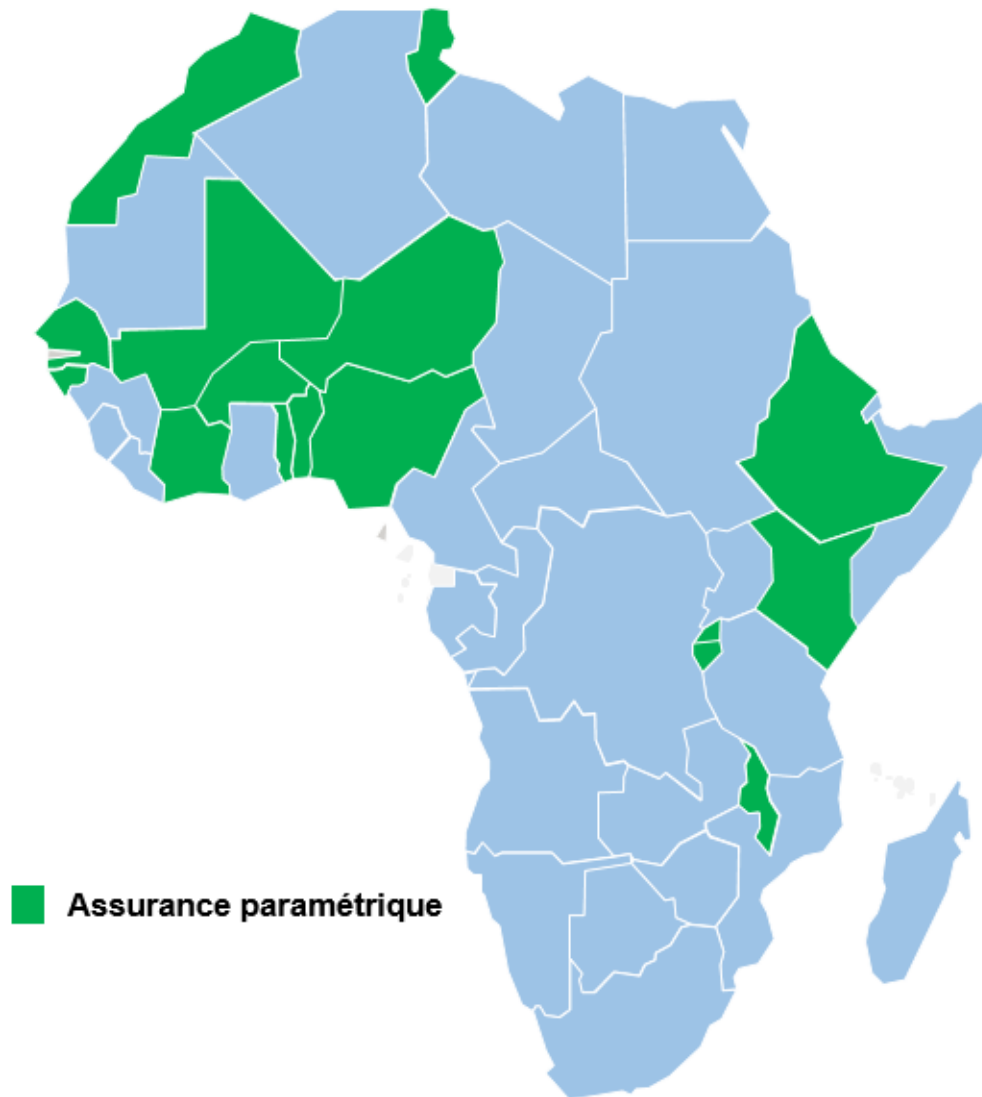


FIGURE 1.2 – Assurance paramétrique en Afrique

Le Groupe FINACTU propose depuis quelques années aussi son expérience sur des projets de mise en place d'assurance paramétrique notamment en zone UEMOA et en Afrique du Nord. Même si la mise en œuvre d'une telle assurance peut s'avérer onéreuse et fort compliquée, de par l'infrastructure nécessaire à la mise en œuvre de la collecte des données, d'autant que ce marché est encore émergent. Avec l'aide des gouvernements et des donateurs, couplé au développement de l'open data et des technologies, l'assurance indicelle semble être une solution prometteuse contre les aléas climatiques.

Pour les acteurs d'assurance comme Axa, l'assurance paramétrique va être le substitut idéal de l'assurance agricole classique dans les pays émergents : dans les pays développés, souvent 4 à 6% de la prime sert à payer les coûts de gestion ; alors que dans les pays en

voie de développement, comme les surfaces sont plus petites, ces coûts sont proportionnellement plus élevés, atteignant 10% , 15% voire plus de 20% de la prime. Avec l'assurance paramétrique, ces coûts seraient beaucoup moins significatifs. Il est intéressant de noter c'est dans les pays en développement qu'on été réalisé les premières tentatives de mise en place d'assurance paramétrique, étant l'un des premiers acteur à se lancer dans le domaine, 40% du chiffre d'affaire du LoB paramétrique d'AXA est généré en Asie

Après ce point sur l'assurance paramétrique en Afrique, l'étape suivante de cette étude vise dans un premier temps à présenter une taxonomie complète des risques agricoles, ensuite à présenter le modèle qui fournit les données pour cette étude, et enfin à présenter une analyse exploratoire de ces données, ce qui conduira à l'élaboration du modèle actuariel de tarification d'assurance paramétrique. La description générale des risques agricoles se présente comme suit :

Risques climatiques : il s'agit de la sécheresse météorologique, stress en ressources hydriques, haute température, vague de froid, gelée, grêle, vent fort, inondation ou excès d'eau, forte hygrométrie, orage.

Risques sanitaires et environnementaux : les risques sanitaires concernent les maladies et ravageurs végétaux, maladies et ravageurs émergents, infections et maladies épizootiques. Pour les risques environnementaux, il s'agit de la pollution industrielle, pollution agricole diffuse, salinité du sol et des eaux, réchauffement climatique, érosion hydrique, érosion éolienne, plante envahissante, déperdition de la biodiversité.

Risques de marché : il s'agit de la hausse des prix des intrants, baisse des prix de marché nationaux et internationaux, augmentation des prix de détail liée à la distribution (structure de marché), risques d'agrégation, inaccessibilité au crédit, changement de réglementation nationale , changement des standards / quotas / pratiques douanières à l'international , changement de fiscalité, fluctuation monétaire, recouvrement des paiements.

1.3 Les outils de gestion de risques agricoles

Face aux risques agricoles présentés ci-dessus, quelle stratégie adopter ? Il existe deux principales stratégies de gestion de risques agricoles : la prévention du risque et le partage du

risque. La prévention du risque consiste à réduire la probabilité des pertes agricoles de telle sorte que la fréquence d'occurrence de l'évènement déclencheur du sinistre soit minimale, prévenir le risque aussi consiste également à agir de manière à limiter l'impact du sinistre. Le traitement du risque est une stratégie à trois leviers d'intervention : l'acceptation du risque, le transfert du risque et le filet de sécurité (pouvoir public).

1.3.1 L'acceptation du risque

Il y a acceptation du risque lorsque le producteur assume seul le risque ou se donne les moyens de le faire. Deux stratégies sont possibles : la constitution de fonds propres (épargne), appelée auto-assurance et la diversification. L'agriculteur peut s'auto-assurer afin de surmonter les difficultés opérationnelles grâce aux réserves constituées. Pour la diversification, le producteur dispose de trois moyens :

- **La diversification temporelle** : il s'agit par exemple d'allonger les périodes de vente de la production pour limiter le risque prix. Si le producteur vend un dixième de sa production chaque semaine pendant 10 semaines, il diminue le risque de vendre l'intégralité de sa production lors d'une chute momentanée des prix.
- **La diversification de l'activité agricole** : chaque type de culture étant exposé à des degrés divers au risque, le risque global est réduit.
- **La diversification financière** : diversification des sources de revenu (artisanat, commerce).

1.3.2 Le transfert du risque

C'est la cession du risque à un tiers contre le paiement d'une prime. Le risque peut être transféré à un assureur ou au marché financier avec des contrats portant sur des aléas spécifiques affectant la production agricole (sécheresse, inondations, etc).

1.3.3 Stratégies publiques et filets de sécurité

Ce sont des stratégies mises en place pour les situations de crise que les deux autres stratégies ne peuvent couvrir. Les pouvoirs publics interviennent donc sur les marchés agricoles afin d'améliorer : la sécurité alimentaire, la durabilité de la production agricole, etc.

Les pouvoirs publics interviennent alors en cas de survenance d'une catastrophe ne pouvant pas être assumée par le producteur, ni par un assureur, ni par le marché. La création de filets de sécurité publics suppose de définir au préalable le seuil à partir duquel le sinistre est considéré comme catastrophique.

Le transfert de risque est la stratégie qui sera envisagée dans cette étude. De quelles stratégies de transfert de risque dispose-t-on pour l'assurance? Il existe deux grandes familles d'assurance agricole : l'assurance agricole classique (AAC) et l'assurance agricole paramétrique ou indicielle (AAP).

1.3.4 L'assurance Agricole Classique

Il s'agit de la forme classique de couverture contre un risque agricole, contre une prime d'assurance payée. L'assuré est indemnisé en cas de survenance d'un sinistre, ceci après expertise sur les déclarations de perte. Comme toutes les branches de l'assurance, l'AAC est caractérisée par : les asymétries de l'information sur la garantie, l'anti-sélection, la fraude, le risque moral, le risque de masse et l'aspect systémique des risques agricoles, le risque de pointe et la dispersion géographique de la production agricole. Parmi ces caractéristiques, celles qui sont susceptibles d'impacter considérablement un contrat d'AAC sont les suivantes.

Les asymétries de l'information sur la garantie : l'assureur et l'agriculteur ne sont pas égaux dans la détention d'information, l'agriculteur a par exemple une idée de la qualité future de sa production, fonction du climat, il dispose donc de cet avantage d'avoir une longueur d'avance sur l'assureur. La charge revient aux assureurs dans ce cas d'être habiles, pour comprendre les processus biologiques et techniques agricoles, non seulement pour déterminer les primes à la mesure des risques, mais également pour évaluer la gérance du risque des producteurs, et aider ces derniers à les améliorer. L'assurance paramétrique se présente comme la solution idéale pour palier au problème d'asymétrie de l'information sur la garantie ; l'utilisation d'un indice indépendant des caractéristiques des cultures (ex : pluviométrie) comme déclencheur d'indemnisation exclut toute inégalité dans la détention de l'information.

Sélection adverse : c'est une conséquence directe de l'asymétrie d'information, une des parties (l'assuré et l'assureur) dispose de plus d'information que l'autre. L'assuré sait qu'il est bon ou mauvais risque, ce qui n'est pas le cas pour l'assureur, les personnes présentant un risque plus élevé ont plus d'avantage à vouloir s'assurer.

L'aléa moral : c'est aussi l'une des conséquences de l'asymétrie d'information, l'agricul-

teur ou l'assuré a tendance à limiter son effort étant donné qu'il sera indemnisé si la récolte est mauvaise.

La dispersion géographique : à la différence des autres branches d'assurance, la dispersion géographique rend les coûts opérationnels et administratifs de l'assurance agricole élevés.

1.3.5 Assurance Agricole paramétrique

L'assurance indicielle est une approche relativement innovante dans le domaine de l'assurance qui indemnise les pertes d'actifs de l'agriculteur ou la baisse de fonds de roulement essentiellement sur la base d'un indice prédéterminé. Pour se couvrir contre un aléa climatique, l'agriculteur paie une prime et attend une indemnisation qui n'est pas fonction des pertes réalisées mais fonction de l'évolution de l'indice.

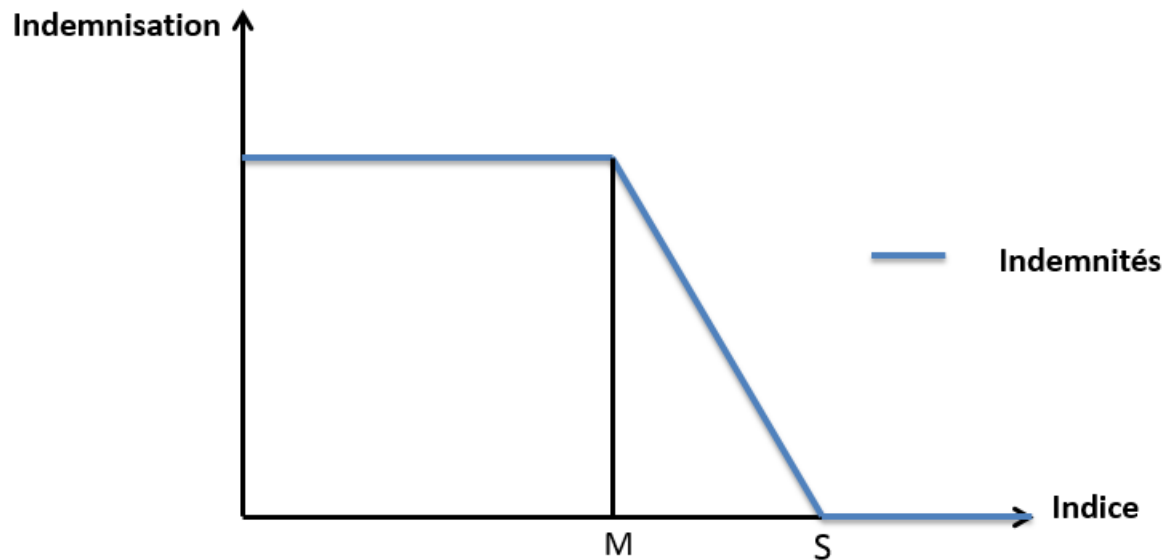


FIGURE 1.3 – Schéma classique d'assurance paramétrique

1.3.5.1 Les Indices

L'étape fondamentale dans la mise en place d'une assurance paramétrique est la construction de l'indice. Puisqu'il n'existe pas d'expertise avant indemnisation, l'indice a pour vocation d'être fortement corrélé aux pertes réellement réalisées, donc d'être clairement identifié, externe, objectif, facilement mesurable et relié à un long historique de données. Parmi les

indices utilisés pour la mise en place d'une assurance paramétrique, on retient :

- **Estimation des rendements** : il s'agit d'estimer sur la base de modèles agronomiques les rendements qui serviront à construire des agrégations qui permettront à leur tour de fixer des seuils déclencheurs des indemnisations. C'est cet indice qui sera utilisé dans cette étude. Ce choix se justifie par le fait que les estimations de rendements disponibles pour l'étude sont issus non seulement d'une combinaison de la plupart des autres indices ci-dessous, mais aussi de modèles agronomiques et météorologiques robustes. Le détail de la construction de cet indice sera présenté dans le chapitre suivant via le modèle CGMS-Maroc et notre modèle de tarification.
- **Le cumul de pluie** : on peut construire un indice de cumul de pluie en ne prenant en compte que les pluies en deçà d'un seuil journalier donné car l'excès d'eau ne peut être capté par la plante et va être éliminé essentiellement par ruissellement. Une autre alternative est de subdiviser l'année en plusieurs phases correspondant aux phases de croissance de la culture assurée car la résistance de la culture au manque de pluie dépend de sa phase de croissance.
- **Le bilan hydrique** : il est établi par comparaison entre les apports et les pertes en eau d'une manière spécifique à une plante et idéalement pour un lieu précis (parcelle, champ ou région). Les apports correspondent aux précipitations et les pertes, les évapotranspirations et ruissellements. Si les apports excèdent les pertes alors il s'agit d'une situation excédentaire en eau, dans le cas contraire il s'agit d'une situation déficitaire.
- **Les indices de stress hydriques** : ils évaluent l'apport réel en eau relativement au maximum potentiellement consommé, et assument une relation entre la croissance de la plante et cette disponibilité réelle de la ressource en eau.
- **Les indices de sécheresse** : ce type d'indice sert à quantifier la sécheresse de l'air sur la base des pluies et des degrés-jours (températures) selon un pas de temps quotidien.
- **Indices de végétation** : Normalised Difference Végétation Index (NDVI), à partir de photographies satellitaires des couverts végétaux, le NDVI mesure l'activité physiologique des végétaux. Il est défini par l'expression : $NDVI = \frac{\rho_2 - \rho_1}{\rho_2 + \rho_1}$ avec ρ_2 : le réflectant proche infrarouge et ρ_1 : le réflectant infrarouge. $NDVI \in [-1, 1]$. Le NDVI

augmente avec la vigueur et la densité végétale, de 0.15 (en moyenne) pour les sols nus à 0.75 (en moyenne) pour les couvers denses.

Aussi à partir d'observation par satellite, on peut construire un autre indice de végétation à savoir : le Leaf Area Index (LAI). Ce indice est tout simplement le ratio de la surface totale recouverte par des feuilles (soit de la végétation) sur la surface totale et est compris entre 0 et 6 (forêt dense).

Par ailleurs, dans la mise en place d'une assurance indicielle, le choix de l'indice ou du paramètre doit se justifier à travers des critères bien spécifiques, l'indice doit donc être :

- Clairement identifié
- Externe et objectif
- Facilement mesurable
- Relié à un long historique de données

L'ensemble de ces indices décrits ci-dessus peut se résumer à travers deux grandes familles d'indices :

Type d'indice	Indice
Indices directs	Rendement
	Revenu
Indices indirects	Pluviométrie, Vent, Gel ...

TABLE 1.1 – Type d'indices en assurance indicielle

Dans le cas d'un indice basé sur le rendement, on réalise généralement une étude sur un panel représentatif de producteurs dans une région donnée, on en déduit un rendement référence dans la région. Les producteurs des zones qui ont un rendement par exemple plus faible que la moyenne des années précédentes sont indemnisés. La démarche serait la même dans le cas du revenu sauf que l'indice n'est pas basé sur le rendement de l'agriculteur mais sur ses revenus. Cependant dans le cas des indices indirects les informations sont obtenues par des relevés climatiques ou données satellite.

1.3.5.2 Le risque de base

Le risque de base est l'entrave fondamentale à la mise en place d'un modèle actuariel d'assurance paramétrique. Il représente les écarts potentiels entre les dommages réels subis par un assuré et les dommages déterminés par l'indice. En pratique c'est le risque d'indemniser un assuré sans sinistre ou de ne pas indemniser un assuré alors qu'il est réellement sinistré. C'est encore l'absence de symétrie certaine entre les indemnités déclenchées et les

pertes effectivement subies. Le risque de base peut se présenter sous plusieurs formes.

Risque de base lié à la modélisation : c'est l'erreur liée soit au traitement des données, soit au choix des variables exogènes.

Le risque de base spatial : il survient lorsque plusieurs zones dépendent de la même station météorologique. Il faut donc installer plusieurs stations à proximité des zones d'étude.

Risque de base temporel : risque dont les conséquences apparaissent tardivement (exemple : maladie attaquant la culture dont les conséquences ne sont pas immédiates).

Risque de base spécifique : il peut résulter d'une mauvaise corrélation entre l'indice retenu et le rendement de la culture.

Après la présentation de ces différents indices, l'étape suivante de cet étude vise dans un premier temps à présenter le modèle qui fournit les données pour cette étude et dans un deuxième temps à présenter une analyse exploratoire de ces données, ce qui conduira à l'élaboration du modèle actuariel de tarification d'assurance paramétrique.

1.4 Référence d'assurance indicielle

L'objectif de cette partie juste avant la présentation des données qui vont servir à la mise en place d'un produit d'assurance paramétrique, est de faire un tour d'horizon de quelques programmes d'assurance indicielle principalement en Éthiopie, au Malawi et en Inde. L'idée dans la présentation de ces programmes est juste de mettre en exergue quelques éléments clés de leur mécanisme à savoir : l'indice, le type d'indice et l'indemnisation ou la non indemnisation du produit.

1.4.1 HARITA en Éthiopie, basée sur des données satellite

C'est un système d'assurance indicielle qui a été lancé dans le nord de l'Éthiopie au cours de la saison 2009. Comme au Maroc, dans cette région, la sécheresse est le principal risque agricole. Ce produit d'assurance indicielle vise essentiellement à couvrir le teff, une céréale éthiopienne, donc de protéger les producteurs d'une perte de revenu liée à une fin précoce de la saison des pluies.

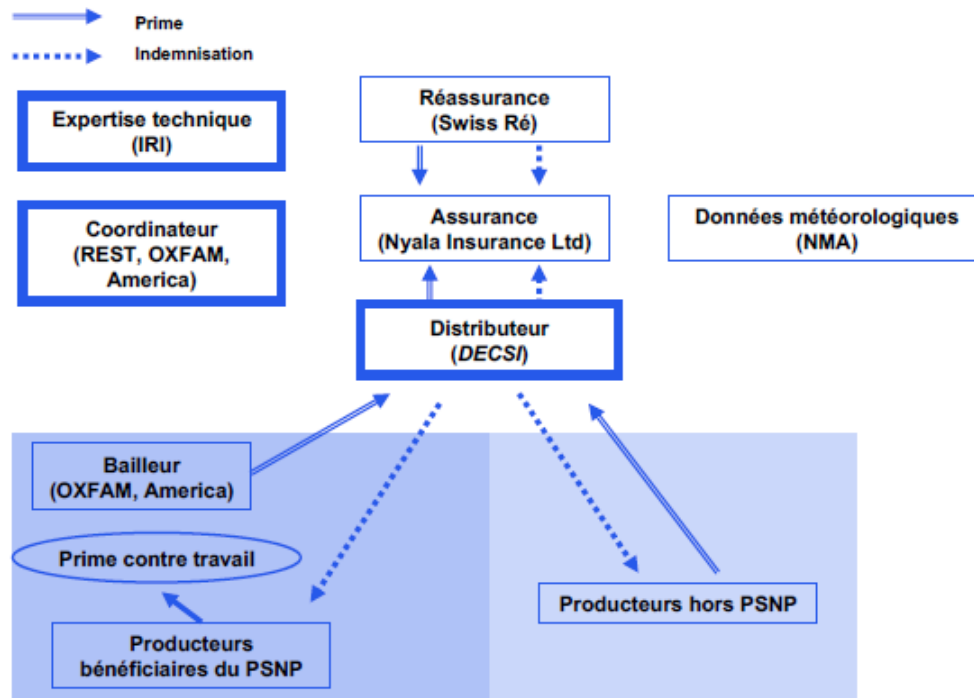


FIGURE 1.4 – Dispositif HARITA

Le Dedebit Credit and Saving Institution (DECSI) collecte les primes d'assurance des producteurs et les transfère à l'assureur du programme, Nyala Insurance Ltd. Cette compagnie d'assurance se réassure ensuite auprès de Swiss Ré. A la fin de la saison, les indemnités sont calculées en utilisant une formule linéaire basée sur la somme des précipitations de chaque décade. Par manque de liquidité et afin d'être assuré, les plus démunis paient leur prime à travers un travail. HARITA n'a pas malheureusement connu de succès à cause du mauvais dimensionnement de l'indice.

