N°__/__/__/__/__/__/__/__/__/__/

THÈSE

pour obtenir le grade de

Docteur

de

l'Institut des Sciences et Industries du Vivant et de l'Environnement (Agro Paris Tech)

Spécialité : Mathématiques Appliquées et Applications des Mathématiques

présentée et soutenue publiquement
par

Amélie CREPET

12 décembre 2007

STATISTIQUE BAYESIENNE ET MONTE-CARLO DE SECOND ORDRE POUR L'EVALUATION DES RISQUES MICROBIOLOGIQUES

LE CAS DE *LISTERIA MONOCYTOGENES* DANS LES PRODUITS DE IV^{EME} GAMME.

Directeur de thèse : ***Frédéric CARLIN***
Codirecteur(s) de thèse : ***Catherine DERVIN***

Travail réalisé : INRA, UR1204 Méthodologies d'analyse de risque alimentaire, F-75231 Paris

Devant le jury :

Mme Marie CORNU, Chargée de Recherche (HDR), **AFSSA** ……..**Rapporteur**
M. Jean-Noël BACRO, Professeur, **UNIVERSITE MONTPELLIER II** ……..**Rapporteur**
Mme Valérie STAHL, Chef de Projet, **AERIAL** ……**Examinateur**
M. Jean-Jacques DAUDIN, Professeur, **AGROPARISTECH**…… ……**Examinateur**
Mme Catherine DERVIN, Chargée de Recherche, **INRA-Mét@risk** ……..**Co-directeur**
M. Frédéric CARLIN, Directeur de Recherche, **INRA-UMR 408**…..……..**Directeur**

L'Institut des Sciences et Industries du Vivant et de l'Environnement (Agro Paris Tech) est un Grand Etablissement dépendant du Ministère de l'Agriculture et de la Pêche, composé de l'INA PG, de l'ENGREF et de l'ENSIA (décret n° 2006-1592 du 13 décembre 2006)

Remerciements

Je tiens à remercier tous ceux qui d'une façon ou d'une autre ont contribué à mon travail de thèse et il y en a !

Tout d'abord un grand merci à mes deux directeurs de thèse qui m'ont accompagnée pendant ces années de travail. Frédéric, cela a été un vrai plaisir d'être ta doctorante. Nos éloignements, aussi bien sur le plan disciplinaire que sur le plan géographique n'ont jamais été des obstacles à un travail efficace, passionant et enrichissant mais bien au contraire ! Catherine, merci de m'avoir toujours conseillée, encouragée et soutenue.

Merci à Marie Cornu et à Jean-Noël Bacro de me faire l'honneur d'être les rapporteurs de ce travail. Je remercie également Jean-Jacques Daudin et Valérie Stahl de se joindre à eux pour la lecture de ce manuscrit.

Cette thèse n'aurait pas pu exister sans la participation financière de l'ACTIA, du Ministère de l'Agriculture et des deux départements MIA et MICA de l'INRA. Merci également aux partenaires : le syndicat de IV^{ème} gamme, Aérial, ADRIA Quimper, ADRIA Normandie et l'ENVA. Je tiens à remercier les personnes représentant ces institutions et qui ont participé à la réalisation de ces travaux notamment au cours des réunions de comités de pilotage. Ces réunions ont été très enrichissantes tant au niveau scientifique qu'au niveau humain. Merci tout particulièrement à Valérie Stahl responsable du projet qui a toujours su mener ces rencontres de façon efficace et pragmatique !

Je remercie tous les membres du comité de thèse qui m'ont accordé leur temps et qui m'ont donné de nombreux conseils. Les séances ont toujours été très animées, les discussions passionnées et l'ambiance très bonne.

Je voudrais alors remercier toutes les personnes qui m'ont conseillée, aiguillée dans mes recherches : Isabelle Albert pour m'avoir bien aidée notamment pour débugger mes programmes en OpenBugs, logiciel qui pour les non initiés "bug" vraiment beaucoup ! Chantal Guihenneuc-Jouyaux et Stéphan Clémençon pour leurs précieux conseils. Un regret cependant : celui de les avoir rencontrés si tardivement. Je remercie également Stéphan de m'avoir

introduite à l'université de Nanterre dans laquelle j'ai effectué une année d'ATER (sans toutefois quitter mon unité préférée, Mét@risk).

Un grand merci à tous les membres (sans oublier ceux qui n'ont fait que passer) de l'unité Mét@risk et ceux de l'unité d'Avignon que je n'ai pas cités mais qui ont beaucoup compté.

Merci également à toutes les personnes avec qui j'ai pu échanger dans des réunions, des congrès... Un clin d'oeil spécial à Olivier Couvert rencontré à Lille lors de nos débuts (pas faciles) avec le logiciel Sym'Previus.

Mais je ne me serais certainement jamais lancée dans une thèse si je n'avais pas rencontré Jean-Charles Leblanc, travaillant alors à l'INRA et qui m'a proposé de faire mon stage de maîtrise avec lui et Patrice Bertail. Je le remercie tout particulièrement pour m'avoir fait connaître le monde de la recherche et de m'en avoir transmis le goût par son enthousiasme. C'est également à cette époque que j'ai rencontré mon cher directeur d'unité Mét@risk (unité qui n'existait d'ailleurs pas encore) Philippe Verger que je remercie pour m'avoir "gardée" si "longtemps". Deux autres personnes ont également beaucoup compté dans ce début d'aventure, Michel et Jessica. Michel, parce que c'est aussi beaucoup grâce à lui que j'écris ces pages aujourd'hui et Jessica parce qu'elle m'a toujours épaulée et que nous avons partagé tant de choses pendant nos années de stages et de thèse : nos différents bureaux, nos problèmes de modélisation (enfin surtout les miens ! !), nos soucis quotidiens (quand on avait encore une vie à l'extérieur !), bref.. MERCI Jessica !

Merci également à Sandrine, il faut toujours avoir une sociologue dans la poche on ne sait pas trop à quoi ça sert au début et puis après on ne peut plus s'en passer ! Une autre personne dont je n'aurais pu me passer, Nadine.

Un grand merci à mes parents, à mes frères Ludovic et Fabien, à Aurel et Isa pour leur soutien inconditionnel.

Merci à tous mes amis qui m'ont soutenue, encouragée et supportée pendant cette période : Milena, Jérôme, Laure, Charlotte, Emilie B., Emilie C., Mounir, Audrey, David, Bibi, Olivia, Antoine, Elise, Nicolas, Nicolo, Patricia, Dave, Eric, Fabien et tous les autres...

Comités de pilotage et de thèse

Membres du comité de pilotage ACTIA 03.13

ALBERT Isabelle, Met@Risk-INRA, Paris
CARLIN Frédéric, UMR-A408-INRA, Avignon
CERF Olivier, ENVA, Maison Alfort
CREPET Amélie, Met@Risk-INRA, Paris
CONTAT GRANPERRIN, représentante syndicat IV^{ème} gamme
DENIS Catherine, ADRIA Normandie, partenaire
DERVIN Catherine, Met@Risk-INRA, Paris
HUCHET Véronique, ADRIA Développement, partenaire
MAJOU Didier, ACTIA
POURQUIE Jacques, DGAL, parrain du projet, ministère de l'agriculture
RIDRIMONT Valérie, représentante syndicat IV^{ème} gamme
STAHL Valérie, Aérial, chef de file du projet
TIXIER Françoise, représentante syndicat IV^{ème} gamme

Membres du comité de thèse

ALBERT Isabelle, Met@Risk-INRA, Paris
CARLIN Frédéric, UMR-A408-INRA, Avignon
CERF Olivier, ENVA, Maison Alfort
CREPET Amélie, Met@Risk-INRA, Paris
DERVIN Catherine, Met@Risk-INRA, Paris
DAUDIN Jean-Jacques, INAP-G, Paris
ROSSO Laurent, AFSSA
STAHL Valérie, Aérial, chef de file du projet

Table des matières

Remerciements **1**

Comités de pilotage et de thèse **3**

Table des matières **4**

Table des figures **9**

Liste des tableaux **11**

Liste d'acronymes et abbréviations **14**

1 Chapitre introductif **15**

1.1 Introduction 15

1.2 Justification du sujet 18

1.2.1 Un danger avéré 18

1.2.2 Nécessité d'évaluer l'exposition à *L. monocytogenes* et de caractériser les risques de listériose par consommation de salade de IV^{ème} gamme 21

1.3 Enjeux scientifiques 23

1.3.1 Construire un modèle d'évaluation de l'exposition à *L. monocytogenes* et de caractérisation des risques de listériose par consommation de salade de IV^{ème} gamme 23

1.3.2 Intégrer dans l'évaluation des risques alimentaires des méthodes statistiques permettant de prendre en compte séparément la variabilité et l'incertitude 25

1.4 Organisation du manuscript 35

2 Estimation de la contamination **38**
2.1 Introduction . . . 38
2.2 La méthode du "nombre le plus probable" dans le cas de dilutions en série . 41
2.2.1 Principe de la méthode des dilutions . . . 41
2.2.2 Estimation de la concentration en micro-organismes . . . 41
2.3 Structure des données de contamination microbiologique et notations . . . 43
2.4 Méthode d'estimation et génération de valeurs de log-concentration . . . 45
2.4.1 Modèle de mélange de deux lois normales . . . 46
2.4.2 Méthode d'estimation des paramètres et des hyperparamètres du modèle 47
2.4.3 Formalisation bayésienne du modèle . . . 49
2.4.4 Génération de valeurs de log-concentration . . . 51
2.5 Application à *Listeria monocytogenes* dans les produits végétaux . . . 53
2.5.1 Description des données . . . 53
2.5.2 Réalisation et contrôle de la convergence . . . 54
2.5.3 Résultats . . . 56
2.6 Discussion . . . 57

3 Estimation des paramètres de croissance **62**
3.1 Introduction . . . 62
3.2 Rappel sur les modèles de croissance bactérienne . . . 65
3.2.1 Modèle primaire de croissance . . . 66
3.2.2 Modèle secondaire de croissance . . . 67
3.3 Description des données . . . 71
3.4 Estimation des distributions des paramètres de croissance . . . 79
3.4.1 Le modèle $M1$ . . . 79
3.4.2 Le modèle $M2$: introduction d'une structure hiérarchique . . . 82
3.5 Validation . . . 86
3.5.1 Comparaison entre les valeurs observées et les valeurs prédites . . . 86
3.5.2 Utilisation d'un critère d'adéquation . . . 86
3.5.3 Sélection de modèles : le Critère d'Information de Déviance (DIC) . 89
3.6 Réalisation pratique et contrôle de la convergence . . . 89
3.7 Résultats . . . 90
3.7.1 Statistiques des distributions *a posteriori* des paramètres des différents modèles . . . 90

3.7.2 Résultats de la validation 93
3.7.3 Résultats de la sélection de modèles par le Critère d'Information de Déviance 97
3.8 Discussion 98

4 Simulations de la contamination 101
4.1 Introduction 101
4.2 Procédé de fabrication et circuit logistique des salades de IV^{ème} gamme . . . 103
4.3 Modèle de simulation de la contamination des sachets à l'instant de la consommation 106
4.3.1 Etapes du modèle 106
4.3.2 Modèle d'estimation de la contamination de la matière première . . . 108
4.3.3 Méthode d'estimation de la réduction en micro-organismes au cours du lavage à l'eau chlorée et du lavage à l'eau non chlorée 109
4.3.4 Modèle de simulation de la croissance de *L. monocytogenes* dans chaque étape du circuit logistique 110
4.4 Nature et origine des données 112
4.4.1 Contamination de la matière première 112
4.4.2 Réduction en micro-organismes au cours du lavage à l'eau chlorée et du lavage à l'eau non chlorée 113
4.4.3 Durées et températures des différentes étapes du circuit logistique des salades en sachet 115
4.5 Distributions des paramètres du modèle 118
4.5.1 Distribution des paramètres du modèle de contamination de la matière première 118
4.5.2 Distribution de la réduction en micro-organismes au cours du lavage à l'eau chlorée et du lavage à l'eau non chlorée 118
4.5.3 Distribution de la quantité de salades contenues dans les sachets . . . 119
4.5.4 Distributions des durées et des températures de chaque étape du circuit logistique 120
4.6 Réalisation des simulations 126
4.7 Résultats 128
4.7.1 Contamination de la matière première entrant dans l'usine 128

4.7.2 Contamination de la matière première après lavage à l'eau chlorée et à l'eau non chlorée . . . 129
4.7.3 Contamination à la sortie de la chaîne de fabrication . . . 130
4.7.4 Contamination à chaque étape du circuit logistique . . . 131
4.7.5 Test de vieillissement . . . 133
4.7.6 Scénarios pour le paramètre y_{max} . . . 134
4.8 Discussion . . . 135
4.8.1 Difficultés de la reproduction d'un procédé de fabrication et d'un circuit logistique . . . 136
4.8.2 De l'influence du volume de données sur le niveau d'incertitude : exemple des données de contamination . . . 138
4.8.3 Effet du chlore sur la contamination des salades de IV^{ème} gamme . . . 139
4.8.4 Peut-on valider le modèle ? . . . 139

5 Evaluation de l'exposition et des risques 148
5.1 Introduction . . . 148
5.2 Les données de consommation disponibles en France . . . 150
5.2.1 L'enquête individuelle nationale sur les consommations alimentaires (INCA) . . . 151
5.2.2 Le panel Sécodip . . . 153
5.3 Utilisation de l'enquête INCA et du panel Sécodip . . . 154
5.3.1 Statistiques de la consommation de salade . . . 154
5.3.2 Proportion de salades en sachet consommées par rapport aux salades non transformées . . . 158
5.3.3 Poids de redressement de la population INCA . . . 159
5.4 Evaluation de l'exposition . . . 160
5.4.1 Définition de l'exposition . . . 160
5.4.2 Evaluation de l'exposition à partir des sachets de salade . . . 160
5.4.3 Evaluation de l'exposition par simulation de la croissance des bactéries directement dans les quantités consommées . . . 162
5.5 Caractérisation des dangers . . . 162
5.6 Caractérisation des risques . . . 164
5.6.1 Risque individuel par acte de consommation . . . 164
5.6.2 Risque individuel annuel . . . 165

5.6.3 Nombre de cas annuel de listériose du à la consommation de salade en sachet . . . 165
5.7 Résultats . . . 166
5.7.1 Exposition estimée à *L. monocytogenes* par consommation de quantités de salades de IV^{ème} gamme contaminées . . . 167
5.7.2 Risque estimé de listériose par acte de consommation de salade de IV^{ème} gamme . . . 168
5.7.3 Risque estimé de listériose annuel . . . 169
5.7.4 Nombre estimé de cas de listériose en France par consommation de salade de IV^{ème} gamme . . . 175
5.8 Discussion . . . 175
5.8.1 Intérêt d'utiliser les données individuelles de consommation et les caractéristiques socio-démographiques des individus . . . 176
5.8.2 Une estimation du risque de listériose par consommation de salade de IV^{ème} gamme relativement élevé . . . 177
5.8.3 Un nombre de cas de listériose plausible ? . . . 179

6 Discussion générale **183**
6.1 Limites des travaux et pistes de recherche . . . 183
6.2 Statistique et évaluation des risques : un dispositif complexe . . . 184
6.2.1 Pluridisciplinarité scientifique . . . 185
6.2.2 Collaboration avec l'industrie . . . 186
6.2.3 Communication avec les gestionnaires . . . 186
Bibliographie . . . 187

Annexes **207**

Table des figures

1.1 Composantes de l'analyse des risques 16
1.2 Modèle d'évaluation de l'exposition et des risques 24
1.3 Formalisme bayésien 27
1.4 Formalisme bayésien pour un modèle hiérarchique 28
1.5 Principe des méthodes de Monte Carlo par Chaîne de Markov 29
1.6 Algorithme de Gibbs 30
1.7 Simulation de Monte-Carlo de second ordre 33
1.8 Simulation de Monte-Carlo à deux dimensions 34

2.1 Fonctions de répartition de la contamination des produits végétaux par *L. monocytogenes* 61

3.1 Courbe de croissance bactérienne 68
3.2 Evolution du taux de croissance en fonction de la température 70
3.3 Graphe acyclique orienté du modèle $M1$ d'estimation des distributions des paramètres de croissance de *L. monocytogenes* dans la salade 81
3.4 Graphe acyclique orienté des modèles $M2$ et $M2_{obs}$ d'estimation des paramètres de croissance de *L. monocytogenes* dans la salade 84
3.5 Tracés des valeurs observées et des valeurs prédites des trois modèles de croissance de *L. monocytogenes* dans la salade 95

4.1 Procédé de fabrication des salades de IV^{ème} gamme 104
4.2 Etapes du modèle d'estimation de la contamination des salades en sachet à l'instant de la consommation, données utilisées et quantité à estimer 107
4.3 Exemple de la réduction en micro-organismes après lavage 110
4.4 Histogramme de la réduction en micro-organismes après lavage 120
4.5 Proportions des poids des sachets de salade vendus dans le commerce au détail 121

4.6 Concentration des salades contenues dans les sachets simulés selon la teneur en chlore de l'eau de lavage (chlorée/non chlorée) 132
4.7 Comparaison entre concentrations observées et concentrations simulées par le modèle selon différents scénarios pour le paramètre $y_{\max}$ 147

5.1 Fonction de répartition du risque annuel de listeriose 173
5.2 Histogrammes des températures moyennes et des durées du circuit logistique des quantités consommées dont la concentration est supérieure à 6 log ufc/g. 182

Liste des tableaux

2.1 Description des données de contamination des produits végétaux par *L. monocytogenes* 55
2.2 Distributions *a posteriori* des paramètres et hyperparamètres du modèle de mélange 56
2.3 Distributions *a posteriori* des paramètres *mc* et *mp* du modèle de mélange . 57
2.4 Contamination par *L. monocytogenes* des produits végétaux 57

3.1 Donnés de croissance 72
3.2 Distributions *a priori* des paramètres du modèle M1 83
3.3 Distributions *a priori* des hyperparamètres des modèles M2 et $M2_{\lambda_{obs}}$ . . . 85
3.4 Distribution *a posteriori* des paramètres de la température minimale de croissance 91
3.5 Distribution *a posteriori* des paramètres de la température optimale de croissance 91
3.6 Distribution *a posteriori* des paramètres de la température maximale de croissance 92
3.7 Distribution *a posteriori* des paramètres de la concentration maximale 92
3.8 Distribution *a posteriori* des paramètres de la constante K 93
3.9 Distribution *a posteriori* des paramètres du taux de croissance optimal . . . 94
3.10 Validation, à l'aide d'un critère d'adéquation partie 1 95
3.11 Validation à l'aide d'un critère d'adéquation partie 2 96

4.1 Description des données de la réduction en micro-organismes par lavage à l'eau chlorée ou non chlorée 114
4.2 Corrélation des rangs de Spearman intra-étapes 117

4.3 Distributions *a posteriori* des paramètres de contamination de la matière première . . . 119
4.4 Distributions des durées et des températures des étapes du circuit logistique de la salade en sachet . . . 122
4.5 Durées des grandes étapes (livraison, vente, consommation) du circuit logistique 123
4.6 Températures des étapes du circuit logistique . . . 123
4.7 Contamination par *L. monocytogenes* de la matière première . . . 129
4.8 Contamination par *L. monocytogenes* de la matière première après lavage . . 129
4.9 Contamination par *L. monocytogenes* des salades des sachets contaminés après lavage . . . 130
4.10 Contamination par *L. monocytogenes* des salades des sachets contaminés après chaque étape du circuit logistique . . . 143
4.11 Contamination par *L. monocytogenes* des salades des sachets contaminés à l'instant de la première consommation . . . 144
4.12 Contamination par *L. monocytogenes* des salades des sachets contaminés après un test de vieillissement . . . 144
4.13 Contamination par *L. monocytogenes* des salades des sachets contaminés à la première consommation selon différents scénarios pour $y_{\max}$ . . . 145
4.14 Comparaison entre concentrations observées et concentrations simulées par le modèle $M2$ selon différents scénarios pour le paramètre $y_{\max}$ . . . 146

5.1 Quantités de salades consommées par les différentes sous-populations de l'enquête INCA . . . 156
5.2 Fréquence par semaine de consommation de salades des différentes sous-populations de l'enquête INCA . . . 156
5.3 Proportions des différentes sous-populations dans INCA et dans la population française et poids de redressement . . . 157
5.4 Paramètre r du modèle dose-réponse exponentiel pour la population sensible et saine . . . 164
5.5 Exposition à *L. monocytogenes* par consommation de quantités contaminées de salades de IV^{ème} gamme . . . 170
5.6 Risque de listériose par acte de consommation de salades de IV^{ème} gamme . 171
5.7 Risque de listériose lié à la consommation annuelle de salades de IV^{ème} gamme 172

5.8 Risque de listériose des différentes sous-populations lié à la consommation annuelle de salades de IV^{ème} gamme 174
5.9 Nombre de cas de listériose estimé dans la population française par consommation de salade de IV^{ème} gamme 176

Liste d'acronymes et abbréviations

Actia : Association de coordination technique pour l'industrie agroalimentaire
AFSSA : Agence française de sécurité sanitaire des aliments
AQR : Analyse quantitative des risques
AQRM : Analyse quantitative des risques microbiologiques
Cemagref : Institut de recherche pour l'ingénierie de l'agriculture et de l'environnement
CI : intervalle de crédibilité
CREDOC : Centre de recherche pour l'étude et l'observation des conditions de vie
DGAL : Direction générale de l'alimentation
DGCCRF : Direction générale de la concurrence, de la consommation et de la répression des fraudes
EFSA : Autorité européenne de sécurité des aliments
FAO : Food agricultural organisation
INCA : enquête individuelle et nationale sur les consommations alimentaires
INRA : Institut national de recherche agronomique
InVS : Institut de Veille Sanitaire
JECFA : Joint FAO/WHO expert committee on food additives and contaminants
JEMRA :Joint FAO/WHO meetings on microbiological risk assessment
JMPR : Joint FAO/WHO meetings on pesticide residues
MCMC : méthodes de Monte-Carlo par Chaîne de Markov
Sécodip : Société d'études de la consommation, de la distribution et de la publicité
ufc : unité formant colonies
UI : intervalle d'incertitude
OMC : Organisation mondiale du commerce
OMS : Organisation mondiale de la santé
vlm : viables *Listeria monocytogenes*

Chapitre 1

Chapitre introductif

1.1 Introduction

La sécurité sanitaire des aliments est devenue une préoccupation politique et sociale majeure en particulier depuis la crise liée à l'épidémie d'encéphalopathie spongiforme bovine (ESB) en 1996. Celle-ci a en effet révélé d'importants dysfonctionnements à la fois de la maîtrise des risques sanitaires par l'industrie bovine mais aussi du contrôle étatique (Borraz et al., 2006). La montée des préoccupations de sécurité alimentaire a suscité des décisions à l'échelon national, avec, notamment, la création de l'Agence française de sécurité sanitaire des aliments (AFSSA) en 1999. Les fonctions de cet établissement public sont la veille sanitaire, l'alerte, l'expertise et la recherche sur les risques sanitaires et nutritionnels des aliments destinés à l'homme et aux animaux. A l'échelon européen, l'Autorité européenne de sécurité des aliments (EFSA) qui a été créé quelques années après, en 2002, fournit en étroite collaboration avec les autorités nationales, des avis scientifiques indépendants ainsi qu'une communication claire sur les risques d'origine alimentaire existants et émergents. Mais, dans des échanges marqués par une tendance croissante à la mondialisation, les échelons national et européen ne sont pas les seuls où se joue la sécurité sanitaire des aliments. Pour que l'Union européenne ne soit pas placée devant un dilemme entre ses exigences dans le domaine alimentaire et la liberté des échanges, elle doit parvenir à ce que ses préoccupations sanitaires soient suffisamment prises en compte au sein de l'Organisation mondiale du commerce (OMC). En effet, dans le cadre des accords de Marrakech (1994) créant l'Organisation Mondiale du Commerce (OMC), l'accord sur l'application des mesures sanitaires et phytosanitaires (SPS) est conclu avec pour objectif spécifique de limiter le recours à des entraves injustifiées aux échanges. L'accord SPS détermine les conditions dans lesquelles un Etat peut

adopter et mettre en œuvre des mesures sanitaires (santé animale, sécurité alimentaire) ou phytosanitaires (protection des végétaux) ayant une incidence directe ou indirecte sur le commerce international. Il recommande que tous les membres de l'OMC s'inspirent des normes et recommandations du Codex Alimentarius. Un Etat peut choisir un niveau de protection, qu'il juge approprié, plus élevé que les normes codex. Afin d'éviter les décisions arbitraires, ce niveau « approprié » de protection est soumis à certains critères définis dans l'article 5 de l'Accord SPS, en particulier la nécessité d'une justification par une évaluation scientifique des risques et la prise en compte des effets négatifs sur le commerce. Ainsi les normes élaborées au sein de la commission du Codex Alimentarius créée en 1963 par l'Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO) et à l'Organisation mondiale de la santé (OMS) sont devenues la principale référence internationale en matière d'innocuité des denrées alimentaires. Les avis sont transmis à la commission du Codex Alimentarius par les comités d'experts. Il existe trois comités d'experts suivant le type de contaminant : le JECFA pour les additifs alimentaires et les contaminants, le JMPR pour les résidus de pesticides et le JEMRA pour les risques microbiologiques. Les experts afin de rendre leurs avis s'appuient sur une analyse des risques, « démarche scientifique » dont le Codex Alimentarius a défini le cadre formel d'organisation.

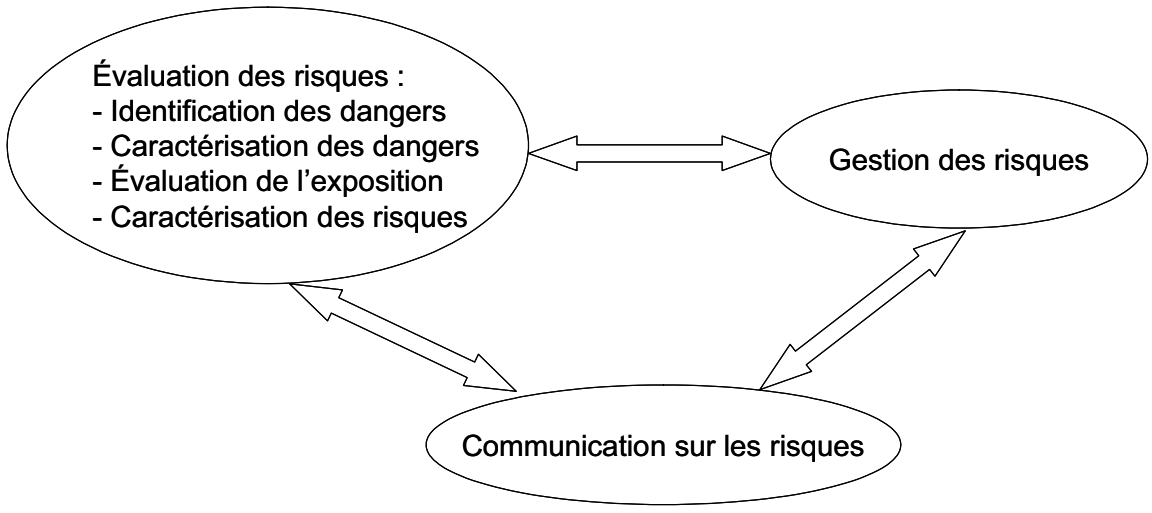

FIG. 1.1: Composantes de l'analyse des risques tel que définie par le Codex Alimentarius.