1.4.2 L'assurance Alliance One au Malawi

En 2005, une assurance indicielle basée sur la pluviométrie a été développée, destinée aux petits agriculteurs. Ici, les agriculteurs signent un contrat avec Alliance One qui est en fait une société d'export de tabac, contrat dans lequel non seulement les producteurs s'engagent à vendre leur production Alliance One, ensuite, un prix minimum d'achat est imposé par le gouvernement à Alliance One qui s'engage à son tour à garantir un prix supérieur au prix exigé l'Autorité.

1.4.3 NAIS, Assurance indicielle basée sur le rendement en Inde

En Inde, l'assurance indicielle a été lancée en 1999 sur le modèle partenaire-agent avec l'assureur national, Agriculture Insurance Company of India Ltd (AIC). Il s'agit d'une assurance basée sur le rendement. La conception du produit s'articule autour des étapes suivantes :

- L'ensemble du territoire est divisé en zones administratives de superficies similaires, comptant environ 300 000 personnes.
- Dans chaque zone, si le rendement estimé est inférieur à la moyenne du rendement des 5 années précédentes, alors l'ensemble de la zone est indemnisé proportionnellement à la perte de rendement subie. Les rendements sont estimés dans chaque zone par des prélèvements aléatoires de $25m^2$. Ils sont réalisés dans les champs par des fonctionnaires déjà présents sur place.

1.4.4 WBCIS, une assurance indicielle basée sur le climat en Inde

En Inde, le WBCIS est un plan d'assurance indicé sur les rendements. L'indice est conçu en estimant le lien entre les conditions climatiques (pluviométrie, température, humidité, gel, ...) et le rendement. Des stations météorologiques sont installées dans chaque zone et relèvent quotidiennement les données. Si l'indice montre qu'au cours de la saison les conditions ont été néfastes à la production, les producteurs sont indemnisés.

1.4.5 BASIS, une assurance indicielle basée sur la pluviométrie

C'est un programme non subventionné en peine qui propose une assurance paramétrique indicée sur la pluviométrie. Aujourd'hui, le modèle BASIS est un produit unique couvrant tout type de cultures. Il est calculé selon les régions en considérant les cultures principales de chaque zone.

1.4.6 Assurance récolte Sahel, une assurance indicielle basée sur l'évapotranspiration

Allianz Mali commercialise un produit d'assurance paramétrique se basant sur un indice satellitaire : l'indice d'évapotranspiration². Ce produit d'assurance couvre les producteurs de

2. C'est la quantité d'eau transférée vers l'atmosphère, par l'évaporation au niveau du sol et par la transpiration des plantes

maïs contre le risque de sécheresse.

Les données collectées auprès de la Direction Nationale de Météorologie Nationale du Mali et les informations fournies par les coopérations de producteurs de maïs ont fortement contribué à calibrer les paramètres de l'indice, afin d'accroître sa pertinence pour les producteurs.

Pays	Programme	Indice	Type d'indice	Type de programme
Ethiopie	HARITA	Intensité nuageuse	Indirect	Public
Malawi	Alliance One	Pluviométrie	Indirect	Public
Inde	NAIS	Rendement	Direct	Public et fortement subventionné
Inde	WCBIS	Pluviométrie	Indirect	Public avec une subvention de primes
Inde	BASIX	Pluviométrie	Indirect	Privé et sans subvention de primes
Mali	Assurance Récolte Sahel	évapotranspiration	Indirect	Privé

TABLE 1.2 – Quelques expériences en assurance indicielle

1.5 Assurance paramétrique agricole en France

A l'instar de quelques pays d'Afrique et d'Asie, la France aussi utilise depuis quelques années l'assurance indicielle agricole comme couverture contre les risques météorologiques notamment le risque de sécheresse. En France, des contrats d'assurance indicielle couvrent également les « risques localisés » comme par exemple la chute des températures dans des petits vignobles, ce que ne font pas les contrats classiques, faute de rentabilité. Mais la mise au point de ces contrats indiciaires requiert un maillage très dense des stations de météorologie pour collecter les données représentant au mieux le risque des producteurs. Quelques expériences réussies sont déjà notables, l'assurance des prairies de PACIFICA filiale de Crédit Agricole et la couverture météo indicielle du courtier Météo Protect.

1.5.1 L'expérience de PACIFICA

Pour faire face aux perpétuels changements climatiques et surtout se couvrir contre le risque de sécheresse, Pacifica a développé un produit d'assurance indicielle. Il s'agit d'une assurance de prairies, première assurance indicielle en France calculée à partir d'images satellite. L'indice de production fourragère est calculé sur la base de douze années d'archives d'images satellite. Les satellites peuvent ainsi mesurer, à l'échelle d'une commune, la baisse de la pousse par rapport à la moyenne des dernières années. En cas de baisse de production de prairies, une indemnisation est versée à l'éleveur assuré, indemnisation devant lui permettre d'acheter de substitution pour ses animaux. La garantie fonctionne comme suit : l'éleveur peut choisir au moment de s'assurer le capital assuré et le niveau de franchise (de 10 à 50%).

En moyenne, le coût brut sans subvention par hectare assuré (dont la production est évaluée à 850 € la tonne de fourrage) avec 25% à 30% de franchise est d'environ 30 € (environ 10 € avec subvention).

1.5.2 L'expérience de METEO PROTECT

Comme tout produit d'assurance indicielle ou paramètres, la couverture météo indicielle de Meteo Protect est une assurance dont le paiement est déclenché lorsque l'indice franchit une valeur prédéfinie et l'indemnisation peut être fixe quand l'indice météo franchit une certaine valeur ou progressive si la perte causée par la météo augmente en fonction de la valeur de l'indice. L'indice peut être une température, un cumul de précipitations, une durée d'ensoleillement, une vitesse de vent ou toute variable météo ou une combinaison de variables qui matérialisent le risque météo auquel votre activité est exposée. D'autres indices sont disponibles sur demande.

Chapitre 2

Étape Data

2.1 Système CGMS-Maroc et historique de l'assurance indicielle au Maroc

Deux principales études de faisabilité d'assurance indicielle ont déjà été menées au Maroc. La première tentative a été réalisée en 2000 en appui au « Programme sécheresse », suite à un accord entre le Gouvernement et la Banque Mondiale, visait à tester la faisabilité d'une assurance sécheresse indicielle en utilisant des modèles statistiques de prévision des rendements des principales céréales à partir d'un indice pluviométrique, à l'échelle provinciale (17 provinces du pays). L'indice avait été élaboré uniquement à partir de la pluviométrie totale de la saison agricole (entre le mois d'Octobre et le mois de Mars), considérée comme le paramètre influençant le plus le rendement des céréales au Maroc. Malgré les bonnes corrélations obtenues entre le paramètre pluviométrique et les rendements des céréales, ce projet n'a jamais pu être mise en œuvre à cause du risque de base considéré comme trop élevé. Le niveau élevé de ce risque pouvait s'expliquer par non seulement la non disponibilité de données sur une longue durée (21 années de données disponibles), mais aussi la faiblesse du réseau de stations synoptiques (1 station par province).

En 2008, une deuxième étude est réalisée, à l'instar de la première, elle vise à l'étude de la faisabilité d'une assurance sécheresse indicielle (avec pour indice : l'indice de Précipitations Normalisé (SPI)) en vue de prédire les rendements céréaliers dans 4 provinces du Maroc. L'indice SPI permet de trouver les saisons sèches, en calculant l'écart statistique d'une saison par rapport aux saisons antérieures disponibles dans l'historique. Malgré la pertinence de ce nouvel indice dans la quantification de la sécheresse météorologique dans le Royaume, ce deuxième projet a également mis en évidence un risque de base bien trop élevé, et donc une

assurance indicielle difficile à mettre en place. Le risque de base constitue donc la principale entrave dans la mise en place d'une assurance paramétrique dans le Royaume.

Un important travail va être réalisé avec l'INRA et ses collaborateurs à travers la mise en place d'un modèle de prévision des rendements (le modèle CGMS-Maroc) des cultures céréalières, prenant initialement en entrées des données météo, des données sur l'utilisation des parcelles, la richesse des sols et des données de rendement. Ayant pour principal but de fournir des prévisions de rendements aux échelles régionale et nationale, le modèle CGMS-Maroc est fonctionnel actuellement à l'échelle administrative des provinces marocaines et ceci à travers les trois niveaux suivants :

2.1.1 Les inputs

Les données d'entrée du modèle CGMS-Maroc sont :

- **Données météorologiques et agro-météorologiques** ($T^{\circ}\text{min}$, $T^{\circ}\text{max}$, pluviométrie, rayonnement, vitesse du vent mesurée par décade), interpolées spatialement et mises dans une grille de référence couvrant tout le pays. Les données météo sont fournies quotidiennement par la DMN¹. Quant aux données agro-météorologiques (Leaf Area Index, exigence en eau, consommation en eau, évaporation du sol), elles sont calculées par le modèle de simulation WOFOST²
- **Données satellitaires** (indice NDVI) : elles sont fournies de façon décadaire et sont agrégées de façon à être mises à l'échelle de la grille de référence. Pour le système CGMS les images NDVI proviennent de deux sources : le capteur SPOT-VEGETATION qui fournit des images depuis 1998 et le capteur Proba-V (Project for On-Board Autonomy -VEGETATION) de meilleure résolution spatiale (1km x 1km), fournissant des images depuis le 1^{er} Octobre 2014. Ce capteur remplace l'ancien capteur SPOT-VEGETATION.

1. Direction de Météorologie Nationale

2. World Food Studies, modèle de croissance et de développement des cultures ; adapté au Maroc

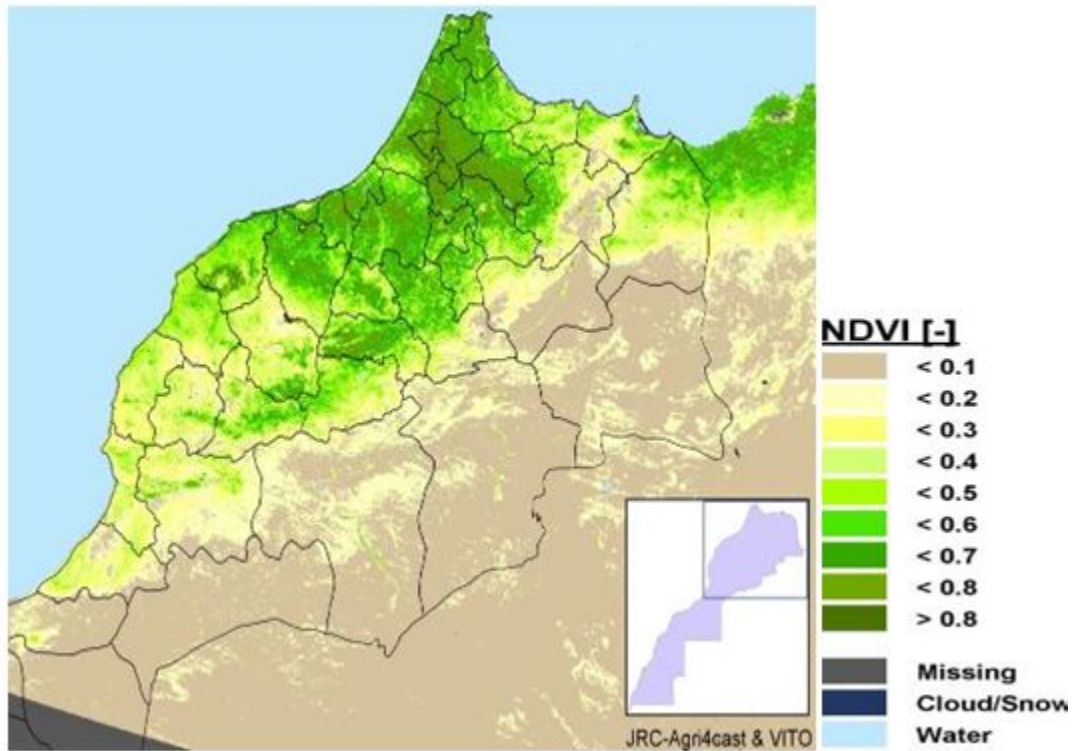


FIGURE 2.1 – Mesure de l'indice NDVI au Maroc

- **Données agricoles** (surfaces et données de rendements) fournies au terme de chaque campagne agricole par la DSS depuis 1980. Il s'agit en effet de statistiques agricoles issus d'enquêtes de terrain réalisé grâce à la méthode d'échantillonnage aréole, cette méthode permet tout simplement d'extrapoler les estimations de superficies et de rendements à partir d'observations de terrain.
- **Données de rendements déclaratifs** : des données de rendements déclaratifs à l'échelle communale issues du Conseil Agricole sont disponibles hormis les données utilisées dans le fonctionnement du CGMS. Il est donc évident que la fiabilité de ces données dépend de la bonne foi de l'agriculteur et la variation de rendements déclaratifs d'une saison à une autre intègre toutes causes (climatiques, sanitaires ou même humaines), alors que le rendement prévu par le modèle CGMS-Maroc n'est calculé que sur des indicateurs objectifs (variation climatique, NDVI,...).

2.1.2 Traitement

Les données météorologiques et agro-météorologiques sont interpolées spatialement et mises dans une maille de référence. En plus des données agricoles (surfaces et rendements),

ces données sont agrégées aux échelles administratives (province et communes).

L'étape suivante consiste à déterminer les indicateurs potentiels qui permettent d'expliquer les rendements des cultures, ces indicateurs sont établis sur chaque phase de croissance de la culture. La dernière étape du traitement consiste à construire le modèle de prévision. Deux méthodes sont utilisées pour estimer les rendements : l'analyse de similarité et le modèle linéaire généralisé. L'analyse de similarité est une méthode destinée à détecter parmi les campagnes passées, celles qui ressemblent le plus à celle qui est en cours. Cette analyse est réalisée sur la base d'indicateurs spécifiques (le cumul de pluviométrie, à partir du 1^{er} septembre de la campagne agricole ; le NDVI moyen à partir de la première décade de la campagne et le taux de satisfaction en eau, à partir du 1^{er} septembre de la campagne agricole). Le modèle linéaire généralisé se construit à travers trois étapes :

Étape 1 : dans un premier temps toutes les combinaisons possibles de 1 à 3 indicateurs potentiels sont testés dans les équations de régression. Les modèles ainsi obtenus sont de la forme :

$$y = b_0 + x_p b_p + \epsilon$$

Avec :

y : le rendement observé

x : les indicateurs de rendement.

b : les coefficients de régression

p : le nombre d'indicateurs (1 à 3)

ϵ : l'erreur résiduelle.

Étape 2 : validation des modèles pour un niveau de risque 5% .

Étape 3 : Pour chacun des modèles retenus à l'étape précédente, une validation croisée est réalisée afin de sélectionner la prévision avec l'erreur moindre.

2.1.3 Outputs :

Le tableau de sortie contient les sorties suivantes : le type de céréale (blé tendre, blé dur, orge).

Le nom de la commune ou de la province.

La superficie emblavée (moyenne des 5 dernières campagnes agricoles).

Le modèle de prévision et le coefficient de détermination R^2 obtenu par régression.

Le nombre de campagnes agricoles utilisées pour élaborer le modèle de prévision.

Un intervalle de confiance du rendement prévu à 95%.

Le rendement prévu.

La production³ totale de la céréale par limite administrative.



FIGURE 2.2 – Dispositif CGMS-Maroc

2.2 Les drivers de risques dans les données

2.2.1 Risque de choc

Dans la réforme solvabilité 2, la prise en compte du facteur choc dans les risques de nature humaine, techniques ou financières et les catastrophes naturelles est fondamentale. La considération du facteur choc dans cette étude va se traduire par la prise en compte des

3. $\text{Production} = \text{Rendement} \times \text{Superficie moyenne}$

événements extrêmes dans le modèle de prévision de données.

En effet il convient d'une part de s'assurer de l'intégration des scénarios de choc dans le modèle de prévision qui fournira les données qui serviront à calibrer le modèle de tarification ou à tester l'effet de scénarios de baisses considérables de rendements sur le coût de l'assurance.

2.2.2 Risque de volatilité et de tendance

La volatilité et la tendance sont des vecteurs directeurs dans la démarche de modélisation, ces deux risques résident ici dans la taille de l'échantillon utilisé. Nous tiendrons aussi compte de ces deux risques dans notre approche de modélisation.

2.3 Data Mining

2.3.1 Démarche de modélisation :

Unité statistique : communes / agriculteurs

Variables cibles : rendements observés et rendements estimés.

Population cibles : toutes les provinces du royaume.

Technique de modélisation : clustering, GEV, modèle area-yield insurance, Monte-Carlo.

Données disponibles : historique de données sur 14 années.

Données fiables : données de l'INRA

Données à jour : données de la campagne agricole 2015-2016.

Agrégation des données : l'agrégation se fait à l'échelle communale.

Vérification de la cohérence des données : analyse des erreurs moyennes et écarts relatifs par céréale.

Filtrage : gestion des valeurs non disponibles, des valeurs extrêmes et des valeurs nulles.

Mise en œuvre des algorithmes de modèle.

Drivers de risque dans les données : tendance, volatilité, choc, et niveau.

2.3.2 Point sur l'existant et choix de la modélisation

Le but de cette partie est d'analyser les données disponibles pour la mise en place du modèle. En vue d'avoir une bonne lisibilité des inputs nous procédons aux calculs des écarts entre les rendements observés et les rendement estimés fournis par le modèle CGMS. Nous disposons de trois bases de données. Pour chaque commune⁴ nous disposons des informations suivantes :

Rendement observé : ce rendement est obtenu par extrapolation des rendements réels constatés sur des points d'observation sur le terrain.

Rendement estimé : le rendement estimé par le modèle CGMS par commune pour chaque céréale par année.

Céréale : il s'agit des trois principales céréales cultivées dans le royaume : blé dur, blé tendre et orge.

Superficie : la surface (en hectare) agrégée pour chaque céréale par commune et par année.

2.3.3 Mesure d'erreur d'estimation du modèle

La fiabilité d'un modèle de tarification réside d'abord dans la granularité des données. Afin d'y parvenir, nous analysons les écarts entre les rendements observés et estimés, données sur lesquels notre modèle sera construit.

Une assurance paramétrique basée sur le rendement est-elle faisable au vu du niveau d'erreurs qui seront évaluées ci-dessous ? nous ne pourrions l'affirmer qu'avec du recul étant donné que c'est le premier produit en son genre ; néanmoins l'évaluation du risque de base nous permettra d'avoir une vision plus claire sur le niveau d'erreur de la modélisation.

2.3.3.1 Erreurs moyennes par céréale et taux de commune par seuil d'erreur relative

Sur les trois bases qui correspondent respectivement aux données arrêtées à la fin des campagnes 2014, 2015 et 2016, nous évaluons les erreurs moyennes dont la méthode de calcul est détaillée en annexe.1.

L'erreur d'estimation du modèle a augmenté au fil des années, de 18% en 2014, nous arrivons à 22% en 2016.

4. données disponibles pour 973 communes

<i>Années</i>	<i>Blé dur</i>	<i>Blé tendre</i>	<i>Orge</i>	<i>Totale</i>
2014	20%	18%	18%	18%
2015	21%	19%	19%	20%
2016	23%	21%	21%	22%

TABLE 2.1 – Erreurs moyennes par céréale

Le tableau 2.1 ci-dessous montre que la tendance du pouvoir d'estimation des rendements est baissière jusqu'en 2015 ; le lecteur comprendra que le modèle paraît plus précis pour l'orge que pour les autres céréales et que le niveau d'erreur pour le blé dur toujours supérieure pour chaque année se répercutera sûrement sur le niveau du coût de sinistre pour cette céréale.

Nous constatons que les erreurs moyennes en 2016 sont plus importantes qu'elles ne le sont pour les années juste précédentes, tendant à retrouver le niveau qu'elles avaient il y a 16 ans. Ce rebondissement pourrait s'expliquer par le fait que la production a été très faible en 2016 avec des rendements faibles, l'erreur relative se retrouve donc très élevée même si l'erreur absolue est presque égale à celle des années passées.

En considérant la tendance d'évolution des erreurs entre 2000 et 2015, nous pouvons affirmer avec du recul que cette tendance sera encore baissière les années à venir, ce qui risque de diminuer considérablement le risque de base.

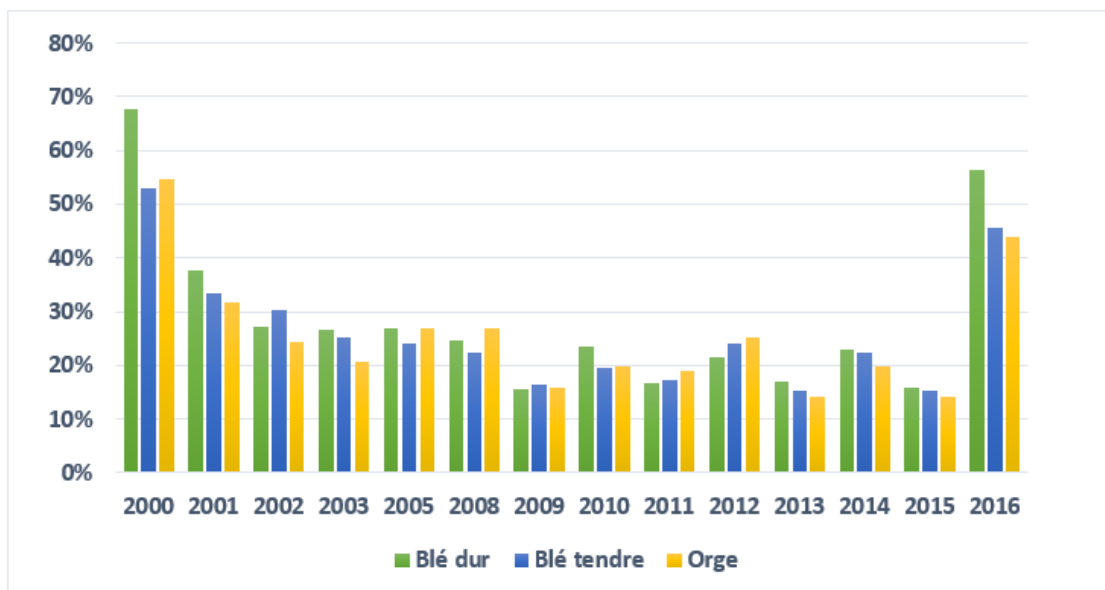


FIGURE 2.3 – Évolution des erreurs moyennes

En observant la figure 2.4, il apparaît plus clairement que pour l'année 2016, il existe globalement une surestimation par le modèle CGMS vue que ces trois graphes représentent

les écarts entre les rendements estimés et les rendements observés. Cela devrait logiquement influencer le risque de base dans la mesure où une très grande surestimation impliquerait a priori une faible sinistralité du portefeuille, traduisant donc un risque de base élevé : risque de ne pas indemniser des assurés réellement sinistrés.

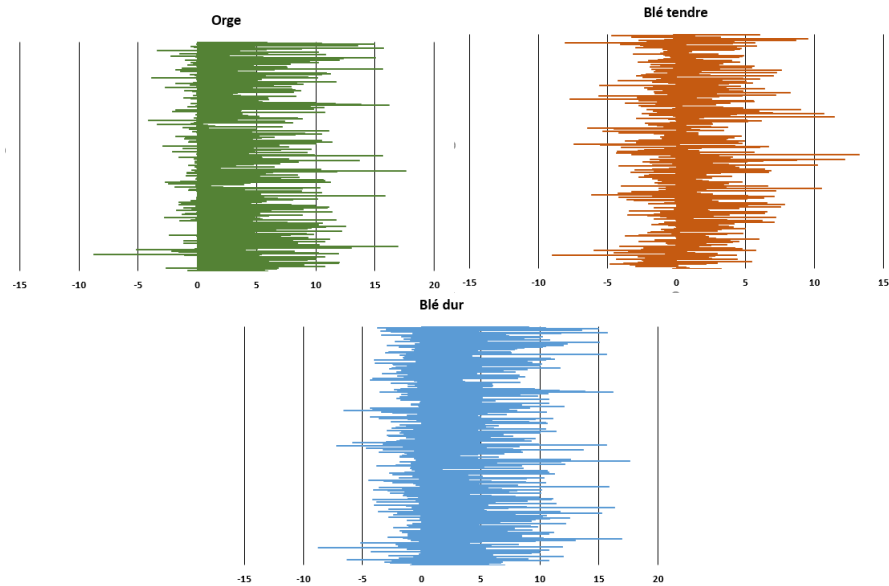


FIGURE 2.4 – Ecart entre les rendements estimés et observés en 2016

Toujours dans le souci d'évaluer la qualité d'estimation du modèle CGMS-Maroc, nous mesurons la part de communes pour chaque céréale pour l'année 2016 par classe d'erreur relative. En fait, cette évaluation vise essentiellement à suivre la variation des rendements estimés par rapport aux rendements de référence (les rendements observés). Une lecture rapide du tableau 2.2 laisse comprendre au lecteur que pour chacune des céréales, plus de 50% des communes présentent une erreur relative dépassant 100%.

	<i>Blé dur</i>	<i>Blé tendre</i>	<i>Orge</i>
$\leq 10\%$	11%	12%	14%
<i>Entre 10% et 50%</i>	22%	28%	30%
<i>Entre 100% et 150%</i>	33%	31%	28%
$\geq 150\%$	34%	29%	27%

TABLE 2.2 – Part de communes par seuil d'erreur relative

Pour la suite du projet et pour le modèle de tarification, nous utiliserons les données arrêtées sur la campagne agricole 2016. Aussi il est fondamental de s'interroger à ce niveau sur l'option de classification à choisir. À cet effet deux options sont envisageables :

1. Déterminer des trois céréales, une céréale de référence et tarifier uniquement sur la base de cette dernière. L'idée ici serait donc d'observer comme l'indique la figure 2.5 l'évolution des superficies exploitées pour les trois céréales et d'en retenir un profil représentatif.

2. Pour chaque céréale, construire des classes homogènes de communes et proposer ensuite un tarif pour le risque sécheresse dans le royaume.

La deuxième option semble la plus plausible pour deux raisons : (i) ne disposant d'information par céréale pour chaque commune sur une durée maximale de **14 années**, l'estimation de l'évolution des rendements/productions pour chaque céréale en vue d'en retenir un profil d'évolution unique serait biaisée. (ii) même si pour chaque céréale, la classification de communes en **K classes** homogènes donne lieu à **K modélisations** ; cela permettrait d'avoir des échantillons d'observations de plus grande taille. Ainsi les méthodes d'estimation seraient plus robustes (si un regroupement : classe, étant homogène, regroupe 200 communes par exemple, nous aurons $200 \times 14 = 2800$ observations issues d'une même loi de distribution, ce qui).

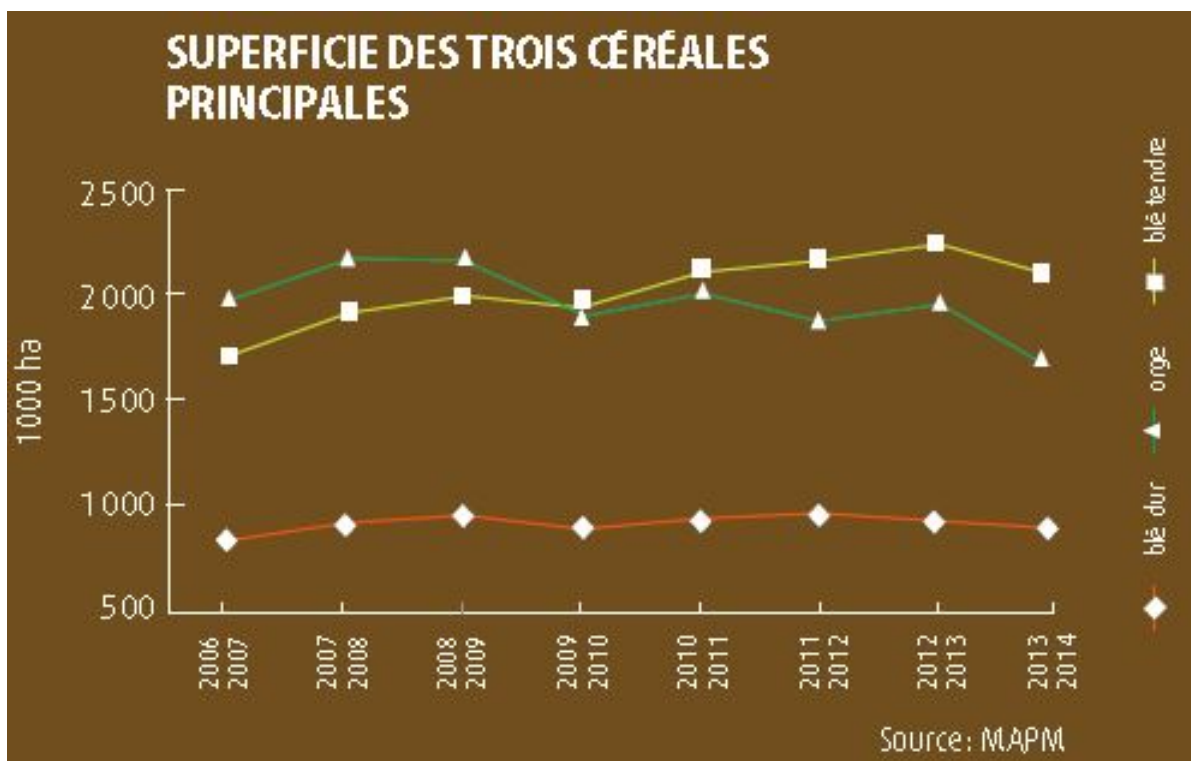


FIGURE 2.5 – Superficies exploitées pour les trois principales céréales

Il est certes évident que dans une commune donnée les rendements de céréales sont fortement corrélés aux conditions climatiques, si les conditions climatiques ne sont pas favorables

pour le blé tendre par exemple dans une commune au cours d'une saison, les conditions climatiques seront aussi défavorable pour les deux autres céréales.

L'objectif est donc de regrouper dans une même classe les communes dont les rendements réagissent de la même façon aux aléas climatiques ; cependant, deux communes peuvent avoir une sensibilité climatique identique mais à partir de rendements qui, dans l'absolu sont différents. L'idée n'est donc pas de regrouper les communes qui auraient des rendements similaires, mais de regrouper les communes dont les rendements ont des réactions similaires aux aléas du climatiques.

Chapitre 3

Évaluation et gestion du risque de base

3.1 Assurance indicielle et approche microéconomique

L'approche « micro » du risque de base consiste à évaluer à l'échelle « agriculteur » les probabilités de perte sans indemnisation ou d'indemnisation en l'absence de sinistre. En effet la consistance de l'évaluation du risque de base « micro » se trouve dans la mesure des fonctions d'utilité des agriculteurs sur le produit d'assurance paramétrique. Dans un souci de minimisation du risque de base « micro » il est envisageable de comparer les pertes réelles des agriculteurs aux statistiques communales sur lesquelles s'élabore notre modèle. Cette comparaison s'avérant la plus objectif, il devient alors clair que la bonne quantification du risque de base micro permettrait un choix optimal des seuils de déclenchement et d'indemnisation de sinistre. Le modèle d'approche micro du risque de base se présente comme suit :

Soit p la probabilité de subir une perte sur son rendement agricole.

Soit Per la perte de l'agriculteur, il est averse au risque si sa fonction d'utilité u est concave.

Soit q la probabilité d'être indemnisé sans avoir subi des pertes et r la probabilité de subir des pertes sans être indemnisé.

La variable *Indice* représentant l'indice est à deux composantes : $Indice(q, r)$, *Indice* vaut 0 si le seuil déclencheur est dépassé, et 1 s'il n'est pas atteint ; et dans la suite de la modélisation du risque de base micro les quatre cas de figure suivants sont envisageables :

1. il n'y a pas de perte et il n'y a pas eu d'indemnisation ($Perte = 0 \& Indice = 0$)
2. il n'y a pas de perte et il n'y a eu d'indemnisation ($Perte = 0 \& Indice = 1$) = q
3. il y a eu de perte et l'indemnisation a eu lieu ($Perte = Per \& Idice = 1$)
4. il y eu perte mais pas d'indemnisation ($Perte = Per \& Indice = 0$) = r

Nous pouvons obtenir les probabilités des cas 1 et cas 3 comme suit :

$$P(Perte = Per) = p = P(Perte = Per \& Indice = 0) + P(Perte = Per \& Indice = 1)$$

$$\text{Donc, } P(Perte = Per \& Indice = 1) = P(Perte = Per) - P(Perte = Per \& Indice = 0)$$

$$\text{Donc, } P(Perte = Per \& Indice = 1) = p - r$$

D'où le résultat.

$$P(Perte = 0) = 1 - p = P(Perte = 0 \& Indice = 0) + P(Perte = 0 \& Indice = 1)$$

$$\text{Donc, } P(Perte = 0 \& Indice = 0) = P(Perte = 0) - P(Perte = 0 \& Indice = 1)$$

$$\text{Donc, } P(Perte = 0 \& Indice = 0) = 1 - p - q$$

D'où le résultat.

On déduit alors tout simplement les deux égalités suivantes :

$$\begin{aligned} P(Indice = 0) &= P(Perte = 0 \& Indim = 0) + P(Perte = Per \& Indice = 0) \\ &= 1 - p - q + r \end{aligned}$$

$$\text{Et, } P(Indim = 1) = 1 - P(Indice = 0) = p + q - r$$

La table de probabilité de ces quatre évènements se présente donc comme suit :

	$Indice = 0$	$Indice = 1$
$Perte = 0$	$1 - p - q$	q
$Perte = Per$	r	$p - r$

TABLE 3.1 – Tableau de contingence du risque de base micro

Soit R_I la richesse initiale d'un agriculteur qui désire maximiser son utilité en souscrivant au produit d'assurance paramétrique. On considère que l'agriculteur souscrit à l'assurance paramétrique pour une prime de valeur s avec un montant assuré M au moment de la souscription ; l'agriculteur devrait déboursier un montant R d'une valeur : $Ms(q + p - r)$. L'utilité

V_u espérée par l'agriculteur dont l'utilité est u se traduit donc par l'équation :

$$V_u(\text{Indemnisation}, M, S) = (1 - q - p)u(R_I - R) + ru(R_I - \text{Per} - R) + qu(R_I + M - R) + (p - r)u(R_I - \text{Per} + M - R)$$

L'objet de notre étude n'est pas d'évaluer tous ces paramètres (p , q , r), de l'équation ci-dessus, autrement de maximiser l'utilité V_u espérée, car la résoudre nous ramènerait à collecter des données auprès des agriculteurs (pertes réelles), ce qui nous renvoie au risque d'anti sélection qui est l'une des entraves principales de l'assurance classique que nous voulons contourner en optant pour un modèle d'assurance paramétrique. Néanmoins, pour notre modèle, nous évaluons le risque de base à l'échelle communale et il serait plus intéressant après le lancement de ce produit d'assurance paramétrique de mesurer le risque de bas « micro » en vue d'améliorer la viabilité des contrats.

3.2 Approche « macro » du risque de base

Comme présenté dans le chapitre précédent, nous rappelons que si l'approche par un indice pour assurer un risque climatique paraît une excellente option (réduction des frais de gestion, accessible aux agriculteurs à faibles revenus, etc.), le risque de base reste le revers de la médaille. Ainsi nous avons évalué sur la base des données disponibles le niveau de décalage des prévisions du modèle CGMS-Maroc au vu de la « réalité ».

Afin de mesurer l'ampleur du risque de base « macro », nous disposons des rendements observés et des rendements estimés par le modèle CGMS-Maroc pour chaque commune sur l'historique. La démarche du calcul du risque de base se décline en deux étapes.

Étape 1 : cette étape consiste pour chaque année de l'historique à calculer les indemnités de chaque commune en considérant les rendements observés réellement. Ce premier calcul se fait donc pour les communes réellement sinistrées.

Étape 2 : à ce niveau, pour chaque année de l'historique nous calculons les indemnités avec le rendement estimé par le CGMS. Cette étape fournit les communes déclarées sinistrées et indemnisées sur la base des rendements estimés.

Si le risque de base est non nul, il y aura certaines communes réellement sinistrées mais non indemnisées sur la base du rendement estimé et inversement, des communes indemnisées sur la base du rendement estimé alors qu'elles ne sont pas réellement sinistrées.

Soient :

Snt : une variable binaire telle que : $Snt = 1$ s'il y a sinistre et $Snt = 0$ sinon.

I une autre variable binaire telle que : l'indemnisation : $I = 0$, s'il n'y pas d'indemnisation et $I = 1$ sinon.

	$Snt = 1$	$Snt = 0$
$I = 0$	a	b
$I = 1$	c	d

TABLE 3.2 – Tableau de contingence du risque macro

Comme l'indique le tableau ci-dessus, quatre cas sont possibles :

Cas 1 : Pas de sinistre et pas d'indemnisation des communes. C'est la proportion b , elle représente la proportion de communes déclarées non sinistrées sur la base de rendements observés et de rendements estimés

Cas 2 : Communes sinistrées mais sans indemnisation. C'est la proportion a , elle représente la part de communes déclarées sinistrées sur la base de rendements observés, mais non sinistrées sur la base de rendements estimés

Cas 3 : Pas de sinistre mais indemnisation des communes. C'est la proportion d , elle représente la part de communes déclarées non sinistrées sur la base de rendements observés, mais sinistrées sur la base de rendements estimés

Cas 4 : Communes sinistrées et indemnisées. C'est la proportion c , elle représente la part de communes déclarées sinistrées sur la base de rendements observés, et déclarées également sinistrées sur base la de rendements estimés

Le cas 4 est celui qui nous intéresse plus, étant plus déterminant, c'est donc sur ce cas que se focalisera notre analyse du risque de base. L'assurance paramétrique agricole ici étant basée sur le rendement, l'indice est un rendement de référence. L'indice construit doit refléter le mieux un niveau équitable pour chaque commune. L'idée est donc de se référer à une statistique (notamment un quantile) des rendements observés sur l'historique disponible, c'est à partir de cette statistique (rendement de référence) que seront déclenchés les sinistres et donc les indemnisations. Pour notre modèle, nous testons la viabilité du contrat sur la base de trois scénarios :

- **Scénario 1** : le rendement de référence est égal à la moyenne historique des rendements observés.
- **Scénario 2** : le rendement de référence est égal au 3^{eme} quartile des rendements observés.
- **Scénario 3** : le rendement de référence est égal au 95^{eme} centile des rendements observés.

Nous ne saurions mesurer le risque de base sans les fonctions d'indemnisation, d'où nous nous basons sur deux fonctions d'indemnisation ; une première avec un seul niveau de déclenchement et un niveau d'indemnisation ; et une seconde fonction avec trois seuils de déclenchement et trois seuils d'indemnisation. Avant de présenter les fonctions d'indemnisation, il convient de définir les deux concepts suivants :

Définition 1 : Le seuil d'indemnisation

C'est le niveau de rendement en dessous duquel il y a déclaration de sinistre. En clair, c'est un niveau de rendement qui sert de repère pour indemniser. Pour notre projet, nous testons les trois seuils tels que décrit ci-dessous.

Définition 2 : Le montant d'indemnisation

Le montant d'indemnisation représente la charge de l'assureur envers tous les assurés d'une commune sinistrée.

Seuil d'indemnisation ζ	ζ_1	ζ_2	ζ_3
Montant d'indemnisation ρ	ρ_1	ρ_2	ρ_3

TABLE 3.3 – Seuils et montants d'indemnisation

Avec :

$$0 < \zeta_1 < \zeta_2 < \zeta_3 < 1$$

et

$$1 > \rho_1 > \rho_2 > \rho_3 > 0$$

D'après le tableau précédent, si le rendement estimé au cours de l'année est inférieur à ζ_1 du rendement de référence ; l'agriculteur est indemnisé à ρ_1 du rendement de référence sinon

si le rendement estimé de l'année en cours est inférieur à ζ_2 , l'agriculteur est indemnisé à ρ_2 du rendement de référence ; sinon si le rendement estimé de l'année en cours est inférieur à ζ_3 , l'agriculteur est indemnisé à ρ_3 du rendement de référence.

Pour notre produit, le calcul du risque de base est le vecteur directeur de la fixation du seuil d'indemnisation.

Définition 3 : La fonction d'indemnisation

C'est grâce à elle que nous pourrons passer des rendements aux pertes financières potentielles.

Si on note $Nbcom$: le nombre de communes.

Les deux fonctions d'indemnisation se présentent comme suit :

Fonction d'indemnisation 1 :

L'indemnisation est déclenchée (il y a sinistre) si le rendement estimé est inférieur à ζ du rendement de référence.

$$\text{Indemnisation}_{i,j} = \sum_{i=1}^3 \sum_{j=1}^{Nbcom} p_i \times \text{superficie}_{i,j} \times \max(0; \zeta \text{ du rendement de référence}_{i,j} - \text{rendement estimé}_{i,j}) \times \rho$$

Fonction d'indemnisation 2 :

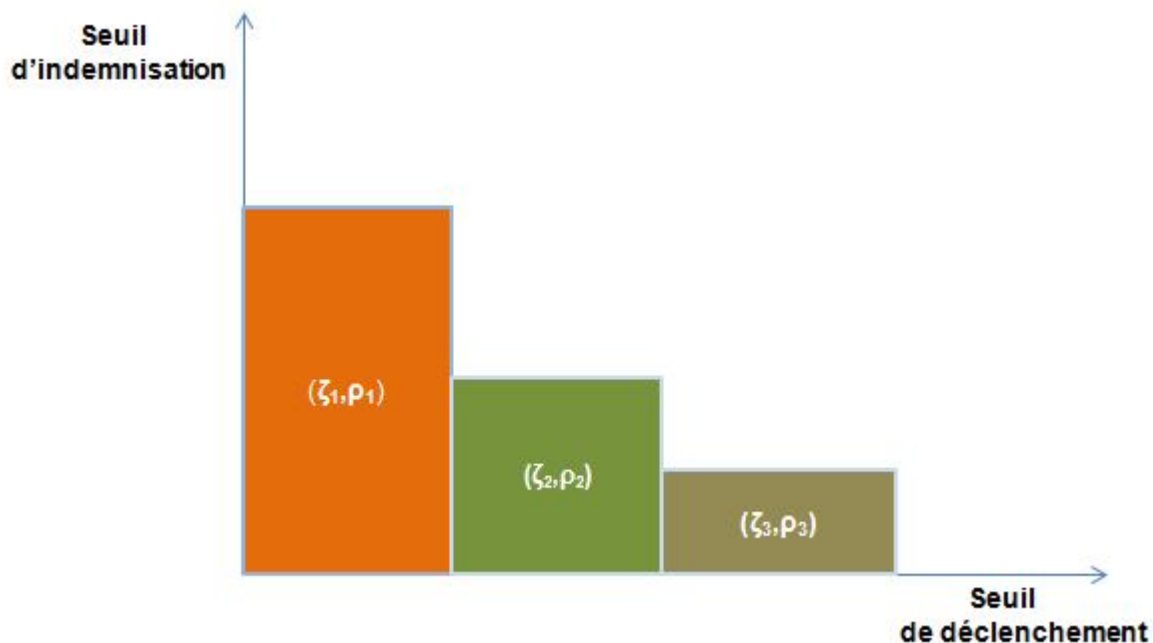


FIGURE 3.1 – fonction d'indemnisation 2

Soient :

- p_i : c'est le prix de la céréale i
- ρ et ς sont définis dans la définition 2 ci-dessus.
- $\text{surface}_{i,j}$: c'est la surface assurée en ha pour la culture i dans la commune j
- ς forfaitaire du rendement de référence $_{i,j}$: c'est le pourcentage forfaitaire accordé selon la garantie.
- rendement estimé $_{i,j}$: c'est le rendement estimé en qx/ha par le modèle CGMS-Maroc pour la culture i dans la commune j .

Alors, l'indemnisation est déclenchée (il y a sinistre) si le rendement estimé est inférieur à ς du rendement de référence. Notons qu'ici, ς correspond à différents seuils comme l'indique la figure 3.1 et on a :

$$\text{Indemnisation}_{i,j} = \sum_i^3 \sum_j^{\text{Nbcom}} p_i \times \text{superficie}_{i,j} \times \varsigma \times \text{forfaitaire du rendement de référence}_{i,j} \times \rho$$

3.2.1 Évaluation du risque de base à partir de la fonction d'indemnisation 1

Sur base de la fonction d'indemnisation 1 décrite pour $\varsigma = 60\%$ et $\rho = 100\%$, nous obtenons les résultats suivants : les proportions de sinistre à partir de la fonction d'indemnisation 1, sur base du scénario 1 ; 21% des cultures sinistrées n'auraient pas été indemnisées. En tenant compte des valeurs extrêmes, on remarque que le risque de base diminue systématiquement d'après le scénario 3, seuls 12% de cultures sinistrées ne sont pas indemnisées. Cela implique qu'en prenant en compte un maximum de cultures, le risque de base s'améliore nettement du scénario 1 au scénario 3, la probabilité d'indemniser les cultures réellement sinistrées augmente de 11%.