L'analyse des risques comporte trois volets (Codex Alimentarius Commission, 2003) (Fig. 1.1). Le premier volet est l'évaluation des risques qui comprend quatre étapes. L'évaluation des risques consiste dans un premier temps à déterminer la nature des contaminants des

aliments (identification des dangers) ainsi que la nature des effets provoqués par ingestion de ces contaminants (caractérisation des dangers). Lorsque le danger est avéré il est nécessaire de déterminer l'exposition (évaluation de l'exposition) et les risques encourus (caractérisation des risques) par la population étudiée. Le deuxième volet de l'analyse des risques est la gestion des risques. Ce processus est distinct de celui de l'évaluation des risques et consiste à mettre en balance les différentes politiques possibles afin de réduire le risque et, au besoin, à choisir les mesures de prévention et de contrôle appropriées. Ces mesures peuvent être de plusieurs formes : introduction de concentrations maximales en contaminant sur certains aliments, retrait du marché de certaines denrées, recommandations au niveau du procédé de fabrication industriel, information des consommateurs, etc. Ce processus s'effectue en consultation avec toutes les parties intéressées en tenant compte de l'évaluation des risques et d'autres facteurs ayant une importance pour la protection de la santé des consommateurs, des intérêts des producteurs et la promotion de pratiques commerciales loyales. Des analyses de coût-bénéfice se développent de plus en plus afin de quantifier en termes monétaires les pertes et les gains des agents liés aux situations de risques sanitaires et aux politiques mises en place pour contrecarrer ces risques. Ces méthodes d'analyse socio-économique renseignent également sur l'efficacité relative des outils de gestion choisis (Boardman et al., 1996). Le dernier volet est la communication sur les risques qui consiste en un échange interactif, tout au long du processus d'analyse des risques, d'informations et d'opinions sur les risques et notamment l'explication des résultats de l'évaluation des risques et des fondements des décisions de gestion, entre les responsables de l'évaluation et de la gestion des risques et les consommateurs, les industriels et les autres parties intéressées. L'analyse des risques alimentaires fait donc appel à de nombreuses disciplines comme la toxicologie, la microbiologie, l'épidémiologie, l'économie, la sociologie et la statistique. Les méthodes statistiques sont de plus en plus employées dans l'analyse de risques alimentaires : développement de nombreux modèles de croissance dans le cadre de la microbiologie prévisionnelle (McMeekin et al., 1993; Rosso et al., 1993), évaluation de la relation entre la dose de contaminant ingérée et l'effet provoqué (Buchanan & Bagi, 1997; Lindqvist & Westöö, 2000), utilisation de la théorie des valeurs extrêmes pour évaluer les risques faibles (Tressou et al., 2004), etc...

Le travail présenté dans ce manuscrit porte sur l'évaluation des risques liés à la présence de *Listeria monocytogenes* dans un produit de IV^{ème} gamme. La IV^{ème} gamme concerne les produits constitués de légumes et de salades crus prêts à consommer (lavés, découpés et emballés dans des sachets). Les travaux sont effectués en particulier pour la salade verte en sachet (laitue, scarole, batavia, roquette, etc...) qui est le produit pour lequel la part de

marché est la plus importante. Les objectifs de ces travaux sont d'apporter des réponses en terme d'exposition et de risques pour une bactérie fréquemment rencontrée dans les végétaux et pour un produit dont la consommation est en constante augmentation depuis les dix dernières années. Les résultats de cette thèse s'intègrent donc totalement dans les enjeux internationaux liés à l'évaluation scientifique des risques alimentaires. En effet, il s'agit de développer des modèles permettant d'améliorer l'évaluation des risques et en particulier l'évaluation des risques microbiologiques qui présente des particularités dues à la nature du contaminant. Ces travaux sont également menés suite à une demande de la part des professionnels de la filière IV^{ème} gamme soucieux des risques sanitaires potentiels de leurs produits et confrontés à la nouvelle législation européenne qui comprend des critères de sécurité spécifiques à *L. monocytogenes* (règlement européen 2073). En outre, il est important pour eux d'anticiper de futures réglementations européennes sur des facteurs pouvant avoir des répercussions sur la contamination et donc sur le risque, tels que le retrait du chlore du procédé industriel qui est actuellement en discussion.

Ce travail présente donc outre des enjeux scientifiques, des enjeux sanitaires et socio-économiques forts.

1.2 Justification du sujet

1.2.1 Un danger avéré

Qu'est-ce qu'un danger ?

Un danger est un agent biologique, chimique ou physique susceptible de provoquer des effets adverses pour la santé et qui peut être présent dans un aliment donné ou un groupe d'aliments. Les dangers chimiques sont par exemple : les additifs alimentaires, les dioxines, les métaux lourds, les résidus de pesticides et de médicaments vétérinaires, les migrants des emballages, etc. L'ingestion répétée de ces dangers chimiques, même de faibles doses s'accumulant dans l'organisme sans s'éliminer complètement, peut provoquer des effets adverses sur la santé se déclarant après plusieurs années (Barlow et al., 2002). On parle alors d'effet et de risque chroniques (Carriquiry et al., 1990; Kroes et al., 2002). Certains de ces contaminants comme les toxines marines présentes dans les mollusques ou encore certains pesticides, s'ils sont présents en grande quantité, peuvent entraîner des effets toxiques aigus, c'est-à-dire des effets qui se manifestent rapidement après l'ingestion du contaminant. Les dangers biologiques sont les bactéries pathogènes, les virus, les moisissures, les parasites uni-

cellulaires, etc. Ces contaminants lorsqu'ils sont ingérés en forte dose peuvent provoquer des effets aigus. Toutefois l'ingestion d'une faible dose de certaines bactéries pathogènes comme *Escherichia coli* O157 :H7 ou *Salmonella* peut entraîner un effet adverse aigu. Les dangers physiques sont la présence de corps étrangers tels que morceaux de verre ou de métal dans les aliments. Il existe aussi des dangers nutritionnels lorsque l'apport en nutriment est déficient (malnutrition) ou à l'opposé lorsqu'il est excédentaire (surnutrition). Dans ce cas, des risques de carences ou d'obésité sont encourus. Un nouveau type de danger chimique est du à l'enrichissement des aliments en nutriments (adjonction de vitamines ou de minéraux), procédé qui se généralise au sein de l'industrie alimentaire et qui est utilisé par les industries comme argument de vente. L'enrichissement contribue à protéger les consommateurs des carences alimentaires mais peut cependant avoir des effets néfastes. Par exemple, une consommation importante de vitamine A provenant de certains suppléments, peu avant ou pendant la grossesse, peut entraîner des malformations congénitales.

Lorsqu'un agent biologique, chimique ou physique est reconnu comme un danger, il est alors procédé à la caractérisation du danger. Cette étape de l'évaluation des risques consiste à évaluer qualitativement et/ou quantitativement la nature des effets néfastes pour la santé associés aux agents biologiques, chimiques et physiques qui peuvent être présents dans un aliment. Ainsi, à cette étape sont déterminées des doses tolérables pour l'organisme ou des relations dose/réponse, c'est-à-dire des relations entre le degré d'exposition (dose) à un agent pathogène et la gravité et/ou la fréquence des effets adverses pour la santé qui en résultent (réponse). Les relations dose/réponse sont obtenues soit à partir d'études épidémiologiques réalisées suite à des toxi-infections alimentaires dues à l'ingestion de contaminants microbiologiques (Buchanan & Bagi, 1997; Lindqvist & Westöö, 2000), soit à partir d'expérimentations animales dont les résultats sont transposés à l'homme en incluant des facteurs de sécurité intra et inter espèces (Dybing et al., 2002).

Le danger *L. monocytogenes*

Le rôle des aliments comme voie de transmission principale de *L. monocytogenes* a été reconnu (Broome et al., 1990; Bill, 1990; Swaminathan et al., 2007) après plusieurs poussées épidémiques importantes qui se sont produites en Amérique du Nord et en Europe dans les années 1980. Les légumes peuvent être les vecteurs de transmission de *L. monocytogenes*. En effet les légumes sont porteurs d'une abondante microflore comprenant des espèces pathogènes. Les sources de contamination sont multiples : les sols de culture qui sont un réservoir naturel, les matières fécales (fumier, lisiers) épandues sur les champs ou les eaux d'irriga-

tion, et parfois les ateliers de transformation. *Listeria monocytogenes*, par ses capacités de survie et de développement en grand nombre à de faibles températures sur des végétaux, est fréquemment impliquée dans des cas de toxi-infections alimentaires (listérioses) (Schlech et al., 1983). Par ailleurs, bien que *Listeria monocytogenes* soit présente dans de nombreux aliments, les poussées épidémiques et les cas sporadiques sont principalement associés aux aliments prêts à consommer. La composition de ces aliments est très hétérogène (saumon fumé, salade feuille, salade composée, viandes fermentées) et varie d'un pays à l'autre selon les habitudes alimentaires. Ces produits sont généralement consommés en l'état c'est-à-dire qu'ils ne subissent aucun processus de lavage ou de cuisson pouvant réduire le nombre de bactéries avant leur consommation. La salade de IV^{ème} gamme peut donc être un vecteur de transmission de *L. monocytogenes,* elle a d'ailleurs déjà été à l'origine de toxi-infections alimentaires (Nguyen-the & Carlin, 2000). Précisons aussi qu'aux Etats-Unis, les légumes de IV^{ème} gamme ont déjà fait l'objet d'une évaluation du risque lié à la présence de *L. monocytogenes* lors d'une étude générale sur les produits prêts à consommer menée par les départements de la santé humaine (DHHS) et de l'agriculture (USDA) américains (FDA-USDA, 2003).

La listériose d'origine alimentaire est une maladie rare mais grave qui touche principalement les populations dites sensibles. En effet, *L. monocytogenes* est un pathogène opportuniste qui atteint plus particulièrement les personnes dont les défenses immunitaires sont déficientes comme les patients immunodéprimés, les malades du SIDA ou les personnes atteintes de maladies chroniques mais aussi les femmes enceintes, les foetus et les nouveaux nés ainsi que les personnes âgées. Les manifestations cliniques associées à la listériose peuvent être groupées en deux catégories : listériose invasive et listériose non invasive. On parle de listériose invasive lorsqu'une infection initiale du tissu intestinal par *L. monocytogenes* se propage à d'autres parties du corps comme l'utérus chez les femmes enceintes, le système nerveux central ou le sang. La listériose peut laisser des séquelles (McLauchlin, 1997) bien que leur incidence soit rarement estimée (Rocourt, 1996). La listériose non invasive (ou gastroentérite fébrile à *Listeria*) a été observée durant un certain nombre de poussées épidémiques où la majorité des patients ont eu des symptômes de gastroentérite, tels que diarrhée, fièvre, céphalée et myalgie après une brève période d'incubation (Dalton et al., 1997; Salamina et al., 1996; Riedo et al., 1994; Aureli et al., 2000). Ces poussées ont généralement été dues à l'ingestion de fortes doses de *L. monocytogenes* par des individus par ailleurs bien portants. Le taux d'incidence et les facteurs qui régissent le début de cette forme non invasive ne sont pas connus si bien que les travaux d'évaluation des risques ne portent en général que sur

la listériose invasive comme conséquence de l'exposition. La listériose touche principalement les pays industrialisés et l'on ignore si les taux d'incidence entre pays développés et pays en développement reflètent de vraies différences géographiques, des différences dans les habitudes alimentaires et de stockage des denrées ou plutôt des différences dans les méthodes de diagnostic et de déclaration de cas.

1.2.2 Nécessité d'évaluer l'exposition à *L. monocytogenes* et de caractériser les risques de listériose par consommation de salade de IV^{ème} gamme

Les éléments précédents montrent qu'un nombre important de travaux de recherche ont été réalisés dans lesquels la pathogénécité de *L. monocytogenes* ainsi que sa présence dans les végétaux ont été mis en évidence. *L. monocytogenes* représente donc un danger pour l'ensemble de la filière IV^{ème} gamme. Pourtant, aucune étude n'a été réalisée en France ou en Europe afin de déterminer les concentrations en *L. monocytogenes* à l'instant de la consommation (exposition) ainsi que le risque de tomber malade par ingestion de *Listeria monocytognes* suite à la consommation de produits de IV^{ème} gamme. Il est donc nécessaire de mettre en place une méthodologie d'évaluation de l'exposition et des risques pour le couple *L. monocytogenes*/salades de IV^{ème} gamme.

Dans le cadre formel de l'analyse de risque, l'évaluation de l'exposition se définit comme l'évaluation qualitative et/ou quantitative de l'ingestion d'agents biologiques, chimiques et physiques par le biais des aliments, ainsi que le cas échéant, par suite de l'exposition à d'autres sources. Cette étape consiste à évaluer les doses de contaminants ingérés à l'instant de la consommation. Les doses peuvent être évaluées soit par analyse directe des aliments consommés (études d'alimentation dupliquée) mais ces études sont rares (Leblanc et al., 2000, 2005), soit en croisant la contamination à l'instant de la consommation avec les quantités consommées en utilisant des outils statistiques plus ou moins complexes (Jaykus, 1996; Kroes et al., 2002). La caractérisation des risques s'effectue en intégrant les trois étapes précédentes (l'identification et la caractérisation du danger ainsi que l'évaluation de l'exposition) afin d'estimer qualitativement et/ou quantitativement, compte tenu des incertitudes inhérentes à l'évaluation, la probabilité de la fréquence et de la gravité des effets adverses connus ou potentiels sur la santé susceptibles de se produire dans une population donnée (Renwick et al., 2003). Ainsi, le risque est déterminé comme la probabilité de dépasser la dose tolérable ou comme la probabilité de tomber malade lors d'un acte de consommation dans le cas d'un

risque aigu.

L'évaluation de l'exposition et la caractérisation des risques pour le couple *L. monocytogenes*/salades de IV^{ème} gamme présentent certaines particularités qui nécessitent la construction d'outils spécifiques. Tout d'abord, les bactéries peuvent se développer lorsque les conditions environnementales leur sont favorables, contrairement à la plupart des contaminants chimiques présents en quantité constante au cours du temps dans les aliments. Par ailleurs, *Listeria monocytogenes* a la faculté de se multiplier même à de faibles températures. C'est pourquoi il est nécessaire dans les analyses de risques relatifs à des dangers microbiologiques d'évaluer les concentrations bactériennes tout au long de la chaîne alimentaire et d'intégrer alors le processus de croissance des bactéries. Ce type d'évaluation a un intérêt majeur pour les gestionnaires du risque (cf. le deuxième volet de l'analyse du risque) qui peuvent ainsi proposer des mesures de gestions à différent niveaux de la chaîne alimentaire et non pas seulement à l'instant de la consommation. Les bactéries et notamment les bactéries pathogènes (néfastes pour l'hôte) comme *Listeria monocytogenes* sont difficilement détectables et quantifiables. De plus, les méthodes analytiques de dénombrements des bactéries sont très coûteuses en terme de temps. Ainsi les données quantifiant le nombre de bactéries par unité d'analyse (g, ml, etc...) sont très peu nombreuses. Notons aussi qu'en comparaison avec certains dangers toxiques ou d'autres couples bactéries pathogènes/aliment, il existe peu de plans de surveillance nationaux de *L. monocytogenes* dans les végétaux et en particulier dans les végétaux frais non transformés. Ainsi les données disponibles sont pour la plupart des données d'origine bibliographique ou de centres techniques et leur recueil est une barrière supplémentaire à la connaissance des concentrations bactériennes. Il faut alors développer des outils permettant d'estimer les concentrations à l'aide des seules données disponibles. De plus, lorsque les concentrations en *L. monocytogenes* dans la salade ont pu être relevées, il s'avère que celles-ci sont très variables d'une analyse à l'autre. Le modèle d'évaluation de l'exposition et des risques proposé doit prendre en compte la variabilité propre aux concentrations microbiennes ainsi que l'incertitude des estimations due au manque d'information et de données. Le modèle doit aussi permettre de propager séparément l'incertitude totale (variabilité et incertitude) lors de la simulation de la croissance/décroissance des bactéries pendant les différentes étapes composant la chaîne alimentaire.

1.3 Enjeux scientifiques

1.3.1 Construire un modèle d'évaluation de l'exposition à *L. monocytogenes* et de caractérisation des risques de listériose par consommation de salade de IV^{ème} gamme

Nous proposons de construire un modèle d'évaluation de l'exposition et de caractérisation des risques pour le couple *L. monocytogenes*/salade de IV^{ème} gamme. Ce modèle est composée de plusieurs variables (ou paramètres) d'entrée et de sortie qui sont présentées dans la figure 1.2. L'objective est d'évaluer les variables de sortie qui sont les contaminations aux différentes étapes de la chaîne alimentaire, l'exposition des consommateurs, le risque de listériose et le nombre de cas. Pour cela, il est nécessaire d'introduire dans le modèle des variables d'entrée, comme la contamination de la matière première entrant dans l'usine, les paramètres de croissance, les températures et les durées de conservation du produit pour les différentes étapes constituant son circuit logistique (non détaillées dans la figure 1.2), la consommation de salade de IV^{ème} gamme, etc. Nous choisissons de réaliser l'évaluation de l'exposition et des risques en deux temps. Dans un premier temps, les distributions des variables d'entrée sont déterminées en utilisant des modèles développés afin de rendre compte de la complexité du comportement des micro-organismes (contamination, croissance, etc). Dans un second temps, les distributions des variables de sortie sont déterminées à l'aide de méthodes de simulation ayant en entrée les distributions des variables établies précédemment. Il s'agit alors de prendre en compte dans l'estimation des variables d'entrée la variabilité et l'incertitude et de les propager séparément au cours des simulations afin d'intégrer ces deux dimensions dans l'évaluation de l'exposition et des risques. Le modèle développé doit aussi permettre d'évaluer les conséquences en terme de concentrations bactériennes, d'exposition et de risque de scénarios comme le retrait du chlore du procédé de fabrication des salades de IV^{ème} gamme ou de scénarios futurs de gestion du risque pouvant être une meilleure gestion de la chaîne du froid. Ce modèle doit aussi être suffisamment générique pour être transposable aux différents produits de la filière IV^{ème} gamme et à d'autres filières alimentaires.

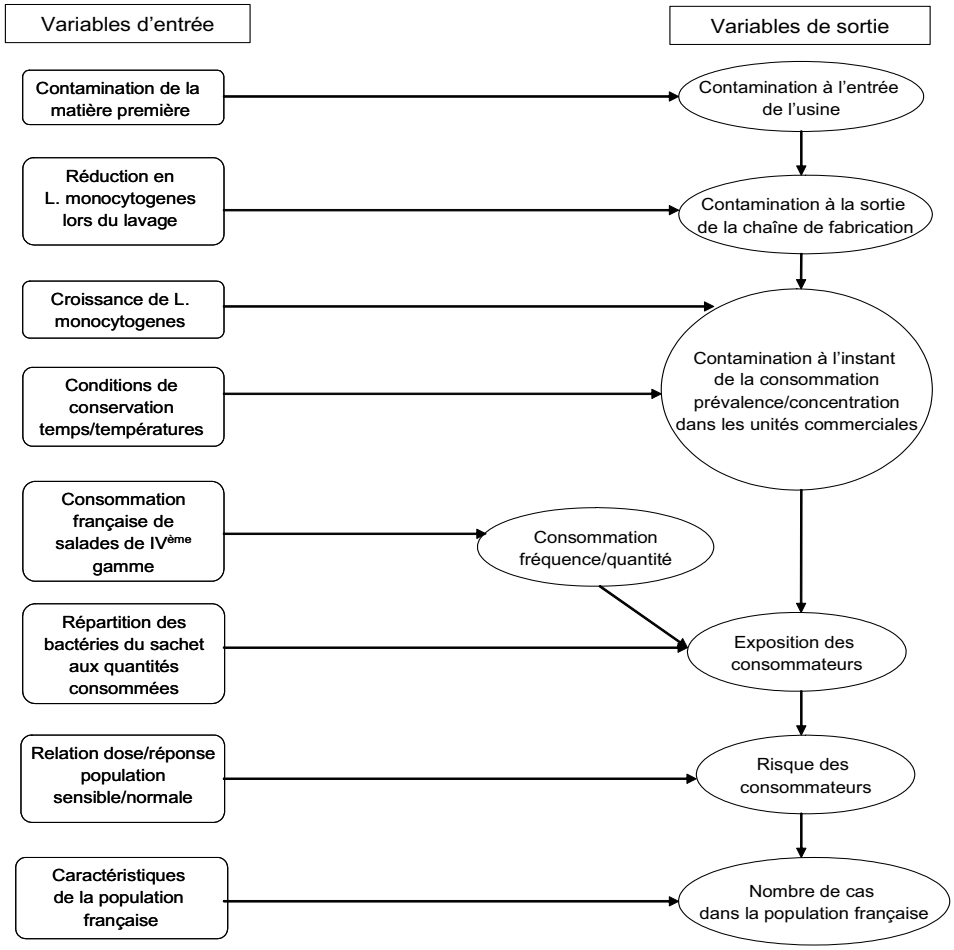

FIG. 1.2: Principaux composants du modèle d'évaluation de l'exposition et des risques lié à la présence de *L. monocytogenes* dans les salades de IV^{ème} gamme

1.3.2 Intégrer dans l'évaluation des risques alimentaires des méthodes statistiques permettant de prendre en compte séparément la variabilité et l'incertitude

La variabilité représente la diversité et l'hétérogénéité de la réponse dans la population étudiée. Elle est intrinsèque à cette population et ne peut être réduite par un complément d'information à la différence de l'incertitude qui représente l'ignorance partielle ou le manque de connaissances de l'évaluateur et qui peut être réduite par des données ou des études supplémentaires. L'incertitude totale est constituée de la variabilité et de l'incertitude. Ainsi tout phénomène biologique est intrinsèquement variable et la connaissance de ce phénomène ne peut être que partielle et cela même dans le cas d'information abondante. La statistique est la science adaptée à l'étude de la variabilité et de l'incertitude. Ainsi le Dr Daniel Schwartz 1994 déclare que "la variabilité est la raison d'être de la statistique" (Schwartz, 1994) ou encore Claudine Vergne dans son postulat pour les nouveaux programmes de statistique au lycée écrit "Prendre conscience de la variabilité d'un phénomène, ce n'est pas seulement constater que les résultats sont sujets à variation, c'est concevoir que, à notre échelle d'observation, les résultats sont nécessairement variables et imprévisibles, c'est accepter de prendre en compte les fluctuations, c'est faire le deuil de la certitude et s'engager dans le monde de l'incertain. Renonçant à des connaissances assurées, on peut alors, par des méthodes statistiques, en suivant des raisonnements de type inductif, accéder à une maîtrise relative de l'incertitude pour estimer, prévoir et prendre des décisions avec risque consenti". Il est alors indispensable lors de la réalisation d'analyse des risques alimentaires d'intégrer une mesure de l'incertitude totale. La commission du codex alimentarius recommande d'ailleurs de prendre en compte la variabilité du risque et d'évaluer les incertitudes autour des estimations du risque (Codex Alimentarius Commission, 2003).

Le moyen de prendre en compte la variabilité d'un phénomène est d'opter pour l'approche probabiliste (ou distributionnelle) qui est l'utilisation de distributions de probabilité en opposition à l'approche déterministe (ou ponctuelle) qui caractérise le phénomène étudié par une seule valeur qui est en général la moyenne. Bien que l'approche probabiliste soit de plus en plus employée lors d'évaluation des risques, peu de travaux prennent en compte tout en les distinguant les deux dimensions de l'incertitude totale (variabilité et incertitude). La prise en compte séparée de la variabilité et de l'incertitude dans l'évaluation des risques alimentaires est l'un des enjeux de cette thèse. Les méthodes statistiques permettant de mener à bien ce travail sont présentées dans cette partie.

L'approche bayésienne : une méthode statistique performante pour l'évaluation des risques et la prise en compte de la variabilité et de l'incertitude

Le paradigme bayésien Considérons un modèle statistique où la loi de probabilité $p(x|\theta)$ qui génère les observations x est donnée par un modèle paramétrique particulier qui dépend d'un paramètre inconnu de dimension $k : \theta \in \Theta \subset \mathbb{R}^k$. Dans l'approche classique de la statistique inférentielle, le paramètre θ, s'il est par nature inconnu, reste néanmoins fixé dans Θ. L'approche bayésienne considère toujours θ inconnu mais ce paramètre est, comme l'observation x, aléatoire (c'est-à-dire issu d'un phénomène aléatoire). Dans ce contexte, le paramètre θ est régi par une certaine loi de probabilité $p(\theta)$, dite loi *a priori* (supposée connue). Cette loi est subjective dans la mesure où elle représente la croyance de l'expérimentateur avant que l'expérience ne soit conduite. L'analyse bayésienne se fait en termes de lois régissant θ, l'inférence sur ce paramètre inconnu se réalisant au travers de ces lois. Plus spécifiquement, elle tire parti de l'observation x pour réactualiser la loi *a priori* $p(\theta)$: on construit sur la base de la loi *a priori* $p(\theta)$ et des observations x, une loi $p(\theta|x)$ qui régit θ, dite loi *a posteriori* (car déterminée après avoir observé x). Le principe est donc de corriger l'*a priori* que l'on suppose sur θ par l'information apportée par x. Une telle modélisation n'est pas universellement acceptée, le choix de la loi *a priori* n'obéissant pas nécessairement à des critères objectifs. Cependant, il est des cas où l'on dispose de résultats d'expériences permettant de conjecturer que le paramètre θ se trouve avec une forte probabilité dans une région déterminée de l'espace des paramètres. On peut ainsi rendre compte de cette information au travers d'une loi *a priori* se concentrant dans cette région. A l'inverse, lorsque aucune information n'est disponible sur le paramètre θ, cela peut être traduit par l'utilisation de lois *a priori* non informatives ou vagues (c'est-à-dire avec une variance élevée). Notons toutefois que le support de la loi *a posteriori* est inclue dans celui de la loi *a priori*. L'approche bayésienne est formalisée dans la figure 1.3.

D'un point de vue pratique, les calculs nécessaires pour déterminer la distribution *a posteriori* sont souvent complexes voire insolubles. En effet, la résolution analytique de l'intégrale (1.3) de la figure 1.3 est particulièrement difficile lorsque la loi $p(\theta)$ n'est pas conjuguée pour le modèle de vraisemblance $p(x|\theta)$, c'est-à-dire que les lois $p(\theta)$ et $p(\theta|x)$ n'appartiennent pas à la même famille de lois, ou lorsque le modèle présente une structure hiérarchique (Fig. 1.4), c'est-à-dire que les lois *a priori* sont découpées en plusieurs niveaux, les paramètres (appelés aussi hyperparamètres) des derniers niveaux possédant également des lois de probabilité. Le découpage hiérarchique donne donc lieu à une explosion de lois marginales et de lois *a posteriori*, suivant le niveau par rapport auquel on conditionne.

La distribution jointe de (x, θ) s'obtient simplement par

$$p(x, \theta) = p(x|\theta)p(\theta). \tag{1.1}$$

La formule de Bayes est basée sur la décomposition inverse de (1.1)

$$p(x, \theta) = p(\theta|x)p(x),$$

on obtient alors la distribution *a posteriori* de θ conditionnellement à l'observation de l'échantillon x

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}, \tag{1.2}$$

où le dénominateur $p(x)$ est la distribution marginale de x et ne dépend pas de θ. On a

$$p(x) = \int_{\Theta} p(x|\theta)p(\theta)d\theta. \tag{1.3}$$

La forme simplifiée de (1.2) s' écrit

$$p(\theta|x) \propto p(x|\theta)p(\theta). \tag{1.4}$$

où $\propto$ signifie "proportionnel à", la constante d'intégration étant donnée en (1.3)
Il est utile de pouvoir générer de futures valeurs de x, appelées aussi valeurs prédites et notées $\widetilde{x}$, étant donnée l'information récoltée sur θ grâce aux observations x. On définit alors la distribution prédictive *a posteriori* suivante

$$p(\widetilde{x}|x) = \int_{\Theta} p(\widetilde{x}|\theta)p(\theta|x)d\theta.$$

FIG. 1.3: Formalisme bayésien

La loi de probabilité $p(x|\theta)$ qui génère les observations x est donnée par un modèle paramétrique particulier qui dépend d'un paramètre inconnu de dimension k : $\theta \in \Theta \subset \mathbb{R}^k$. La loi de probabilité $p(\theta|\psi)$ du paramètre θ est elle aussi décrite par un modèle paramétrique qui dépend d'un paramètre (hyperparamètre) inconnu de dimension l : $\psi \in \Psi \subset \mathbb{R}^l$. L'hyperparamètre ψ est alors régi par la loi *a priori* $p(\psi)$. La distribution *a posteriori* de ψ conditionnellement à l'observation de l'échantillon x s'écrit alors :

$$p(\psi|x) \propto \int_{\Theta} p(x|\theta)p(\theta|\psi)p(\psi)d\theta. \tag{1.5}$$

On peut aussi obtenir la distribution *a posteriori* de θ

$$p(\theta|x) \propto \int_{\Psi} p(\theta|\psi)p(\psi|x)d\psi.$$

La distribution prédictive *a posteriori* dans ce cas est

$$p(\widetilde{x}|x) = \int_{\Theta}\int_{\Psi} p(\widetilde{x}|\theta,\psi)p(\theta,\psi|x)d\psi d\theta.$$

FIG. 1.4: Formalisme bayésien pour un modèle hiérarchique

Ainsi, dans le cas où les calculs sont complexes, le recours à des intégrations numériques est nécessaire pour approcher les distributions *a posteriori*. Pour cela, de nombreux algorithmes fondés sur les méthodes de Monte Carlo par Chaîne de Markov (MCMC) ont été développés et implémentés dans des logiciels, comme par exemple OpenBugs (variante de Winbugs), cf. Spiegelhalter et al. (2005). Le principe des algorithmes de MCMC est d'approcher les distributions *a posteriori* des paramètres (et/ou des hyperparamètres dans le cas d'un modèle hiérarchique), comme expliqué dans la figure 1.5. Les deux algorithmes de MCMC les plus utilisés sont l'algorithme de Hasting-Metropolis (Metropolis et al., 1953; Hastings, 1970) et l'échantillonnage de Gibbs (Geman & Geman, 1984; Gelfand & Smith, 1990). Le logiciel OpenBugs utilise principalement l'algorithme de Gibbs (ou une version hybride). Un exemple de cet algorithme est donné dans la figure 1.6. Pour un approfondissement des connaissances sur les méthodes bayésiennes et les méthodes MCMC, on peut par exemple se reporter aux ouvrages de Robert (2001); Gelman et al. (2004a) et Robert (1996).