Dans le souci de réduction du risque de base, il serait donc optimal de s'en tenir au scénario 3 pour indemniser. Nous allons toutefois mener la modélisation selon les trois niveaux de référence.

	Scénario 1	Scénario 2	Scénario 3
	$Snt = 1$	$Snt = 1$	$Snt = 1$
$I = 0$	21%	18%	12%

TABLE 3.4 – Risque de base à partir de la fonction d'indemnisation 1

3.2.2 Évaluation du risque de base à partir de la fonction d'indemnisation 2

Sur base de la fonction d'indemnisation 2 décrite ci-dessus, nous évaluons le risque de base avec les niveaux suivants :

seuil de déclenchement ς	10%	30%	50%
niveau d'indemnisation ρ	50%	30%	10%

TABLE 3.5 – Paramètres de la fonction d'indemnisation 2

	Scénario 1	Scénario 2	Scénario 3
	$Snt = 1$	$Snt = 1$	$Snt = 1$
$I = 0$	24%	19%	13%

TABLE 3.6 – Risque de base à partir de la fonction d'indemnisation 2

Le risque de base diminue avec le niveau du rendement de référence, ici le taux de « bons classements » ($Snt = 1$, $I = 1$) progresse de 76% à 87% du scénario 1 au scénario 3. Cette analyse nous permet d'affirmer que la fonction d'indemnisation 1 semble être mieux adaptée dans un souci de minimisation du risque de base. Même si un taux de sinistres non indemnisés variant entre 12% et 24% n'est pas assez faible pour soutenir l'hypothèse de robustesse de nos deux fonctions d'indemnisation, ce niveau reste néanmoins assez satisfaisant pour mettre en place un modèle de tarification d'assurance indiciaire basée sur le rendement.

De toute l'analyse précédente, nous pouvons présenter ces proportions de sinistres indemnisés pour les deux fonctions conjointement, cette analyse est décrite à travers la figure 3.2

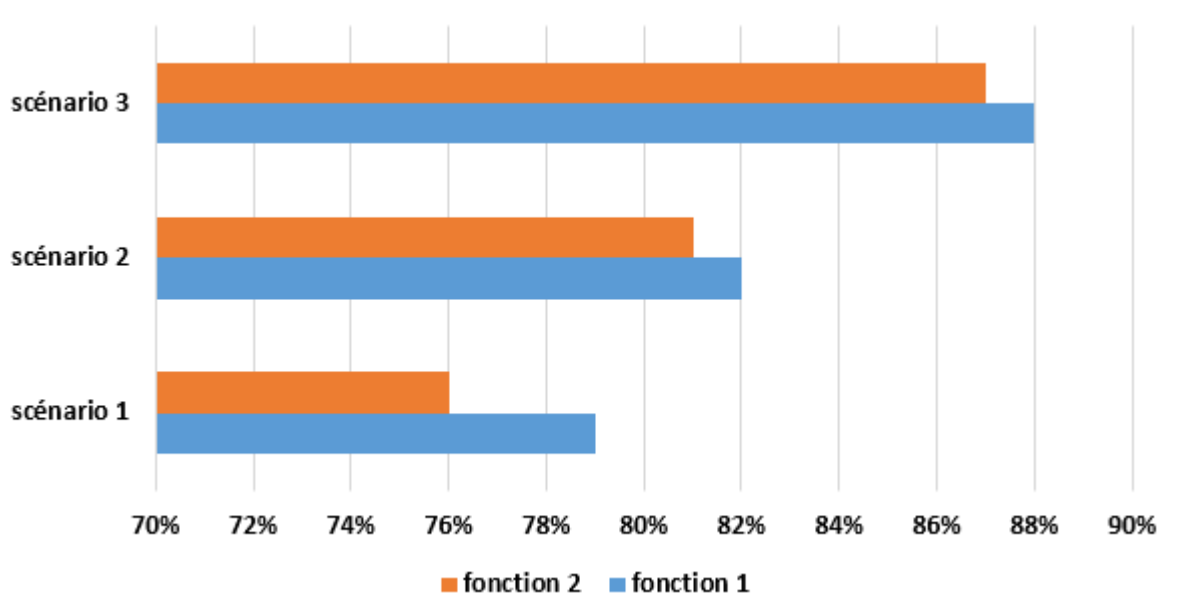


FIGURE 3.2 – Taux de sinistres indemnisés

En se basant sur la fonction d'indemnisation 1, le taux de sinistres indemnisés varie de 79% à 88% du scénario 1 au scénario 3. Comme noté plus haut, ce taux est moins performant sur la base de la fonction 2 d'autant plus que sa variation entre le scénario 1 et le scénario 3 est de 11%. Ces taux de sinistres indemnisés confirment tout simplement que pour la mise en place d'un produit viable d'assurance paramétrique basé sur le rendement même s'il est évident que l'on ne saurait réduire le risque de base à 0, il paraît plus optimal de se référer aux quantiles supérieurs, se référer aux quantiles supérieurs en ce sens où la modélisation permettrait de mieux estimer le coût d'une sécheresse avec les quantiles supérieurs.

Par ailleurs, il est clair que quelque soit le scénario et la fonction d'indemnisation à adopter, le risque de base reste significative (plus de 10% pour la mise en place d'une assurance indiciaire), ce qui mettrait en péril la viabilité d'un produit d'assurance paramétrique. Il est donc souhaitable que le risque de base a priori tourne autour du taux de chargement (entre 0% et 10%).

Chapitre 4

Modèle actuariel d'assurance indicielle

4.1 Rappel des modèles d'assurance dommage et de catastrophe naturelle

Définition 1 : Évènement rare

Le caractère exceptionnel d'un évènement se détermine au travers de sa fréquence : un évènement qui se produit une fois par siècle est ainsi associé à une probabilité de survenance dans l'année de 1 %. Si on considère un évènement dont la probabilité d'occurrence est p dans un espace de temps discret, la distribution de l'instant de première occurrence est modélisée par une loi géométrique $geo(1 - p)$ tel que $P(X = k) = p(1 - p)^{k-1}$, le temps moyen de la première occurrence de cet évènement est donné par $E(X) = \frac{1}{p}$.

Théorème 1 : Loi forte des grands nombres (LFGN)

Soit $(X_i, i \geq 1)$ une suite de variables aléatoires *iid* tel que $E(|X_1|) < +\infty$. Alors :

$$\lim_{n \rightarrow \infty} \frac{1}{n} (X_1 + X_2 + \dots + X_n) = E(X_1) \text{ p.s.}$$

Théorème 2 : Théorème limite centrale (TLC)

Soit $(X_i, i \geq 1)$ une suite de variables aléatoires *iid* telles que :

$$E(|X_1|^2) < +\infty, \text{ et } V(X_1) = \sigma^2 :$$

Si on pose :

$$\varepsilon_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$$

Alors : $\frac{\sqrt{n}}{\sigma}\varepsilon_n$ converge en loi vers G où G est une gaussienne de moyenne nulle et de variance 1.

Théorème 3 : Théorème de Berry-Essen¹

Soit $(X_i, i \geq 1)$, une suite de variables aléatoire *iid* de moyenne nulle et d'écart type σ et on suppose que :

$$E(|X_1|^3) < +\infty$$

Soit F_N :

$$F_N(x) = P\left(\frac{\sum_{j=1}^N X_j}{\sigma\sqrt{N}} \leq x\right).$$

et ϕ la fonction de distribution d'une gaussienne de moyenne nulle et d'écart type σ :

$$\phi(x) = \left(\frac{1}{\sqrt{2\pi}}\right) \int_{-\infty}^x \exp\left(-\frac{u^2}{2}\right) du$$

Alors on a :

$$\begin{aligned} \epsilon_N &= \sup_{x \in \mathbf{R}} |F_N(x) - \phi(x)| \\ &\leq \frac{CE(|X_1|^3)}{\sigma^3\sqrt{N}} \end{aligned}$$

Où $0.398 \leq C \leq 0.8$

Le théorème de Berry-Essen a pour but de maximiser et de préciser l'erreur commise en substituant à la fonction de répartition F_N la fonction ϕ .

1. Andrew Berry en 1941 et Carl-Gustav Esséen en 1942

4.1.1 Modèle fréquence-coût

On considère un risque ou un groupe de risques, pour un portefeuille d'assurance, et K sinistres survenant sur une période donnée (une année par exemple). Le montant total X de sinistre pour cette période est :

$X = \sum_{i=1}^K Z_i$ où Z_i correspond au coût du i^{eme} sinistre survenu pendant la période observée.

Le coût du risque qui correspond à la prime pure est obtenu par l'évaluation de l'espérance des sinistre, *i.e.* $E(X)$.

Pour des montants de sinistre Z_1, Z_2, \dots indépendants et de même loi, et indépendants de nombre K , on a :

$$\begin{aligned} E(X) &= E\left(\sum_{i=1}^K Z_i\right) \\ &= E\left(E\left(\sum_{i=1}^{K=k} Z_i\right) / K = k\right) \\ &= E(K)E(Z) \end{aligned}$$

L'étape fondamentale de cette méthode est d'estimer les deux composantes de cette prime pure : la fréquence du sinistre qui correspond à $E(K)$ et le coût moyen qui correspond à $E(Z)$. Puisque K et Z sont aléatoires, évaluer ces deux quantités revient à déterminer les lois de distribution de K et de Z . La loi de fréquence, celle de K est le plus souvent une loi de Poisson ou une loi binomiale négative.

Pour une loi de Poisson la probabilité d'avoir k sinistres est :

$$P(K = k) = \frac{e^{-\lambda} \lambda^k}{k!}.$$

Et dans ce cas :

$$E(K) = \lambda = V(K).$$

Pour une loi binomiale négative avec pour paramètres p et m la probabilité d'avoir k sinistres est :

$$P(K = k) = C_{m+k-1}^{k-1} p^m (1-p)^k.$$

Et dans ce cas :

$$E(K) = m \frac{1-p}{p}$$

et

$$V(K) = m \frac{1-p}{p^2}$$

Si la loi de distribution des coûts correspond à celle de : Weibull, Log Normal, Pareto et Pareto généralisé, alors à partir de la fonction de répartition $F(z) = \int_{-\infty}^z f(z)dz$ et pour une troncature de niveau u , tel que : $\xi(z) = \frac{F(z)-F(u)}{1-F(u)}$, les fonctions de répartition tronquées de ces quatre lois sont données dans le tableau ci-dessous.

Lois	$\xi(z)$
Log-normale	$\frac{\phi(\frac{\ln z - m}{\sigma}) - \phi(\frac{\ln u - m}{\sigma})}{1 - \phi(\frac{\ln u - m}{\sigma})}$ avec $\phi(z) = (\frac{1}{\sqrt{2\pi}}) \int_{-\infty}^z \exp(-\frac{u^2}{2}) du$
Weibull	$1 - \frac{\exp(-\frac{z}{a})^\alpha}{\exp(-\frac{u}{a})^\alpha}$
Pareto	$1 - (\frac{u}{z})^\alpha$
Pareto généralisé	$1 - (\frac{1+\alpha \frac{z}{a}}{1+\alpha \frac{u}{a}})^{-\frac{1}{\alpha}}$

TABLE 4.1 – Caractéristiques de quelques lois usuelles

4.1.2 Titrisation des catastrophes naturelles par les Cat bonds

En cas de survenance de catastrophes naturelles : d'ampleur majeure, les pertes économiques sur les biens assurés sont généralement très élevées, dépassant souvent largement les capacités de couverture sur fonds propres des compagnies d'assurance. Le mécanisme de Cat Bonds permet de faire porter le risque directement aux marchés financiers, selon un mécanisme de titrisation qui se traduit par le schéma suivant.

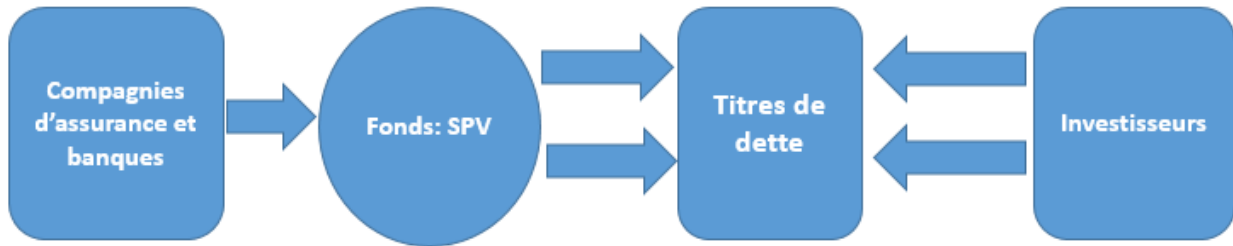


FIGURE 4.1 – Fonctionnement des Cat bonds

La compagnie d'assurance émet des titres de dette obligataires par le biais d'un fonds (Special Purpose Vehicle) et avec le soutien d'une banque d'affaires. Ces titres sont vendus aux investisseurs et les fonds perçus sont placés sur un compte accessible à l'assureur en cas de réalisation de la catastrophe naturelle déterminée. En contrepartie, les investisseurs perçoivent un coupon annuel correspondant à la rémunération de leur placement assorti d'une prime de risque. Si au bout de la maturité (3 à 5 ans) des obligations, il n'y a pas eu survenance de catastrophe l'ensemble des fonds est reversé aux investisseurs.

Certes onéreuses tout comme la mise en place des assurances paramétriques, les Cat bonds ont le mérite d'être dans un premier temps, un substitut aux contrats de réassurance (transfert du risque de ré(assurance) aux marchés financiers) pour la couverture des risques catastrophes naturels notamment dans des cas climatiques extrêmes.

D'un point de vue systémique, il est donc bien plus sûr de partager ces risques de pointe avec un nombre important d'investisseurs que de les concentrer massivement chez quelques réassureurs sur-exposés en cas de catastrophe majeure. Ils permettent aux compagnies d'assurance d'augmenter significativement leurs engagements sans avoir à les porter dans leur bilan ni à entamer leur solvabilité. Contrairement aux contrats de réassurance, les Cat bonds sont très liquides et ne présentent pas de risque de contrepartie.

Cependant, par rapport aux contrats classiques de réassurance où les indemnisations de sinistre se basent sur les pertes effectives liées au sinistre, les Cat Bonds, tout comme tous les produits d'assurance paramétrique, présentent souvent un seuil de déclenchement lié à un indice ou à un modèle paramétrique. Le payoff peut donc être largement différent des pertes enregistrées par l'assureur.

4.1.3 Modèle de tarification de catastrophes naturelles

L'approche catastrophe naturelle du risque sécheresse dans le Royaume nous permet d'appréhender la gestion de la sécheresse par les modules suivants :

Module exposition : il comprend le risque (perte sur rendements) et sa localisation géographique : ici il s'agit donc de toutes les communes agricoles dans le royaume, il s'agira pour le modèle de permettre de quantifier les pertes potentielles de rendement qui sont associés à des risques particuliers météorologiques à divers étapes du cycle de culture.

Module Aléa : il permet de générer la loi de distribution de la fréquence du risque (pertes sur rendements dues à la sécheresse) et de fournir donc la LEC.²

2. Loss Exceedance Curve : c'est la fonction de répartition inverse des coûts de sinistre.

Module vulnérabilité & finance : il consiste à évaluer dans un premier temps la fréquence et l'intensité du risque, afin de déterminer dans un second temps l'impact financier de ce dernier sur l'assuré. Le seuil de déclenchement d'indemnisation est le vecteur directeur de ce module un choix non optimal de ce seuil conduit à une sur-tarification du risque ou à des probabilité de survenance très faibles.

4.2 Modèle AMRC

L'assurance multirisque climatique (MRC) au Maroc est un produit qui couvre les principaux aléas climatiques auxquels sont exposés les cultures céréalières, légumineuses et oléagineuses. Ce produit d'assurance s'étend à l'ensemble des communes rurales du Royaume et couvre principalement les risques suivants : sécheresse, excès d'eau, gel, grêle, vent violent, vent de sable.

Les agriculteurs peuvent souscrire aux niveaux de garantie déterminés ci-après en fonction des zones, des cultures, de la superficie et de la récolte assurées :

- zones défavorables en Bour, la garantie est de 600 dhs / Ha, avec une subvention de primes à 90% où la prime payée par l'agriculteur est d'une valeur de 16 dhs / Ha.
- Zones moyennement favorables en Bour : 900 dhs / Ha avec une subvention de primes à 90% où la prime payée par l'agriculteur est d'une valeur de 22 dhs / Ha.
- Zones favorable en Bour et périmères irriguées option 1 : 1450 dhs / Ha, avec une subvention de primes à 90% où la prime payée par l'agriculteur est d'une valeur de 26 dhs / Ha.
- Zones favorable en Bour et périmères irriguées option 2 : 2900 dhs / Ha, avec une subvention de primes à 68% où la prime payée par l'agriculteur est d'une valeur de 199 dhs / Ha.
- Zones favorable en Bour et périmères irriguées option 3 : 4350 dhs / Ha, avec une subvention de primes à 57% où la prime payée par l'agriculteur est d'une valeur de 398 dhs / Ha.

Une commune est considérée comme défavorable en bour si le rendement historique du blé tendre (ou à défaut l'orge) est inférieur ou égal à 6qx/ha ; une commune est considérée comme moyennement favorable en bour si le rendement historique du blé tendre (ou à défaut l'orge) est strictement supérieur à 6qx/ha et inférieur ou égal à 12 qx/ha ; enfin, une commune est considérée comme favorable en bour si le rendement historique du blé tendre (ou à défaut

l'orge) est strictement supérieur à 12 qx/ha.

Le montant du capital assuré est obtenu de la manière suivante :

$$\text{Capital assuré} = \text{monant du niveau de garantie} \times \text{superficie assurée}$$

Le versement de la subvention aux cotisations sera effectué comme suit :

- Une subvention aux cotisations calculée sur la base des superficies assurées selon les taux de subvention arrêtés par niveau de garantie tel que mentionné plus haut.
- Un montant d'égalisation calé sur 95 MDh pour une superficie assurée de 300.000 Ha. Si la superficie assurée est inférieure à 300.000 Ha, le montant d'égalisation est diminué proportionnellement à la superficie assurée. Si non si la superficie assurée dépasse 300.000 Ha, le montant d'égalisation est revu à la hausse, selon la formule :

$$\text{Montant d'égalisation} = 95\text{Mdh} + \left[95\text{Mdh} \times \frac{\text{superficie assurée} - 300.000\text{Ha}}{\text{superficie assurée}} \right]$$

Le déclenchement de la garantie est lié à une déclaration officielle de l'état de sinistre par commune rurale. Le rendement moyen réalisé dans la commune concernée est inférieur au rendement de référence R_{ref} tel que défini ci-dessous.

$$R_{ref} = 60\% \text{ de la moyenne sur les 10 dernières années des rendements de la commune}$$

L'indemnisation des assurés relevant des communes rurales sinistrées sera effectuée, après déclaration officielle de l'état de sinistre, selon la formule suivante :

$$\text{Indemnisation} = \text{Capital Assuré} \times \text{Pourcentage de perte}$$

$$\text{Pourcentage de perte} = 1 - (\text{Rendement réel constaté} / \text{Rendement de référence})$$

4.3 Modèle d'assurance indicielle basée sur le rendement

4.3.1 Approche Burning Cost (déterministe)

C'est l'approche la plus simple pour le calcul de la prime. Il s'agit d'évaluer les indemnités pour chacune des années de la chronique de données disponibles. Le coût moyen du risque est simplement exprimé par :

$$CM = \frac{1}{N} \sum_{i=1}^N Val_i$$

avec Val_i la valeur de l'indice historique pour l'année i .

Comme limites, cette démarche ipso facto non seulement ne tient pas compte du caractère aléatoire du risque mais également de l'aspect solvabilité et probabilité de ruine. L'approche déterministe est aussi très limitée dans un contexte de profondeur d'historique très réduite, les données disponibles pour le risque étudié couvrent généralement moins de 30 années, ce qui en réalité n'est pas propice pour des estimations non biaisées. L'alternative est l'approche stochastique que nous présentons ci-après.

4.3.2 Approche actuarielle et Monte Carlo

L'approche stochastique consiste à caler une loi de distribution statistique continue avec celle des valeurs historiques de l'indice, il est possible ensuite d'utiliser la loi de distribution pour prévoir les réalisations possibles de cet indice. Même si la simulation Monte Carlo paraît moins aisée à mettre en place, elle est plus sophistiquée et nous l'utilisons dans notre modèle de tarification pour deux raisons : (i) l'approche déterministe est très limitée du fait qu'elle ne prend pas en compte l'aléa dans la modélisation ; (ii) nous disposons d'une profondeur d'historique assez limitée pour chaque commune (14 années au maximum), l'approche Monte Carlo reposant sur la loi forte des grands nombres, le fait de répéter un grand nombre de tirages aléatoires permet de converger presque sûrement vers la valeur de la prime pure la plus juste possible. Le coût moyen dans ce cas est exprimé par :

$$CM = \int Val(x) d\mu(x)$$

où μ est la fonction de distribution de l'indice à déterminer. La suite du détail de l'évaluation de cette expression par démarche Monte-Carlo est présentée en annexe 5.

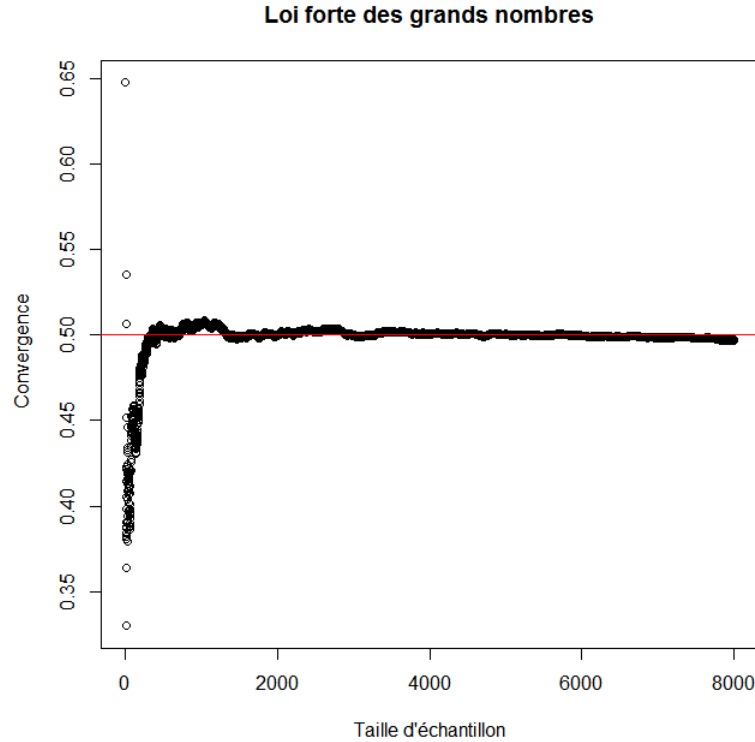


FIGURE 4.2 – Convergence d'après la loi forte des grands nombres

4.3.3 Area-yield insurance

Le modèle d'assurance paramétrique adopté dans ce projet est l'assurance indicielle basée sur les rendements. L'indice se construit à travers une agrégation communale des agriculteurs, ce qui permettra de définir un rendement de référence. Les assurés sont indemnisés indépendamment de leur rendements réels chaque fois que le rendement global de la zone n'atteint pas les seuils prédéfinis. L'idée est donc de déterminer la loi de distribution des rendements puis à l'aide de méthodes stochastiques de déterminer les primes.

4.4 Cadre de la Théorie sur les valeurs extrêmes

Lorsque l'on s'intéresse à la partie centrale d'un échantillon, le résultat clé est le Théorème de la Limite Centrale (abrégé en TLC) donnant la loi asymptotique de la somme des observations. Par contre, si l'on souhaite étudier les valeurs extrêmes de cet échantillon, le TLC ne présente que peu d'intérêt. On utilise plutôt un résultat analogue établissant la loi asymptotique du maximum de l'échantillon. Il permet de classer la plupart des lois en trois domaines d'attraction.

4.4.1 Convergence en loi du maximum d'un échantillon

Soit X_1, \dots, X_n un échantillon *iid*. Si F_X est la fonction de répartition de X , alors :

$$\begin{aligned} P(M_n \leq x) &= P(\max(X_1, \dots, X_n) \leq x) \\ &= P(X_1 \leq x, \dots, X_n \leq x) \\ &= (F_X(x))^n \end{aligned}$$

Avec $M_n = \max(X_1, \dots, X_n)$

Conformément à ce qui a été énoncé plus haut, lorsque $n \rightarrow \infty$, la loi de M_n est dans le domaine d'attraction de la loi de Fréchet, soit de Gumbel ou de Weibull. Et ces trois lois peuvent se synthétiser en une seule loi, GEV (Generalized Extremes Values) dont la distribution est la suivante :

$$\begin{aligned} H_{\gamma, \mu, \psi}(x) &= \exp \left[- \left(1 + \gamma \frac{x - \mu}{\psi} \right)^{-1/\gamma} \right], \\ 1 + \gamma \frac{x - \mu}{\psi} &> 0, \end{aligned}$$

Avec

$$H_{0, \mu, \psi}(x) = \exp \left[- \left(\frac{x - \mu}{\psi} \right) \right], x \in R$$

Le résultat ci-dessous établit la loi asymptotique du maximum $M_n = \max(X_1, \dots, X_n)$ de l'échantillon.

Théorème 4 : S'il existe deux suites réelles $(a_n > 0)$, (b_n) et un réel γ tels que

$$P \left\{ \frac{M_n - b_n}{a_n} \leq x \right\} \rightarrow H_\gamma(x)$$

lorsque $n \rightarrow \infty$ alors

$$H_\gamma(x) = \begin{cases} \exp[-(1 + \gamma x)^{\frac{-1}{\gamma}}]_+ & \text{si } \gamma \neq 0 \\ \exp(-e^{-x}) & \text{si } \gamma = 0, \text{ sinon} \end{cases}$$

La fonction de répartition de $H_\gamma(\cdot)$ est la fonction de répartition de la loi des valeurs extrêmes. Cette loi dépend du seul paramètre γ appelé indice des valeurs extrêmes. Selon le signe de γ .

Domaines d'attraction	Gumbel $\gamma = 0$	Fréchet $\gamma > 0$	Weibull $\gamma < 0$
-----------------------	---------------------	----------------------	----------------------

TABLE 4.2 – Caractéristiques des GEV

On définit trois domaines d'attraction :

si $\gamma > 0$, on dit que $H(\cdot)$ appartient au domaine d'attraction de Fréchet. Il contient les lois dont la fonction de survie décroît comme une fonction puissance. On parle aussi de lois à queue lourde. Dans ce domaine d'attraction, on trouve les lois de Pareto, Student, Cauchy, etc ...

si $\gamma = 0$, $H(\cdot)$ est dans le domaine d'attraction de Gumbel qui regroupe les lois ayant une fonction de survie à décroissance exponentielle. C'est le cas des lois normale, gamma, exponentielle, etc.

si $\gamma < 0$, $H(\cdot)$ appartient au domaine d'attraction de Weibull. Ce domaine contient les lois dont le point terminal $x_F = \inf\{x, F(x) \geq 1\}$ est fini. C'est le cas par exemple des lois uniformes, lois Beta etc.

Ci-dessous la densité associée pour différentes valeurs de γ .

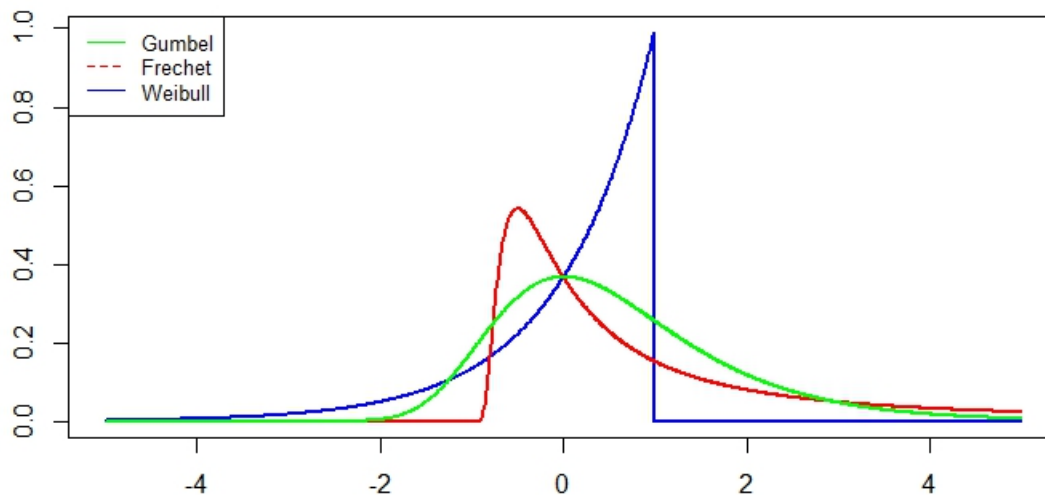


FIGURE 4.3 – Distribution des lois de Gumbel, Fréchet et de Weibull

4.4.2 Détermination de la loi de distribution

Nous allons donc tester deux distributions : la loi Log-normale (domaine d'attraction de Gumbel), et la loi de Weibull (Domaine d'attraction de Weibull) en vue de voir laquelle ajuste le mieux nos données ; l'idée est d'utiliser cette loi d'ajustement dans la suite pour la simulation Monte-Carlo qui nous permettra d'obtenir le coût des indemnisations dans un premier temps et les primes à payer par les agriculteurs dans un second temps.

Nous décrivons en annexe ces deux lois et la méthode d'estimation de leur paramètre. Le choix à priori de ces deux lois se réfère au fait que parmi les lois GEV, ces lois sont à support positif et fini tout comme la variable cible de notre étude : le rendement estimé.

Les graphiques d'ajustement dans un premier temps nous permettront d'observer le degré d'adéquation entre ces lois et nos données, ensuite les tests d'adéquation nous permettront de conclure et donc d'en retenir la plus adéquate.

4.4.3 Test du « Q-Q Plot »

L'ajustement Q-Q Plot est un test graphique permettant de vérifier visuellement la compatibilité des données avec une hypothèse de loi.

Notons $x_{(i)}$ la statistique d'ordre i et F la fonction de répartition de la loi à tester. Sous l'hypothèse d'adéquation, on doit avoir : $F(x_{(i)}) = \frac{i}{n} \quad \forall i \in (1, 2, \dots, n)$.

Graphiquement, cela signifie que les points $[x_{(i)}, F^{-1}(\frac{i}{n})]$ doivent être proches de la première bissectrice, ce qui est facile à contrôler visuellement.

4.4.4 Tests d'adéquation

L'objectif étant d'ajuster au mieux nos données à une loi de distribution théorique et étant donné que la fonction de répartition caractérise la loi, nous procédons comme suit :

Étape 1 : Estimations³ par maximum de vraisemblance des paramètres.

Étape 2 : Afin de corroborer les résultats des deux tests, nous ajoutons les qq-plots qui sont une approche graphique permettant de juger l'adéquation des données avec une loi de distribution théorique donnée.

Étape 3 : Tests basés sur la fonction de répartition : les test de Kolmogorov-Smirnov et de Cramer Von Mises.

3. Estimateurs sans biais

Chapitre 5

De la calibration à la tarification

5.1 Techniques de clustering sur la variable cible

5.1.1 Technique de classification : classification hiérarchique ascendante

Afin d'obtenir un ajustement maximal des primes, nous allons construire des classes de communes selon l'évolution de leur rendement au fil des années. Concrètement nous allons regarder à une année t comment varie les rendements d'une commune par rapport aux années précédentes. Sur la base de ce critère, la classification permet de regrouper par classe les communes semblables. Ainsi pour chaque céréale, nous procédons comme suit :

Étape 0 : soit $x_{c,i}$ le rendement de la commune c à l'année i .

Étape 1 : on calcule $\widehat{x}_{c,i} = \frac{x_{ci}}{\widetilde{x}}$ avec $\widetilde{x} = \frac{1}{T} \sum_{k=1}^T x_{c,k}$
où T représente l'horizon (14 années d'après nos données).

Étape 2 : nous construisons des classes homogènes à partir des $\widehat{x}_{c,i}$ à travers l'évaluation des indices de validation interne et de stabilité dans un premier temps, puis la CAH proprement dite dans un second temps.

Cette démarche a deux intérêts. Elle permet primo de prendre en compte la volatilité des rendements sur l'historique ; en effet si la pseudo statistique $\widehat{x}_{c,i} \simeq 1$, ceci indique que l'année i de la commune c se comporte en moyenne de la même manière que les autres années. Secondo elle donne la possibilité à partir des $\widehat{x}_{c,i}$ de construire des variables aléatoires pour

chaque classe ; en effet vu que l'historique de données (14 années) n'est pas très long pour des résultats significatifs d'après la loi forte des grands nombres, nous avons l'avantage d'avoir après cette démarche par classe n (n plus significatif, $n = 14 \times$ nombre d'éléments de la classe) observations d'une variable aléatoire, les résultats de cette classification serviront donc à l'ajustement de distribution par classe.

5.1.2 Résultats et validation de la classification

La réalisation d'une classification hiérarchique ascendante comme décrite annexe 2, permet d'avoir le nombre de classes de communes pour laquelle la partition a une perte d'inertie maximale. Concrètement, nous réalisons une classification hiérarchique sur les composantes principales d'une ACP, le choix du nombre de composantes principales de l'ACP repose sur le critère de Kaiser (1960) d'après lequel le nombre d'axes principaux correspond au nombre de composantes principales dont la valeur propre est supérieure à 1. Cette classification donne lieu à trois classes de communes homogènes pour le blé dur, quatre classes communales pour le blé tendre et trois classes pour l'orge.

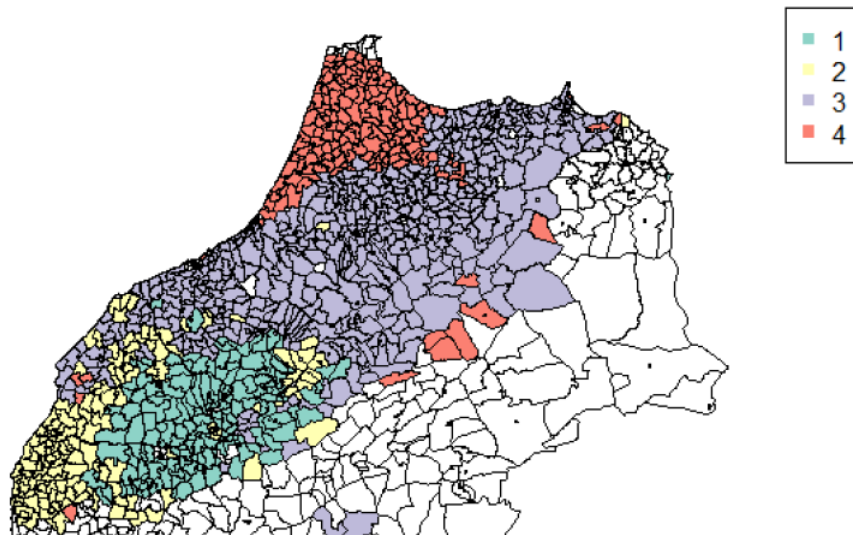


FIGURE 5.1 – Classification du blé tendre

Céréale	blé dur	blé tendre	Orge
Nombre de classes	3	4	3

TABLE 5.1 – Résultats de la classification

Les résultats de cette classification sont explicités par les cartes factorielles des individus en annexe 5. Nous avons réalisé les tests de validation interne et de stabilité sur la CAH et les résultats pour l'orge par exemple sont donnés par :

	2	3	4
Dunn	0.1080	0.1080	0.1285
ADM	0.0155	0.0022	0.1156
AD	2.0212	2.0062	2.0187

TABLE 5.2 – Tests de validation et de stabilité de la classification sur l'orge

Ces résultats confirment bien le nombre de classes et la stabilité. On constate en effet que l'indice de Dunn est proches de zéro, cela soutient la bonne stabilité des classes, les valeurs aussi proches de zéro des distances moyennes (Average Distance) conduit à la bonne stabilité des classes. Les résultats de test de stabilité des deux autres céréales sont détaillés en annexe 2.

A partir des résultats de la CAH réalisée ci-dessus, nous allons calibrer les lois de chaque classe pour chaque céréale ensuite grâce à la méthode Monte-Carlo nous parviendrons à proposer des primes pour chaque classe. Afin d'ajuster les lois, nous allons utiliser dans notre démarche de modélisation les qq-plots, la fonction de répartition et les tests non paramétriques.

Nous observons avant tout un ajustement par les fonctions de réparation afin de nous donner une idée sur la distribution la plus proche de la distribution empirique.

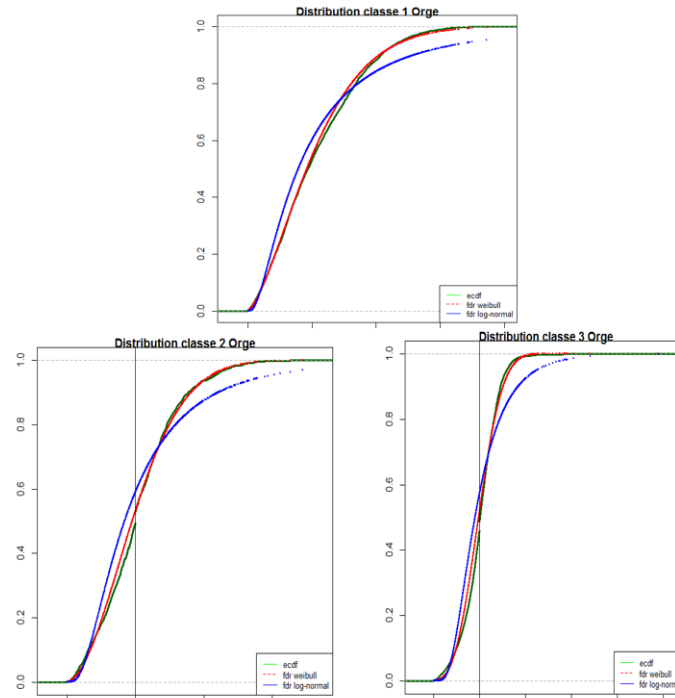


FIGURE 5.2 – Fonctions de répartition

L'analyse de ces trois graphiques nous permet d'observer que comparativement à la loi log-normale en bleu, la loi de Weibull en rouge paraît ajuster au mieux nos données par classe. L'observation des qq-plots ci-dessus confirment cette première tendance, par analyse d'ajustements graphiques ; nous rejetons à cette étape l'hypothèse H_0 pour la log-normale et allons nous limiter dans la suite à la loi de Weibull . Pour corroborer ces ajustements graphiques, nous réalisons tels que décrits en annexe, les tests d'adéquation mentionnés plus haut avec la loi de Weibull.

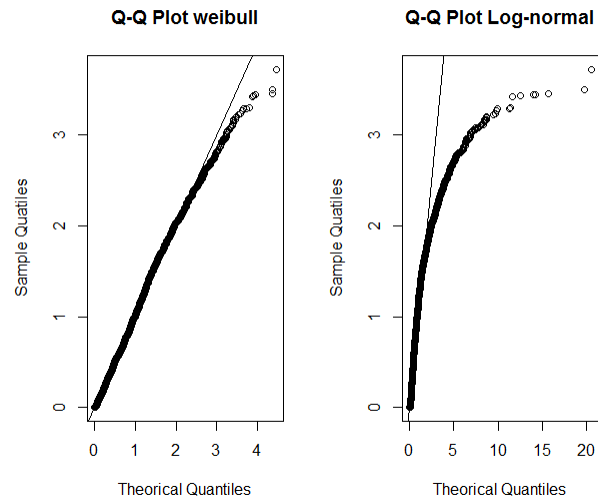


FIGURE 5.3 – Q-Q plot orge classe 1

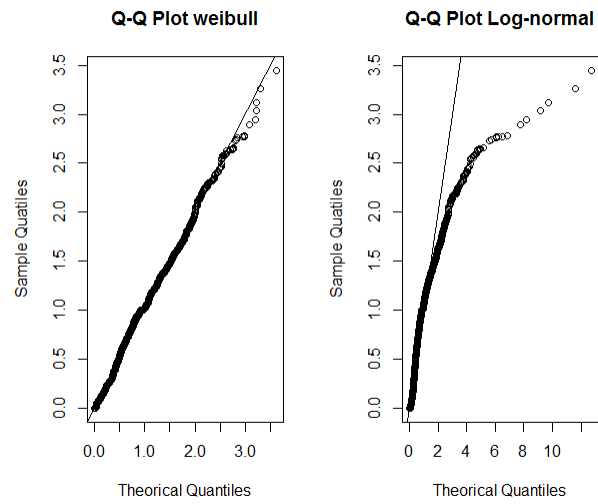


FIGURE 5.4 – Q-Q plot orge classe 2

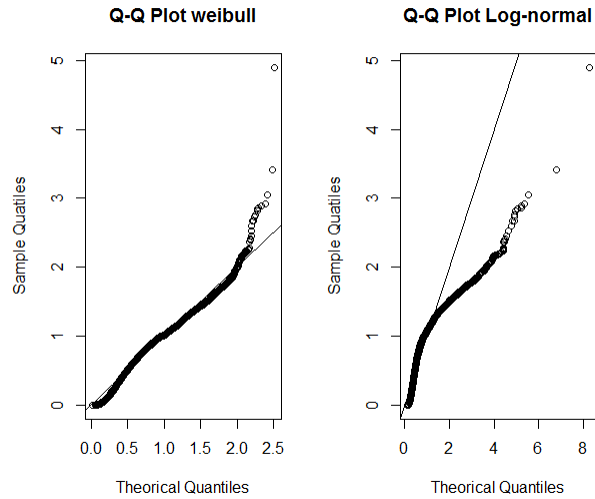


FIGURE 5.5 – Q-Q plot orge classe 3

Les résultats du test d'adéquation de Cramer Von Mises présentés dans le tableau 5.3 nous conduisent à accepter l'hypothèse H_0 (confer Annexe 3.) seuil de 1%. Nous rappelons que les tests d'adéquations sont très exigeants surtout pour des échantillons de grandes tailles. Toutes les classes communales construites étant de taille supérieure à 1000, même si les ajustements graphiques montrent qu'il y a une parfaite adéquation, il est difficile d'accepter l'hypothèse H_0 avec un niveau de 5% .

Orge Weibull	classe 1	classe 2	classe 3
Cramer Von Mises	0.54991	0.87836	0.7623
p-value	0.035	0.057	0.046

TABLE 5.3 – Test d'adéquation de Cramer Von Mises

Cette étape d'ajustement de loi des classes nous laisse comprendre que la loi de Weibull paraît la mieux adaptée à toutes les classes ; et d'après le test de Cramer Von Mises et de Kolmogorov-Smirnov (les résultats sont présentés en annexe 7 et annexe 8), réalisés seulement sur hypothèse de distribution de Weibull, nous acceptons l'hypothèse d'adéquation de nos données avec une loi de Weibull.

5.2 Paramétrisation du modèle

Notre modèle de tarification se base sur une approche par simulation comme souligné dans la partie 4.2.2, nous calculons la prime pure grâce à la méthode Monte Carlo.

5.2.1 Algorithme de la simulation

Après avoir ajusté des lois de distribution dans la partie précédente pour chaque classe de commune par céréale, l'algorithme de simulation des indemnisations passe par les étapes suivantes :

Pour une céréale c :

Étape 0 : à partir de l'estimation par maximum de vraisemblance des paramètres des lois de distribution $W(a, \alpha)$ des classes obtenues par classification dans le traitement précédent, nous simulons N réalisations (rendements « normés ») par classe.

On réalise les étapes suivantes N fois en vue d'obtenir la fonction de répartition des indemnisations, ce qui permettra d'en déduire la prime pure.

Étape 1 : pour chaque classe on tire une observation (tirage uniforme).

Étape 2 : on extrapole le rendement tiré à l'étape précédente à toutes les communes de la classe. Cette étape fournit des rendements simulés pour chaque commune.

Étape 3 : les données de l'étape 2 nous permettent de calculer le coût selon le modèle d'indemnisation : pour une céréale i cultivée dans une commune c , les fonctions d'indemnisation sont celles fixées au chapitre 2 lors du calcul du risque de base.

5.2.2 Prime d'assurance indicielle

Par définition la prime pure correspond au montant du sinistre moyen auquel devra faire face l'assureur pour le risque. La simulation Monte-Carlo nous permet d'obtenir des indemnisations à partir desquelles nous déduisons simplement la prime pure par burning cost. Il revient donc de procéder au chargement de cette dernière afin d'obtenir la prime qui sera commercialisée. Le chargement à faire doit tenir compte des différents risques liés au risque de base.

Ainsi en suivant la démarche Monte-Carlo décrite en annexe 5, les primes pures sont données par :

La prime pure d'une classe i d'une céréale j s'obtient par :

$$PP_i = \frac{1}{N} \left(\sum_{k=1}^N \frac{\text{Indemnisation}_k \text{ de la classe } i}{\text{Surface totale de la classe } i} \right)$$

La prime pure PP_j , pour une céréale j :

$$PP_j = \frac{1}{N} \left(\sum_{k=1}^N \frac{\text{Indemnisation}_k \text{ de la céréale } j}{\text{Surface totale de la céréale}} \right)$$

Et donc la prime pure globale pour les trois céréales s'obtient par :

$$PP_T = \left(\frac{1}{N} \sum_{j=1}^3 \sum_{k=1}^N \frac{\text{Indemnisation}_k}{\text{Surface totale assurée}} \right)$$

Les chargements se présentent comme suit :

Chargement de gestion : ce chargement est destiné à couvrir les frais de l'assureur dans la gestion du nouveau produit : Informatique, direction générale, fonctions centrales, masse salariale, etc.