Les méthodes MCMC permettent de simuler un échantillon $x_1, ..., x_n$ sans avoir recours à la distribution f (ici f peut-être la distribution *a posteriori* de l'équation (1.5)) et ainsi d'approcher l'intégrale $E_f[h(x)] = \int h(x)f(x)dx$. Le principe est d'utiliser une chaîne de Markov de loi stationnaire f : partant d'une valeur arbitraire $x^{(0)}$, une chaîne $x^{(t)}$ est générée à partir d'un noyau de transition k de loi stationnaire f, qui garantit la convergence en loi vers f. Pour T "assez grand", on peut considérer $x^{(T)}$ comme distribué suivant f et obtenir ainsi un échantillon $x^{(T)}, x^{(T+1)}, ...$, qui est effectivement distribué suivant f, même si les $(x^{(T+t)})_{t\geq 1}$ ne sont pas indépendants. Toute méthode produisant une chaîne de Markov $(x^{(t)})$ de loi stationnaire la distribution d'intérêt est appelée algorithme MCMC. L'utilisation d'une chaîne $(x^{(t)})$ produite par un algorithme MCMC est semblable à celle d'un échantillon indépendant et identiquement distribué suivant f au sens où le théorème ergodique (appellation de la loi des grands nombres dans ce cadre) garantit la convergence de la moyenne empirique vers son espérance théorique, c'est-à-dire

$$\frac{1}{B}\sum_{t=T}^{T+B} h(x^{(t)}) \underset{B\to\infty}{\to} E_f[h(x)].$$

La convergence vers la loi stationnaire est donc atteinte à partir d'un certain nombre T d'itérations. Ce nombre est à déterminer à l'aide des outils de contrôle de la convergence. La convergence peut être contrôlée graphiquement en identifiant le nombre d'itérations devant être réalisées afin que les valeurs produites, à partir de plusieurs chaînes ayant des valeurs initiales distinctes, se rejoignent. Une fois la convergence vers la loi stationnaire atteinte, il est possible de simuler B valeurs. En général, l'inférence bayésienne est fondée sur l'estimation des distributions marginales *a posteriori* des paramètres et des hyperparamètres. A partir des B dernières réalisations de la chaîne de Markov sont déduits, pour chaque paramètre et hyperparamètre, la moyenne, l'écart-type et certains percentiles ($2.5^{ème}$, $50^{ème}$, $97.5^{ème}$, etc). La loi jointe *a posteriori* des paramètres, formée par les B valeurs obtenues, peut également être utilisée pour simuler des valeurs de paramètres d'entrée d'un modèle de simulations de Monte Carlo.

FIG. 1.5: Principe des méthodes de Monte Carlo par Chaîne de Markov

L'algorithme de Gibbs pour approcher la loi jointe *a posteriori* $p(\theta, \psi|x)$ d'un modèle hiérarchique (Fig. 1.4) où x est le vecteur des observations, $\theta = (\theta_1, ..., \theta_i, ..., \theta_k)$ le vecteur des paramètres et $\psi = (\psi_1, ..., \psi_h, ..., \psi_l)$ le vecteur des hyperparamètres s'effectue comme suivant :

- Partir de valeurs initiales arbitraires $\psi^{(0)} = (\psi_1^{(0)}, ..., \psi_h^{(0)}, ..., \psi_l^{(0)})$ et générer des valeurs initiales $\theta^{(0)} = (\theta_1^{(0)}, ..\theta_i^{(0)}, ..\theta_k^{(0)})$ à partir de la distribution conditionnelle $p(\theta|\psi^{(0)}, x)$.
- Pour $t = 1, ..., T + B$, générer chaque $\theta_i^{(t)}$ et chaque $\psi_h^{(t)}$ en utilisant respectivement les distributions conditionnelles suivantes $p(\theta_i|\theta_{-i}^{(t-1)}, \psi^{(t-1)}, x)$ et $p(\psi_h|\theta^{(t)}, \psi_{-h}^{(t-1)}, x)$, où $\theta_{-i}^{(t-1)} = (\theta_1^{(t)}, ..., \theta_{i-1}^{(t)}, \theta_{i+1}^{(t-1)}, ..., \theta_k^{(t-1)})$ et $\psi_{-h}^{(t-1)} = (\psi_1^{(t)}, ..., \psi_{h-1}^{(t)}, \psi_{h+1}^{(t-1)}, ..., \psi_l^{(t-1)})$ sont respectivement les vecteurs en θ sauf la $i^{ème}$ composante et en ψ sauf la $h^{ème}$ composante. Les valeurs de ces vecteurs correspondent aux valeurs prises à l'instant t ou $t-1$ selon si la mise à jour a été effectuée ou non.

FIG. 1.6: Algorithme de Gibbs

Modèles bayésiens hiérarchiques et prise en compte de la variabilité et de l'incertitude L'inférence bayésienne permet de quantifier l'incertitude sur les paramètres des modèles. Notons qu'en statistique classique, l'incertitude peut se quantifier en ayant recours au bootstrap (Efron & Tibshirani, 1993) cependant, ce type de problème est plus facile à appréhender par une approche bayésienne dans le sens où tous les paramètres (et hyperparamètres) sont traités comme des variables aléatoires. Dans le cas d'un modèle bayésien hiérarchique, il est possible d'interpréter la distribution de vraisemblance ainsi que la distribution des paramètres en terme de variabilité et la distribution des hyperparamètres en terme d'incertitude (Pouillot, 2006). De plus, l'effet de l'incertitude globale sur l'ensemble des hyperparamètres inconnus est prise en compte par la loi jointe *a posteriori* des hyperparamètres fournie par l'inférence bayésienne, contrairement aux méthodes de statistique classique où la relation de dépendance entre les paramètres n'est souvent pas considérée.

Méthodes bayésiennes et évaluation des risques microbiologiques Les méthodes bayésiennes sont de plus en plus employées dans l'évaluation des risques pour leur capacité à résoudre de nombreux problèmes propres à ce domaine. Les exemples sont nombreux pour les évaluations liées à des risques microbiologiques (Pouillot et al., 2007; Ranta et al., 2005; Barker et al., 2005). Certains auteurs comme Paulo et al. (2005) les ont aussi employées pour l'évaluation des risques liés aux pesticides. Enfin, ces méthodes sont utilisées dans d'autres domaines comme la science de l'eau (cf. par exemple la thèse de Pasanisi, 2004), l'épidémio-

logie avec par exemple les travaux de Richardson & Gilks (1993), ou encore l'écologie (cf. par exemple les travaux de Rivot et al., 2004). Nous relevons ci-dessous les différentes problématiques que pose l'évaluation des risques microbiologiques et pour lesquelles les méthodes bayésiennes peuvent apporter des solutions.

L'évaluation des risques microbiologiques nécessite l'intégration d'informations externes afin de compléter l'information apportée par les seules données disponibles. En effet, les données indispensables à la quantification de l'exposition et des risques microbiologiques sont souvent incomplètes, peu adaptées au problème ou peu représentatives du phénomène étudié. Il est alors nécessaire d'intégrer de l'information supplémentaire. Cette information peut être issue de l'opinion personnelle du statisticien, de techniques d'élicitation auprès d'experts (Kadane & Wolfson, 1998; O'Hagan & Oakley, 2007) ou d'études antérieures. L'approche bayésienne permet d'intégrer de l'information externe par l'utilisation de distributions de probabilité *a priori* sur le paramètre inconnu θ (ou ψ dans le cas d'un modèle hiérarchique). Notons que l'indépendance entre l'information *a priori* et les données étudiées est nécessaire afin que l'information ne soit pas redondante. L'utilisation de distribution *a priori* permet également de travailler sur des modèles éventuellement non identifiables. En effet, l'appui sur les distributions *a priori* permet de procéder à l'inférence même si les données utilisées ne permettent pas d'estimer certains paramètres du modèle. Dans ce cas, les distributions *a posteriori* sont calculées uniquement à partir des distributions *a priori*. Pour donner un exemple, dans cette thèse, les données rassemblées pour estimer les paramètres de croissance de *L. monocytogenes* dans la salade ne couvraient pas l'ensemble des températures auxquelles *L. monocytogenes* est susceptible de se développer. Ainsi, les températures optimales et maximales de croissance qui sont des paramètres de ce modèle n'étaient pas identifiables avec ces seules données. L'utilisation de distributions établies à partir d'une étude antérieure réalisée sur un autre aliment a permis de passer outre ce problème. Cela a toutefois pu être réalisé car les paramètres en question sont des paramètres propres à la bactérie étudiée et ne varient que très peu selon l'aliment. Il est donc conseillé de bien choisir les distributions *a priori* avant d'avoir recours à ce genre de pratique.

Les modèles de microbiologie prévisionnelle modélisant la croissance ou encore celui développé dans cette thèse afin d'évaluer la distribution de la contamination des végétaux par *L. monocytogenes* sont des modèles complexes c'est-à-dire comportant plusieurs niveaux hiérarchiques et un nombre important de paramètres. Le calcul analytique des distributions *a posteriori* est alors irréalisable et le recours à des intégrations numériques est obligatoire. L'inférence bayésienne a donné naissance aux algorithmes de MCMC (Monte Carlo par

Chaînes de Markov) très performants pour approcher les distributions *a posteriori* de ces modèles complexes.

Les méthodes de simulation de Monte-Carlo de second ordre : propagation séparée de la variabilité et de l'incertitude

Lorsque le modèle est complexe, l'utilisation de distributions de probabilité pour un grand nombre de variables d'entrée rend le calcul analytique des distributions des variables de sortie insoluble. C'est pourquoi les calculs sont menés numériquement par simulations de Monte-Carlo. Les méthodes de simulation de Monte-Carlo sont largement utilisées en évaluation des risques alimentaires (Nauta, 2000). En effet, ces méthodes permettent de travailler avec des modèles stochastiques (approche probabiliste) et de propager le hasard (la variabilité et l'incertitude) tout au long de la chaîne alimentaire. Les simulations de Monte-Carlo nécessitent un grand nombre d'itérations. A chaque itération, une valeur est tirée aléatoirement dans chaque distribution de probabilité des paramètres d'entrée du modèle. Chaque itération représente une situation réelle potentielle et une distribution de probabilité est obtenue pour chacune des variables de sortie.

Il est important et conseillé d'évaluer séparément la variabilité des sorties et l'incertitude des estimations produites. L'intérêt de séparer la variabilité de l'incertitude est de pouvoir déterminer laquelle des deux prédomine afin d'analyser les sorties en fonction de cette information. La méthode de Monte-Carlo à deux dimensions également appelée méthode de Monte-Carlo de second ordre (Nauta, 2000; Vose, 2000; Pouillot et al., 2007) permet de transférer séparément la variabilité et l'incertitude des paramètres dans le modèle. Le principe est de dissocier, lors de la simulation de Monte-Carlo, la simulation des paramètres incertains des paramètres variables, c'est-à-dire de réaliser deux simulations de Monte-Carlo imbriquées : une modélisation de la variabilité imbriquée dans une modélisation de l'incertitude (Fig. 1.7). Pour ce faire, il faut au préalable déterminer ce que représente les distributions utilisées en entrée du modèle : la variabilité du paramètre ou l'incertitude autour de l'estimation du paramètre. Il n'est pas toujours évident de qualifier de variabilité ou d'incertitude chaque distribution, cependant nous avons vu que l'utilisation de modèles bayésiens hiérarchiques facilite cette interprétation.

La simulation de Monte-Carlo à deux dimensions consiste à tirer aléatoirement un nombre n_u de valeurs dans chacune des distributions $p(\psi)$ des (hyper)paramètres, $\psi \in \mathbb{R}^l$, reflétant l'incertitude (voir le schéma illustratif Fig. 1.8). Notons que si la distribution de ψ a été établie par inférence bayésienne alors les n_u valeurs sont tirées dans la loi jointe *a posteriori* $p(\psi|x)$ de ψ. Puis, pour chaque valeur fixe des paramètres d'incertitude, un nombre n_v de valeurs est tiré aléatoirement dans chacune des distributions $p(\theta|\psi_u)$ des paramètres $\theta \in \mathbb{R}^k$ reflétant la variabilité (ou dans la distribution *a posteriori* $p(\theta|\psi_u, x)$ dans le cas où celle-ci a été établie par inférence bayésienne). Ces opérations sont effectuées pour chacune des variables d'entrée du modèle de simulation. Ensuite, par simulation de Monte-Carlo, pour chaque valeur fixée d'incertitude, une distribution empirique constituée des n_v valeurs est obtenue pour chaque variable de sortie (par exemple : concentrations, exposition, risque...). Cette distribution reflète la variabilité de chacune des variables de sortie et différentes statistiques de chaque variable sont calculées comme la moyenne, l'écart-type ou différents percentiles. Ainsi un nombre n_u de valeurs est obtenu pour chaque statistique de chaque variable et les percentiles de ces statistiques sont calculés. Le $50^{ème}$ percentile est proposé comme estimateur des différentes statistiques calculées, le $2.5^{ème}$ et le $97.5^{ème}$ percentiles sont proposés comme bornes de l'intervalle d'incertitude (UI, Uncertainty Interval) des estimations. La comparaison entre les intervalles d'incertitude des estimations et les percentiles représentant la variabilité du paramètre de sortie étudié permet de voir laquelle de la variabilité et de l'incertitude prédomine.

FIG. 1.7: Simulation de Monte-Carlo de second ordre

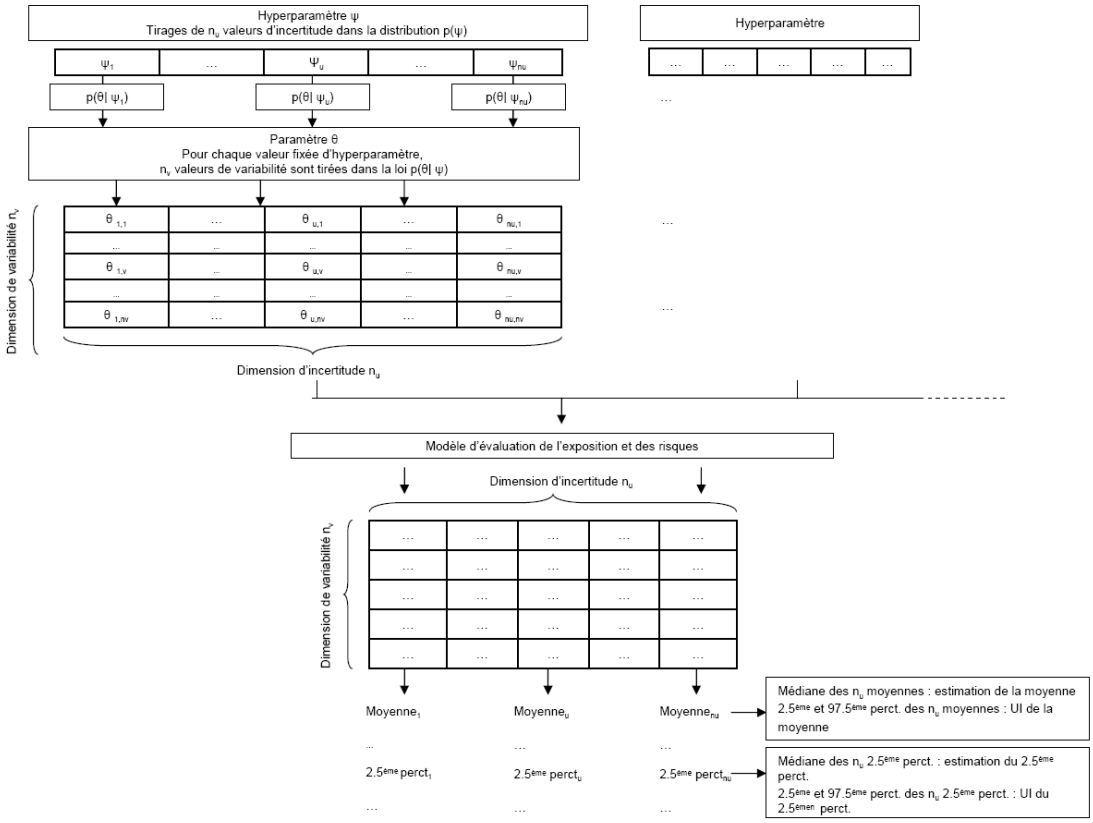

FIG. 1.8: Marche de la simulation de Monte-Carlo à deux dimensions

1.4 Organisation du manuscript

Cette thèse s'articule autour de six chapitres, le premier étant celui-ci et le dernier est une discussion générale sur les travaux présentés. Chacun des quatre chapitres intermédiaires est composé d'une introduction, d'une partie méthodologique, d'une description des données utilisées, d'une partie résultats et d'une discussion. Les chapitres 2 et 3 sont consacrés aux développements de modèles spécifiques afin de déterminer la distribution de la contamination microbiologique d'un aliment et les distributions des paramètres de croissance. Les deux chapitres suivants (chapitres 4 et 5) portent sur la construction du modèle de simulation des risques liés à la présence de *Listeria monocytogenes* dans les salades de IV^{ème} gamme.

Dans le chapitre 2, je me suis intéressée à un des problèmes majeurs de l'évaluation quantitative des risques microbiologiques qui est le manque de données de contamination adaptées. En effet, les données quantitatives de contamination microbiologique (les concentrations) sont peu nombreuses, et en particulier pour *L. monocytogenes* qui est généralement présente à de très faibles concentrations ce qui rend sa détection et sa quantification difficile. Ainsi, un modèle permettant d'estimer la distribution de contamination microbiologique d'aliments à l'aide des deux types de données disponibles (prévalence et concentration) a été développé. Ce modèle permet aussi de quantifier les concentrations inférieures à la limite de détection des méthodes analytiques qui sont rarement prises en compte dans les évaluations des risques alors que de faibles concentrations peuvent atteindre des valeurs élevées lorsque le produit est conservé assez longtemps dans des conditions favorables à la croissance. Ce chapitre, en proposant cette modélisation, présente un intérêt au niveau de l'évaluation des risques mais aussi au niveau statistique. En effet, on montre comment il est possible d'estimer les paramètres d'un mélange de deux lois normales lorsqu'une partie des données est absente mais que l'on dispose d'un estimateur de l'espérance conditionnelle. Les méthodes d'inférence bayésienne sont utilisées afin d'obtenir les distributions *a posteriori* des paramètres d'intérêt. Ce modèle est appliqué à *L. monocytogenes* dans les légumes toutes catégories confondues (transformés et non transformés) et de toutes origines géographiques afin d'avoir le plus de données possible.

Le troisième chapitre présente trois modèles permettant d'estimer les distributions des paramètres de croissance de *L. monocytogenes* dans les salades feuilles à partir de courbes de croissance obtenues par expérimentation. Ces modèles diffèrent légèrement au niveau de leur structure hiérarchique mais leur construction générale repose sur la méthodologie développée par Pouillot et al. (2003). Cette méthodologie consiste à estimer au sein d'un même

processus les paramètres des modèles de croissance primaire et secondaire de microbiologie prévisionnelle. L'estimation des paramètres s'effectue également par inférence bayésienne ce qui permet d'intégrer de l'information *a priori* issue de travaux antérieurs et indispensable pour estimer l'ensemble des paramètres. Une méthode permettant de valider les modèles proposés et de sélectionner le ou les plus performant(s) est développée.

Le quatrième chapitre est consacré à la modélisation de l'évolution des concentrations en *L. monocytogenes* sur l'ensemble de la chaîne alimentaire et tout au long de la vie du produit. La modélisation s'effectue en trois étapes :

1. l'évaluation, à l'aide du modèle développé dans le second chapitre, de la contamination de la matière première par *L. monocytogenes*.
2. l'évolution des concentrations pendant le processus de transformation de la matière première en produit de IV^{ème} gamme. La principale cause de variation des concentrations en *L. monocytogenes* étant le lavage des feuilles de salades à l'eau chlorée, cette étape est résumée par la modélisation de la réduction en micro-organismes pendant le lavage.
3. l'évolution des concentrations lors du stockage des produits de leur sortie de la chaîne de fabrication jusqu'à leur consommation. Cette étape intègre les résultats obtenus dans le chapitre précédent sur les distributions des paramètres de croissance. Les distributions des durées et des températures de stockage sont estimées à partir de diverses sources d'information comme des dires d'experts, des données bibliographiques ou des rapports d'étude.

Ainsi le pourcentage de sachets contenant des salades contaminées et la distribution de leur concentration à l'instant de leur consommation sont établis à l'aide de simulations de Monte Carlo intégrant les étapes précédentes. Différents scénarios fondés sur des hypothèses méthodologiques comme le choix du modèle de croissance ou sur des hypothèses liées au procédé de fabrication comme le retrait du chlore de l'eau de lavage seront testés afin d'évaluer les effets sur les concentrations bactériennes. Le traitement de l'eau de lavage par le chlore est effectué par les industriels afin d'améliorer la qualité microbiologique des aliments bien qu'aucune étude n'ait montré que le chlore ait un effet virucide. Son utilisation a une mauvaise image auprès des consommateurs même si celle-ci ne présente pas de risque sanitaire pour le consommateur (Afssa Saisine n°2003-SA-0015) dans le cas où les doses ajoutées à l'eau de lavage sont respectées[1]. Ainsi certains pays comme l'Allemagne l'ont déjà banni du procédé de

[1] **Guide des bonnes pratiques d'hygiène des produits végétaux crus prêts à l'emploi, 1996**

fabrication et son utilisation tolérée en France est remise en cause en vue d'une harmonisation des pratiques des pays européens. Enfin, les résultats obtenus sont comparés avec les données de contamination de salade de IV^{ème} gamme disponibles.

Dans le cinquième chapitre, l'exposition des consommateurs à *L. monocytogenes* est calculée en croisant la contamination à l'instant de la consommation avec les quantités consommées. L'enquête de consommation individuelle INCA et le panel des ménages Sécodip répertoriant les achats de ces derniers sur une année sont les deux enquêtes permettant d'évaluer la consommation des français. Le panel d'achat des ménages SECODIP est utilisé afin de déterminer la part de marché des salades en sachet par rapport à l'ensemble des salades achetées (transformées et non transformées), information non dissociée dans l'enquête INCA. Le nombre de bactéries présentes dans les quantités consommées est estimé à partir de la distribution de contamination des sachets et en utilisant un coefficient de groupement afin de prendre en compte la répartition hétérogène des bactéries. Une simulation de la croissance directement dans les quantités consommées est réalisée afin de comparer les deux méthodes. Puis, à l'aide d'une courbe dose-réponse de référence internationale (FAO/WHO, 2004) dont le paramètre diffère selon le type de population (sensible ou normale), le risque d'être atteint de la listériose invasive par consommation de salade en sachet est calculé. Le risque est calculé pour différentes sous-populations : enfants, adultes, personnes âgées et femmes enceintes et pour l'ensemble de la population française. Puisque certaines populations sont sous ou surreprésentées dans l'enquête INCA, le calcul du risque pour l'ensemble de la population française nécessite d'utiliser des poids de redressement afin de caler la répartition des populations de l'enquête INCA sur celle de la population française. Le nombre de cas de listériose invasive par an dans la population française est également simulé et comparé avec celui déclaré à l'institut de veille sanitaire (InVS) chargé de surveiller l'incidence de cette maladie.

Il est important de souligner que ces travaux ont été réalisés à partir d'un grand nombre de données. Le recueil de données a nécessité un travail considérable de prospection, d'harmonisation, de gestion et d'actualisation de base de données. Ces données ont été extraites de la bibliographie existante avec la collaboration de Frédéric Carlin ou m'ont été fournies par la DGCCRF et par les centres Actia chargés de mener en parallèle de mes travaux de modélisation des études de terrain en collaboration avec les entreprises afin que ce travail soit constamment alimenté, autant que possible, en données pertinentes et de qualité.

Chapitre 2

Modèle d'estimation de la contamination microbiologique d'un aliment

2.1 Introduction

Une appréciation quantitative des risques microbiologiques ou chimiques nécessite de connaître avec précision le niveau de contamination des aliments. Les données de contamination chimique ou microbiologique proviennent de sources variées. En effet, les données de contamination sont issues de plans de contrôle effectués par les institutions nationales en charge du contrôle sanitaire des produits (DGAL ou DGCCRF) ou sont fournies par les offices nationaux interprofessionnels des filières agro-alimentaires, par les institutions de recherche spécialisées ou par des centres techniques. Lorsque le couple aliment/contaminant a suscité peu d'intérêt, l'évaluateur peut avoir recours à des valeurs extraites de la bibliographie sur le sujet pour compléter les données disponibles. Lorsqu'elles sont extraites de revues à comité de lecture, ces données peuvent être considérées de bonne qualité, car elles reflètent l'expertise de leurs auteurs et elles sont soumises à une évaluation de leur pertinence scientifique et technique.

Le principal problème des données de contamination microbiologique est que pour la plupart des aliments, les analyses relevant le nombre de bactéries présentes dans une quantité donnée d'aliment (données de concentration) sont peu nombreuses. En effet, comme ces analyses sont coûteuses en terme de temps et de moyens, seule l'information sur la présence

ou l'absence de micro-organismes dans le produit est relevée. Ainsi, au cours d'une étude, lorsque plusieurs produits correspondant au même aliment sont analysés, le plus souvent, seule la prévalence de la contamination (i.e. le pourcentage de produits détectés contaminés (positifs) parmi les produits analysés) est disponible. De plus, les données de contamination microbiologique ainsi que les données de contamination chimique présentent la même difficulté qui est la présence d'une limite de détection en dessous de laquelle la quantité de contaminant ne peut être détectée. Dans la plupart des travaux, les produits dont la contamination n'a pas été détectée sont considérés non contaminés ou alors une valeur leur est attribuée, comme la valeur de la limite de détection divisée par deux (par exemple, Gill et al., 1997) ou encore la valeur de la limite de détection elle-même (par exemple, Garriga et al., 2002). Les travaux de Lorimer & Kiermeier (2007) mettent en évidence la surestimation de la moyenne des concentrations dans ce dernier cas. Ils soulignent aussi une surestimation de la moyenne lorsque celle-ci est calculée à partir des seules valeurs supérieures à la limite de détection. Ces traitements peuvent entraîner des biais importants sur l'évaluation du risque, notamment dans le cas où les concentrations microbiologiques non détectées sont considérées comme nulles. En effet, une quantité initiale même faible de bactéries présentes sur un produit peut atteindre une dose critique pour la santé humaine si les conditions de stockage ultérieures du produit sont favorables à une croissance rapide des bactéries. Des méthodes ont été développées afin de traiter ce problème de censure (voir par exemple l'ouvrage de Helsel, 2004). C'est le cas des travaux de Tressou (2006) qui appliquent l'estimateur non-paramétrique de type Kaplan & Meier (Kaplan & Meier, 1958) à la censure à gauche dans le cadre d'une analyse de risque lié à la présence d'Ochratoxine A, ou encore des travaux de Shorten et al. (2006) ainsi que ceux de Lorimer & Kiermeier (2007) qui développent une méthode d'estimation paramétrique par maximum de vraisemblance en modélisant le nombre de bactéries par une loi log-normale (Gill, 1996; Reinders et al., 2003; Jarvis, 2000). Néanmoins, ces méthodes sont performantes seulement dans le cas où les données de contamination avec des concentrations au-dessus de la limite de détection sont disponibles en grand nombre.

Ce chapitre présente une méthode permettant d'estimer la contamination microbiologique d'un aliment en tenant compte des concentrations inférieures à la limite de détection, à partir d'études pour la majorité de prévalences. Les données et en particulier les données de concentration étant peu nombreuses, différentes sources d'information ont été explorées (bibliographiques, des centres techniques et des plans de contrôle nationaux). Les concentrations en logarithme décimal sont modélisées par un mélange de lois normales. La première composante du mélange représente la distribution de contamination des produits dont les

concentrations sont supérieures à la limite de détection et la seconde composante représente la distribution de contamination des produits dont les concentrations sont inférieures à la limite de détection. Lindqvist et al. (2002) considèrent aussi qu'il existe deux niveaux de concentration en *Staphylococcus aureus* dans les fromages frais et proposent de modéliser les niveaux faibles par une loi uniforme. Afin de pallier le manque de données de concentration et d'estimer les concentrations inférieures à la limite de détection, la modélisation proposée intègre la relation entre la prévalence et la concentration des produits définie au début du siècle (Halvorson & Ziegler, 1933). Les chercheurs de cette époque mettent en place une méthode appelée la méthode "du nombre le plus probable" (NPP) afin de déterminer la concentration moyenne en micro-organismes à partir de sa prévalence dans différentes dilutions réalisées à partir de la quantité à analyser. Dans notre cas, cette relation permet d'estimer le nombre moyen de micro-organismes contaminant un aliment à partir de leur prévalence dans les différents produits analysés. Dans notre modèle, elle nous permet d'inférer sur l'espérance conditionnelle du logarithme des concentrations. L'inférence sur cette espérance conditionnelle ainsi que l'observation des concentrations supérieures à la limite de détection permettent d'estimer les paramètres des deux lois normales composant le mélange. Le modèle a une forme hiérarchique permettant la décomposition de la variance totale des log-concentrations en la variance inter-études et la variance intra-études. Ainsi la structure hiérarchique du modèle et l'estimation par inférence bayésienne facilitent la prise en compte de la variabilité de la contamination microbiologique et de l'incertitude sur l'estimation des paramètres.

Ce chapitre est organisé de la manière suivante. Une première partie est consacrée à la présentation de la méthode dite "du nombre le plus probable" (NPP) permettant de quantifier le nombre de micro-organismes à partir de prévalences lors de dilutions en série. Dans une deuxième partie, la structure des données de contamination microbiologique est présentée et l'utilisation de la méthode du NPP pour évaluer la concentration des aliments est explicitée. Le modèle de mélange et sa formalisation bayésienne sont présentés dans une troisième partie. La méthode permettant de générer des valeurs de log-concentration est également expliquée dans cette section. Enfin, la dernière partie est consacrée à l'application du modèle à la contamination des produits végétaux par *Listeria monocytogenes.* Pour plus de clarté, la présentation est faite pour *Listeria monocytogenes* dans les produits végétaux mais la méthodologie proposée s'adapte à toutes les espèces bactériennes pour lesquelles les données disponibles ont les mêmes caractéristiques (principalement des données de prévalence et peu de données de concentration). Précisons que dans ce chapitre la contamination est

estimée pour l'ensemble des légumes transformés (de IV^{ème} gamme) et non transformés et que la distinction serait faite uniquement à partir du chapitre 4.

2.2 La méthode du "nombre le plus probable" dans le cas de dilutions en série

2.2.1 Principe de la méthode des dilutions

La méthode des dilutions est une méthode biologique de quantification utilisée à l'origine pour estimer la concentration en micro-organismes dans un liquide (McCrady, 1915; Halvorson & Ziegler, 1933; Fisher, 1922; Cochran, 1950). Elle est employée dans plusieurs domaines tels que ceux de la bactériologie (Crainiceanu et al., 2003; Clough et al., 2005; Blodgett & Garthright, 1998), de l'immunologie et de la virologie (Bloch & Chavance, 1998). Cette méthode permet de quantifier le nombre de micro-organismes quand leur comptage direct est impossible mais lorsqu'il est possible de détecter l'absence ou la présence d'au moins un micro-organisme dans la quantité de liquide prélevée pour l'analyse. La méthode consiste à diluer successivement le prélèvement initial dont on veut déterminer la concentration en micro-organismes et de constituer plusieurs réplicats de volume identique pour chaque dilution. Pour chaque réplicat, on cherche à déterminer s'il y a présence ou absence de micro-organismes, le nombre de réplicats détectés positifs dans chacune des séries de dilutions apportant de l'information sur la quantité de micro-organismes présents dans le prélèvement initial.

2.2.2 Estimation de la concentration en micro-organismes

L'estimation de la concentration en micro-organismes repose sur deux hypothèses principales :

- les micro-organismes sont supposés répartis de façon homogène dans les réplicats, c'est-à-dire qu'un micro-organisme a la même probabilité de se trouver dans n'importe quelle partie du réplicat et qu'il n'existe pas de regroupement de micro-organismes dans une partie particulière du réplicat.
- chaque réplicat issu du prélèvement initial peut être testé positif dès lors qu'au moins un micro-organisme est présent.

La spécificité (l'absence de micro-organisme induit bien une réponse négative) et la sensibilité (l'existence d'au moins un micro-organisme donne une réponse positive) de la réponse sont importantes pour la validité du modèle car si certains réplicats sont mal classés (par exemple un réplicat classé négatif alors qu'il contient un ou plusieurs micro-organismes), l'estimation de la concentration sera biaisée.

Si l'hypothèse de répartition homogène des micro-organismes dans les réplicats est vérifiée, pour un niveau de dilution k, le nombre Y_k de micro-organismes présents dans un réplicat est distribué selon une loi de Poisson de paramètre $q_k\lambda$, où q_k est le facteur de dilution et λ la concentration du prélèvement initial à estimer (Halvorson & Ziegler, 1933; Cochran, 1950; Bloch, 1998). Ainsi la probabilité que le réplicat contienne y micro-organismes est égale à

$$P(Y_k = y) = \frac{(q_k\lambda)^y}{y!}e^{-q_k\lambda}. \tag{2.1}$$

Etant donnée la dichotomisation de la réponse en présence/absence, seules comptent les deux probabilités complémentaires de n'avoir aucun micro-organisme dans le réplicat ou d'en avoir au moins un. La probabilité qu'il y ait au moins un micro-organisme dans le réplicat s'écrit donc

$$P(Y_k \neq 0) = p_k = 1 - e^{-q_k\lambda}.$$

Le nombre R_k de réplicats positifs parmi n_k réplicats analysés pour chaque niveau de dilution k obéit à une loi binomiale de paramètre n_k et p_k, ainsi

$$P(R_k = r_k) = \left(\frac{n_k!}{r_k!(n_k - r_k)!}\right)(p_k)^{r_k}(1 - p_k)^{n_k - r_k}.$$

Dans le cas d'une seule dilution et de plusieurs réplicats (n), l'estimateur du maximum de vraisemblance de p est $\widehat{p} = \frac{r}{n}$ où r est le nombre de réplicats positifs parmi les n analysés, ce qui produit la relation suivante

$$\frac{r}{n} = 1 - e^{-q\lambda} \tag{2.2}$$

et l'on obtient ainsi une estimation de la concentration λ, appelée le "nombre le plus probable" (NPP) de micro-organismes par unité de mesure du réplicat (ml, cm^3 ou g),

$$\widehat{\lambda} = \frac{\ln(\frac{n}{n-r})}{q}. \tag{2.3}$$

Dans le cas d'un nombre K de niveaux de dilution, où pour chaque niveau k de facteur de dilution q_k, il y a n_k réplicats analysés dont r_k positifs, la méthode du maximum de vraisemblance est aussi utilisée pour estimer la concentration λ. L'estimateur du maximum de vraisemblance $\widehat{\lambda}$, est la solution de l'équation qui annule la dérivée de la log-vraisemblance. La log-vraisemblance s'écrit, au facteur $\frac{n_k!}{r_k!(n_k-r_k)!}$ près

$$\ln\left(\prod_{k=1}^{K}\left(1-e^{-q_k\lambda}\right)^{r_k}\left(e^{-q_k\lambda}\right)^{n_k-r_k}\right).$$

2.3 Structure des données de contamination microbiologique et notations

Les données de contamination microbiologique peuvent être recensées sous forme "d'études", chaque étude étant constituée d'une série d'analyses. Les études peuvent être classées en deux types :

- soit l'étude recense uniquement la prévalence de la contamination, c'est-à-dire la proportion de produits contaminés. Les données sont alors le nombre de produits contaminés r parmi le nombre total de produits analysés n et la quantité en gramme q prélevée par produit pour les analyses. Cette quantité peut être prélevée, suivant l'étude, soit sur les différentes parties d'un seul aliment (par exemple pour la salade feuille : feuilles externes, coeur, etc...) soit sur l'ensemble des n représentants de cet aliment (plusieurs salades). Pour simplifier, le terme "produit" est employé pour désigner chacune des n quantités de taille q analysées. Les concentrations des r produits contaminés n'ont pas été relevées. La seule information disponible lorsqu'un produit est détecté contaminé est que sa concentration en micro-organismes est supérieure à la limite de détection. La limite de détection en microbiologie dépend de la quantité q analysée. Celle-ci est en général de 25g pour *Listeria monocytogenes*. Les $(n-r)$ produits restants sont déclarés non-contaminés, leur concentration est inconnue et la seule information disponible est que leur concentration est inférieure à la limite de détection.
- soit l'étude relève en plus de la prévalence, la concentration bactérienne, obtenue par dénombrement, de chacun des r produits détectés contaminés. La concentration de ces produits est généralement exprimée en unité formant colonie (ufc) par gramme. Ainsi les valeurs de concentration des r produits de ces études sont observées et supérieures à la limite de détection. Les concentrations des produits contaminés par *L. monocyto-*

genes peuvent parfois être très élevées. Les valeurs sont alors très éloignées de celle de la limite de détection. Les concentrations des $(n - r)$ produits restant sont inconnues et inférieures à la limite de détection.

Pour détecter la présence de bactéries dans une quantité q, la quantité prélevée est broyée et homogénéisée avec un bouillon d'enrichissement. L'homogénat ainsi obtenu est placé pour une durée variable, le plus souvent 24h ou 48h à une température favorable à la croissance de la bactérie étudiée. Ensuite une fraction aliquote (le plus souvent comprise entre 50 et 200 μl) de l'homogénat est étalée sur une ou plusieurs boîtes de Petri adaptées à la bactérie étudiée. Celle-ci est placée à une température favorable pendant plusieurs jours permettant la croissance et l'identification de la bactérie si elle est présente. Un test de confirmation est réalisé pour vérifier que la bactérie identifiée appartient bien à l'espèce considérée.

Pour estimer la concentration moyenne de bactéries présentes dans l'ensemble des n produits analysés d'une étude, la méthode du nombre le plus probable peut être utilisée en considérant que les n produits sont n réplicats d'un même aliment (la condition d'homogénéité étant respectée au sein des réplicats), qu'il n'y a qu'une seule dilution et que le volume analysé est la quantité q (Carlin et al., 2004; Dodds, 1993).

La présentation du modèle d'estimation de la contamination microbiologique des aliments utilise les notations suivantes. Chaque étude est référencée par un indice i $(i = 1, ..., N)$ et chaque produit d'une étude i est référencé par un indice j $(j = 1, ..., n_i)$. Pour un nombre N_0 d'études parmi les N études disponibles, en plus de la prévalence, les concentrations des r_i produits contaminés par étude i $(i = 1, ..., N_0)$ sont observées et relevées. Pour les $(N - N_0)$ études restantes les concentrations sont non observées pour tous les produits analysés et seule la prévalence de la contamination par étude est disponible. La limite de détection d'une étude i est notée d_i. Sa valeur dépend de la quantité q_i de produit analysé qui est différente d'une étude à l'autre. Toutefois, pour simplifier les notations dans la suite, une unique valeur d est considérée pour la limite de détection des différentes études. On note $\log(c_{ij})$, le logarithme en base 10 de la concentration ou log-concentration du produit j $(j = 1, ..., n_i)$ de l'étude i $(i = 1, ..., N)$. $\log(c_{ij})$ est donc observé ou non observé. En résumé, pour les études $i = 1, ..., N_0$,

- la concentration des $j = 1, ..., r_i$ produits détectés contaminés est observée. Les données de concentration sont disponibles et leur valeur est supérieure à la limite de détection d, ce qui implique que $\log(c_{ij})_{\substack{j=1,...,r_i \\ i=1,...,N_0}} > d$.
- la concentration des $j = r_i + 1, ..., n_i$ produits détectés non contaminés est non observable et donc inférieure à d, ainsi $\log(c_{ij})_{\substack{j=r_i+1,...,n_i \\ i=1,...,N_0}} < d$.

Pour les études $i = N_0 + 1, ..., N$, les concentrations des n_i produits analysés sont non observées,

- les $j = 1, ..., r_i$ produits sont détectés contaminés. La seule information sur leur concentration est qu'elle est supérieure à la limite de détection, ainsi $\log(c_{ij})_{\substack{j=1,...,r_i \\ i=N_0+1,...,N}} \geq d$,
- les $j = r_i + 1, ..., n_i$, produits sont détectés non-contaminés, les concentrations sont inférieures à la limite de détection d, $\log(c_{ij})_{\substack{j=r_i+1,...,n_i \\ i=N_0+1,...,N}} < d$.

Par exemple, dans une étude extraite de l'article de (Sagoo et al., 2003b) analysant la contamination des produits végétaux par *L. monocytognes* , sur $n = 3\ 849$ produits analysés, $r = 90$ produits ont été détectés contaminés c'est-à-dire que leur concontration est supérieure à $d = \log(1/25)$ log ufc/g et $n - r = 3\ 759$ ont été déclarés négatifs, leur concentration est donc inférieure à d. De plus, la concentration de chacun des 90 produits détectés contaminés a été mesurée par dénombrement et les valeurs varient entre 0.04 ufc/g et 1000 ufc/g. Dans une autre étude extraite de l'article de Heisick et al. (1989b), sur $n = 132$ produits analysés, $r = 28$ produits ont été détectés contaminés et $n - r = 104$ ont été détectés non contaminés. Les concentrations des 28 produits contaminés n'ont pas été relevées. La seule information disponible est que leur concentration est supérieure à $d = log(1/25)$ log ufc/g et que la concentration des 104 autres produits est inférieure à d.

Dans la suite, la notation "log" ou l'expression "logarithme" fera référence au logarithme en base 10. Comme les calculs se feront à partir du logarithme du nombre de *L. monocytogenes* exprimé en ufc/g ou en NPP/g, l'unité employée sera le logarithme du nombre de *L. monocytogenes* viables par gramme soit log vlm/g.

2.4 Méthode d'estimation et génération de valeurs de log-concentration

Les produits contiennent des concentrations en *L. monocytogenes* de niveau très différent. En effet, les produits détectés contaminés peuvent présenter des concentrations très élevées, nettement supérieures à la limite de détection. A contrario, les concentrations des produits déclarés non contaminés sont non détectables car leur concentration microbienne est faible, inférieure à la limite de détection. Il est important de pouvoir prendre en compte les concentrations élevées de contamination qui sont un danger pour la santé humaine mais aussi les faibles concentrations qui lors de la conservation des produits par exemple peuvent augmenter et atteindre des valeurs élevées. Compte tenu de leurs valeurs très différentes, Lindqvist

et al. (2002) ont proposé de modéliser les concentrations inférieures et supérieures à la limite de détection selon deux distributions. Les données étant disponibles pour les concentrations supérieures à la limite de détection, ils utilisent la fonction de répartition empirique de ces concentrations. Par contre, les concentrations inférieures à la limite de détection n'étant pas observées, les auteurs utilisent une loi uniforme et testent plusieurs scénarios pour les valeurs des paramètres de cette loi.

Partant d'un point de vue similaire à Lindqvist et al. (2002), nous proposons un modèle de mélange de deux distributions afin de modéliser les concentrations inférieures et supérieures à la limite de détection. Ces modèles sont adaptés à notre problèmatique car ils permettent de modéliser la variable d'une population présentant plusieurs sous-populations. Ici, les produits peuvent être répartis en deux catégories selon leur niveau de concentration (faible ou élevé). La concentration des produits de chacune des catégories est modélisée par une distribution qui est propre à cette catégorie. Les distributions choisies sont des lois normales. La loi normale est classiquement utilisée dans la littérature pour modéliser le logarithme des concentrations en micro-organismes (Jarvis, 2000; Lorimer & Kiermeier, 2007). Cependant, l'absence d'observations de concentration inférieure à la limite de détection pose problème pour l'estimation des paramètres de la loi normale représentant les faibles concentrations. La solution que nous proposons ici est de considérer qu'un estimateur de l'espérance conditionnelle des concentrations peut être obtenu par la méthode du NPP. Ainsi, en connaissant les concentrations supérieures à la limite de détection et l'espérance conditionnelle de l'ensemble des concentrations, il est alors possible d'estimer la totalité des paramètres du modèle. Le modèle de mélange ainsi que son développement dans le cas particulier de l'absence de données pour une des composantes et de l'observation en partie des données de l'autre composante sont présentés dans cette section. La formalisation bayésienne du modèle ainsi que la méthode d'estimation des distributions des paramètres et celle de génération de valeurs de log-concentration sont également décrites.

2.4.1 Modèle de mélange de deux lois normales

Les produits sont répartis en deux catégories selon leur niveau de concentration. La première correspond aux produits dont la log-concentration est supérieure ou égale à la limite de détection d. La deuxième catégorie de produits comprend les produits dont la log-concentration est inférieure à la limite de détection d. Les proportions respectives de ces deux niveaux de concentration sont β et $(1-\beta)$. Le logarithme de la concentration des produits est

modélisé par un mélange de deux lois normales dont les paramètres dépendent du niveau de concentration (inférieur ou supérieur à d). Ainsi la densité du logarithme de la concentration $\log(c_{ij})$ d'un produit $j = 1, ..., n_i$ d'une étude $i = 1, ..., N$ s'écrit

$$\left[\log(c_{ij}) \,|\, \beta, mc_i, mp_i, \sigma_3^2\right] \sim \beta N(mc_i, \sigma_3^2) + (1-\beta)N(mp_i, \sigma_3^2), \tag{2.4}$$

où les deux composantes du mélange sont interprétées comme

$$\begin{aligned} \left[\log(c_{ij}) \,|\, \log(c_{ij}) \geq d, mc_i, \sigma_3^2\right] &\sim N(mc_i, \sigma_3^2), \\ \left[\log(c_{ij}) \,|\, \log(c_{ij}) < d, mp_i, \sigma_3^2\right] &\sim N(mp_i, \sigma_3^2), \end{aligned} \tag{2.5}$$

et les moyennes mc_i et mp_i des log-concentrations des deux catégories de produits présents dans chaque étude i sont respectivement distribuées selon une loi normale

$$\begin{aligned} \left[mc_i \,|\, \mu_1, \sigma_1^2\right] &\sim N(\mu_1, \sigma_1^2), \\ \left[mp_i \,|\, \mu_2, \sigma_2^2\right] &\sim N(\mu_2, \sigma_2^2), \end{aligned}$$

où σ_1^2 et σ_2^2 sont les variances inter-études et σ_3^2 est la variance intra-études supposée égale pour les deux catégories de produits.

Ce modèle est hiérarchique c'est-à-dire qu'il est composé de plusieurs niveaux. Le premier niveau de la hiérarchie est le modèle ajusté aux données (2.4). Le deuxième niveau est constitué des distributions des paramètres mc, mp, β et σ_3. Le troisième et dernier niveau de la hiérarchie est composé des distributions des hyperparamètres $\mu_1, \sigma_1, \mu_2, \sigma_2$.

2.4.2 Méthode d'estimation des paramètres et des hyperparamètres du modèle

Le principe de la méthode développée pour estimer les paramètres et les hyperparamètres du modèle de mélange est d'utiliser les données de prévalence apportant de l'information sur la concentration moyenne de l'ensemble des produits de chaque étude $i = 1, ..., N$ ainsi que les données de concentration observées des produits $j = 1, ..., r_i$ des études $i = 1, ..., N_0$, précisant plus particulièrement la forme de la deuxième composante du mélange.

A partir de la relation dérivée de la méthode du NPP, définie en (2.2), on sait que la prévalence p_i et le niveau moyen de concentration λ_i des produits de l'étude i satisfont

l'égalité suivante lorsqu'une quantité q_i a été analysée

$$p_i = 1 - \exp(-q_i \times \lambda_i).$$

A partir de l'hypothèse de mélange de deux lois normales on peut écrire, sachant les paramètres du mélange, l'espérance conditionnelle des $\log(c_{ij})$

$$\mathbb{E}(\log(c_{ij}) \,|\, \beta, mc_i, mp_i) = \beta mc_i + (1-\beta)mp_i. \tag{2.6}$$

En considérant que λ_i est la valeur naturelle de l'espérance conditionnelle des $\log(c_{ij})$, on obtient finalement que

$$p_i = 1 - \exp(-q_i \times 10^{\beta mc_i + (1-\beta)mp_i}).$$

Ceci permet d'exprimer la loi de r_i conditionnellement aux moyennes des différentes composantes du mélange. En effet, le nombre de produits contaminés par étude suit une loi binomiale de paramètres n_i, le nombre total de produits analysés par étude, et p_i la probabilité que les produits de l'étude i soient contaminés par la concentration moyenne $10^{\beta mc_i + (1-\beta)mp_i}$. Ainsi le modèle hiérarchique suivant est construit

$$\begin{aligned}
[r_i \,|\, n_i,\, q_i,\, mc_i, mp_i, \beta] &\sim Bin\left(n_i, 1 - \exp\left(-q_i \times 10^{\beta mc_i + (1-\beta)mp_i}\right)\right), && (2.7)\\
\left[\log(c_{ij}) \,|\, \beta, mc_i, mp_i, \sigma_3^2\right] &\sim \beta N(mc_i, \sigma_3^2) + (1-\beta)N(mp_i, \sigma_3^2), && (2.8)\\
\left[mc_i \,|\, \mu_1, \sigma_1^2\right] &\sim N(\mu_1, \sigma_1^2), && (2.9)\\
\left[mp_i \,|\, \mu_2, \sigma_2^2\right] &\sim N(\mu_2, \sigma_2^2), && (2.10)
\end{aligned}$$

$\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_3^2$ et β étant des paramètres du modèle pour lesquels, dans le cadre d'une approche bayésienne, des lois *a priori* seront déterminées.

Le fait que seulement une partie des concentrations supérieures à la limite de détection d soit observée, c'est-à-dire uniquement pour les N_0 premières études, implique que la loi conditionnelle des log-concentrations définie en (2.8) ne sera que partiellement utilisée dans le cadre de l'estimation des paramètres de ce modèle. Ainsi, pour $i = 1, ..., N_0$; $j = 1, ..., r_i$, on utilisera simplement le sous modèle suivant :

$$[\log(c_{ij}) \,|\, \log(c_{ij}) > d,\, mc_i,\, \sigma_3^2] \;\sim\; N(mc_i, \sigma_3^2), \text{ avec } [mc_i \,|\, \mu_1,\, \sigma_1^2] \;\sim\; N(\mu_1, \sigma_1^2). \tag{2.11}$$

L'information manquante sur les concentrations inférieures à la limite de détection d sera apportée par la relation définie en (2.7) entre le nombre de produits contaminés r_i d'une étude i et la concentration moyenne de l'ensemble des produits de cette étude explicitée par l'espérance conditionnelle des $\log(c_{ij})$ (2.6).

La loi (2.8) sera par contre utilisée pour générer des valeurs de log-concentrations. La combinaison de (2.8) avec (2.9) et (2.10) permet d'en décrire la distribution sans utiliser les moyennes des deux types de produits (mc_i et mp_i) ou les deux types de données (r_i et $\log(c_{ij})$), ainsi

$$\log(c_{ij})|\beta, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \sigma_3^2 \sim \beta N(\mu_1, \sigma_1^2 + \sigma_3^2) + (1-\beta)N(\mu_2, \sigma_2^2 + \sigma_3^2). \tag{2.12}$$

L'intégration des paramètres du second niveau hiérarchique mp et mc équivaut à un modèle bayésien non hiérarchique classique.

2.4.3 Formalisation bayésienne du modèle

Dans cette partie, la vraisemblance des observations est déterminée de manière explicite, les différentes lois *a priori* choisies sont données et les calculs des lois *a posteriori* d'intérêts sont réalisés.

Vraisemblance des observations

Afin d'écrire la loi des observations, il est nécessaire d'expliciter la relation entre les deux types de données (nombres de produits contaminés et log-concentrations) sachant que le nombre de produits contaminés est observé pour chaque étude $i = 1, ..., N$ mais que les log-concentrations ne sont observées que pour les r_i produits contaminés des N_0 premières études. Notons $lc_i = (\log(c_{ij}))_{j \leq r_i}$ pour $i \leq N_0$.

D'après (2.11) et en notant ϕ la densité de la loi normale centrée réduite, la loi des log-concentrations observées ($i \leq N_0$) sachant le nombre de produits contaminés et les hyperparamètres du modèle peut s'écrire

$$L(lc_i \mid r_i, \beta, \mu_1, \sigma_1, \mu_2, \sigma_2, \sigma_3) = \prod_{j=1}^{r_i} \phi \left\{ \frac{\log(c_{ij}) - \mu_1}{\sqrt{\sigma_1^2 + \sigma_3^2}} \right\},$$

ceci, en intégrant la concentration moyenne mc_i des produits de l'étude i contaminés à une

valeur supérieure à d.

La loi du nombre de positifs sachant les hyperparamètres du modèle peut être obtenue en combinant (2.7) avec (2.9) et (2.10), ainsi

$$\begin{aligned} & L(r_i|\beta,\mu_1,\sigma_1,\mu_2,\sigma_2,\sigma_3) \\ \propto & \int\int [1-\exp(-q_i\times 10^{\beta mc+(1-\beta)mp})]^{r_i}\left[\exp(-q_i\times 10^{\beta mc+(1-\beta)mp})\right]^{n_i-r_i} \\ & \times\phi\left\{\frac{mc-\mu_1}{\sigma_1}\right\}\phi\left\{\frac{mp-\mu_2}{\sigma_2}\right\}d(mc)d(mp) \end{aligned}$$

Ainsi, la vraisemblance des observations est donc proportionnelle au produit des deux lois précédentes pour les N_0 premières études et seulement à la loi ci-dessus pour les autres études

$$\begin{aligned} & L(lc_1,...,lc_{N_0},r_1,...,r_N \mid \beta,\mu_1,\sigma_1,\mu_2,\sigma_2,\sigma_3) \\ \propto & \prod_{i=1}^{N_0}\left[\prod_{j=1}^{r_i}\phi\left(\frac{\log(c_{ij})-\mu_1}{\sqrt{\sigma_1^2+\sigma_3^2}}\right)\right]\prod_{i=1}^{N}L(r_i \mid \beta,\mu_1,\sigma_1,\mu_2,\sigma_2,\sigma_3). \end{aligned}$$

Choix des distributions *a priori* des paramètres et des hyperparamètres

La loi jointe des paramètres et des hyperparamètres du modèle est à déterminer. Les distributions marginales *a priori* des paramètres β et σ_3 et des hyperparamètres $\mu_1,\sigma_1,\mu_2,\sigma_2$ sont supposées indépendantes ainsi

$$p(\beta,\mu_1,\sigma_1,\mu_2,\sigma_2,\sigma_3)=p(\mu_1)p(\sigma_1)p(\mu_2)p(\sigma_2)p(\beta)p(\sigma_3).$$

Des lois vagues sont utilisées comme prior sur ces paramètres, la loi $Normale(0,10^6)$ pour les moyennes μ_1 et μ_2, et la loi $Uniforme(0,10^3)$ pour les écart-types σ_1,σ_2 et σ_3. La loi $Beta(\sum_{i=1}^{N}r_i+1,\sum_{i=1}^{N}n_i-\sum_{i=1}^{N}r_i+1)$ informative est attribuée *a priori* au paramètre β.

Distributions *a posteriori*

La loi jointe *a posteriori* des paramètres et des hyperparamètres $\beta,\mu_1,\sigma_1,\mu_2,\sigma_2$ et σ_3 sachant les observations est proportionnelle à

$$L(lc_1,...,lc_{N_0},r_1,...,r_N \mid \beta,\mu_1,\sigma_1,\mu_2,\sigma_2,\sigma_3)p(\beta,\mu_1,\sigma_1,\mu_2,\sigma_2,\sigma_3).$$

Ainsi il est possible de déterminer les lois *a posteriori* de chaque paramètre et hyperparamètre.