Chargement de distribution : il est destiné à couvrir les frais de distribution du nouveau produit (publicité, commerciaux, etc.), il est fixé tout simplement en ligne avec l'expérience du produit existant.

Chargement d'incertitude ou uncertainty loading : ce chargement a pour objectif de couvrir les erreurs liées au modèle d'estimation des rendements (éventuels écarts entre les rendements observés et les rendements estimés)

Coût du capital et LEC : La Loss Exceedance Curve (LEC) associe à chaque niveau de perte totale annuel la probabilité que cette perte soit dépassée à une année donnée. Grâce à la représentation du LEC, un assureur peut mieux se situer par rapport aux caractéristiques du risque qu'il assure, donc de voir si ses capacités lui permettent d'être solvable faces aux caractéristiques du risque assuré.

Pour un risque X de fonction de répartition F_X :

$$LEC_X(x) \sim 1 - F_X(x) = P(X > x)$$

Nous intégrons donc cette composante à travers le « coût du capital » dans le chargement de la prime pure en vue de tenir compte de la solvabilité de l'assureur. Ainsi pour une probabilité de ruine de l'assureur égale à 0.5%, il ne reste qu'à déterminer le niveau de la $LEC(X, 99.5\%)$.

	Blé dur	Blé tendre	Orge
% de surface	20%	45%	35%
Prix (MAD/quintal)	243	240	167

TABLE 5.4 – Paramètres de calibrage

Nous avons effectué $N = 10^5$ simulations de la garantie du produit 1 (seuil de déclenchement fixé à 60% du rendement de référence et niveau de garantie de 100%) avec les paramètres (répartition de la surface par céréale et prix des céréales), cette simulation va donc permettre de calculer cette fonction de pertes (montant d'indemnisation et probabilité correspondante) que nous présentons à travers la figure 5.6; sur base du scénario 1 par exemple, à 90% la survenance d'une sécheresse engendrerait une indemnisation supérieure à 1.7 milliard de MAD.

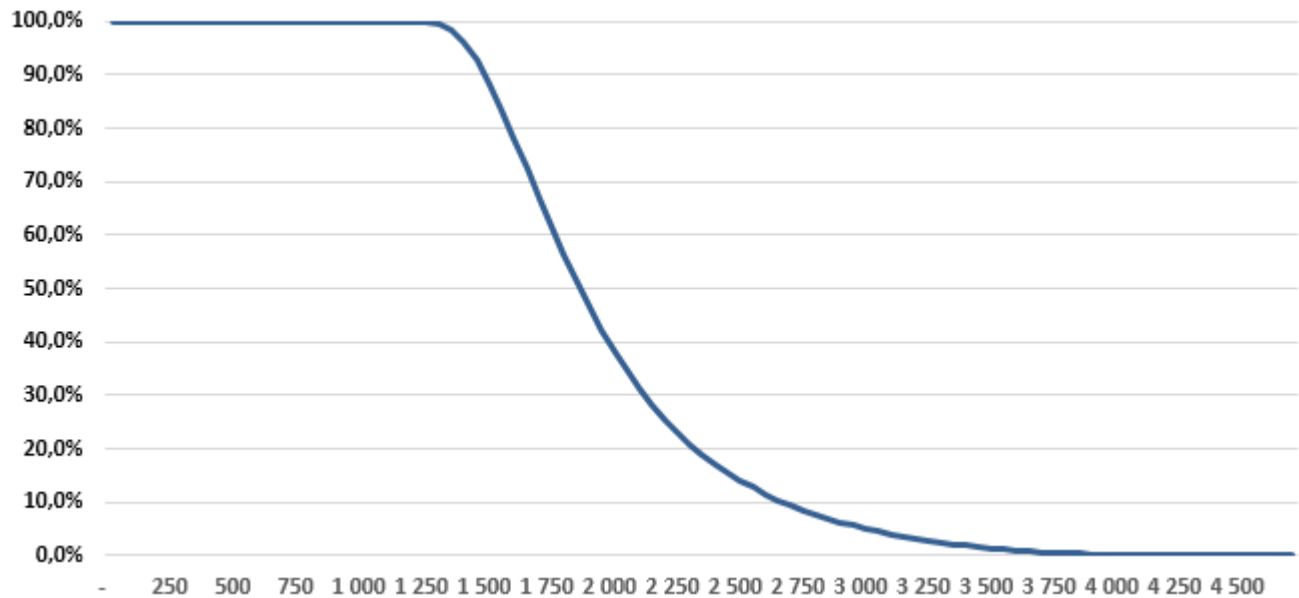


FIGURE 5.6 – LEC à partir du scénario 1 avec la fonction d'indemnisation 1

Taux de produits financiers : ce taux correspond à la rémunération du placement des primes. Il est fixé par rapport au niveau de taux règlementaire local de placement des produits d'assurance.

Pour notre modèle, on a donc les hypothèses suivantes :

Chargement	Valeurs
coût du capital	$\psi(99.5^{eme} \text{ centile des indemnisations})$
Charges de gestion	β des prime pure
Chargement de distribution	γ des prime pure
Uncertainty loading	δ des prime pure
Taux de produits financiers	λ de prime pure

TABLE 5.5 – Chargement de la prime pure

La prime chargée totale pour les trois céréales s'obtient par :

$$\text{Prime commerciale totale} = (1 - \lambda + \beta + \gamma + \delta) \times \text{Prime pure totale} + \text{coût du capital}.$$

Pour la calibration de la prime commerciale, les taux de chargements (taux de produits financiers et le chargement de distribution) sont fixés par rapport avec l'expérience du marché de l'assureur des risques agricoles.

Les charges de gestion quant à elles sont de l'ordre de 5% des primes, ces charges comme souligné dans le chapitre 1 (point sur l'assurance paramétrique en Afrique) vont jusqu'à atteindre 20% en assurance classique et sont de l'ordre de 5% voire nulles en assurance paramétrique.

Le coût du capital inséré dans le modèle est le coût indiqué (20%) diminué du taux des produits financiers afin de tenir compte de la rémunération du placement de ces capitaux.

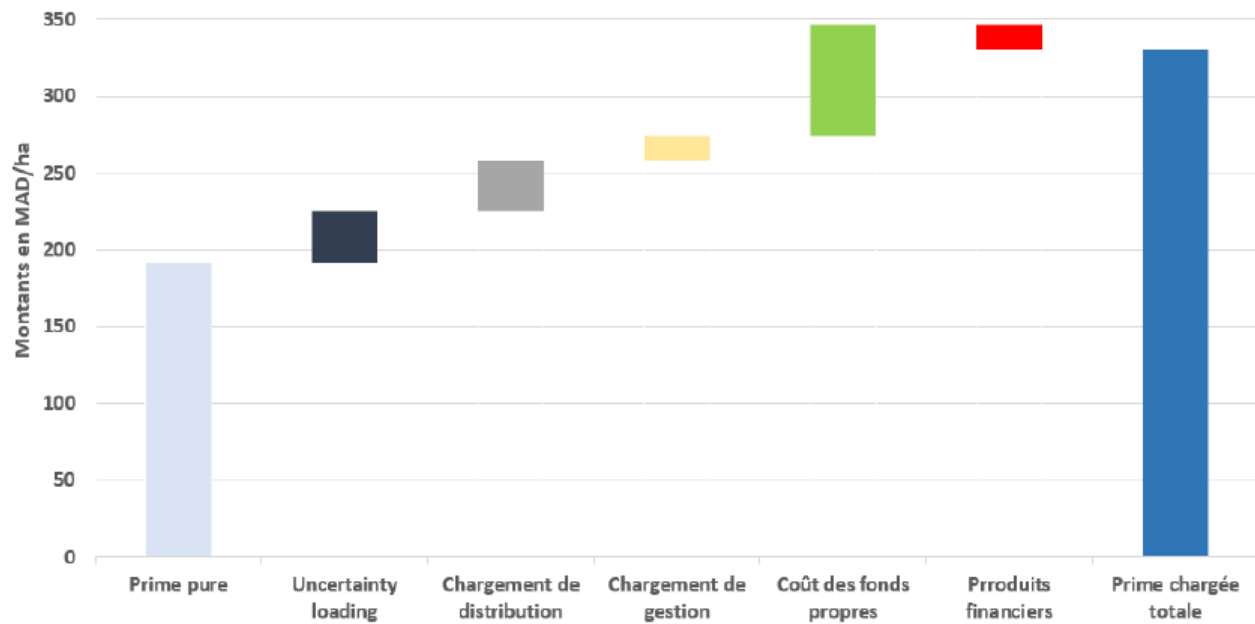


FIGURE 5.7 – Tarification à partir de la fonction d'indemnisation 1

En assurant conjointement les 3 céréales, le chargement le plus important (38% de la prime pure) correspond au coût de rémunération du capital à mobiliser qui est de 72MAD/ha et son optimisation pourrait se faire à travers la réassurance. Le chargement devant tenir compte de l'ensemble des incertitudes qui affectent le risque sécheresse, l'« uncertainty loading », vaut 33 MAD/ha soit 17% de la prime pure.

L'« uncertainty loading » reste toutefois difficile à déterminer avec précision et dépend à la fois du contexte, de l'assureur et de la conjoncture. Les charges de distribution et de gestion valent respectivement 33 MAD/ha et 16 MAD/ha. Les produits financiers correspondant aux bénéfices engrangés par l'assureur dans le placement des primes, ils valent selon nos hypothèses, 16 MAD/ha, soit 8,57% de la prime pure.

Les rémunérations de placement des primes (produits financiers) correspondant à un gain pour les assurés, elles réduisent donc le montant de la prime pour l'assuré, d'où le signe négatif dans le tableau 5.6.

Les résultats de la tarification à partir du scénario 1 sur la base de la fonction d'indemnisation 1 se présentent dans le tableau 5.7 : pris conjointement, la prime pure pour les trois céréales vaut 192 MAD/ha alors que si on devait assurer chaque céréale séparément, la prime pure aurait valu 216 MAD/ha pour le blé tendre, 229 MAD/ha pour le blé dur et 141 MAD/ha pour l'Orge.

Chargements	Prime pure totale	% de la Prime Pure
Prime Pure (en MAD/ha)	192	100%
Coût des fonds propres	72	38%
Uncertainty Loading	33	17%
Charges de distribution	33	17%
Charges de gestion	17	9%
Produits financiers	-17	-9%
Prime commerciale (en MAD/ha)	331	172%

TABLE 5.6 – Tarifs sur base du scénario 1

	Blé tendre	Blé dur	Orge	Total
Prime Pure (en MAD/ha)	216	229	141	192
Prime commerciale (en MAD/ha)	412	423	261	331

TABLE 5.7 – Tarifs par type de céréale sur base du scénario 1

5.2.3 Analyse du tarif sur les trois scénarios

Dans l'hypothèse d'assurer les trois céréales conjointement, la figure ci-dessous montre que le coût moyen du risque (prime pure) sur la base du scénario 3 s'élève presque au triple de celui qui est évalué sur la base du scénario 1.

Le scénario 3 présente des niveaux de prime assez élevés et révèle la sévérité du risque ; on retient que le coût de l'assurance augmente considérablement si l'on ne se limite pas seulement à la moyenne historique comme rendement de référence. Tarifier sur la base du scénario 3 permet donc à l'assureur de tenir compte des valeurs extrêmes et donc de mieux gérer le risque.

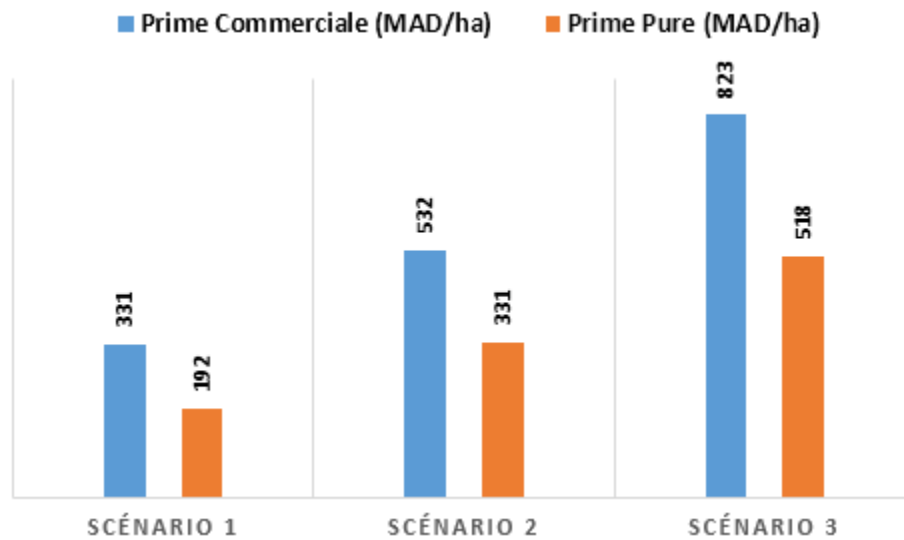


FIGURE 5.8 – Primes sur les trois scénarios à partir de la fonction d'indemnisation 1

5.3 Mesure du risque de tarification : scénario de baisse de rendements

En vue de finaliser le modèle, évaluer la viabilité du produit, il est possible de déterminer l'effet de changement climatique sur les primes. Nous allons donc suivre l'impact d'une variation de rendement sur les primes. Il s'agit de voir comment varie le niveau des primes pour une variation de rendement, notamment pour une baisse de niveau α de rendement. Concrètement on intègre dans le calcul des fonctions d'indemnisation une diminution des rendements.

Dans un scénario de baisse de 5% du rendement, la prime augmente exactement de 5%, ceci n'est pas le cas dans les situations les plus extrêmes ; dans un scénario de baisse de rendement de 50%, la prime du risque double par rapport à son niveau initial (Prime sans

aucune hypothèse de baisse de rendement). Ces scénarios de baisse de rendement rendent bien compte de l'effet d'un choc climatique (forte sécheresse) sur le coût de l'assurance, en cas de survenance de forte sécheresse donc catastrophe naturelle si le risque doit continuer à être assuré, le coût du risque pour l'assuré devient très élevé. C'est ici qu'intervient réellement les filets de sécurité qui doivent être mis en place pour sauvegarder et garantir un minimum de revenu aux agriculteurs.

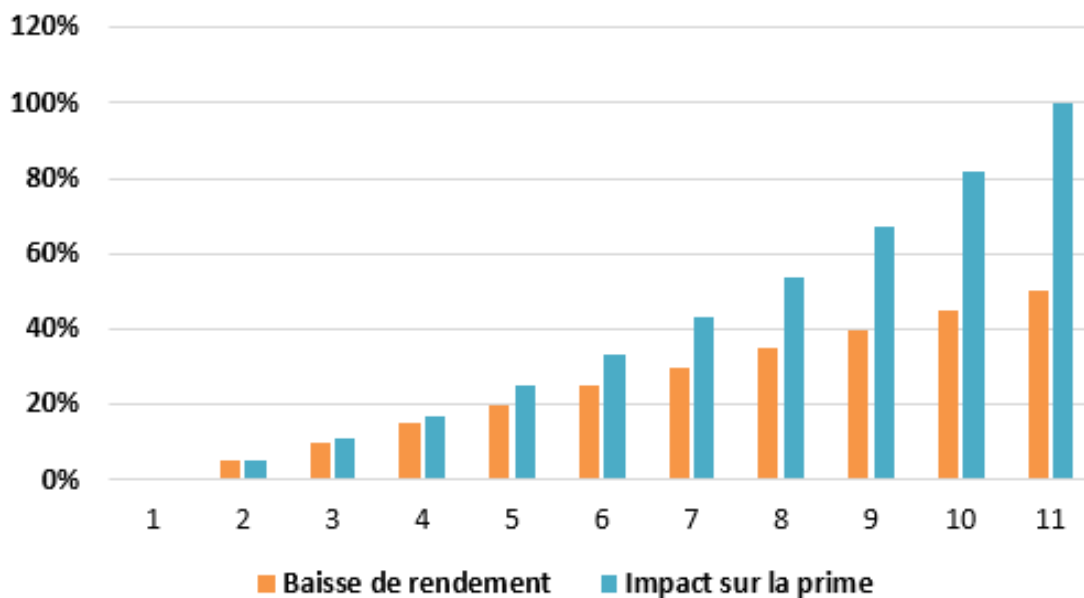


FIGURE 5.9 – Impact d'une baisse de rendement sur le niveau des primes

5.4 Comparaison des deux garanties : l'assurance paramétrique et la MRC

Après cette tentative de mise en place d'une garantie d'assurance paramétrique qui a aussi pour vocation de compléter voire de remplacer la MRC, il est intéressant de présenter un tableau comparatif de ces deux garanties. L'idée serait tout simplement de juger la générosité d'une garantie par rapport à l'autre en comparant les ratios S/P des deux garanties.

Nous rappelons que la garantie AMRC est un produit dont le seuil de déclenchement et le montant d'indemnisation dépendent du type de sol alors le modèle développé dans le cadre de notre projet est plus orienté vers le type de culture. La comparaison est délicate à mener ; si nous voulons comparer l'indemnisation de l'assurance paramétrique à celle de la MRC,

cette comparaison ne peut se faire en gardant le même seuil de déclenchement pour les deux types d'assurance. A cause du risque de base, si la comparaison est effectuée avec le même seuil de déclenchement (60% du rendement historique), l'indemnisation sera différente.

Ce risque de base, lui-même induit par le niveau d'erreur observé doit donc être pris en compte. Vue qu'une erreur est présente dans les prévisions du modèle CGMS-Maroc, l'idée serait d'avoir un seuil déclencheur (pour l'assurance paramétrique) plus large pour essayer de prendre en compte cette erreur.

Ainsi une fois le seuil de déclencheur élargi, il revient tout simplement de faire une comparaison des garanties commune par commune, nous allons donc choisir la fonction d'indemnisation à un seul seuil de déclenchement, pousser plus le trigger (70% rendement historique à 80% rendement historique) afin que dans la comparaison, les agriculteurs couverts par les deux garanties soient traités de la même manière.

Avec les niveaux de garantie de la MRC selon le type de zones, nous allons tester les fonctions d'indemnisation suivantes :

- Nous supposons que la garantie accorde la totalité (100%) de la différence entre le rendement estimé et le rendement seuil, la fonction d'indemnisation de l'assurance paramétrique est donnée par :

$$\begin{aligned} & \text{Indemnisation}_{\text{blé tendre}} \\ & = \\ & \text{prix}_{\text{blé tendre}} \times \max(0; 75\% \text{ du rendement de référence} - \text{rendement estimé}) \end{aligned}$$

- La fonction d'indemnisation de la MRC pour une zone j est donnée par :

$$\text{Indemnisation}_j = \text{garantie}_j \times \max(0; 1 - \frac{\text{rendement estimé}_j}{60\% \text{ rendement historique}})$$

On fait toujours l'hypothèse que le prix du blé tendre est 240 MAD /qx, nous testons une commune par type de zone avec zone DEB pour zone défavorable en bour, zone MFEB pour zone moyennement favorable en bour et zone FEB pour zone favorable en bour.

Type de zone	Commune	Classe de la commune	Rdt moyen
Zone DEB	Commune de DEB	1	5.42
Zone MFEB	commune de MFEB	3	7.63
Zone FEB	Commune de FEB	4	15.2

TABLE 5.8 – Caractéristiques de communes pour la comparaison des garanties

En soulignant que les primes présentées dans le tableau 5.9 sont des primes y compris la subvention de l'État, les résultats de calculs se présentent comme suit :

$\frac{\text{rendement estimé}}{\text{rendement moyen observé}}$	Commune	Rdt estimé	AMRC			Ass. Paramétrique		
			S	P	S/P	S/P	S	P
0%	Commune de DEB	-	600	160	3.75	3.48	975	280
25%		1.35	350	160	2.18	2.34	651	280
60%		3.25	-	160	-	0.69	195	280
0%	Commune de MFEB	-	900	220	4.09	3.05	1373	450
25%		1.59	595	220	2.7	2.2	992	450
60%		4.7	-	220	-	0.54	245	450
0%	Commune de FEB option 1	-	1450	260	5.57	6.84	2736	400
25%		3.8	898	260	3.45	4.56	1824	400
60%		9.12	-	260	-	1.36	547	400
0%	Commune de FEB option 2	-	2900	622	4.66	6.84	2736	400
25%		3.8	1691	622	2.71	4.56	1824	400
60%		9.12	-	622	-	1.36	547	400
0%	Commune de FEB option 3	-	4350	926	4.69	6.84	2736	400
25%		3.8	2537	926	3.73	4.56	1824	400
60%		9.12	-	926	-	1.36	547	400

TABLE 5.9 – Comparaisons de la MRC à l'assurance paramétrique

En analysant les résultats du tableau 5.9 nous remarquons que, dans la commune défavorable en bour pour un rendement estimé de 1.23 qx/ha, ce qui représente 25% du rendement moyen observé blé tendre de la commune, l'assurance paramétrique indemniserait 651 MAD/ha pour une prime de 280 MAD/ha tandis que la MRC indemniserait 350 MAD/ha pour une prime de 160 MAD/ha, soit avec des ratios très proches (respectivement 2.34 et 2.18). Pour un niveau de rendement estimé égal à 60% du rendement de référence, l'assurance paramétrique présente un ratio S/P de 0.69 donc inférieur à 1, ce qui signifie que l'agriculteur est indemnisé moins qu'il en a cotisé.

Dans la zone favorable en bour, pour les trois options de garantie d'après la MRC et dans toutes les proportions (0%, 25%, 60%), le ratio S/P de l'assurance paramétrique est supérieur à celui de la MRC.

Il est certes difficile de comparer les deux garanties mais la comparaison des ratios S/P nous permet de réaliser que l'assurance paramétrique est globalement plus généreuse que l'assurance traditionnelle MRC. Si l'assurance paramétrique semble revenir moins couteuse pour l'agriculteur ou encore qu'elle est plus intéressante pour les agriculteurs, ces résultats sont à prendre toutefois avec du recul compte de la difficulté à égaliser les deux assurances.

En tenant même compte du fait que l'assurance MRC (primes) est fortement subventionnée par l'État et que les primes présentées dans le tableau présentés sont des primes y compris la subvention, dans le cas des zones fortes en bour option 1 par exemple, la prime payée par l'agriculteur dans le cas d'une assurance paramétrique est plus élevée que celle qu'il devrait payer en assurance MRC.