Par exemple, la loi *a posteriori* de μ_1 sachant $lc_1, ..., lc_{N_0}, r_1, ..., r_N$ est proportionnelle à

$$p(\mu_1) \int_{\Psi} L(lc_1, ..., lc_{N_0}, r_1, ..., r_N \mid \mu_1, \psi) p(\psi) d\psi,$$

où ψ désigne l'ensemble des paramètres autres que μ_1 et Ψ désigne l'espace de toutes les valeurs possible de ψ.

La loi jointe *a posteriori* de (μ_1, σ_1) sachant $lc_1, ..., lc_{N_0}, r_1, ..., r_N$ est de la même manière proportionnelle à

$$p(\mu_1)p(\sigma_1) \int_{\Psi'} L(lc_1, ..., lc_{N_0}, r_1, ..., r_N \mid \mu_1, \sigma_1, \psi') p(\psi') d\psi',$$

ou ψ' désigne l'ensemble des paramètres autres que μ_1 et σ_1 et Ψ' désigne l'espace de toutes les valeurs possibles de ψ'.

La loi marginale *a posteriori* des log-concentrations moyennes peut aussi être calculée : par exemple la loi *a posteriori* de mc sachant l'ensemble des observations est proportionnelle à

$$\int_{\Sigma} \int_{\Omega} \phi \left\{ \frac{mc - \mu_1}{\sigma_1} \right\} p(\mu_1, \sigma_1) \int_{\Psi'} L(lc_1, ..., lc_{N_0}, r_1, ..., r_N \mid \mu_1, \sigma_1, \psi') p(\psi') d\psi' d\mu_1 d\sigma_1,$$

où Ω et Σ désignent respectivement l'espace de toutes les valeurs possibles de μ_1 et σ_1.

En théorie, il est possible de calculer toutes les distributions d'intérêts. Cependant la forme finale des distributions n'est pas toujours explicite comme par exemple celle de $\int \int [1 - \exp(-q_i \times 10^{\beta mc + (1-\beta)mp})]^{r_i} \left[\exp(-q_i \times 10^{\beta mc + (1-\beta)mp})\right]^{n_i - r_i} \phi \left\{ \frac{mc - \mu_1}{\sigma_1} \right\} \phi \left\{ \frac{mp - \mu_2}{\sigma_2} \right\} d(mc) d(mp)$. C'est pourquoi la plupart des estimations réalisées par inférence bayésienne sont effectuées en utilisant les méthodes MCMC (Markov Chain Monte Carlo) présentées dans le chapitre introductif, afin d'obtenir les distributions *a posteriori* des paramètres et des hyperparamètres.

2.4.4 Génération de valeurs de log-concentration

Il est nécessaire de pouvoir générer des valeurs de concentration à partir de notre modèle afin de représenter graphiquement leur distribution, de déterminer des statistiques de cette distribution ou encore d'injecter ces valeurs comme valeurs d'entrée dans un modèle

général destiné à simuler la croissance des bactéries présentes dans un aliment en vue d'une appréciation quantitative d'un risque microbiologique.

Pour chaque paramètre et hyperparamètre $\beta, \mu_1, \sigma_1, \mu_2, \sigma_2$ et σ_3, $b = 1, ..., B$ valeurs *a posteriori* sont générées par les algorithmes MCMC et ainsi une matrice $(B \times 6)$ est construite. A partir de cette matrice qui contient les B réalisations de la loi jointe *a posteriori* des paramètres et des hyperparamètres, il est possible de simuler des valeurs de log-concentrations. Le principe est de tirer $b = 1, ..., B$ lignes dans la matrice des valeurs *a posteriori* et de simuler directement des valeurs de log-concentrations (notées $\widetilde{\log(c)}$) à l'aide du mélange des deux lois normales ayant comme paramètres les valeurs *a posteriori* tirées de chaque ligne b,

$$\left[\widetilde{\log(c)}|lc_1, ..., lc_{N_0}, r_1, ..., r_N\right] \sim \beta_b N(\mu_{1,b}, \sigma^2_{1,b} + \sigma^2_{3,b}) + (1 - \beta_b)N(\mu_{2,b}, \sigma^2_{2,b} + \sigma^2_{3,b}).$$

Notons qu'il n'est pas nécessaire d'utiliser les distributions des paramètres mc et mp (cf. équation 2.12).

Il est intéressant d'évaluer séparément la variabilité de la contamination microbiologique d'un aliment, dûe à l'hétérogénéité et à la diversité des contaminations, de l'incertitude des estimations produites, dûe au modèle d'estimation et aux données utilisées. A cette fin, la simulation des paramètres incertains est dissociée de la simulation des paramètres variables et ainsi une simulation de Monte-Carlo de second ordre est réalisée (cf. chapitre 1). Les distributions *a posteriori* des hyperparamètres μ_1, μ_2, σ_1 et σ_2 et des paramètres β et σ_3 représentent une incertitude autour de l'estimation de ces paramètres, la distribution de mc, respectivement de mp, représente la variabilité de la contamination moyenne des produits dont la concentration est supérieure, respectivement inférieure à d, et la distribution des $\log(c)$ représente la variabilité de la contamination.

En pratique, n_u valeurs sont tirées dans la loi jointe *a posteriori* de $\beta, \mu_1, \sigma_1, \mu_2, \sigma_2$ et σ_3, reflétant l'incertitude. Puis pour chaque valeur fixée des paramètres d'incertitude, un nombre n_v de valeurs est tiré aléatoirement dans la distribution des log-concentrations $\log(c)$ et des statistiques comme la moyenne et certains percentiles sont calculés. Ainsi un nombre n_u de valeurs est obtenu pour chaque statistique et les $50^{ème}$, $2.5^{ème}$, et $97.5^{ème}$ percentiles de ces statistiques sont calculés afin d'obtenir un estimateur de chaque statistique ainsi qu'un intervalle d'incertitude (UI). Les statistiques descriptives des distributions prédites ainsi que les intervalles d'incertitude des paramètres mc et mp sont aussi calculés.

2.5 Application à *Listeria monocytogenes* dans les produits végétaux

2.5.1 Description des données

A partir de 51 publications issues de la littérature scientifique internationale, de rapports du service d'inspection français (DGCCRF, 1993-1996 and 1997-2001) et de rapports de centres techniques (Stahl et al., 2002, 2005), des données concernant la contamination par *Listeria monocytogenes* des légumes frais transformés et non transformés ont été recensées. Ces publications et rapports couvrent une période d'environ 20 ans, de 1988 à nos jours, et informent sur la prévalence de *Listeria monocytogenes* ainsi que sur le niveau de concentration (unité : ufc/g). Les publications présentant des méthodes de détection mal utilisées ou pour lesquelles le nombre de produits analysés était ambigu n'ont pas été retenues. Un résumé des données est présenté dans le tableau 3.1. Pour chaque publication et rapport, la contamination a été regroupée par produit (ex : salade feuille, chou, persil, tomate, etc...). Chaque regroupement donne lieu à la création d'une étude. Au total N =165 études ont été créées. Parmi ces 165 études, N_0 =15 d'entre elles présentent de l'information sur la concentration des produits trouvés positifs. Cette information est décrite par le nombre de produits présentant une contamination se situant dans un intervalle de concentration c'est-à-dire entre une valeur minimale (a) et une valeur maximale (b) de concentration. Le format par intervalle des données de concentration est principalement dû au regroupement des données dans des tableaux pour publication. Ce format n'étant pas intrinsèque aux données, des valeurs de concentration c_{ij} ont été créées pour chaque produit $j = 1, ..., r_i$ de chaque étude $i = 1, .., N_0$, en procédant à un tirage aléatoire dans une distribution triangulaire[1] (Vose, 2000), $Triang(\min_{ij}, \mathrm{mod}_{ij}, \max_{ij})$ où $\min_{ij} = a_{ij}$, $\mathrm{mod}_{ij} = a_{ij}$ et $\max_{ij} = b_{ij}$. La distribution triangulaire semble la mieux appropriée car ses paramètres peuvent être déterminés sans ambiguïté à partir des données disponibles. De plus, les intervalles étant grands, il est nécessaire d'introduire une décroissance de la probabilité des niveaux de concentration élevés sur un intervalle. Le format par intervalle est également du à la présence d'une limite de quantification au-dessous de laquelle les concentrations, trop faibles, ne peuvent être quantifiées. Les concentrations sont alors comprises entre la limite de détection (borne a de l'intervalle) et la limite de quantification (borne b de l'intervalle). Le nombre total de pro-

[1] $Triang(a, b, c)$, Densité : $f(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)} & \text{si } a \leq x \leq b \\ \frac{2(c-x)}{(c-a)(c-b)} & \text{si } b < x \leq c \end{cases}$

duits contaminés est $\sum_{i=1}^{N} r_i = 754$ parmi $\sum_{i=1}^{N} n_i = 25078$ analysés. La limite de détection d est pour 91% des études égale à 0.04 ufc/g, ce qui correspond à une quantité de produit prélevé de 25g. Les quantités prélevées pour les autres études correspondent à des valeurs de d plus élevées de 0.07 et 0.1 ufc/g.

2.5.2 Réalisation et contrôle de la convergence

L'estimation des paramètres du modèle est réalisée par inférence bayésienne à l'aide du logiciel OpenBugs (Spiegelhalter et al., 2005). Deux chaînes sont générées à partir de valeurs initiales différentes afin de contrôler visuellement la convergence des paramètres vers leur distribution *a posteriori* respective. La convergence est atteinte lorsque les deux chaines se chevauchent. La statistique de Gelman et Rubin, modifiée par Brooks & Gelman (1998) est également utilisée pour chaque paramètre afin de vérifier la convergence. Cette statistique fournit une mesure de la similitude des lois sous-jacentes des dernières valeurs simulées de chaque chaîne. Elle s'appuie sur la comparaison entre la variance inter-chaînes (c'est-à-dire entre les valeurs des différentes chaînes) et la variance intra-chaîne (c'est-à-dire entre les valeurs d'une même chaîne). Quand le rapport entre ces deux grandeurs s'approche de la valeur asymptotique de 1, la convergence est considérée atteinte pour toutes les chaînes. Après convergence, il est nécessaire d'effectuer B itérations afin d'obtenir des valeurs des distributions *a posteriori*. Plus le nombre d'itérations B est élevé, meilleure est l'estimation des distributions *a posteriori*. Un moyen de vérifier la précision des estimations *a posteriori* est de calculer l'erreur de Monte-Carlo pour chaque paramètre. L'erreur de Monte-Carlo est une estimation de la différence entre la moyenne des valeurs simulées, c'est-à-dire celle qui sert d'estimation de la moyenne de la distribution *a posteriori* de chaque paramètre et la moyenne réelle. Une régle est que le nombre d'itérations devant être réalisées doit correspondre pour chaque paramètre à une erreur de Monte-Carlo inférieure à 5% de la valeur de l'écart-type simulé. La convergence est atteinte pour tous les paramètres après 5 000 itérations. Il est nécessaire d'effectuer 40 000 itérations supplémentaires pour satisfaire la régle sur l'erreur de Monte-Carlo. Des statistiques descriptives sont calculées empiriquement pour chaque paramètre et hyperparamètre à partir des valeurs *a posteriori* ainsi obtenues. Des valeurs de log-concentrations sont générées à partir de la distribution prédictive par méthode de simulation de Monte-Carlo de second-ordre. A cette fin, $n_v = 10\ 000 \times n_u = 1000$ itérations sont réalisées. Le choix du nombre d'itérations est expliqué dans la partie réalisation des simulations du chapitre 4.

TAB. 2.1: Nombre d'études extraites des 51 articles et rapports sur la contamination des légumes transformés et non transformés par *L. monocytogenes* selon le type de produit ainsi que le nombre total minimal et maximal de produits analysés

Produit	Nb d'études			Nb de produits analysés			
	Total	Prévalence seule	Prévalence et concentration	Total	Min	Max	Sources
Légumes transformés[a,d]	48	42	6	8 511	1	2 934	A_1
Feuillus transformés[b]	15	10	5	7 511	3	3 849	A_2
Graines germées[c]	12	10	2	785	6	276	A_3
Légumes non transformés[a]	67	66	1	3 934	1	425	A_4
Feuillus non transformés[b]	23	22	1	4 337	10	2 966	A_5
Total	165	150	15	25 078	1	3 849	

[a] betterave, carotte, céleri, chou, chou-fleur, concombre, racine, bulbe, tomate, épinard, brocolis, poireau, herbe diverse, mélange de crudité

[b] laitue, roquette, frisée, mâche, endive

[c] graine germée d'haricot mungo, d'alfalfa et autres graines germées

[d] coupés, lavés, désinfectés, ensachés, etc...

A_1 : Arnold & Coble (1995), Arumugaswamy et al. (1994), Beaufort et al. (1992), Beckers et al. (1988), Bind (1992), Cho et al. (2004), Curtis et al. (2002), DGCCRF (1993-1996 and 1997-2001), Guerra et al. (2001), Harvey & Gilmour (1993), Kaneko et al. (1999), Levre et al. (1995), Little et al. (1997), MacGowan et al. (1994), McLauchlin & Gilbert (1990), Messi et al. (2000), Monge & Arias (1996), Norrung et al. (1999), Previdi et al. (2000), Sagoo et al. (2001), Sagoo et al. (2003a), Stahl et al. (2005), Stahl et al. (2002)

A_2 : Beaufort et al. (1992), Cho et al. (2004), DGCCRF (1993-1996 and 1997-2001), Garcia-Gimeno et al. (1996a), Gombas et al. (2003), Levre et al. (1995), Little et al. (1997), Sagoo et al. (2001), Sagoo et al. (2003b), Stahl et al. (2002), Stahl et al. (2005), Szabo et al. (2000), Tiwari & Aldenrath (1990)

A_3 : Arumugaswamy et al. (1994), Cho et al. (2004), DGCCRF (1993-1996 and 1997-2001), Murase et al. (2002), Thunberg et al. (2002)

A_4 : Arnold & Coble (1995), Arumugaswamy et al. (1994), Beuchat (1996), Bind (1992), Breer & Schopfer (1989), Breer & Baumgartner (1992), de Simon et al. (1992), Dhanashree et al. (2003), Farber et al. (1989), Gohil et al. (1995), Guerra et al. (2001), Heisick et al. (1989a), Inoue et al. (2000), Johannessen et al. (2002), Kaneko et al. (1999), Levre et al. (1995), MacGowan et al. (1994), McLauchlin & Gilbert (1990), Messi et al. (2000), Gunasena et al. (1995), Murase et al. (2002), Pacini et al. (1993), Porto & Uboldi Eiroa (2001),Prazak et al. (2002), Sizmur & Walker (1988), Soriano et al. (2001), Stahl et al. (2002), Stahl et al. (2005), Tiwari & Aldenrath (1990), Villari et al. (1997), Walsh et al. (1998), Wong et al. (1990)

A_5 : Bind (1992), Breer & Baumgartner (1992), Farber et al. (1989), Gombas et al. (2003), Gunasena et al. (1995), Heisick et al. (1989b), Johannessen et al. (2002), Lainé & Michard (1988), Lainé & Michard (1988), Levre et al. (1995), Little et al. (1997), Porto & Uboldi Eiroa (2001), Soriano et al. (2001),Stahl et al. (2002), Stahl et al. (2005), Thunberg et al. (2002), Walsh et al. (1998)

TAB. 2.2: Statistiques descriptives des distributions *a posteriori* des paramètres du modèle de mélange appliqué à la contamination par *L. monocytognes* des produits végétaux (unité : log ufc/g excepté β)

Paramètre	Moyenne	Ecart-type	$P_{0.025}$	$P_{0.50}$	$P_{0.975}$
β	0.030	0.001	0.028	0.030	0.032
μ_1	1.01	0.28	0.46	1.01	1.58
μ_2	-3.65	0.14	-3.95	-3.65	-3.40
σ_1	1.02	0.23	0.69	0.98	1.55
σ_2	1.14	0.12	0.93	1.13	1.39
σ_3	0.57	0.02	0.54	0.57	0.61

2.5.3 Résultats

Le tableau 2.2 présente les statistiques descriptives des distributions *a posteriori* des paramètres et des hyperparamètres du modèle de mélange appliqué à la contamination des végétaux par *L. monocytogenes*. La valeur moyenne *a posteriori* du paramètre β est 3% et les bornes de son intervalle de crédibilité *a posteriori* (CI) de niveau 95% sont données par les $2.5^{ème}$ et $97.5^{ème}$ percentiles. Ainsi 95% des valeurs possibles de ce paramètre se situent dans l'intervalle [2.8 ; 3.2]%. La valeur moyenne *a posteriori* du paramètre μ_1 représentant la moyenne des log-concentrations supérieures à la limite de détection est 1.01 log vlm/g. 95% des valeurs de ce paramètre fluctuent entre 0.46 log vlm/g et 1.58 log vlm/g. L'écart-type σ_1 associé à ce paramètre a une valeur moyenne *a posteriori* de 1.02 log vlm/g et un CI à 95% de [0.69 ; 1.55] log vlm/g. Le paramètre μ_2 représentant la moyenne des log-concentrations inférieures à la limite de détection d a une valeur moyenne *a posteriori* de -3.65 log vlm/g et un CI à 95% de [-3.95 ; -3.40] log vlm/g. Son écart-type associé σ_2 a pour moyenne *a posteriori* une valeur de 1.14 log vlm/g et un CI à 95% de [0.93 ; 1.39] log vlm/g. Enfin l'écart-type intra-étude σ_3 a une valeur moyenne *a posteriori* de 0.57 log vlm/g et un CI à 95% de [0.54 ; 0.61] log vlm/g.

Le modèle prévoit pour les produits contaminés à un niveau supérieur à la limite de détection une contamination moyenne (mc) égale à 1.02 log vlm/g et un intervalle d'incertitude à 95% de [0.30 ; 1.97] log vlm/g (Tab. 2.3). Le $2.5^{ème}$ percentile des log-concentrations des produits fortement contaminés est -1.03 log vlm/g en moyenne avec un UI à 95% de [-2.41 ; 0.01] log vlm/g et le $97.5^{ème}$ percentile a une valeur de 3.06 log vlm/g avec un UI à 95% de [2.03 ; 4.70] log vlm/g. Pour les produits contaminés à un niveau inférieur à la limite de détection, la log-concentration moyenne (mp) est -3.65 log vlm/g avec un UI à 95% de

TAB. 2.3: Statistiques descriptives des paramètres *mc* et *mp* du modèle de mélange appliqué à la contamination par *L. monocytogenes* des produits végétaux (unité : log ufc/g)

		Moyenne	Ecart-type	$P_{0.025}$	$P_{0.5}$	$P_{0.975}$
mc	Estimation	1.02	1.19	-1.03	1.00	3.06
	UI_{inf}	0.30	0.83	-2.41	0.23	2.03
	UI_{sup}	1.97	1.72	0.01	1.98	4.70
mp	Estimation	-3.65	1.26	-6.11	-3.65	-1.18
	UI_{inf}	-3.96	1.09	-6.79	-3.95	-1.53
	UI_{sup}	-3.39	1.50	-5.57	-3.37	-0.76

UI_{inf} et UI_{sup} borne inférieure et borne supérieure de l'intervalle d'incertitude à 95%

TAB. 2.4: Statistiques descriptives de la contamination par *L. monocytogenes* des produits végétaux (unité : log ufc/g)

	Moyenne	Ecart-type	$P_{0.025}$	$P_{0.5}$	$P_{0.975}$
Estimation	-3.51	1.50	-6.10	-3.60	0.07
UI_{inf}	-3.81	1.31	-6.84	-3.91	-0.56
UI_{sup}	-3.25	1.73	-5.54	-3.33	0.82

UI_{inf} et UI_{sup} borne inférieure et borne supérieure de l'intervalle d'incertitude à 95%

[-3.96 ; -3.39] log vlm/g. Le $2.5^{ème}$ percentile des log-concentrations des produits fortement contaminés est -6.11 log vlm/g en moyenne avec un UI à 95% de [-6.79 ; -5.57] log vlm/g et le $97.5^{ème}$ percentile a une valeur de -1.18 log vlm/g avec un UI à 95% de [-1.53 ; -0.76] log vlm/g.

Le modèle prévoit une contamination moyenne en *L. monocytogenes* des produits végétaux de -3.51 log vlm/g et un intervalle d'incertitude à 95% de [-3.81 ; -3.25] log vlm/g (Tab. 2.4). La contamination médiane prédite est du même ordre de grandeur que la moyenne : -3.60 log vlm/g avec un UI à 95% de [-3.91 ; -3.33] log vlm/g. Le $2.5^{ème}$ percentile des log-concentrations prédites est -6.10 log vlm/g avec un UI à 95% de [-6.84 ; -5.54] log vlm/g et le $97.5^{ème}$ percentile a une valeur de 0.07 log vlm/g avec un UI à 95% de [-0.56 ; 0.88] log vlm/g.

2.6 Discussion

La plupart des travaux d'évaluation du risque microbiologique n'intégrent pas les valeurs de contamination inférieures à la limite de détection (Bemrah et al., 2003; Cassin et al., 1998;

Lindqvist & Westöö, 2000; Nauta et al., 2003a). De récentes publications (par exemple, Lorimer & Kiermeier, 2007) ont montré les biais que pouvait entraîner cette absence de prise en compte des faibles niveaux de concentration. Le modèle présenté ici répond à la nécessité d'évaluer les faibles valeurs de concentration. Le développement d'un tel modèle semblait indispensable en vue d'une évaluation pertinente du risque lié à la présence de *L. monocytogenes* dans les salades de IV^{ème} gamme. En effet, *L. monocytogenes* est présente en faible quantité dans les produits végétaux, il est donc difficile de la détecter et de la quantifier. Certains auteurs ont proposé des méthodes pour prendre en compte les faibles concentrations, comme l'ajustement d'une loi log-normale (Lorimer & Kiermeier, 2007) qui nécessite cependant une bonne connaissance des valeurs situées au dessus de la limite de détection ou encore les scénarios testés par Lindqvist & Westöö (2000) sur les paramètres de la distribution uniforme qu'il utilise pour modéliser les concentrations inférieures à la limite de détection. Nous avons voulu proposer un modèle mieux adapté au pathogène *L. monocytogenes* pour lequel les données de concentration sont rares. Notons que ce modèle est générique et peut donc être appliqué à toute espèce bactérienne présentant la même structure de données. Le principe était de combiner les données de concentrations diponibles et l'information sur la concentration moyenne de plusieurs produits que pouvait nous fournir la prévalence de la contamination dans ces produits afin détablir la distribution complète de la contamination des produits végétaux par *L. monocytogenes*. Ce chapitre est alors intéressant d'un point de vue méthodologique. En effet, il est montré comment il est possible d'estimer les distributions des paramètres de la deuxième composante d'un modèle de mélange lorsque les données nécessaires à cette estimation sont absentes et lorsque l'on dispose par contre d'un estimateur de l'espérance conditionnelle. Ce travail montre aussi la performance des techniques d'inférence par méthode MCMC pour l'estimation de modèles complexes. L'estimation de ce modèle avec la technique usuelle du Maximum de Vraisemblance aurait nécessité de mettre en place un algorithme EM complexe. Notons toutefois que l'information apportée par les données ne permet pas l'actualisation d'un prior vague sur le paramètre β et une loi informative a du être utilisée comme loi *a priori* pour ce paramètre. Ce paramètre représente la proportion de produits dont le niveau de concentration est supérieur à la limite de détection, c'est-à-dire la proportion de produits qui ont été déterminés contaminés. Une loi naturelle pour ce paramètre est donc la loi $Beta(\sum_{i=1}^{N} r_i + 1, \sum_{i=1}^{N} n_i - \sum_{i=1}^{N} r_i + 1)$ de moyenne $\frac{\sum_{i=1}^{N} r_i+1}{\sum_{i=1}^{N} n_i+2}$. Dans l'article de Crépet et al. (2007a), β est fixé et différentes valeurs pour ce paramètre ont été testées. Un des objectifs de ce travail était d'évaluer l'impact de l'incapacité des mé-

thodes de détection à détecter certains produits pourtant contaminés par *L. monocytogenes*. L'utilisation d'un coefficient de sensibilité des méthodes de détection jouait sur la valeur du paramètre β. La conclusion était que la sensibilité des méthodes de détection doit être assez grande (c'est-à-dire détecter au moins 80% des positifs réels) afin de limiter le nombre de faux-négatifs qui, dans le cas d'un faible taux de sensibilité (40%), peut altérer la précision de l'estimation de la contamination.

La prise en compte de la variabilité de la contamination microbiologique d'un aliment et de l'incertitude des estimations produites est importante en particulier lorsque les distributions obtenues sont destinées à servir de paramètres d'entrée dans un modèle global d'évaluation des risques (Nauta, 2000; Vose, 1998). La structure hiérarchique du modèle permet d'introduire dans la modélisation la variance intra-étude et la variance inter-étude ainsi que l'incertitude. La variabilité de la contamination par *L. monocytogenes* des produits végétaux est plus grande entre les différentes études que la variabilité existant entre les produits d'une même étude. Ainsi, pour estimer la variabilité globale de la contamination des produits végétaux par *L. monocytogenes*, il est nécessaire de disposer d'un grand nombre d'études de contextes différents (période, origine et produits). Notons que la structure des données ne permet pas d'estimer la variance intra-étude des produits dont la concentration est inférieure à la limite de détection. Il est donc nécessaire de considérer l'égalité entre la variance intra-étude de ces produits et la variance des produits dont la concentration est supérieure à la limite de détection, estimée à partir des observations de concentration. L'incertitude est prise en compte par l'utilisation d'hyperparamètres et ainsi leurs distributions *a posteriori* reflétent l'incertitude des estimations. L'étude des écart-types des hyperparamètres ou encore des distributions des paramètres mc et mp montre que l'incertitude est plus grande pour les paramètres μ_1 et σ_1 associés à la contamination des produits dont le niveau de concentration est supérieur à la limite de détection.

Il est expliqué dans ce chapitre comment il est possible à l'aide des techniques de Monte-Carlo de second ordre de dissocier la variabilité de l'incertitude lors de la génération de valeurs de contamination. Cette technique permet de déterminer laquelle de la variabilité ou de l'incertitude prédomine dans l'estimation de la contamination. La figure 2.1 représente le tracé de 50 fonctions de répartition de la contamination par *L. monocytogenes* des produits végétaux, chacune ayant été réalisée à partir de $n_v = 10\ 000$ valeurs de log-concentrations obtenues à partir d'une valeur fixée des paramètres reflètant l'incertitude. L'allure générale du graphique montre que la variabilité de la contamination est supérieure à l'incertitude. Le cas contraire aurait entraîné une plus grande dispersion des courbes le long de l'axe des

abscisses (Vose, 2000).

Nous donnons ici les principaux résultats intéressants au niveau microbiologique. Nous les comparons également aux résultats obtenus lors d'une étude réalisée récemment par Francis & O'Beirne (2006) sur des produits collectés dans les meubles de vente de supermarchés irlandais. La contamination des produits végétaux transformés et non transformés, prévue par le modèle se situe en moyenne entre 2 et 6 *L. monocytogenes* par 10 kg. Le modèle prévoit que 1.56% (avec un UI à 95% de [0.92 ; 2.23]%) des produits aient des concentrations supérieures à 1 log vlm/g. De même, le pourcentage de produits ayant des concentrations supérieures à 2 log vlm/g et 3 log vlm/g est estimé respectivement à 0.61% avec un UI à 95% de [0.23 ; 1.41]% et à 0.15% avec un UI à 95% de [0.02 ; 0.67]%. Les pourcentages correspondants calculés avec les données de concentration utilisées dans la modélisation sont respectivement 1.87% et 0.61%. Francis & O'Beirne (2006) trouvent une prévalence de 2.9% soit 21 produits contaminés par *L. monocytogenes* parmi 720 analysés. Dans cette étude le pourcentage de produits dont la concentration est supérieure à 3 log vlm/g est 0.14%. Ce pourcentage est équivalent à celui obtenu avec notre modèle (0.15% avec un UI à 95% de [0.02 ; 0.67]%). Enfin, l'exemple suivant illustre l'importance d'utiliser un modèle intégrant les faibles niveaux de concentration. Avec notre modèle le pourcentage de quantités de 200g contenant au moins 1 *L. monocytogenes* est 16.99% avec un UI à 95% de [13.22 ; 21.10]%. Ce pourcentage serait environ 6 fois inférieur (3%) si seules les données de concentration disponibles (c'est-à-dire supérieures à la limite de détection) avait été utilisées.

Finalement, les données de prévalence utilisées comme information sur l'espérance conditionnelle permettent de reconstruire la distribution des concentrations inférieures à la limite de détection. Cependant, ces données sont d'un intérêt réduit pour estimer les contaminations les plus élevées sur lesquelles règnent les incertitudes les plus fortes. Les données de concentration étant un élément déterminant aussi bien dans l'établissement d'une distribution complète, que dans les estimations des forts niveaux de contamination initiale, il est clair que les études combinant la recherche et le dénombrement, certes beaucoup plus lourdes à réaliser, sont d'un apport bien supérieur pour la qualité des estimations réalisées. Dans le cadre d'évaluations quantitatives des risques, elles sont même à privilégier.

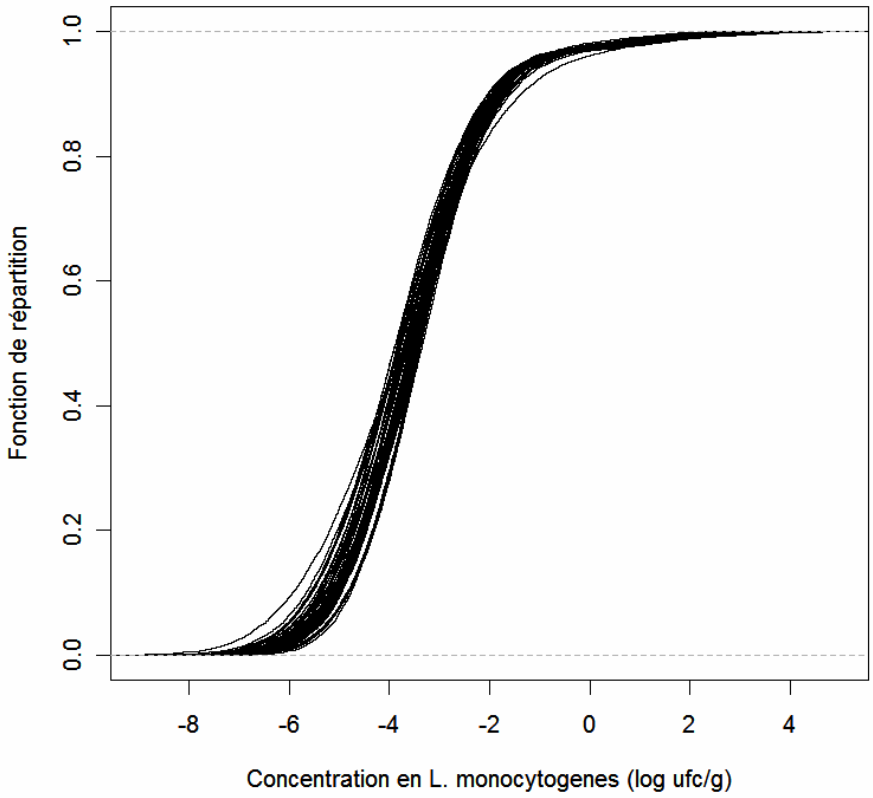

FIG. 2.1: Tracé de 50 fonctions de répartition de la contamination des produits végétaux par *L. monocytogenes* . Chaque fonction de répartition est obtenue à partir de 10 000 valeurs tirées dans la distribution des log-concentrations, dont les paramètres reflétant l'incertitude sont fixés à une valeur prise au hasard dans leur distribution a posteriori.

Chapitre 3

Méthodes d'estimation des paramètres de croissance de *Listeria monocytogenes* dans les salades

3.1 Introduction

Ce chapitre est consacré à la détermination des distributions des paramètres de croissance de *L. monocytogenes* dans les salades feuilles. En effet, même si de nombreuses expériences ont été mises en place afin d'étudier le comportement de *L. monocytogenes* dans les salades feuilles en fonction du temps et de la température, aucune valeur permettant de simuler la croissance de *L. monocytogenes* dans cet aliment n'a été publiée jusqu'à ce jour. Il est donc nécessaire pour réaliser une appréciation quantitative du risque de listériose par consommation de salade de IV^{ème} gamme de déterminer certains paramètres essentiels à la modélisation de la croissance de cette bactérie dans cet aliment. L'étude et la prévision de l'évolution d'une population bactérienne sont les objectifs principaux de la microbiologie prévisionnelle. Ainsi des modèles dits primaires et secondaires ont été développés. Les modèles primaires décrivent l'évolution des concentrations bactériennes en fonction du temps. Un des paramètres de ces modèles est le taux de croissance maximum. Les modèles secondaires décrivent l'influence des paramètres environnementaux (température, pH, activité de l'eau) et propres à la bactéries (type de souche étudiée) sur les paramètres des modèles primaires tels que le taux de croissance maximum et le temps de latence. La méthodologie présentée dans ce chapitre s'appuie sur les travaux de Pouillot et al. (2003) et Delignette-Muller et al. (2006) qui estiment

au sein d'une modélisation globale les distributions des paramètres des modèles primaires et secondaires à partir de plusieurs courbes expérimentales de croissance. L'estimation des distributions des paramètres est réalisée par inférence bayésienne. L'inférence bayésienne facilite l'estimation globale et simultanée des distributions des paramètres des modèles ou des hyperparamètres dans le cas de modèles hiérarchiques. En effet, les procédures d'inférence bayésienne par simulations de Monte-Carlo par Chaîne de Markov, sont suffisamment souples pour pouvoir faire des inférences sur les paramètres de modèles non-linéaires aussi complexes que le sont les modèles secondaires de croissance, à partir de données de courbes de croissance, en une seule étape. En microbiologie prévisionnelle, la procédure s'effectue généralement en deux étapes distinctes : ajustement d'un modèle primaire sur les données expérimentales de croissance, puis ajustement d'un modèle secondaire sur les valeurs du taux de croissance maximum ainsi obtenues. Une autre raison de l'utilisation de l'approche bayésienne est que celle-ci permet l'intégration d'information externe aux données. En effet, bien que les données utilisées proviennent d'une large analyse des articles scientifiques et de rapports techniques étudiant la croissance de *Listeria monocytogenes* dans les salades feuilles, l'information nécessaire à la détermination de certains paramètres comme la température optimale ou maximale de croissance n'est pas disponible. Il est donc nécessaire de pouvoir intégrer dans la modélisation des dires d'experts ou de l'information provenant d'autres études comme celles de Pouillot et al. (2003) ou de Augustin & Carlier (2000a) complétant ainsi les données analysées. De plus, dans ce chapitre nous nous attacherons également à prendre en compte la variabilité de la croissance et l'incertitude sur les estimations. En effet, une première analyse des données montre que pour une température donnée, certaines courbes présentent une absence de croissance tandis que pour d'autres la croissance est très élevée. Cette observation a également été confirmée par des discussions avec des experts. La variabilité et l'incertitude seront donc modélisées dans un second temps par l'utilisation d'un modèle hiérarchique, introduisant ainsi des hyperparamètres dont les distributions *a posteriori* rendront compte de l'incertitude des estimations. L'approche bayésienne permet une interprétation intuitive de la distinction entre la variabilité et l'incertitude. De nombreux travaux montrent l'intérêt de prendre en compte la variabilité et aussi l'incertitude sur les paramètres de croissance comme ceux de Delignette-Muller & Rosso (2000) qui calculent l'exposition à *Bacillus cereus* suite à la consommation de lait pasteurisé.

La première partie de ce chapitre présente quelques généralités sur la microbiologie prévisionnelle et détaille plus particulièrement les modèles primaire et secondaire de croissance qui sont utilisés dans notre modélisation. Les données récoltées pour réaliser ce travail sont

présentées dans une deuxième partie. La troisième partie est consacrée à la présentation de la modélisation. Un premier modèle d'estimation des paramètres de croissance $M1$ est construit. Les résultats obtenus avec ce modèle mettent en évidence une sous-estimation des temps de latence certainement due au faible nombre de points expérimentaux par courbe de croissance. Un deuxième modèle est donc proposé car une sous-estimation des temps de latence engrendre une sous-estimation des taux de croissance maximums $\mu_{\max}$ et donc une sous-estimation du taux de croissance optimal μ_{opt} qui est un des paramètres essentiels pour la simulation de la croissance en conditions réelles de conservation. Le deuxième modèle appelé $M2$ est hiérarchique, un niveau supplémentaire est ajouté sur les paramètres du modèle secondaire. Cette structure hiérarchique définie par rapport à la souche et aux conditions de culture et de préculture de *Listeria monocytogenes* permet de prendre en compte la variabilité pouvant exister au sein d'une même population bactérienne ainsi que celle due aux conditions expérimentales. L'incertitude est aussi prise en compte par l'utilisation de distributions de probabilité sur les hyperparamètres. De plus, la structure hiérarchique de ce modèle permet d'intégrer une contrainte sur les paramètres λ et $\mu_{\max}$ (Delignette-Muller, 1998) afin d'améliorer l'estimation des temps de latence. Cependant, les temps de latence obtenus avec ce deuxième modèle ne sont toujours pas satisfaisants. Pour remédier à ce problème, pour chaque courbe le temps de latence est au préalable calculé comme étant le point d'intersection entre la droite horizontale passant par la concentration initiale et la tangente à la courbe au point de croissance maximum. Ce point est visible sur la figure 3.1. Ensuite, les valeurs obtenues sont introduites dans le modèle. Ainsi le modèle $M2$ est reproduit en considérant que les temps de latence sont des observations, c'est le modèle $M2_{\lambda_{obs}}$. La quatrième partie de ce chapitre présente différentes méthodes de validation, permettant d'évaluer l'adéquation des modèles aux données et de sélectionner celui ou ceux qui seront utilisés dans la suite du travail. Un critère original d'adéquation des modèles aux données est alors développé. Ce critère est la prévision avec un faible taux d'erreur du nombre moyen de *Listeria monocytogenes* s'étant développé en un temps fini et à une température donnée. Enfin les résultats décrivant les distributions des paramètres de croissance selon le modèle utilisé ainsi que les résultats de la validation sont présentés dans une cinquième partie. Une discussion sur la méthodologie et les résultats obtenus clôt ce chapitre.

3.2 Rappel sur les modèles de croissance bactérienne

L'objectif de la microbiologie prévisionnelle est de prédire à l'aide de modèles mathématiques l'évolution d'une population microbienne en fonction du temps, selon des paramètres propres à la bactérie (β_B) et des paramètres environnementaux (β_E). Notons que ces modèles ont d'abord été développés dans un objectif de description de courbes de croissance, puis directement utilisés dans un objectif de prédiction. On distingue deux types de modèles de croissance :

1. les modèles de croissance dits "primaires" décrivant la cinétique d'évolution de la concentration bactérienne au cours du temps, pour un type de micro-organismes et pour des conditions environnementales données. Ces modèles sont de la forme

$$y(t) = f(\beta(\phi), t) + \varepsilon$$

où
- $y(t)$ est le logarithme décimal de la concentration bactérienne à l'instant t (log ufc/g ou log ufc/ml),
- f est une fonction de régression décrivant la relation entre le temps et la concentration bactérienne,
- $\beta(\phi)$ est un vecteur de paramètres comme par exemple le taux de croissance maximum et le temps de latence, fonction des conditions expérimentales ϕ qui sont les conditions environnementales de croissance, la souche étudiée et les conditions de pré-incubation de la souche avant sa mise en culture dans le milieu étudié (appelées conditions de pré-culture),
- ε est l'erreur associée au modèle, combinant l'erreur d'approximation du modèle et l'erreur de mesure de l'observation

2. les modèles de croissance dits "secondaires" permettant de caractériser les paramètres du modèle primaire selon l'aptitude de croissance des bactéries et les conditions environnementales subies. Ces modèles sont de la forme

$$\beta(\phi) = g(\beta_E, \beta_S) + \varepsilon$$

où
- β_S est un vecteur de paramètres propres à la souche,

- β_E est un vecteur de covariables environnementales comme par exemple les conditions de pré-cultures, la température, le pH, la concentration d'acide ou l'activité de l'eau,
- ε est l'erreur associée au modèle.

3.2.1 Modèle primaire de croissance

La cinétique de croissance bactérienne est classiquement composée de cinq phases (Fig. 3.1)

- une phase de latence, durant laquelle la variation de la concentration bactérienne au cours du temps est quasi-nulle ou supposée nulle dans certains modèles. Cette phase est une période d'adaptation de la bactérie entre un état physiologique initial et l'état physiologique permettant la croissance. Cette phase peut-être absente lorsque les conditions de culture sont proches des conditions de celles de pré-incubation ;
- une phase d'augmentation de la vitesse de croissance ;
- une phase de croissance dite "exponentielle" durant laquelle la vitesse de croissance passe par un maximum. Ce maximum correspond physiologiquement à un optimum de condition de croissance de la bactérie ;
- une phase de diminution de la vitesse de croissance, devenant progressivement nulle. Cette phase correspondrait physiologiquement à une phase de saturation de la croissance liée au milieu ;
- une phase stationnaire, la concentration bactérienne est alors à son maximum ;
- une phase de décroissance qui traduit une perte de viabilité cellulaire consécutive à l'appauvrissement du milieu.

Les principaux paramètres utilisés par les microbiologistes avant même l'utilisation des modèles mathématiques sont

- la taille initiale de l'inoculum, y_0 (log ufc/g), définie comme le logarithme décimal de la concentration initiale en bactéries lors de la mise en culture dans le milieu étudié. Ce paramètre correspond à $f(0)$ et dépend donc de la courbe de croissance étudiée ;
- la durée de la phase de latence, $\lambda(\phi)$ (h), classiquement définie comme le moment de l'intersection de la droite correspondant à la phase exponentielle de la croissance avec la droite horizontale passant par la concentration initale y_0. Ce paramètre a été également défini comme le point d'accélération maximale de la croissance (Buchanan & Cygnarowicz, 1990). Ce paramètre est fonction principalement de la souche utilisée et des conditions de pré-culture ;

- le taux de croissance maximum, $\mu_{\max}(\phi)$ (h^{-1}), qui est la pente maximale atteinte durant la croissance dans un graphique en coordonnées logarithmiques. Ce paramètre dépend principalement des conditions environnementales ;
- la concentration maximale atteignable par la population bactérienne en logarithme décimal, $y_{\max}$ (log ufc/g), correspondant à la phase stationnaire.

De nombreux modèles de croissance primaires ont été décrits et utilisés dans la littérature (Zwietering et al., 1990; Baranyi et al., 1993; Baranyi & Roberts, 1994; Buchanan et al., 1996). Le modèle retenu et décrit ici est le modèle logistique avec délai et rupture proposé par Kono (1968) et repris par Rosso (1995). Ce modèle a pour hypothèse une absence totale de croissance durant la phase de latence et la transition entre la phase de latence et la phase exponentielle est considérée instantanée. Dans la littérature, les concentrations bactériennes sont généralement exprimées en logarithme décimal, le modèle logistique avec délai et rupture exprimé dans cette unité s'écrit (Delignette-Muller et al., 2006)

$$y(t) = \begin{cases} y_0 & t \leq \lambda(\phi) \\ y_{\max} - \log_{10}(1 + (10^{(y_{\max}-y_0)} - 1)\exp(-\mu_{\max}(\phi)(t - \lambda(\phi)))) & t > \lambda(\phi) \end{cases} \tag{3.1}$$

où y_0, $y(t)$ et $y_{\max}$ sont respectivement la taille de l'inoculum, la concentration à l'instant t et la concentration maximale atteignable exprimées en logarithme décimal (log ufc/g). Le taux de croissance maximum $\mu_{\max}(\phi)$ et le temps de latence $\lambda(\phi)$ sont les solutions de l'équation des modèles secondaires pour les conditions expérimentales ϕ.

3.2.2 Modèle secondaire de croissance

Ces modèles permettent de décrire l'influence des paramètres environnementaux et propres à l'espèce bactérienne sur le taux de croissance maximum $\mu_{\max}(\phi)$ et le temps de latence $\lambda(\phi)$. Ces paramètres sont en général la température, le pH, l'activité de l'eau, le taux d'oxygène et de dioxyde de carbone, les conditions de pré-culture et la souche étudiée. De nombreux modèles ont été proposés prenant en compte l'influence des différents paramètres séparément (Ratkowsky et al., 1982; Rosso et al., 1993) ou prenant en compte l'interaction entre les facteurs (Augustin & Carlier, 2000b). Dans ce travail, la température à laquelle est conservé le produit sera considérée comme l'unique paramètre ayant une influence sur le taux de croissance maximum et le temps de latence. Cette hypothèse semble valide dans le sens où le pH présente peu de variation selon la variété de la salade et les valeurs de pH relevées ne correspondent pas à celles rencontrées par la bactérie à la surface du produit. De même,

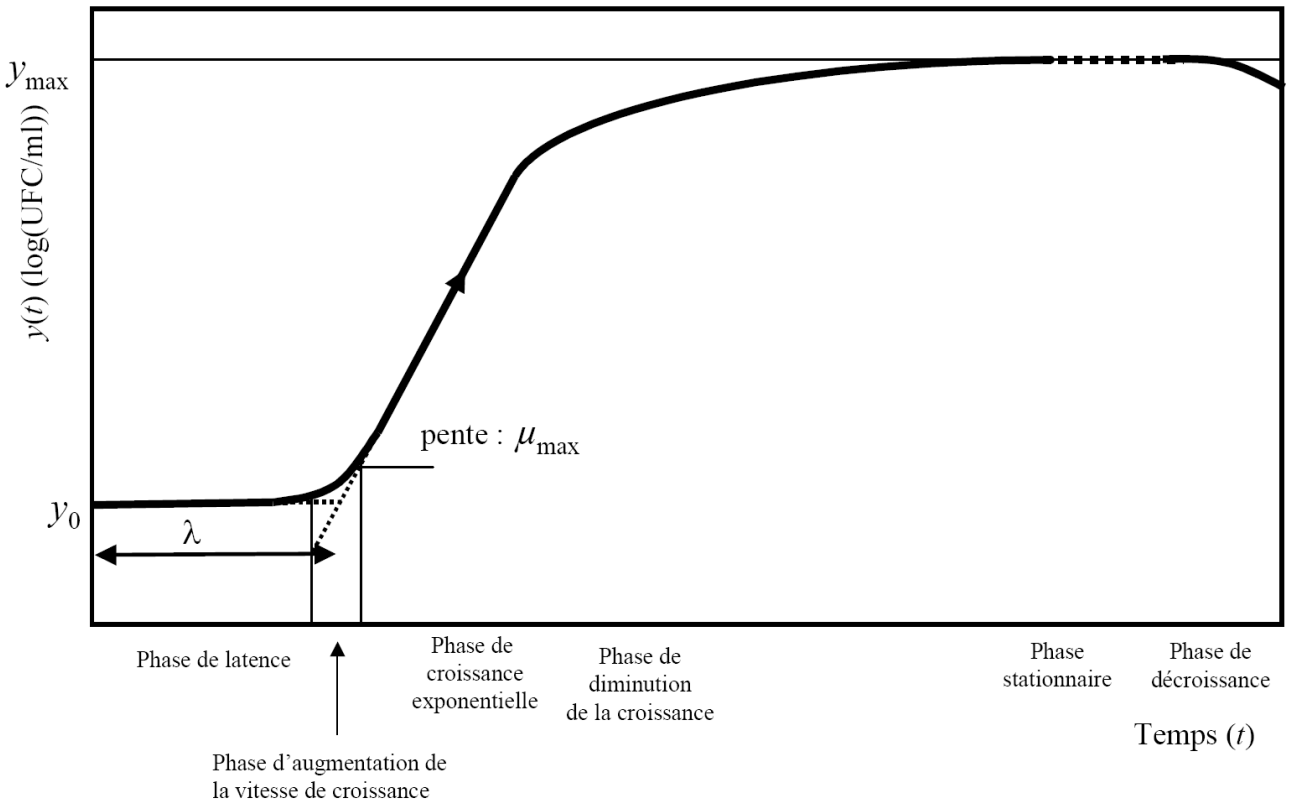

FIG. 3.1: Courbe de croissance bactérienne (Source : Pouillot (2006))

l'activité de l'eau est difficilement mesurable à la surface du produit. Sa valeur est proche de 1 en tenant compte de la très forte humidité du produit et des conditions d'humidité saturante dans les sachets. Cette hypothèse est aussi posée d'un point de vue pratique car seule la température est un paramètre systématiquement étudié lors des expérimentations.

Modèle secondaire sur le taux de croissance maximum $\mu_{\max}$

Le modèle utilisé pour décrire l'influence de la température sur le taux de croissance maximum est le modèle à valeurs cardinales développé par Rosso (1995), qui a pour particularité qu'une signification biologique peut être attribuée à tous ses paramètres (cf. Figure 3.2). Lorsque seule la température est prise en compte il s'écrit :

$$\sqrt{\mu_{\max}(T)} = \sqrt{g(T)} + \varepsilon$$

où ε est l'erreur du modèle et g est définie par

$$g(T) = \begin{cases} 0 & T \notin [T_{\min}, T_{\max}] \\ \frac{\mu_{opt}(T-T_{\max})(T-T_{\min})^2}{(T_{opt}-T_{\min})[(T_{opt}-T_{\min})(T-T_{opt})-(T_{opt}-T_{\max})(T_{opt}+T_{\min}-2T)]} & T \in [T_{\min}, T_{\max}] \end{cases} \tag{3.2}$$

où $T_{\min}$ et $T_{\max}$ sont respectivement la température théorique minimale et la température théorique maximale de croissance (unité : °C) i.e. que les bactéries ne se développent pas lorsqu'elles sont exposées à des températures inférieures ou supérieures à celles-ci. T_{opt} est la température théorique pour laquelle le taux de croissance maximum théorique $\mu_{\max}(T)$ est optimal et égal à μ_{opt} lorsque l'erreur du modèle est négligée. En d'autres termes, T_{opt} est la température idéale pour le développement des bactéries, i.e. la température pour laquelle les bactéries ont le taux de croissance le plus élevé. μ_{opt} est donc le taux de croissance optimal (unité :h^{-1}) atteignable dans les conditions environnementales autre que la température. Ce paramètre dépend de la souche étudiée mais aussi de la matrice alimentaire dans laquelle se développe cette bactérie. Pour les raisons énoncées précédemment, l'influence des autres facteurs comme le pH ou l'activité de l'eau ou l'influence des interactions entre plusieurs facteurs sur la valeur du taux de croissance maximum $\mu_{\max}(T)$ ne sont pas directement prises en compte par l'introduction de paramètres spécifiques mais sont indirectement prises en compte dans la valeur du taux de croissance optimale μ_{opt}. La transformation racine-carré pour $\mu_{\max}(T)$ est utilisée pour stabiliser la variance (Zwietering et al., 1990; Schaffner, 1994).

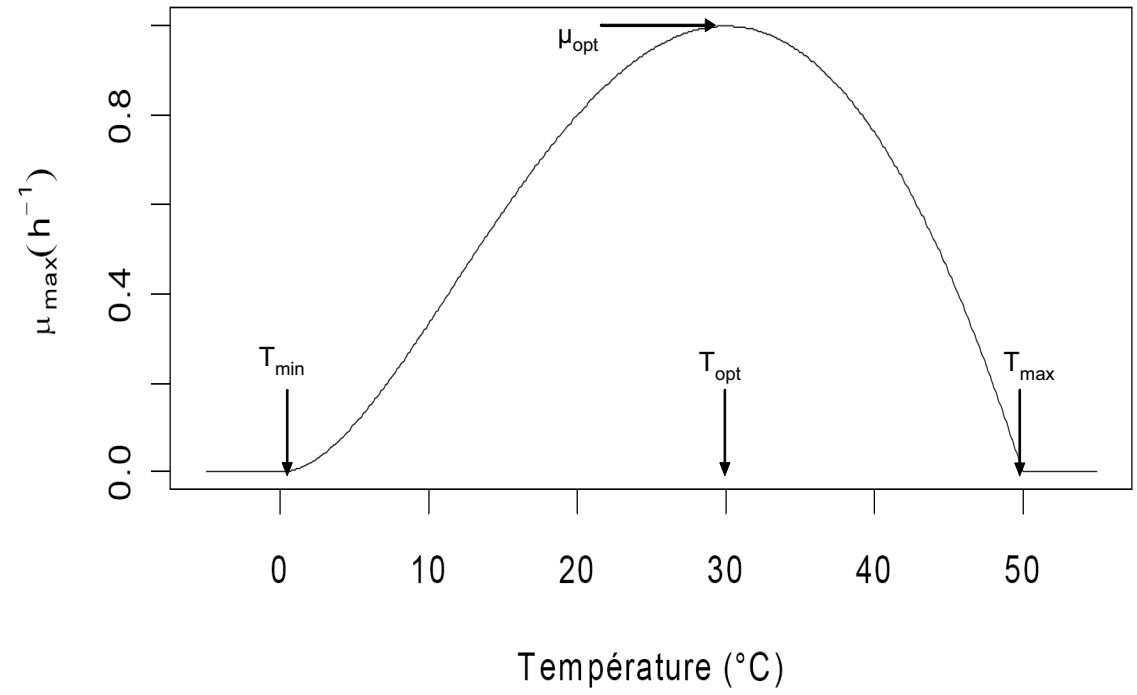

FIG. 3.2: Evolution du taux de croissance $\mu_{\max}$ en fonction de la température tel que proposé par le modèle à valeurs cardinales (Source : Pouillot (2006))

Modèle secondaire sur le temps de latence λ

Soit $T_g = \frac{\ln(2)}{\mu_{\max}(T)}$ le temps de génération, qui est le temps nécessaire pour que la concentration initiale en bactérie soit multipliée par deux durant la phase exponentielle de croissance. Plusieurs auteurs (Zwietering et al., 1994; Rosso, 1995; Delignette-Muller, 1998) ont montré que quelle que soit la température T, il existe une relation de proportionalité entre le temps de latence $\lambda(T)$ et le temps de génération T_g pour une souche ayant subi des conditions de pré-culture données. Ainsi, on peut écrire la relation suivante

$$\lambda(T) \times \mu_{\max}(T) = K + \varepsilon \tag{3.3}$$

où K est une constante propre à la souche bactérienne et dépendant des conditions de pré-cultures et ε l'erreur associée.

3.3 Description des données

Les données ont été sélectionnées parmi celles existant dans la littérature scientifique ou proviennent de rapports des centres techniques (Stahl et al., 2002). Une analyse étendue des articles scientifiques étudiant la croissance de *Listeria monocytogenes* dans les salades feuilles de 1991 à nos jours a donc été réalisée. Certaines courbes de croissance proviennent de la base de données Symprevius (www.symprevius.net).Ces courbes concernent la croissance de *L. monocytogenes* sur toutes variétés de salades feuilles (laitue, salade iceberg, scarole, mâche et frisée) à différentes températures. Ainsi, 206 courbes de croissance comportant au total 1134 valeurs de concentrations (ou points expérimentaux) sur salades feuilles sur des durées d'incubation pouvant aller jusqu'à 18 jours ont été rassemblées et sont présentées dans le tableau 3.1. Les températures auxquelles ont été étudiés les niveaux de croissance varient de 1°C à 25°C.

Les 206 courbes sont classées en 43 groupes correspondant aux différentes souches de *L. monocytogenes* étudiées ou aux différentes conditions expérimentales mises en place. Ainsi deux courbes sont classées dans deux groupes différents si les souches étudiées sont différentes ou si ces courbes de croissance ne sont pas issues de la même publication ou du même rapport, ce qui suppose les conditions expérimentales différentes.

Il est important de noter que les températures auxquelles ont été effectuées les expériences ne couvrent pas entièrement les valeurs nécessaires au modèle secondaire à valeurs cardinales. En effet, Augustin & Carlier (2000a) déterminent des valeurs médianes respectivement de

37°C et 45°C pour les paramètres de température optimale T_{opt} et de température maximale $T_{\max}$. Toutefois, ce manque de données sera comblé par l'utilisation de distributions *a priori* informatives pour ces paramètres.

Notons aussi que le nombre de points expérimentaux par courbe est variable et le plus souvent faible, entre 2 et 29 points par courbe (Tab. 3.1). De plus, certaines courbes effectuées à de basses températures présentent une décroissance en *L. monocytogenes*. Pour ces courbes, le taux de croissance maximum sera considéré comme nul, $\mu_{\max} = 0\text{h}^{-1}$.