Conclusion générale

Avec pour objectif de gérer le risque sécheresse, nous avons proposé dans ce projet un modèle actuariel d'assurance indicielle visant à couvrir les pertes agricoles contre le risque de sécheresse dans le Royaume, produit d'assurance dont les grands avantages se déclinent autour des trois points suivants :

- Moindre consommation de données liées aux sinistres : une fois le modèle calibré, l'assurance paramétrique n'exige plus de données liées aux sinistres pour le déclenchement de l'assurance et la gestion des indemnités. Tout repose alors sur une formule objective et définie à l'avance, précisant les montants versés aux assurés en fonction du niveau d'un indice mesuré.
- Suppression de l'aléa moral et de l'anti-sélection : dans l'assurance paramétrique, l'agriculteur est indemnisé en fonction du paramètre, qu'il ait ou non subi un sinistre. Il n'a donc non seulement aucun intérêt à augmenter son sinistre pour maximiser l'indemnisation, mais s'il gère correctement sa récolte, il peut même espérer percevoir l'indemnité tout en tirant bénéfice de sa récolte. De ce fait, cette forme d'assurance intéresse et attire aussi bien les bons agriculteurs que les mauvais agriculteurs.
- Réduction des coûts d'expertise : l'assurance indicielle a le mérite de réduire considérablement les coûts d'expertise, puisqu'il n'est plus nécessaire d'envoyer des experts sur le terrain pour évaluer la valeur réelle du sinistre .

Nous pouvons retenir qu'en l'état actuel l'amélioration des pouvoirs d'estimation du modèle CGMS-Maroc développé avec l'objectif d'aider à la construction d'un indice d'assurance paramétrique contribuerait à réduire le risque de base. L'étude a confirmé que le risque de base reste la principale entrave d'un produit d'assurance paramétrique, mais une étude pratique du risque de base « micro » développé dans ce projet permettrait d'améliorer la viabilité du produit et donc une meilleure gestion du risque sécheresse.

En outre même s'il est difficile d'obtenir un risque de base nul dans la mise en place d'une assurance paramétrique, l'étude menée dans ce projet nous permet de conclure que le risque de base est nettement réduit en prenant pour rendement de référence les quantiles supérieurs. La prise en compte des deux cas extrêmes : approches de tarification à partir du scénario 3 (rendement de référence égal au 95^{ème} centile des rendements observés) et la mise en œuvre de l'effet d'une baisse de rendement sur les tarifs, nous permet de remarquer que plus la variation à la baisse des rendements est élevée, plus les tarifs d'assurance seront élevés, lourds à supporter pour les agriculteurs.

En attendant que le pouvoir prédictif du modèle CGMS-Maroc s'améliore encore, il est possible d'envisager la mise en œuvre d'un projet pilote, projet qui consisterait à lancer un produit d'assurance paramétrique à partir d'une étude sur les provinces les plus représentatives (une erreur moyenne d'estimation des rendements faible, un nombre important de communes, une diversité des sols, avec des superficies agricoles élevées, tout ceci pouvant permettre une mutualisation, nécessaire au mécanisme d'assurance). Le choix en effet de ces zones pilotes tenant surtout compte de l'erreur d'estimation des rendements permettra une fois encore de mieux évaluer le risque de base, ce qui aidera dans l'extension du produit à l'échelle nationale.

Enfin il serait intéressant dans la suite de cette étude d'aborder la réassurance du risque que nous tentons de gérer. Il s'avère en effet capital de réassurer le risque d'autant plus que nous avons pu réaliser l'effet des scénarios extrêmes sur les primes donc sur les montants du capital à disposer par l'assureur en vue d'assurer parfaitement le risque sécheresse dans le Royaume.

Bibliographie

- [1] CHETAÏLLE, A., DUFFAU, A., HORREARD, G., LAGANDRE, D., OGGIERI, B., ROZENKOPF, I. [2011], GRET. Gestion des risques agricoles par les petits producteurs Focus sur l'assurance récolte indiciaire et le warrantage.
- [2] FALL, M., MULLET, B., MAHUL, O., DICK, W., WADE, I., AFFHOLDER, F., L'assurance agricole : un outil potentiel d'appui au développement en Afrique de l'Ouest soudano-sahélienne, Colloque : Agir en situation d'incertitude, novembre 2010, Montpellier, France.
- [3] HARMIGNIE, O., HENRI DE FRAHAN, B., POLOME, P., GASPART, F. [2004], Gestion des risques : Perspectives pour l'agriculture wallonne, Décembre 2004.
- [4] HAZELL, P., ANDERSON, J., BALZER, N., HASTRUP, A., CLEMMENSEN, HESS, U., RISPOLI, F. [2010], FIDA, L'assurance basée sur un indice climatique : potentiel d'expansion et de durabilité, Rome.
- [5] KOCH, E. [2011], Étude de faisabilité d'une assurance rendement basée sur indice climatique, Université Paris-Dauphine, Mémoire d'actuariat.
- [6] MAHUL, O. [1998]. Vers une redéfinition du rôle de l'assurance agricole dans la gestion des risques sur récoltes.
- [7] MAICHANOU, A., La micro-assurance agricole indiciaire : raisons et conditions d'exercice au Niger, cahiers de GREThA N° 2016-05 février.
- [8] RITLENG, F. [2014] & NGUYENC, C. [2014] Étude d'un produit d'assurance paramétrique contre le risque de pluie torrentielle en Jamaïque. Mémoire d'actuariat, ENSAE ParisTech.
- [9] SENAY, G. [2004]. Crop Water Requirement Satisfaction Index (WRSI) Model Description.
- [10] The World Bank [2011], Weather Index Insurance For Agriculture, Guidance for Development Practitioners, agriculture and rural development discussion 50 (page 15 à 65).

- [11] ZIMMERLI, P.[2003], en collaboration avec différents membres de l'équipe Cat Perils Swiss Re, Catastrophes naturelles et réassurance Swiss Re.
- [12] http://antoine.leblois.free.fr/Revue_fr.pdf (Les assurances agricoles basees sur des indices meteorologiques : bilan des connaissances et agenda de recherche)
- [13] http://eric.univ-lyon2.fr/~ricco/cours/cours/Comp_Pop_Tests_Nonparametriques.pdf : page 26 à page 34
- [14] <https://www.insurancespeaker-wavestone.com/2015/07/pacifica-airbus-defence-and-space-sallient-contre-risques-climatiques-agricoles>

ANNEXES

1. Méthode de calcul d'erreur moyenne par céréale.

Pour l'année j et la communes c on a :

$$Erreur(j,c) = | rendement estimé(j,c) - rendement observé(j,c) |$$

l'écart moyen relatif sur toutes les communes pour l'année j est donné par :

$$\frac{\sum_c superficie(j,c) \times ecart(j,c)}{\sum_c superficie(j,c) \times rendement observé(j,c)}$$

l'écart moyen relatif sur toutes les communes pour la commune c est donné par :

$$\frac{\sum_j superficie(j,c) \times ecart(j,c)}{\sum_j superficie(j,c) \times rendement observé(j,c)}$$

l'écart moyen relatif sur toutes les communes pour toutes les années est donné par :

$$\frac{\sum_{j,i} superficie(j,c) \times ecart(j,c)}{\sum_{j,c} superficie(j,c) \times rendement observé(j,c)}$$

2. Classification

Soient :

n : la taille de la population,

d : le nombre de variables,

x_1, x_2, \dots, x_n les vecteurs d'observations à classer,

$C = \{C_1, \dots, C_n\}$ la classification à tester,

$C(j)$: la classe contenant x_j ,

Pour tout x_j et $i \in \{1, \dots, n\} \setminus \{j\}$, $x_j^{(i)}$ est la i^{ieme} observation la plus proche de x_j .

$$d(x_j, x_{(1)}) \leq d(x_j, x_{(2)}) \leq \dots \leq d(x_j, x_{(i)}) \leq \dots \leq d(x_j, x_{(n-1)})$$

$$\forall i, j \in \{1, \dots, n\}; \quad \delta(x_j, x_j^{(i)}) = \begin{cases} 0 & x_j, x_j^{(i)} \in C(j) \\ 1/i & \text{sinon} \end{cases}$$

Algorithme de la classification hiérarchique ascendante

Soit k le nombre de classes

Étape 1 : Chaque individu est considéré comme formant à lui seul une classe, on a donc $k = n$ classes, chaque individu forme à lui seul une classe.

Étape 2 : Calcul de la matrice de distance entre les classes.

Étape 3 : Pour deux classes distincts : C^i et C^j , on veut avoir $d(C^i, C^j) = \min(C^{i'}, C^{j'})$

Étape 3 : Agrégation C^i et C^j , dans une même classe C^{ij} , donc $k = k - 1$

Étape 4 : Si $k = 1$ stop ; si non reprendre l'étape 2.

Indices de validité interne

L'indice de connectivité : il est défini par : $Conn(\mathcal{C}) = \sum_{i,j=1,\dots,n} \delta(x_j, x_j^{(i)})$

avec

$Conn(\mathcal{C}) \in \mathbb{R}_+$, si $Conn(\mathcal{C}) \rightarrow 0$, bonne connectivité.

L'indice de Dunn : il correspond au ratio entre la plus petite distance entre les individus n'appartenant pas aux mêmes classes par rapport à la plus grande distance intra-classe et il est défini par :

$$D(\mathcal{C}) = \frac{\min_{C, C' \in \mathcal{C}, C \neq C'} (\min_{x \in C, x' \in C'} d(x, x'))}{\max_{C \in \mathcal{C}} diam(C)}$$

où

$$diam(C) = \max_{x, x' \in C} d(x, x')$$

$D(\mathcal{C}) \in \mathbb{R}_+$ et doit être minimisé.

Indices de stabilité :

Distance moyenne : Average Distance, cet indice est défini par :

$$AD(K) = \frac{1}{nd} \sum_{j=1}^n \sum_{l=1}^d \frac{1}{|C(i')||C(j)|} \left(\sum_{x \in C(j), x \in C(i)'} dist(x, x') \right)$$

$AD(K) \in \mathbb{R}_+$, si $AD(K) \rightarrow 0$, il y a forte consistance.

Distance moyenne entre les moyennes : l'indice de distance moyenne entre les moyennes est défini par :

$$ADM(K) = \frac{1}{nd} \sum_{j=1}^n \sum_{l=1}^d dist(\bar{x}_{C(j)'}, \bar{x}_{C(j)^0})$$

$ADM(K) \in \mathbb{R}_+$, si $ADM(K) \rightarrow 0$, il y a forte consistance.

	2	3	4
Dunn	0.12	0.13	0.12
ADM	0.0055	0.0008	0.44
AD	2.08	2.04	2.08

TABLE 10 – Résultats des test de stabilité interne et de validité de la classification pour le blé dur

	2	3	4
Dunn	0.11	0.11	0.11
ADM	0.28	0.23	0.18
AD	2.39	2.15	1.99

TABLE 11 – Résultats des test de stabilité interne et de validité de la classification pour le blé tendre.

3. Test d'adéquation

Test de Kolmogorov-smirnov (KS)

Soit X_1, \dots, X_n *i.i.d* de fonction de répartition F . Le test de K-S permet de comparer la distribution observée avec une autre, le principe est simple :

on mesure l'écart maximum qui existe entre la fonction de répartition F observée et la fonction de répartition théorique F_0 .

On veut tester l'hypothèse $H_0 : F = F_0$ contre $F \neq F_0$.

On rejette H_0 si $D_n \geq d_{n,\alpha}$ avec $D_n = \sup_{x \in R} |F_n(x) - F_0(x)|$ où $F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{X_i \leq x}$ et $d_{n,\alpha}$ correspond la valeur tabulée.

Test d'Anderson-Darling (AD)

Soit X_1, \dots, X_n *i.i.d* de fonction de répartition F . L'AD test est un test d'adéquation qui vise à tester l'hypothèse : $H_0 : F = F_0$ contre $F \neq F_0$.

Avec la statistique de test : $D_n = n \int_{-\infty}^{+\infty} (F_n(x) - F(x))^2 \frac{f_0(x)}{F_0(x)(1-F_0(x))} dx$.

L'hypothèse H_0 est rejetée si $D_n \geq d_{n,\alpha}$ et $d_{n,\alpha}$ correspond à la valeur tabulée.

Test de Cramer Von Mises (CVM)

Tout comme les deux tests précédents, il vise à comparer deux distributions et repose sur la somme des carrés des écarts en valeur absolue entre les deux fonctions de répartitions. En notant S_d^2 cette somme, la statistique de ce test est donnée par :

$$T = \frac{nmS_d^2}{n+m}$$

avec n et m les nombres d'observations des deux groupes. A noter que les deux groupes ont la même taille dans notre cas. Il vise à comparer des distributions. Étant donnée qu'il n'existe pas directement de valeurs tabulées pour des échantillons de grandes tailles, il convient de se référer à des approximations

4. Estimation par Maximum de Vraisemblance

Soit $X_1 \dots X_n$ iid de loi X dont la densité est : $f(x, \theta)$. La fonction de vraisemblance s'écrit :

$\mathcal{L}(x_1 \dots x_n, \theta) = \prod_{i=1}^n f_{X_i}(x_i, \theta) = \prod_{i=1}^n f(x_i, \theta)$, l'estimation de θ est donnée par la maximisation de la fonction : $\mathcal{L}(x_1 \dots x_n, \theta)$ ou de $\ln(\mathcal{L}(x_1 \dots x_n, \theta))$

EMV pour la loi de Weibull

On rappelle la densité de probabilité de la loi de Weibull à deux paramètres :

$W(a, \alpha) \rightsquigarrow f(x, a, \alpha) = \frac{a}{\alpha} \left(\frac{x}{\alpha}\right)^{a-1} e^{-\left(\frac{x}{\alpha}\right)^a}$, avec $x, \alpha, a > 0$, le logarithme de la fonction de vraisemblance s'écrit :

$$\ln(\mathcal{L}(x_1, \dots, x_n, a, \alpha)) = n \ln(a) - n \ln(\alpha) - \sum_{i=1}^n \left(\frac{x_i}{\alpha}\right)^a + (a-1) \sum_{i=1}^n \ln\left(\frac{x_i}{\alpha}\right) = \ln\left(\frac{a}{\alpha}\right)^n - \left(\frac{1}{\alpha}\right)^a \sum_{i=1}^n x_i^a + (a-1) \sum_{i=1}^n \ln\left(\frac{x_i}{\alpha}\right)$$

$$\begin{cases} \frac{\partial}{\partial a} \ln(\mathcal{L}(x_1, \dots, x_n, a, \alpha)) = 0 \\ \frac{\partial}{\partial \alpha} \ln(\mathcal{L}(x_1, \dots, x_n, a, \alpha)) = 0 \end{cases} \iff \begin{cases} n \frac{1}{a} - \frac{1}{\alpha^a} \ln(\alpha) \sum_{i=1}^n x_i^\alpha - \frac{1}{\alpha^a} \sum_{i=1}^n x_i^\alpha \ln(x_i) + \sum_{i=1}^n \ln\left(\frac{x_i}{\alpha}\right) = 0 \\ -n \frac{1}{\alpha} - \frac{a}{\alpha^a} \sum_{i=1}^n x_i^a + (1-a) n \frac{1}{\alpha} = 0 \end{cases}$$

$$\iff \begin{cases} a = \left[\left(\sum_{i=1}^n x_i^a \ln x_i \right) \left(\sum_{i=1}^n x_i^a \right)^{-1} - \frac{1}{n} \sum_{i=1}^n x_i \right]^{-1} \\ \alpha = \left(\frac{1}{n} \sum_{i=1}^n x_i^a \right)^{1/a} \end{cases}$$

La suite de la résolution de ces deux équations se fait par itération, pour y parvenir, nous avons utilisé le package `fitdistrplus` du logiciel R, ce même package nous permet aussi d'obtenir les estimateurs par maximum de vraisemblance de la loi log-normale.

La loi log-normale

Une variable aléatoire X suit une loi log-normale quand son logarithme suit une loi normale

c'est à dire que $Y = \ln X$ suit une loi $\mathcal{N}(\mu, \sigma^2)$

$$\text{Log} - \mathcal{N}(\mu, \sigma^2) \rightsquigarrow f(x, \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), \text{ avec } x > 0$$

$$E(X) = E'(e^Y) = e^{(\mu + \sigma^2/2)}$$

$$V(X) = E(X^2) - (E(X))^2 = (e^{\sigma^2} - 1)e^{(2\mu + \sigma^2)}.$$

5. Monte Carlo

Soit g la fonction de densité d'une variable aléatoire X . L'espérance de $g(X)$, est donné par :

$$G = E(g(X)) = \int g(x)f_X(x)dx$$

où f_X est une fonction de densité sur un support $[a, b]$. Si on prend une distribution uniforme sur $[a, b]$:

$$f_X(x) = \frac{1}{b-a}$$

L'idée est de produire un échantillon (x_1, x_2, \dots, x_N) de loi X , d'après la densité de f_X sur le support $[a, b]$, on devrait pouvoir calculer un nouvel estimateur G appelé estimateur de Monte-Carlo. A partir de cet échantillon, la LFGN suggère de construire cet estimateur à partir de la moyenne empirique :

$$\tilde{g}_N = \frac{1}{N} \sum_{i=1}^N g(x_i),$$

et cet estimateur est sans biais dans le sens où :