TAB. 3.1: Description des données disponibles pour l'étude de la croissance de *L. monocytogenes* dans les salades

N° de courbe	Variété	Souche ou sérotype	Temp. (°C)	Nb de points /courbe	Temps d'incubation (h)	N° de groupe	Source
1	Laitue	4b	5	4	168	1	Ohsone et al. (1999)
2	Laitue	4b	10	5	120	1	Ohsone et al. (1999)
3	Laitue	4b	25	3	72	1	Ohsone et al. (1999)
4*	Laitue	Scott A	10	4	168	2	Carlin et al. (2004)
5*	Frisée	Scott A	10	4	168	2	Carlin et al. (2004)
6	Mâche	Scott A	10	4	168	2	Carlin et al. (2004)
7	Scarole	Scott A	10	4	168	2	Carlin et al. (2004)
8	Scarole	4b	4	2	300	3	Garcia-Gimeno et al. (1996b)
9	Laitue	ATCC 7644	5	3	336	4	Steinbruegge et al. (1988)
10	Laitue	ATCC 7644	12	3	336	4	Steinbruegge et al. (1988)
11	Laitue	ATCC 7644	25	3	8	4	Steinbruegge et al. (1988)
12	Laitue	ATCC 7644	25	3	8	4	Steinbruegge et al. (1988)
13	Laitue	4b	4	2	336	5	Kakiomenou et al. (1998)
14	Laitue	4b	4	2	336	5	Kakiomenou et al. (1998)
15	Laitue	Mixte	7	3	168	6	Jacxsens et al. (1999)
16*	Laitue	Mixte	7	3	168	6	Jacxsens et al. (1999)
17	Scarole	Scott A	3	6	336	7	Carlin et al. (1996a)
18	Scarole	Scott A	3	6	336	7	Carlin et al. (1996a)
19	Scarole	Scott A	3	6	336	7	Carlin et al. (1996a)
20	Scarole	Scott A	3	6	336	7	Carlin et al. (1996a)
21*	Scarole	Scott A	10	4	168	7	Carlin et al. (1996a)

suite page suivante

Tab. 3.1 – suite

N° de courbe	Variété	Souche ou sérotype	Temp. (°C)	Nb de points /courbe	Temps d'incubation (h)	N° de groupe	Source
22*	Scarole	Scott A	10	4	168	7	Carlin et al. (1996a)
23	Scarole	Scott A	10	4	168	7	Carlin et al. (1996a)
24*	Scarole	Scott A	10	4	168	7	Carlin et al. (1996a)
25*	Scarole	Scott A	20	7	70	8	Carlin et al. (1995)
26*	Scarole	Scott A	20	7	70	8	Carlin et al. (1995)
27	Scarole	Scott A	3	5	336	8	Carlin et al. (1995)
28*	Scarole	Scott A	6	6	336	8	Carlin et al. (1995)
29*	Scarole	Scott A	10	4	168	8	Carlin et al. (1995)
30	Scarole	Scott A	3	5	336	8	Carlin et al. (1995)
31*	Scarole	Scott A	6	6	336	8	Carlin et al. (1995)
32*	Scarole	Scott A	10	4	168	8	Carlin et al. (1995)
33	Scarole	LCDC 81-1087	3	4	240	9	Carlin et al. (1996a)
34	Scarole	LCDC 81-1087	3	4	240	9	Carlin et al. (1996a)
35	Scarole	LCDC 81-1087	3	4	240	9	Carlin et al. (1996a)
36	Scarole	LCDC 81-1087	3	4	240	9	Carlin et al. (1996a)
37*	Scarole	LCDC 81-1087	10	4	168	9	Carlin et al. (1996a)
38*	Scarole	LCDC 81-1087	10	4	168	9	Carlin et al. (1996a)
39*	Scarole	LCDC 81-1087	10	4	168	9	Carlin et al. (1996a)
40	Scarole	LCDC 81-861	3	4	240	10	Carlin et al. (1996a)
41	Scarole	LCDC 81-861	3	4	240	10	Carlin et al. (1996a)
42	Scarole	LCDC 81-861	3	4	240	10	Carlin et al. (1996a)
43	Scarole	LCDC 81-861	3	4	240	10	Carlin et al. (1996a)
44*	Scarole	LCDC 81-861	10	4	168	10	Carlin et al. (1996a)
45*	Scarole	LCDC 81-861	10	4	168	10	Carlin et al. (1996a)
46*	Scarole	LCDC 81-861	10	4	168	10	Carlin et al. (1996a)
47*	Scarole	LCDC 81-861	10	4	168	10	Carlin et al. (1996a)
48	Scarole	Scott A	3	4	240	7	Carlin et al. (1996a)
49	Scarole	Scott A	3	4	240	7	Carlin et al. (1996a)
50	Scarole	Scott A	3	4	240	7	Carlin et al. (1996a)
51	Scarole	Scott A	3	4	240	7	Carlin et al. (1996a)

suite page suivante

Tab. 3.1 – suite

N° de courbe	Variété	Souche ou sérotype	Temp. (°C)	Nb de points /courbe	Temps d'incubation (h)	N° de groupe	Source
52*	Scarole	Scott A	10	4	168	7	Carlin et al. (1996a)
53*	Scarole	Scott A	10	4	168	7	Carlin et al. (1996a)
54	Scarole	Scott A	10	4	168	7	Carlin et al. (1996a)
55*	Scarole	Scott A	10	4	168	7	Carlin et al. (1996a)
56	Scarole	Scott A	3	6	336	7	Carlin et al. (1996a)
57	Scarole	Scott A	3	6	336	7	Carlin et al. (1996a)
58	Scarole	Scott A	3	6	336	7	Carlin et al. (1996a)
59	Scarole	Scott A	3	6	336	7	Carlin et al. (1996a)
60	Scarole	Scott A	10	4	168	7	Carlin et al. (1996a)
61	Scarole	Scott A	10	4	168	7	Carlin et al. (1996a)
62	Scarole	Scott A	10	4	168	7	Carlin et al. (1996a)
63*	Scarole	Scott A	10	4	168	7	Carlin et al. (1996a)
64	Scarole	LCDC 81-1087	10	3	168	11	Carlin et al. (1995)
65	Scarole	LCDC 81-1087	10	4	168	11	Carlin et al. (1995)
66	Scarole	LCDC 81-861	10	3	168	12	Carlin et al. (1995)
67	Scarole	LCDC 81-861	10	4	168	12	Carlin et al. (1995)
68*	Scarole	Scott A	10	3	168	13	Carlin et al. (1995)
69*	Scarole	Scott A	10	4	168	13	Carlin et al. (1995)
70	Scarole	Scott A	10	4	168	13	Carlin et al. (1995)
71	Scarole	Scott A	10	3	168	13	Carlin et al. (1995)
72	Scarole	Scott A	10	3	168	13	Carlin et al. (1995)
73	Scarole	Scott A	10	4	168	13	Carlin et al. (1995)
74	Scarole	Scott A	10	3	168	13	Carlin et al. (1995)
75	Scarole	Scott A	10	4	168	13	Carlin et al. (1995)
76	Scarole	Scott A	10	3	168	13	Carlin et al. (1995)
77	Scarole	Scott A	10	3	168	13	Carlin et al. (1995)
78*	Scarole	Scott A	10	4	168	13	Carlin et al. (1995)
79*	Scarole	Scott A	10	3	168	13	Carlin et al. (1995)
80	Scarole	Scott A	10	4	168	13	Carlin et al. (1995)
81	Salade iceberg	Scott A	5	3	360	14	Beuchat & Brackett (1990)

suite page suivante

Tab. 3.1 – suite

N° de courbe	Variété	Souche ou sérotype	Temp. (°C)	Nb de points /courbe	Temps d'incubation (h)	N° de groupe	Source
82	Salade iceberg	Scott A	5	3	360	14	Beuchat & Brackett (1990)
83	Salade iceberg	Scott A	5	2	360	14	Beuchat & Brackett (1990)
84	Salade iceberg	Scott A	5	2	360	14	Beuchat & Brackett (1990)
85	Salade iceberg	Scott A	5	3	360	14	Beuchat & Brackett (1990)
86	Salade iceberg	Scott A	5	3	360	14	Beuchat & Brackett (1990)
87	Salade iceberg	Scott A	5	2	360	14	Beuchat & Brackett (1990)
88	Salade iceberg	Scott A	5	2	360	14	Beuchat & Brackett (1990)
89*	Salade iceberg	Scott A	10	4	240	14	Beuchat & Brackett (1990)
90*	Salade iceberg	Scott A	10	4	240	14	Beuchat & Brackett (1990)
91	Salade iceberg	Scott A	10	3	240	14	Beuchat & Brackett (1990)
92*	Salade iceberg	Scott A	10	3	240	14	Beuchat & Brackett (1990)
93*	Salade iceberg	Scott A	10	4	240	14	Beuchat & Brackett (1990)
94*	Salade iceberg	Scott A	10	4	240	14	Beuchat & Brackett (1990)
95*	Salade iceberg	Scott A	10	4	240	14	Beuchat & Brackett (1990)
96	Salade iceberg	Scott A	10	4	240	14	Beuchat & Brackett (1990)
97	Salade iceberg	LCDC 81-861	5	3	360	15	Beuchat & Brackett (1990)
98	Salade iceberg	LCDC 81-861	5	3	360	15	Beuchat & Brackett (1990)
99	Salade iceberg	LCDC 81-861	5	2	360	15	Beuchat & Brackett (1990)
100	Salade iceberg	LCDC 81-861	5	2	360	15	Beuchat & Brackett (1990)
101	Salade iceberg	LCDC 81-861	5	3	360	15	Beuchat & Brackett (1990)
102	Salade iceberg	LCDC 81-861	5	3	360	15	Beuchat & Brackett (1990)
103	Salade iceberg	LCDC 81-861	5	2	360	15	Beuchat & Brackett (1990)
104*	Salade iceberg	LCDC 81-861	10	4	240	15	Beuchat & Brackett (1990)
105*	Salade iceberg	LCDC 81-861	10	4	240	15	Beuchat & Brackett (1990)
106	Salade iceberg	LCDC 81-861	10	4	360	15	Beuchat & Brackett (1990)
107*	Salade iceberg	LCDC 81-861	10	4	240	15	Beuchat & Brackett (1990)
108*	Salade iceberg	LCDC 81-861	10	4	240	15	Beuchat & Brackett (1990)
109*	Salade iceberg	LCDC 81-861	10	4	240	15	Beuchat & Brackett (1990)
110*	Salade iceberg	LCDC 81-861	10	4	240	15	Beuchat & Brackett (1990)
111*	Salade iceberg	LCDC 81-861	10	4	240	15	Beuchat & Brackett (1990)

suite page suivante

Tab. 3.1 – suite

N° de courbe	Variété	Souche ou sérotype	Temp. (°C)	Nb de points /courbe	Temps d'incubation (h)	N° de groupe	Source
112	Laitue	Mélange	4	4	216	16	Farber et al. (1998)
113	Laitue	Mélange	10	4	216	16	Farber et al. (1998)
114	Laitue	Mélange	25	5	48	16	Farber et al. (1998)
115*	Salade iceberg	ATCC 19114	8	6	288	17	Francis & O'Beirne (2001)
116	Salade iceberg	ATCC 19114	4	6	288	17	Francis & O'Beirne (2001)
117	Salade iceberg	NCTC 11994	8	6	288	18	Francis & O'Beirne (2001)
118	Salade iceberg	NCTC 11994	4	6	288	18	Francis & O'Beirne (2001)
119	Salade iceberg	Mélange	4	3	336	19	Szabo et al. (2003)
120	Salade iceberg	Mélange	8	3	336	19	Szabo et al. (2003)
121	Salade iceberg	LCDC-81-861	1	3	336	20	Delaquis et al. (2002)
122	Salade iceberg	LCDC-81-861	10	3	336	20	Delaquis et al. (2002)
123	Salade iceberg	LCDC-81-861	1	3	336	20	Delaquis et al. (2002)
124	Salade iceberg	LCDC-81-861	10	3	336	20	Delaquis et al. (2002)
125	Scarole	Mélange	2	2	216	21	Jacxsens et al. (2002)
126	Scarole	Mélange	4	2	168	21	Jacxsens et al. (2002)
127	Scarole	Mélange	7	2	120	21	Jacxsens et al. (2002)
128	Scarole	Mélange	10	2	72	21	Jacxsens et al. (2002)
129	Laitue	ATCC 19114	3	6	336	22	Francis & O'Beirne (1997)
130	Laitue	NCTC 11994	3	6	336	23	Francis & O'Beirne (1997)
131	Laitue	ATCC 19114	8	6	336	22	Francis & O'Beirne (1997)
132	Laitue	NCTC 11994	8	6	336	23	Francis & O'Beirne (1997)
133	Laitue	ATCC 19114	3	6	336	22	Francis & O'Beirne (1997)
134	Laitue	NCTC 11994	3	6	336	23	Francis & O'Beirne (1997)
135*	Laitue	ATCC 19114	8	6	336	22	Francis & O'Beirne (1997)
136*	Laitue	NCTC 11994	8	6	336	23	Francis & O'Beirne (1997)
137	Laitue	ATCC 19114	3	6	336	22	Francis & O'Beirne (1997)
138	Laitue	NCTC 11994	3	6	336	23	Francis & O'Beirne (1997)
139	Laitue	ATCC 19114	8	6	336	22	Francis & O'Beirne (1997)
140	Laitue	NCTC 11994	8	6	336	23	Francis & O'Beirne (1997)
141*	Laitue	ATCC 19114	8	7	336	17	Francis & O'Beirne (2001)

suite page suivante

Tab. 3.1 – suite

N° de courbe	Variété	Souche ou sérotype	Temp. (°C)	Nb de points /courbe	Temps d'incubation (h)	N° de groupe	Source
142*	Laitue	ATCC 19114	8	7	336	17	Francis & O'Beirne (2001)
143*	Laitue	ATCC 19114	8	5	240	24	Francis & O'Beirne (2005)
144	Laitue	NCTC 11994	8	5	240	25	Francis & O'Beirne (2005)
145*	Laitue	F5817 (4b)	8	5	240	26	Francis & O'Beirne (2005)
146*	Laitue	1210 (4b)	8	5	240	27	Francis & O'Beirne (2005)
147*	Laitue	ATCC 19118	8	5	240	28	Francis & O'Beirne (2005)
148	Laitue	ATCC 19177	8	5	240	29	Francis & O'Beirne (2005)
149*	Laitue	269 (1\2a)	8	5	240	30	Francis & O'Beirne (2005)
150	Laitue	NCTC 7973	8	5	240	31	Francis & O'Beirne (2005)
151	Laitue	1263 (1\2b)	8	5	240	32	Francis & O'Beirne (2005)
152	Laitue	SLCC 1694 (1\2b)	8	5	240	33	Francis & O'Beirne (2005)
153	Laitue	996 (1\2c)	8	5	240	34	Francis & O'Beirne (2005)
154	Laitue	ATCC 19112	8	5	240	35	Francis & O'Beirne (2005)
155*	Laitue	ATCC 19115	8	5	240	36	Francis & O'Beirne (2005)
156	Laitue	4604 (3b)	8	5	240	37	Francis & O'Beirne (2005)
157	Laitue	SLCC 2479 (3c)	8	5	240	38	Francis & O'Beirne (2005)
158	Salade iceberg	Mélange	5	7	432	39	Li et al. (2002)
159	Salade iceberg	Mélange	5	7	432	39	Li et al. (2002)
160	Salade iceberg	Mélange	5	7	432	39	Li et al. (2002)
161	Salade iceberg	Mélange	15	4	168	39	Li et al. (2002)
162*	Salade iceberg	Mélange	15	4	168	39	Li et al. (2002)
163*	Salade iceberg	Mélange	15	4	168	39	Li et al. (2002)
164	Salade iceberg	Mélange	5	7	432	39	Li et al. (2002)
165	Salade iceberg	Mélange	5	7	432	39	Li et al. (2002)
166*	Salade iceberg	Mélange	15	4	168	39	Li et al. (2002)
167*	Salade iceberg	Mélange	15	4	168	39	Li et al. (2002)
168	Laitue	Scott A	10	4	168	2	Carlin, unpublished data
169	Frisée	Scott A	10	4	168	2	Carlin, unpublished data
170	Mâche	Scott A	10	4	168	2	Carlin, unpublished data
171	Scarole	Scott A	10	4	168	2	Carlin, unpublished data

suite page suivante

Tab. 3.1 – suite

N° de courbe	Variété	Souche ou sérotype	Temp. (°C)	Nb de points /courbe	Temps d'incubation (h)	N° de groupe	Source
172	Laitue	Scott A	10	4	168	2	Carlin, unpublished data
173	Frisée	Scott A	10	4	168	2	Carlin, unpublished data
174	Mâche	Scott A	10	4	168	2	Carlin, unpublished data
175	Scarole	88/7137 (4b)	4	3	192	40	Ringlé et al. (1991)
176	Scarole	88/7137 (4b)	8	3	192	40	Ringlé et al. (1991)
177	Scarole	88/7137 (4b)	4	5	240	40	Ringlé et al. (1991)
178	Scarole	88/7137 (4b)	8	5	240	40	Ringlé et al. (1991)
179	Scarole	88/7137 (4b)	4	5	240	40	Ringlé et al. (1991)
180*	Scarole	88/7137 (4b)	8	5	240	40	Ringlé et al. (1991)
181	Scarole	88/7137 (4b)	4	4	240	40	Ringlé et al. (1991)
182*	Scarole	88/7137 (4b)	8	4	240	40	Ringlé et al. (1991)
183	Scarole	ScottA	3	6	240	41	Nguyen-The et al. (1996)
184	Scarole	ScottA	6	6	240	41	Nguyen-The et al. (1996)
185*	Scarole	ScottA	9	6	240	41	Nguyen-The et al. (1996)
186	Salade iceberg	Mélange	5	8	0	42	Koseki & Isobe (2005a)
187*	Salade iceberg	Mélange	10	10	163	42	Koseki & Isobe (2005a)
188	Salade iceberg	Mélange	15	11	36	42	Koseki & Isobe (2005a)
189	Salade iceberg	Mélange	20	12	36	42	Koseki & Isobe (2005a)
190	Salade iceberg	Mélange	25	11	36	42	Koseki & Isobe (2005a)
191	Scarole	Scott A	10	3	168	7	Carlin et al. (1996b)
192	Scarole	Scott A	10	3	168	7	Carlin et al. (1996b)
193	Scarole	Scott A	10	4	168	7	Carlin et al. (1996b)
194	Scarole	1011/1410	14	14	73	43	Stahl et al. (2002)
195	Scarole	1011/1410	14	16	78	43	Stahl et al. (2002)
196	Scarole	1011/1410	14	24	240	43	Stahl et al. (2002)
197	Scarole	1011/1410	14	24	240	43	Stahl et al. (2002)
198	Scarole	1011/1410	7	18	240	43	Stahl et al. (2002)
199	Scarole	1011/1410	7	17	240	43	Stahl et al. (2002)
200	Scarole	1011/1410	7	18	240	43	Stahl et al. (2002)
201	Laitue	1011/1410	15	14	69	43	Stahl et al. (2002)

suite page suivante

Tab. 3.1 – suite

N° de courbe	Variété	Souche ou sérotype	Temp. (°C)	Nb de points /courbe	Temps d'incubation (h)	N° de groupe	Source
202	Laitue	1011/1410	15	28	245	43	Stahl et al. (2002)
203	Laitue	1011/1410	15	29	245	43	Stahl et al. (2002)
204	Laitue	1011/1410	9	7	76	43	Stahl et al. (2002)
205	Laitue	1011/1410	9	7	76	43	Stahl et al. (2002)
206	Laitue	1011/1410	9	8	94	43	Stahl et al. (2002)

3.4 Estimation des distributions des paramètres de croissance

3.4.1 Le modèle $M1$

Soient L courbes de croissance en environnement constant disponibles, à chaque courbe $l = 1, ..., L$ correspond une température T_l constante. A des temps différents $(t_k^{(l)})_{k=1,...,K_l}$, où K_l représente le nombre d'observations pour chaque courbe, des observations de concentration en logarithme décimal de *L. monocytogenes* $y_l(t_k^{(l)})$ sont disponibles. Les concentrations d'une même courbe sont supposées indépendantes et distribuées selon une loi normale. Le modèle peut s'écrire pour chaque temps $t_k^{(l)}$ correspondant à une observation de la concentration bactérienne

$$y_l(t_k^{(l)}) \sim N(f_l(t_k^{(l)}), \sigma_y^2)$$

où σ_y^2 est la variance supposée identique pour tous les points des différentes courbes. La fonction f est celle du modèle logistique avec délai et rupture (3.1), et son paramètre, $\mu_{\max,l}$ est distribué selon une loi normale tronquée

$$\sqrt{\mu_{\max,l}(T_l)} \sim N(\sqrt{g_l(T_l)}, \sigma^2_{\mu_{\max}})I(0, \infty)$$

où $\sigma^2_{\mu_{\max}}$ est la variance supposée identique pour les différentes courbes et $g_l(T_l)$ est donnée par l'équation (3.2) du modèle secondaire à valeurs cardinales. $I(0, \infty)$ signifie que la loi normale est tronquée et est définie sur l'intervalle $[0, \infty]$, cette contrainte sur $\sqrt{\mu_{\max,l}(T_l)}$ est naturelle.

Le graphique acyclique orienté de la figure 3.3 correspond au modèle $M1$.

Lois de distribution *a priori* des paramètres et des hyperparamètres du modèle primaire

$y_{0,l}$, représente le niveau de concentration au temps $t = 0$ pour la courbe l. Cette valeur est définie par la quantité de bactéries inoculées lors de chaque expérience, aux erreurs de mesure près. Une distribution pour ce paramètre peut être une loi normale tronquée et centrée autour de la valeur observée à $t = 0$, notée $Y_{0,l}$

$$y_{0,l} \sim N(Y_{0,l}, \sigma_y^2) I(0, \infty).$$

$y_{\max,l}$, correspond à la concentration bactérienne maximale atteinte lors de la phase stationnaire de la courbe l. Il est admis que la concentration maximale dépend principalement du milieu de croissance, on peut donc supposer que les $y_{\max,l}$ sont issus d'une distribution normale de moyenne $M_{y_{\max}}$ et de variance $\sigma^2_{y_{\max}}$

$$y_{\max,l} \sim N(M_{y_{\max}}, \sigma^2_{y_{\max}}).$$

Une distribution uniforme sur l'intervalle $[5; 8.5]$ log ufc/g est utilisée comme distribution *a priori* du paramètre $M_{y_{\max}}$

$$M_{y_{\max}} \sim U(5, 8.5).$$

Selon les experts participant au projet, cette distribution semble raisonnable compte tenu de l'hétérogénéité des valeurs maximales atteintes sur les salades.

Le temps de latence de chaque courbe est estimé et modélisé par une loi normale

$$\lambda_l \sim Normale(M_\lambda, \sigma^2_\lambda).$$

Une loi peu informative est choisie pour l'hyperparamètre M_λ

$$M_\lambda \sim Normale(10, 10^4).$$

Ainsi 95% des valeurs *a priori* des temps de latence moyens se situent dans l'intervalle $[-186, 206]$h. Une valeur négative de λ est envisageable et correspond à une bactérie en phase de croissance lors de sa mise en culture et ne subissant pas de phase de latence. Notons que

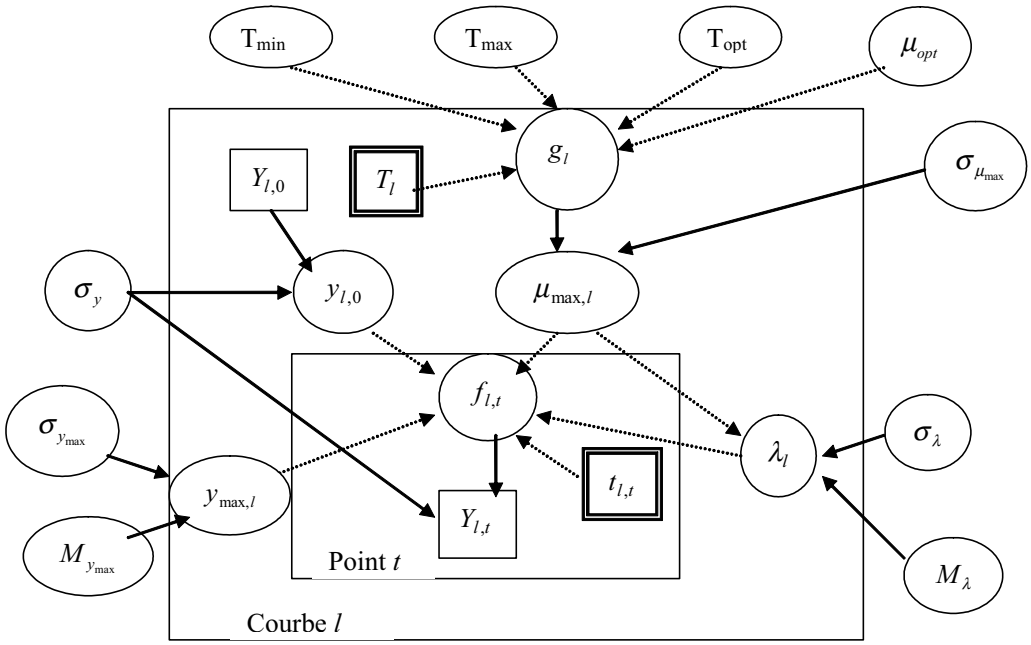

FIG. 3.3: Graphe acyclique orienté du modèle $M1$ d'estimation des distributions des paramètres de croissance de *L. monocytogenes* dans la salade. Les paramètres sont représentés par des "noeuds" selon les conventions suivantes : a) un rectangle représente les données observées, un double carré représente une covariable et une ellipse représente les paramètres et hyperparamètres ; b) les flèches pleines représentent une relation probabiliste. Elles indiquent que le noeud fils a une distribution de paramètre(s) le(s) noeud(s) parent(s). Les flèches en pointillées représentent une relation fonctionnelle. Elles indiquent que le noeuds fils est relié au(x) noeud(s) parent(s) par une fonction mathématique ou logique. Chaque noeud est indépendant, conditionnellement à ses parents, de tous les autres noeuds excepté de ses descendants. Les cadres indiquent l'échelle par rapport à laquelle sont définis les noeuds : les points expérimentaux, les courbes ou s'ils sont uniques.

les hyperparamètres M_λ et σ^2_λ n'ont aucune signification biologique et sont utilisés seulement pour la modélisation. En effet, un temps de latence moyen M_λ estimé à partir de toutes les courbes disponibles n'a aucune signification puisque la valeur du temps de latence dépend de la température à laquelle a été réalisée l'expérience.

Aucune information *a priori* n'est disponible sur les variances suivantes : $\sigma^2_y, \sigma^2_{\mu_{\max}}, \sigma^2_{y_{\max}}$ et σ^2_λ. La loi vague $G(10^{-3}, 10^3)$ est attribuée à la précision (inverse de la variance) de chacune de ces variances comme le recommande Smith et al. (1995) en l'absence d'information *a priori.*

Lois de distribution *a priori* sur les paramètres du modèle secondaire

Les ditributions *a priori* attribuées aux paramètres du modèle secondaire sont présentées dans le tableau 3.2. Les paramètres de températures du modèle secondaire $T_{\min}, T_{opt}, T_{\max}$ sont des paramètres propres à la bactérie, c'est-à-dire que leurs valeurs dépendent principalement de la bactérie étudiée et non de la matrice alimentaire. Augustin & Carlier (2000a) a calculé les valeurs médianes de ces paramètres à partir de l'examen de 1865 cinétiques de croissance, issues de la littérature de *Listeria monocytogenes* dans des milieux de culture spécifiques, de la viande, des oeufs, des produits laitiers et des produits de la mer. Ces valeurs ont été utilisées pour injecter de l'information a priori dans notre modèle : la distribution *a priori* attribuée à chaque paramètre est une loi normale ayant pour moyenne la valeur estimée par Augustin & Carlier (2000a) et pour variance la valeur 1. Notons que cette information *a priori* permettra pour les paramètres T_{opt} et $T_{\max}$ de pallier le manque de courbes réalisées à ces températures.