$$E(\tilde{g}_N) = G = E(g(X))$$

6. Résultats de classifications et de tests d'ajustement

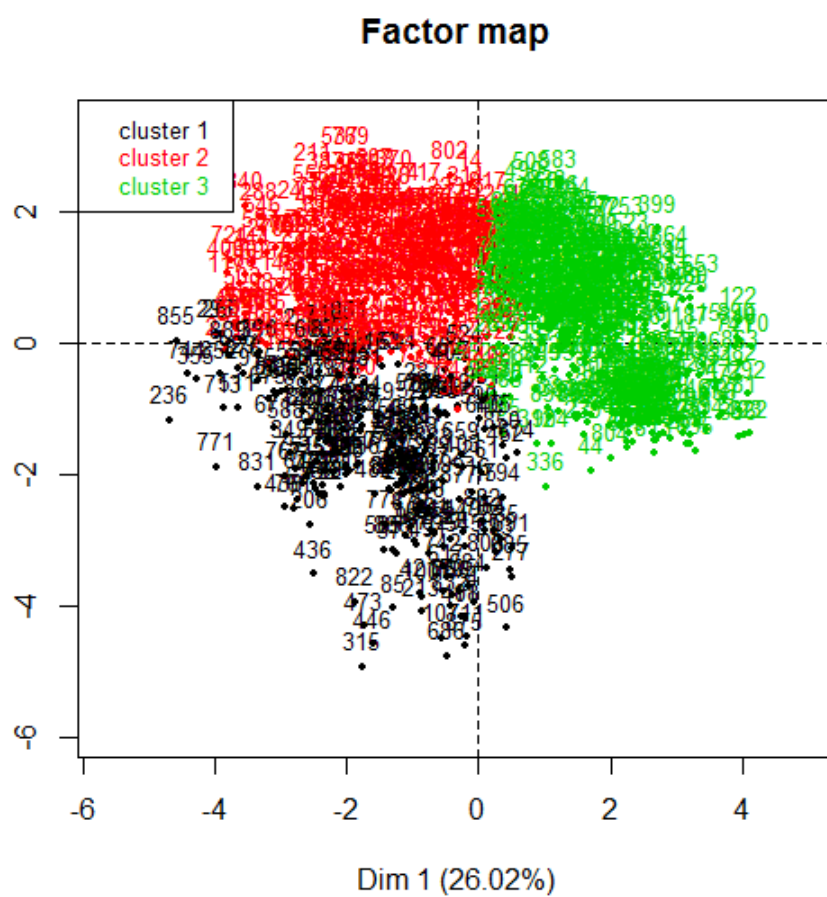


FIGURE 10 – Carte factorielle du blé dur

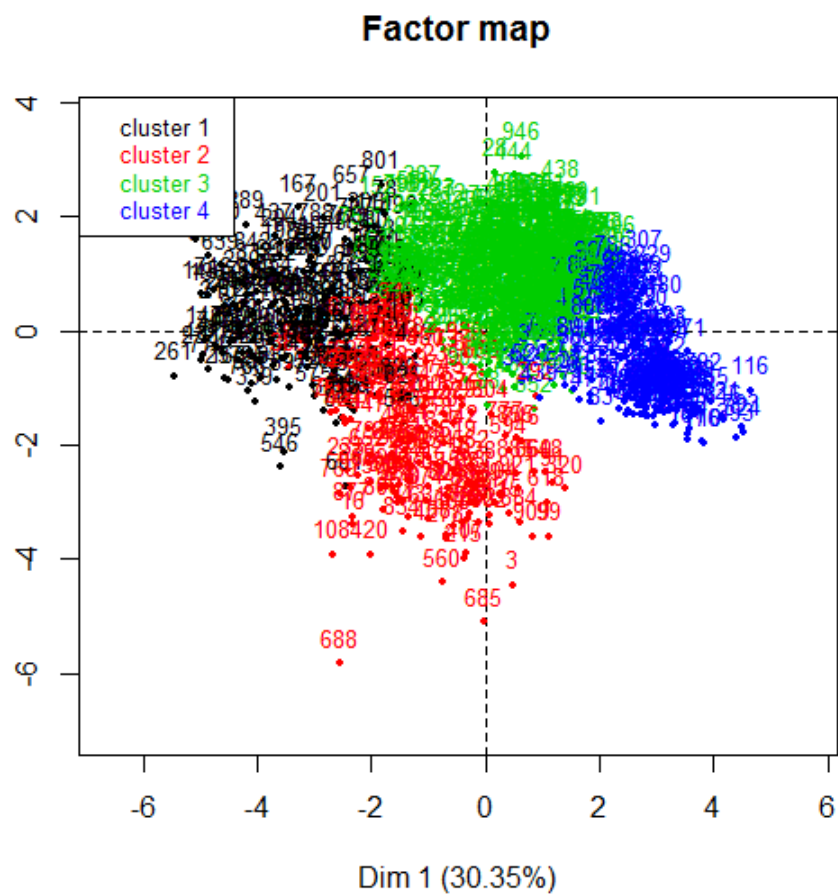


FIGURE 11 – Carte factorielle du blé tendre

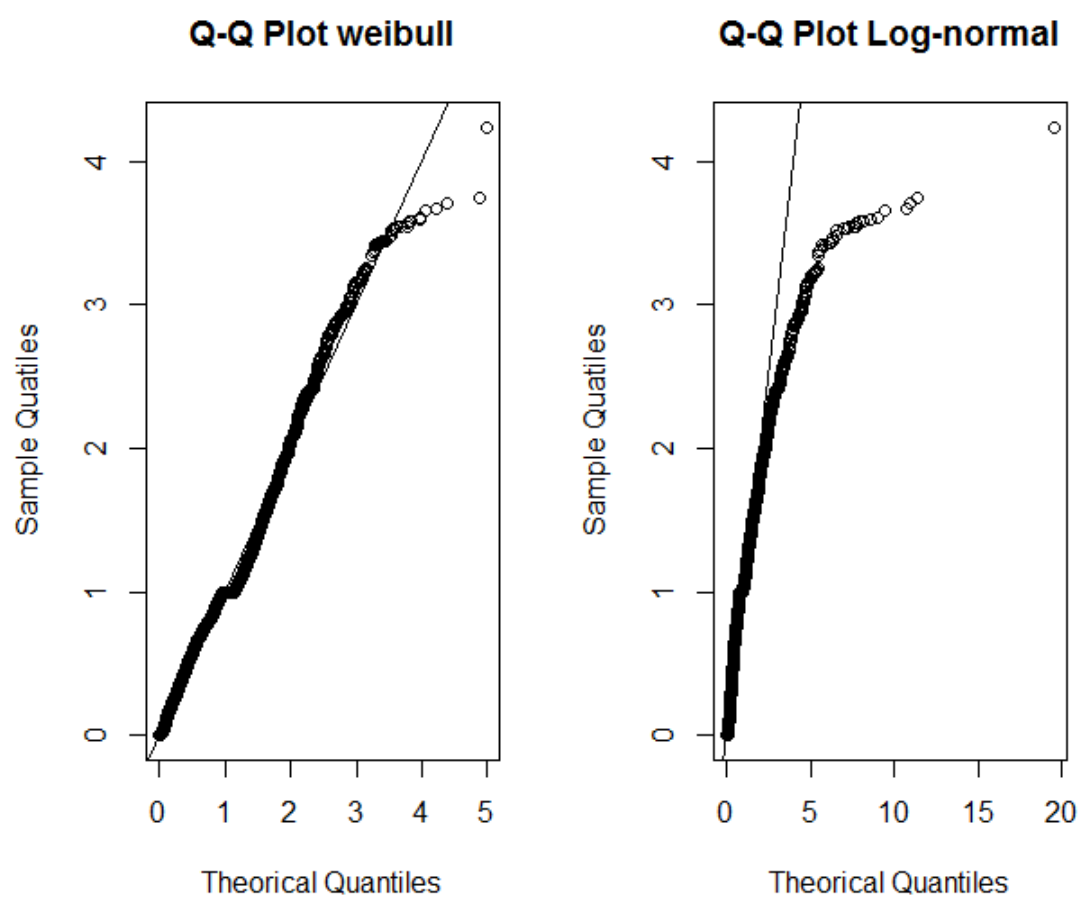


FIGURE 13 – Q-Q plot blé dur : classe 1

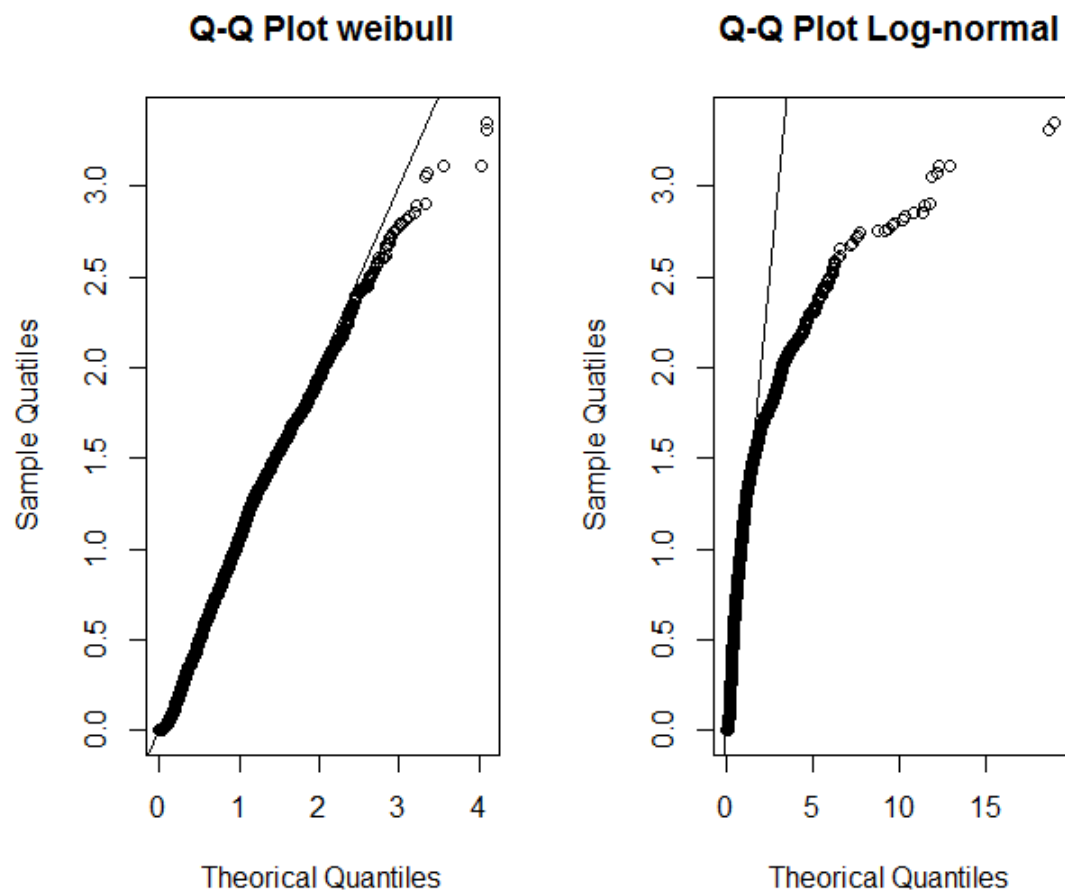


FIGURE 14 – Q-Q plot blé dur : classe 2

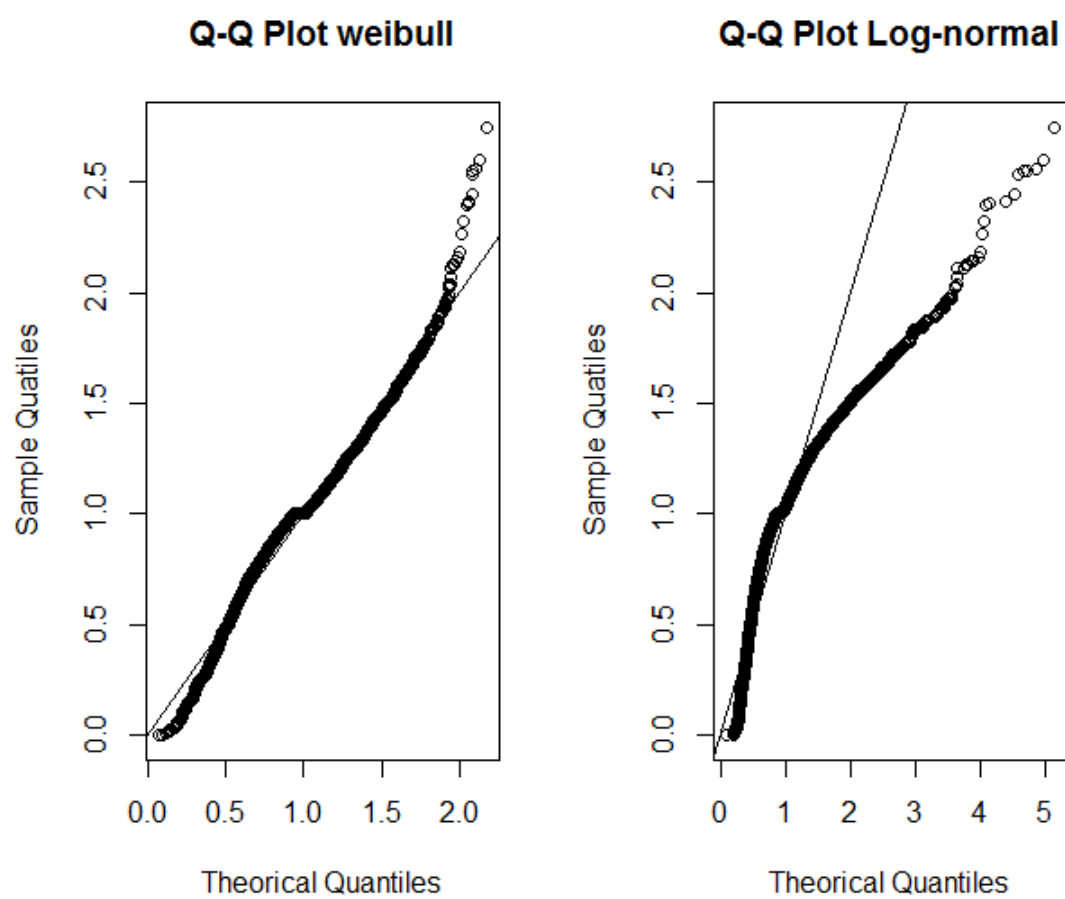


FIGURE 15 – Q-Q plot blé dur : classe 3

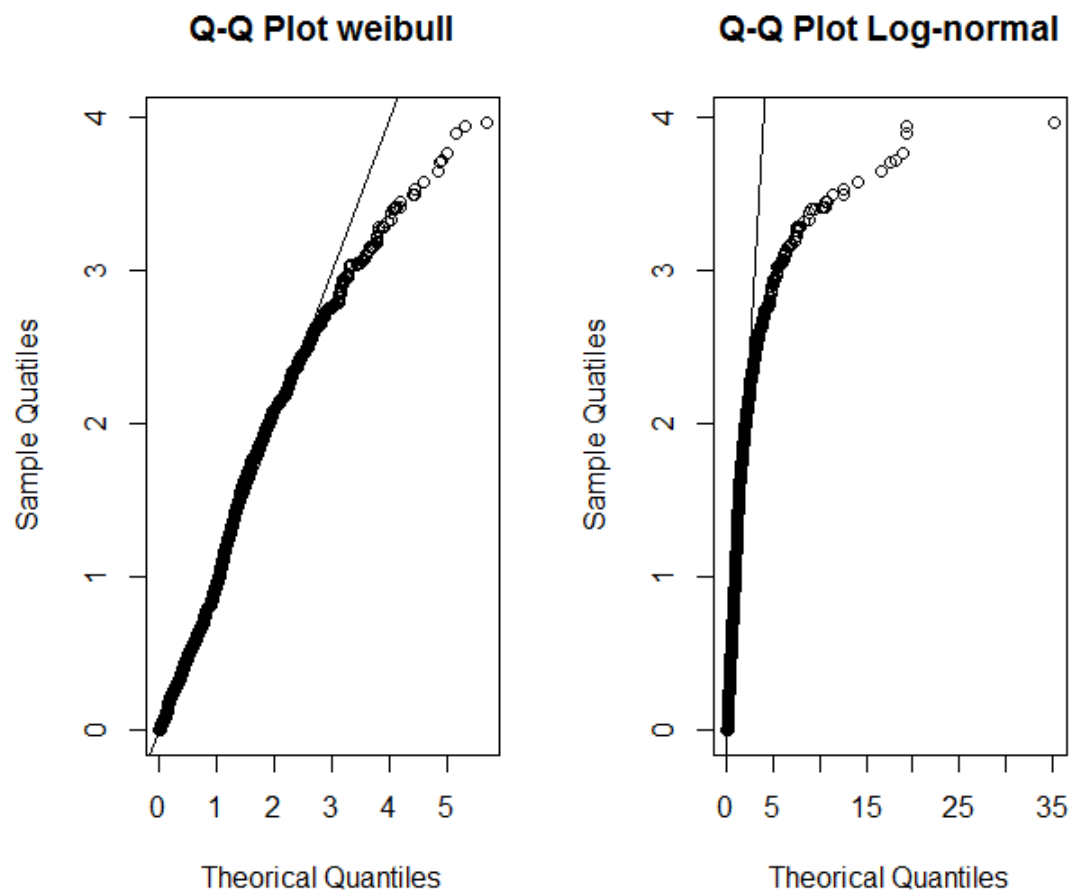


FIGURE 16 – Q-Q plot blé tendre : classe 1

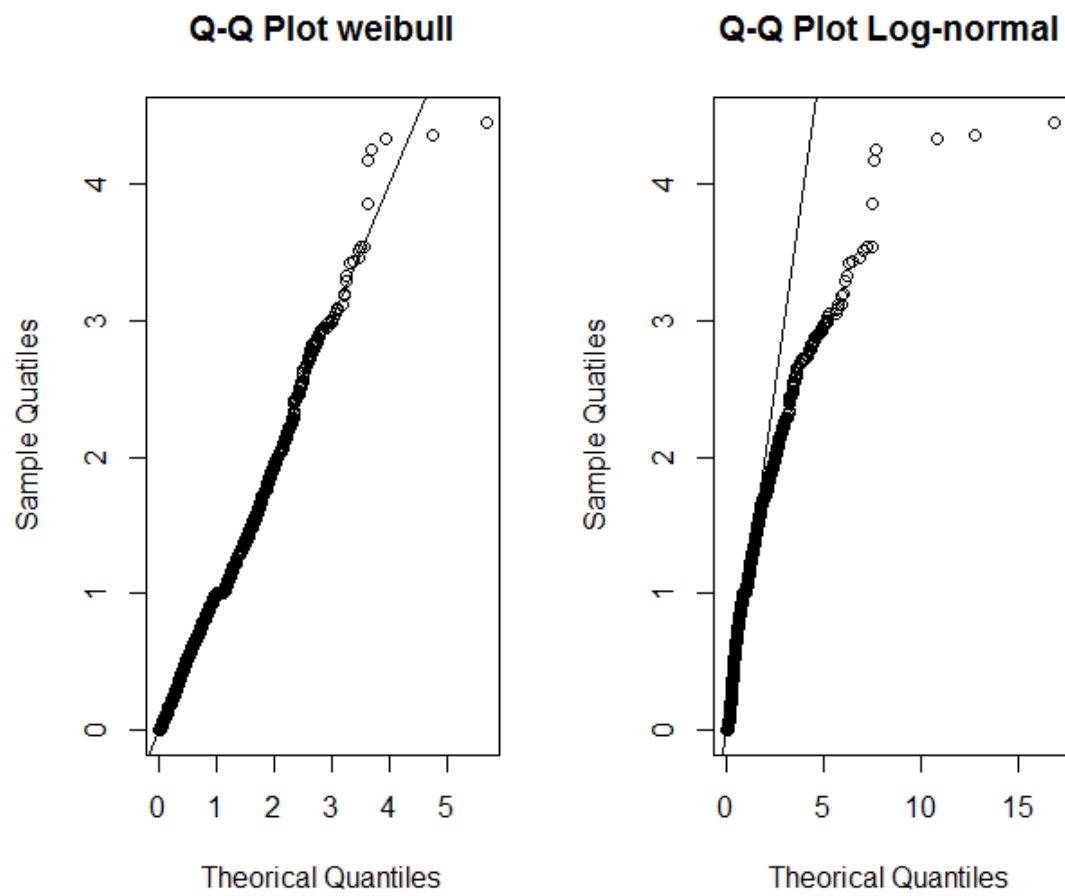


FIGURE 17 – Q-Q plot blé tendre : classe 2

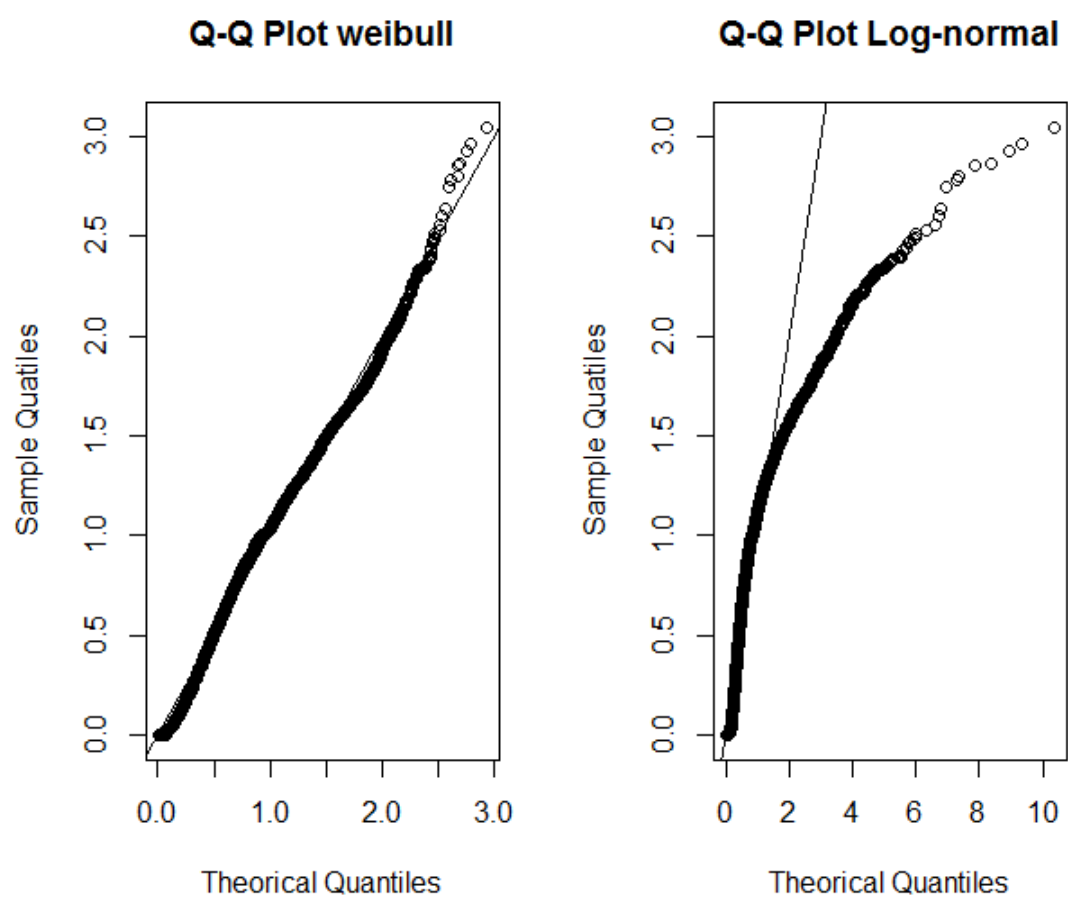


FIGURE 18 – Q-Q plot blé tendre : classe 3

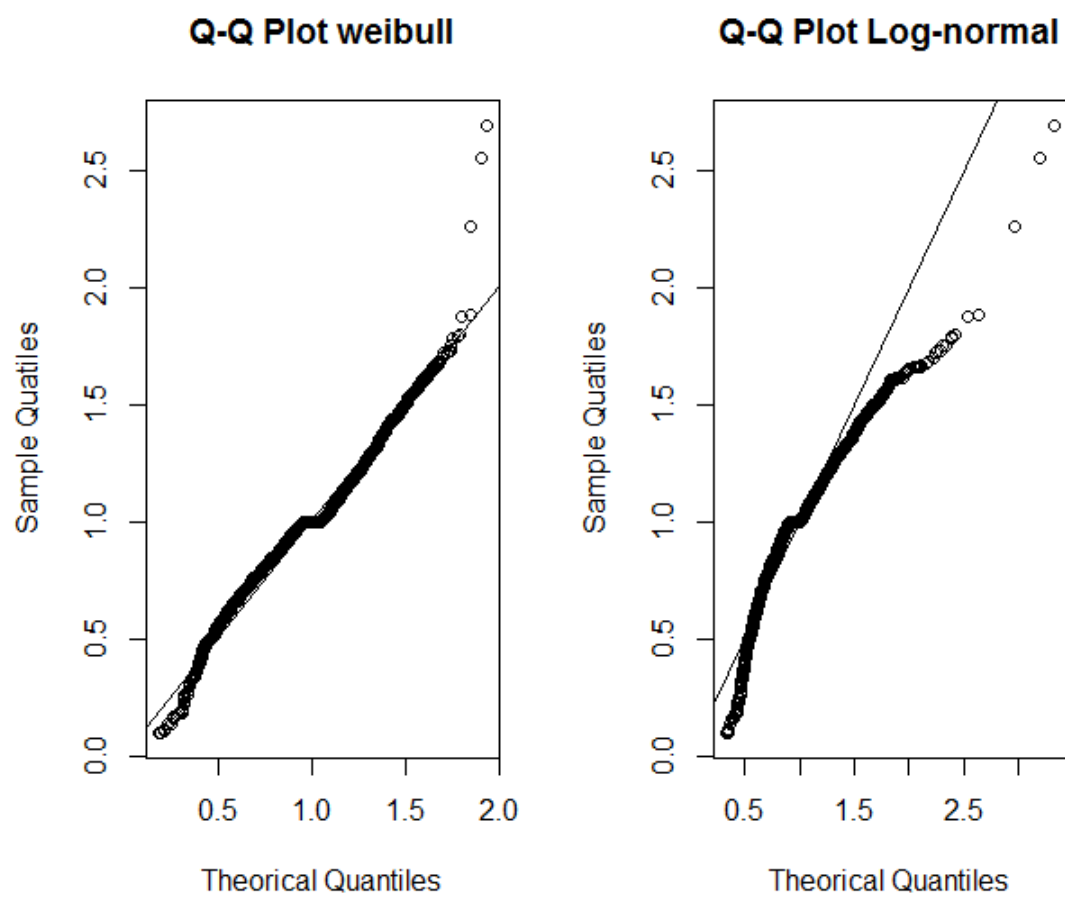


FIGURE 19 – Q-Q plot blé tendre : classe 4

7. Résultats des CVM test pour le blé dur et le blé tendre

Blé dur Weibull	classe 1	classe 2	classe 3
D	2.795	1.862	8.2261
p-value	0.025	0.023	0.015

TABLE 12 – Test de CVM pour le blé dur

Blé tendre Weibull	classe 1	classe 2	classe 3	classe 4
D	1.056	0.990	5.4173	5.2829
p-value	0.026	0.015	0.042	0.056

TABLE 13 – Test de CVM pour le blé dur

8. Résultats de K-S test

Blé dur Weibull	classe 1	classe 2	classe 3
D	0.070	0.044	0.068
p-value	0.032	0.045	0.035

TABLE 14 – Test de KS pour le blé dur

Blé tendre Weibull	classe 1	classe 2	classe 3	classe 4
D	0.050	0.040	0.060	0.075
p-value	0.025	0.046	0.052	0.043

TABLE 15 – Test KS pour le blé tendre

Orge Weibull	classe 1	classe 2	classe 3
D	0.029	0.050	0.067
p-value	0.03	0.034	0.03

TABLE 16 – Test de KS pour Orge :