A contrario, étant donné que la valeur du taux de croissance optimal μ_{opt} dépend principalement de la matrice alimentaire, une variance permettant d'utiliser une loi *a priori* peu informative a été attribuée à ce paramètre. Une variance de valeur 1 autorise un intervalle de valeurs *a priori* assez grand pour ce paramètre. Les valeurs du taux de croissance optimal μ_{opt} devant être positives, une loi normale tronquée et définie sur l'intervalle $[0, \infty]$ est utilisée comme distribution *a priori.*

3.4.2 Le modèle $M2$: introduction d'une structure hiérarchique

Le modèle $M2$ intégre une structure hiérarchique sur les paramètres du modèle secondaire afin de prendre en compte la variabilité due à la présence de différentes souches n'ayant pas le même comportement face aux différentes températures ou à la variabilité due aux conditions

TAB. 3.2: Distributions *a priori* utilisées pour les paramètres du modèle M1

Paramètre	Unité	Distribution	$P_{0.025}$	$P_{0.50}$	$P_{0.975}$
$T_{\min}$	°C	$N(-2.7, 1^2)$	-4.66	-2.7	-0.74
T_{opt}	°C	$N(37, 1^2)$	35.04	37.00	38.96
$T_{\max}$	°C	$N(45.5, 1^2)$	43.54	45.5	47.46
μ_{opt}	h^{-1}	$N(1.016, 1^2)I(0, \infty)$	0.085	1.21	3.05

$X \sim N(m, s^2)I(a, b)$, distribution normale de moyenne m, d'écart-type s, tronquée et définie sur l'intervalle [a,b].

de préculture de la bactérie. Les courbes ont été classées en $i = 1, ..., 43$ groupes correspondant aux différentes souches de *L. monocytogenes* étudiées ou aux différentes conditions expérimentales mises en place (cf. le paragraphe de description des données)

Les paramètres $T_{\min,i}, T_{opt,i}, T_{\max,i}$ et $\mu_{opt,i}$ propres à chaque groupe i sont supposés issus de distributions normales

$$\begin{aligned} T_{\min,i} &\sim N(M_{T_{\min}}, S^2_{T_{\min}}), \\ T_{opt,i} &\sim N(M_{T_{opt}}, S^2_{T_{opt}}), \\ T_{\max,i} &\sim N(M_{T_{\max}}, S^2_{T_{\max}}), \\ \mu_{opt,i} &\sim N(M_{\mu_{opt}}, S^2_{\mu_{opt}})I(0, \infty). \end{aligned}$$

La structure hiérarchique de ce modèle facilite l'intégration de la contrainte supplémentaire sur les paramètres λ et $\mu_{\max}$ définie par la relation 3.3 permettant de modéliser le temps de latence. Comme les bactéries d'un même groupe i ont subi les mêmes conditions pré-incubatoires, la relation

$$\lambda_l \times \mu_{\max,l} = K_i + \varepsilon_l,$$

doit être vérifiée pour chaque courbe l appartenant au groupe i quelles que soient les températures auxquelles ont été effectuées les observations. Notons que l'indice l comporte implicitement l'information de l'appartenance de la courbe à un groupe i. Il n'est donc pas nécessaire d'indicer par exemple $\mu_{\max}$ par i et l.

La constante K est introduite dans le modèle par

$$K_i \sim N(M_K, S^2_K),$$

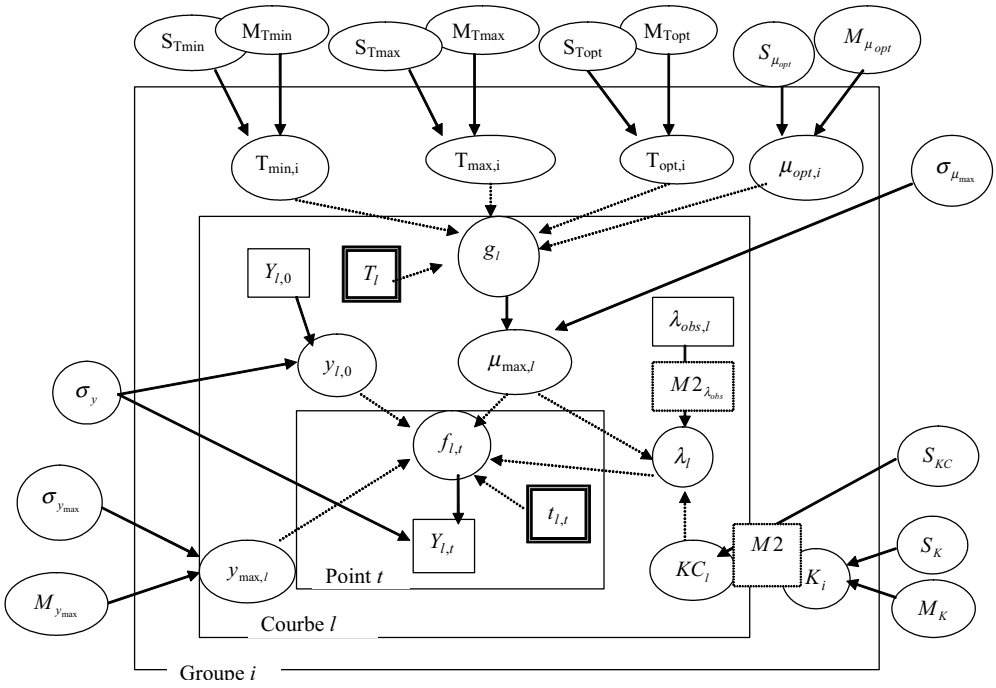

FIG. 3.4: Graphe acyclique orienté des modèles $M2$ et $M2_{obs}$ d'estimation des paramètres de croissance de *L. monocytogenes* dans la salade. Les carrés en pointillés représentent les différences entre les deux modèles au niveau de la modélisation du temps de latence.

et

$$KC_l \sim N(K_i, S^2_{KC}).$$

L'utilisation du paramètre KC_l permet de prendre en compte la variabilité entre les courbes et produit une meilleure estimation de M_K.

Le modèle $M2$ est représenté par le graphique acyclique de la figure 3.4.

Lois de distribution *a priori* du modèle $M2$

Les distributions a priori utilisées sont présentées dans le tableau 3.3. Les distributions *a priori* attribuées aux hyperparamètres $M_{T_{\min}}, M_{T_{opt}}, M_{T_{\max}}$ et $M_{\mu_{opt}}$ de ce modèle sont celles des paramètres $T_{\min}, T_{opt}, T_{\max}$ et μ_{opt} du modèle $M1$. La distribution *a priori* attribuée à chacun des hyperparamètres de présicion $S^{-2}_{T\min}, S^{-2}_{Topt}$ et $S^{-2}_{T\max}$ est la distribution $G(1.68, 2.82)$

TAB. 3.3: Distributions *a priori* utilisées pour les hyperparamètres des modèles M2 et $M2_{\lambda_{obs}}$

Paramètre	Unité	Distribution	$P_{0.025}$	$P_{0.50}$	$P_{0.975}$
$M_{T_{\min}}$	°C	$N(-2.7, 1^2)$	-4.66	-2.7	-0.74
$M_{T_{opt}}$	°C	$N(37, 1^2)$	35.04	37	38.96
$M_{T_{\max}}$	°C	$N(45.5, 1^2)$	43.54	45.5	47.46
$M_{\mu_{opt}}$	h^{-1}	$N(1.016, 1^2)I(0, \infty)$	0.085	1.21	3.05
$S_{T_{\min}}$	°C	$S_{T_{\min}}^{-2} \sim G(1.68, 2.82)$	0.266	0.510	1.53
$S_{T_{opt}}$	°C	$S_{T_{opt}}^{-2} \sim G(1.68, 2.82)$	0.266	0.510	1.53
$S_{T_{\max}}$	°C	$S_{T_{\max}}^{-2} \sim G(1.68, 2.82)$	0.266	0.510	1.53
$S_{\mu_{opt}}$	h^{-1}	$S_{\mu_{opt}}^{-2} \sim G(3.26, 14.6)$	0.0948	0.153	0.306
M_K		$N(3.09, \sqrt{10}^2)$	-3.11	3.09	9.29
S_K		$S_K^{-2} \sim G(10^{-3}, 10^3)$	13270	$1.4\ 10^{149}$	∞
S_{KC}		$S_{KC}^{-2} \sim G(3.26, 1.31)$	0.316	0.510	1.02

$X \sim G(a, b)$, *distribution gamma de paramètre de forme a, de paramètre d'échelle b, d'espérance* $E(X) = a \times b$ *et de variance* $V(X) = a \times b^2$.

$X \sim N(m, s^2)I(a, b)$, distribution normale de moyenne m, d'écart-type s, tronquée et définie sur l'intervalle [a,b].

de moyenne 4.74°C et d'écart-type 3.66°C, obtenue par Pouillot et al. (2003) à partir de dires d'experts de l'AFSSA. De même la distribution Gamma $G(3.26, 14.6)$ de moyenne 47.6 (h^{-1}) et d'écart-type 26.36 est utilisée comme distribution *a priori* de l'hyperparamètre $S_{\mu_{opt}}^{-2}$. La distribution *a priori* du paramètre M_K est une loi normale centrée autour de la valeur médiane déterminée par Augustin & Carlier (2000a), $N(3.09, 10)$. La loi vague $G(10^{-3}, 10^3)$ est choisie comme *a priori* pour la précision S_K^{-2} de K. La distribution $G(3.26, 1.31)$ de moyenne 4.27 et d'écart-type 2.37, établie par Pouillot et al. (2003) à partir des dires d'experts, est utilisée pour la précision S_{KC}^{-2} de KC_l.

Le modèle $M2_{\lambda_{obs}}$: calcul au préalable des temps de latence

La valeur du temps de latence de chaque courbe est estimée au préalable à l'aide d'un algorithme développé sous le logiciel R (Team, 2007). Cet algorithme calcule le temps de latence pour chaque courbe, comme étant le temps d'intersection entre la droite horizontale passant par la concentration initale $Y_{0,l}$ et la tangente à la courbe au point où $\mu_l(t) = \mu_{\max,l}$. En pratique, la tangente est déterminée par l'équation de la droite passant par les deux points successifs dont la différence de concentration est la plus grande.

Ensuite les valeurs obtenues sont injectées dans le modèle en supposant que le temps de latence est distribué selon une loi normale centrée autour des valeurs des temps de latence

calculées à l'aide de l'algorithme. Ainsi,

$$\lambda_l \sim N(\lambda_{obs,l}, \sigma_\lambda^2).$$

Enfin, la loi vague $G(10^{-3}, 10^{-3})$ est utilisée comme distribution *a priori* pour la précision σ_λ^{-2}. Les autres paramètres sont modélisés de la même manière que dans le modèle $M2$ (3.3) et le graphique acyclique de ce modèle $M2_{\lambda_{obs}}$ est présenté sur la même figure 3.4.

3.5 Validation

Cette section présente trois méthodes permettant de valider les modèles développés dans ce chapitre. La première méthode consiste à tracer les données prédites en fonction des données observées. La deuxième méthode est le calcul de la déviance (DIC). Enfin la troisième méthode est l'utilisation d'un critère d'adéquation afin de déterminer l'écart entre les valeurs prédites par chaque modèle et les données observées. Ces méthodes permettront de sélectionner un ou plusieurs modèles qui seront ensuite utilisés pour modéliser la croissance de *L. monocytogenes* dans les salades en sachets dans le chapitre 4.

3.5.1 Comparaison entre les valeurs observées et les valeurs prédites

Une méthode simple pour vérifier si un modèle est compatible avec les données observées est de tracer les valeurs observées en fonction des valeurs prédites par le modèle et de vérifier graphiquement l'adéquation entre le modèle et les données observées. En pratique, pour chaque courbe et pour chaque temps t, la moyenne des concentrations en *Listeria monocytogenes* prédite par le modèle est tracée en fonction de la concentration correspondante observée. Le modèle sera en adéquation avec les données si les valeurs se trouvent sur la bissectrice de l'angle formé par l'axe des abscisses et l'axe des ordonnées.

3.5.2 Utilisation d'un critère d'adéquation

L'objectif est de définir un critère permettant de valider l'adéquation de chaque modèle aux données observées. Ce critère pour les modèles Bayésiens appelé "test quantity" par Gelman (Gelman et al., 2004b) est le pendant du test statistique pour les modèles classiques. Dans une modèlisation Bayésienne, ce critère d'adéquation noté $T(y, \theta)$ est fonction

du paramètre θ (θ peut être un vecteur de paramètres et d'hyperparamètres dans le cas d'un modèle hiérarchique) et des données utilisées y. Ce critère est utile pour calculer l'écart entre le modèle et les données.

La p-valeur Bayésienne

Pour évaluer l'adéquation aux données d'un modèle Bayésien, les données observées y peuvent être comparées aux données répliquées y^{rep}. Les valeurs répliquées correspondent à celles qui pourraient être observées si l'expérience qui a produit les données y était refaite avec le même modèle, les mêmes covariables (ici la durée et la température) et les mêmes valeurs de θ ayant produit les données y. A la différence des valeurs prédites qui seraient notées $\widetilde{y}$ pour reprendre les notations du second chapitre, et qui représenteraient toute valeur observable et future de y, y^{rep} est une réplication exacte de y et possède donc les mêmes covariables que y. Les données y^{rep} sont générées en utilisant la distribution prédictive

$$p(y^{rep}|y) = \int_{\Theta} p(y^{rep}|\theta)p(\theta|y)d\theta,$$

où Θ est l'espace de toutes les valeurs possibles du paramètre θ.

La p-valeur Bayésienne est définie comme la probabilité que le critère d'adéquation calculé à partir des données répliquées y^{rep} soit plus grand que celui calculé à partir des données observées y

$$p_{valeur} = \Pr(T(y^{rep}, \theta) \geq T(y, \theta)|y).$$

En théorie cette probabilité est calculée par rapport à la distribution *a posteriori* de θ et par rapport à la distribution jointe $p(y^{rep}, \theta|y) = p(y^{rep}|\theta)p(\theta|y)$, ainsi la p_{valeur} se réécrit sous la forme

$$p_{valeur} = \int_{\Theta}\int_{\Upsilon} I_{[T(y^{rep},\theta)\geq T(y,\theta)]}p(y^{rep}|\theta)p(\theta|y)dy^{pred}d\theta,$$

où I est la fonction indicatrice et Υ l'espace de toutes les valeurs possibles de y^{rep}.

En pratique, on tire $b = 1, ..., B$ valeurs dans la distribution *a posteriori* de θ puis, pour chaque valeur θ^b de θ et à partir de la distribution prédictive, on calcule une valeur $y^{rep,b}$. Ensuite on compare le critère calculé à patir des données observées $T(y, \theta^b)$ avec le critère calculé à partir des valeurs répliquées $T(y^{rep,b}, \theta^b)$. La p-valeur est la proportion de ces B simulations pour lesquelles la valeur du critère est supérieure ou égale à la valeur du critère

calculé avec les données observées

$$T(y^{rep,b}, \theta^b) \geq T(y, \theta^b), b = 1, ..., B.$$

Lorsque la p_{valeur} prend des valeur proches de 0.5, le modèle est en adéquation avec les données. *A contrario* quand la p_{valeur} est proche de 0 ou de 1, le modèle proposé ne reproduit pas, de manière satisfaisante, la structure des données.

Définition d'un critère d'adéquation approprié aux modèles d'estimation des paramètres de croissance

Plusieurs critères peuvent être étudiés. Ces critères doivent être déterminés en fonction de ce que l'on attend du modèle. Dans ces travaux, les modèles ont été développés afin de simuler la croissance de *L. monocytogenes* dans les salades en conditions réelles de stockage. La prévision par les modèles, avec un taux d'erreur faible, du nombre de *Listeria monocytogenes* s'étant développées en un temps fini et à une température donnée est donc un critère pertinent. C'est ce critère que nous avons défini et qui sera calculé pour chaque modèle.

Soit $\Delta y = y(t) - y_0$, le nombre de bactéries s'étant développées après une durée $\Delta t = t - t_0$ avec $t_0 = 0$, passé à une température D. Soit L_D le nombre total de courbes effectuées à la température D et pour lesquelles le nombre de bactéries a été relevé après une durée Δt. Pour chaque courbe $l = 1, ..., L_D$ dont la croissance a été étudiée à la température D, un nombre Δy_l peut être calculé, et on peut ainsi obtenir un nombre moyen de bactéries se développant en un temps t à une température D

$$\overline{\Delta y}_{t,D} = \frac{1}{L_D} \sum_{l=1}^{L_D} (y_l(t) - y_{0,l}).$$

Le minimum et le maximum de bactéries se développant en un temps t, à la température D peuvent aussi être calculés.

Le critère d'adéquation $T(y, \theta)$ est le nombre moyen $\overline{\Delta y}_{t,D}$ de *Listeria monocytogenes* s'étant développées en un temps t et à une température D. Le paramètre θ représente les paramètres et hyperparamètres de chaque modèle, par exemple pour le modèle M_1,

$$\theta = (T_{\min}, T_{\max}, T_{opt}, \mu_{opt}, y_{\max}, M_{y_{\max}}, \sigma^{-2}_{y_{\max}}, \sigma^{-2}_{y}, \sigma^{-2}_{\mu_{\max}}).$$

A partir de B valeurs tirées aléatoirement dans la distribution *a posteriori* de θ, B

réplications du nombre moyens de bactéries se développant en un temps t à une température D peuvent être calculés, ainsi pour $b = 1, ..., B$

$$T(y^{rep,b}, \theta^b) = \overline{\Delta y_{t,D}^{rep,b}} = \frac{1}{L_D} \sum_{l=1}^{L_D} (y_l^{rep,b}(t) - y_{0,l}).$$

La moyenne sur les B simulations est ensuite calculée

$$T(y^{rep}, \theta) = \overline{\Delta y_{t,D}^{rep}} = \frac{1}{B} \frac{1}{L_D} \sum_{b=1}^{B} \sum_{l=1}^{L_D} (y_l^{rep,b}(t) - y_{0,l}),$$

ainsi que l'intervalle d'incertitude à 95% de la moyenne. Les valeurs obtenues sont ensuite comparées à celles observées.

La p-valeur associée à ce critère d'adéquation est la proportion de ces B réplications pour lesquelles la valeur du critère est supérieure ou égale à la valeur du critère calculée avec les données observées

$$\frac{1}{L_D} \sum_{l=1}^{L_D} (y_l^{rep,b}(t) - y_{0,l}) \geq \frac{1}{L_D} \sum_{l=1}^{L_D} (y_l(t) - y_{0,l}), b = 1, ..., B.$$

3.5.3 Sélection de modèles : le Critère d'Information de Déviance (DIC)

Le Critère d'Information de Déviance est utilisé pour comparer différents modèles et sélectionner le meilleur (Spiegelhalter et al., 2002). Il est composé d'un critère évaluant l'ajustement du modèle aux données et d'un critère prenant en compte la compléxité du modèle c'est-à-dire le nombre de paramètres composant le modèle. Une augmentation de la complexité du modèle s'accompagnant d'une meilleure adéquation, les modèles sont comparés en arbitrant entre ces deux critères en utilisant des tests du rapport de vraisemblance. Le DIC est une généralisation du Critère d'Information d'Akaike, l'AIC (Akaike's Information Criterion, Akaike (1973)). Le modèle le plus performant est celui qui minimise le DIC.

3.6 Réalisation pratique et contrôle de la convergence

L'inférence bayésienne pour les différents modèles est réalisée à l'aide du logiciel OpenBugs (Spiegelhalter et al., 2005). Pour chaque modèle, deux chaînes ont été générées à partir

de valeurs initiales différentes des paramètres afin de contrôler visuellement la convergence vers les distributions *a posteriori* des paramètres et des hyperparamètres. La statistique de Gelman et Rubin, modifiée par Brooks & Gelman (1998) a été également utilisée pour vérifier la convergence (cf. paragraphe 2.5.2). La convergence vers les distributions *a posteriori* de tous les paramètres et hyperparamètres est atteinte après 20 000 itérations pour les modèles $M1$ et $M2$. La convergence est beaucoup plus rapide pour le modèle $M2_{\lambda_{obs}}$, elle est obtenue à partir de 2 000 itérations. Les statistiques des distributions *a posteriori* des paramètres des différents modèles ont été calculées après vérification de la convergence, à partir de 50 000 itérations supplémentaires provenant des deux chaînes afin de respecter pour chaque paramètre la régle sur l'erreur de Monte-Carlo (cf. paragraphe 2.5.2). La robustesse des sorties a été testée en partant de valeurs initiales différentes pour les chaînes.

3.7 Résultats

3.7.1 Statistiques des distributions *a posteriori* des paramètres des différents modèles

Les statistiques descriptives des distributions *a posteriori* des paramètres et hyperparamètres des différents modèles sont présentées dans les tableaux 3.4 à 3.9. La température minimale de croissance de *L. monocytogenes* obtenue avec le modèle $M1$ est en moyenne de -5.3°C, avec le modèle $M2$ de -4.5°C et de -4.4°C avec le modèle $M2_{\lambda_{obs}}$ (Tab. 3.4). L'intervalle de crédibilité a posteriori (CI) de niveau 95% de la température minimale est [-6.96 ; -3.65]°C avec le modèle $M1$, [-6.84 ; -2.02]°C avec le modèle $M2$ et [-6.70 ; -1.81]°C avec le modèle $M2_{\lambda_{obs}}$. Les valeurs de la distribution *a posteriori* du paramètre $T_{\min}$ du modèle $M1$ et du paramètre $M_{T\min}$ des modèles $M2$ et $M2_{\lambda_{obs}}$ sont plus faibles respectivement d'environ -3°C et -2°C par rapport aux valeurs de leur distribution *a priori* (Tab. 3.2 et Tab. 3.3).

Les valeurs obtenues *a posteriori* pour les paramètres T_{opt} et $T_{\max}$ du modèle $M1$ et $M_{T_{opt}}$ et $M_{T_{\max}}$ des modèles $M2$ et $M2_{\lambda_{obs}}$ sont proches de la gamme de températures de leur distribution *a priori*. Ainsi la température moyenne optimale de croissance est d'environ 37°C et la température moyenne maximale d'environ 45.5°C (Tab. 3.5 et Tab. 3.6). Le fait que les valeurs *a posteriori* soient proches de celles *a priori* peut signifier soit que l'information entrée dans le modèle par l'intermédiaire des distributions *a priori* est pertinente soit que les données n'ont pas permis d'actualiser l'information *a priori* pour ces paramètres. C'est cette dernière hypothèse qui semble être la raison de la proximité des valeurs *a priori* et

TAB. 3.4: Statistiques descriptives de la distribution *a posteriori* des paramètres de la température minimale de croissance de *L. monocytogenes* dans les salades (unité : °C)

Modèle	Paramètre	Moyenne	Ecart-type	$P_{0.025}$	$P_{0.50}$	$P_{0.975}$
$M1$	T_{min}	-5.29	0.85	-6.96	-5.32	-3.65
$M2$	$M_{T_{min}}$	-4.52	0.88	-6.35	-4.54	-2.83
	$S_{T_{\min}}$	0.72	0.63	0.28	0.55	2.24
	T_{min}	-4.52	1.29	-6.84	-4.55	-2.02
$M2_{\lambda_{obs}}$	$M_{T_{min}}$	-4.38	0.83	-6.12	-4.36	-2.76
	$S_{T_{\min}}$	0.76	0.54	0.28	0.59	2.24
	T_{min}	-4.38	1.24	-6.70	-4.40	-1.81

TAB. 3.5: Statistiques descriptives de la distribution *a posteriori* des paramètres de la température optimale de croissance de *L. monocytogenes* dans les salades (unité : °C)

Modèle	Paramètre	Moyenne	Ecart-type	$P_{0.025}$	$P_{0.50}$	$P_{0.975}$
$M1$	T_{opt}	36.63	1.05	34.62	36.64	38.74
$M2$	$M_{T_{opt}}$	37.05	1.04	35.01	37.08	39.04
	$S_{T_{opt}}$	0.63	0.39	0.27	0.52	1.63
	T_{opt}	37.05	1.27	34.52	37.07	39.48
	$M_{T_{opt}}$	36.94	1.01	34.95	36.96	38.89
$M2_{\lambda_{obs}}$	$S_{T_{opt}}$	0.64	0.51	0.27	0.52	1.70
	T_{opt}	36.94	1.30	34.49	36.95	39.35

a posteriori puisque aucune courbe n'a été réalisée à des températures proches de celle du taux de croissance optimal ou de la température maximale.

L'intervalle de confiance de niveau 95% du paramètre $M_{y_{\max}}$ est plus petit que celui donné en *a priori* (Uniforme[5 ; 8.5]) pour les trois modèles : [6.16 ; 6.78] log ufc/g pour $M1$, [5.76 ; 6.29] log ufc/g pour $M2$ et [5.74 ; 6.20] log ufc/g pour $M2_{\lambda_{obs}}$ (Tab. 3.7). L'écart-type étant en moyenne environ égal à 1 log ufc/g, les valeurs *a posteriori* de la concentration bactérienne maximale $y_{\max}$ fluctuent pour 95% d'entre elles entre 4.5 et 8.5 log ufc/g pour le modèle $M1$ et entre 4 et 8 log ufc/g pour les modèles $M2$ et $M2_{\lambda_{obs}}$.

Les valeurs *a posteriori* du paramètre M_K du modèle $M2$ sont éloignées de celles entrées comme priors dans le modèle. L'intervalle à 95% *a priori* était [-3.11 ; 9.29]. La moyenne a posteriori de M_k est égale à 0.64 et le CI à 95% est [0.43 ; 0.90] (Tab. 3.8). Les valeurs de la constante K varient à 95% dans l'intervalle [0.08 ; 1.44]. Ce décalage ne semble pas étonnant étant donné que le produit étudié (la salade) diffère de ceux pris en compte par Augustin & Carlier (2000a) (viande, oeuf, poisson). Les données ont donc permis d'affiner l'estimation

TAB. 3.6: Statistiques descriptives de la distribution *a posteriori* des paramètres de la température maximale de croissance de *L. monocytogenes* dans les salades (unité : °C)

Modèle	Paramètre	Moyenne	Ecart-type	$P_{0.025}$	$P_{0.50}$	$P_{0.975}$
$M1$	T_{max}	45.81	1.02	43.86	45.81	47.78
$M2$	$M_{T_{max}}$	45.52	1.02	43.36	45.55	47.43
	$S_{T_{\max}}$	0.61	0.38	0.27	0.51	1.59
	T_{max}	45.52	1.25	42.95	45.54	47.89
$M2_{\lambda_{obs}}$	$M_{T_{max}}$	45.49	1.04	43.43	45.52	47.60
	$S_{T_{\max}}$	0.62	0.36	0.27	0.52	1.58
	T_{max}	45.49	1.26	43.01	45.51	47.97

TAB. 3.7: Statistiques descriptives de la distribution *a posteriori* des paramètres de la concentration maximale de *L. monocytogenes* dans les salades (unité : log ufc/g)

Modèle	Paramètre	Moyenne	Ecart-type	$P_{0.025}$	$P_{0.50}$	$P_{0.975}$
$M1$	$M_{y_{max}}$	6.46	0.16	6.16	6.46	6.78
	$\sigma_{y_{\max}}$	0.94	0.10	0.75	0.94	1.15
	y_{max}	6.45	0.96	4.57	6.44	8.34
$M2$	$M_{y_{max}}$	6.02	0.14	5.76	6.02	6.29
	$\sigma_{y_{\max}}$	1.04	0.09	0.88	1.04	1.22
	y_{max}	6.03	1.05	3.96	6.03	8.09
$M2_{\lambda_{obs}}$	$M_{y_{max}}$	5.97	0.12	5.74	5.96	6.20
	$\sigma_{y_{\max}}$	0.95	0.08	0.81	0.95	1.12
	y_{max}	5.96	0.96	4.07	5.96	7.85

TAB. 3.8: Statistiques descriptives de la distribution *a posteriori* des paramètres de la constante K de *L. monocytogenes* dans les salades

Modèle	Paramètre	Moyenne	Ecart-type	$P_{0.025}$	$P_{0.50}$	$P_{0.975}$
	M_K	0.64	0.12	0.43	0.63	0.90
	S_K	0.04	0.02	0.02	0.04	0.09
$M2$	K	0.65	0.36	0.08	0.60	1.44
	S_{KC}	0.37	0.07	0.26	0.36	0.52
	KC	0.65	0.37	0.07	0.60	1.45

du paramètre K. Cependant les valeurs *a posteriori* obtenues pour les temps de latence des cinétiques ne sont toujours pas satisfaisantes avec ce modèle. En effet, en comparant avec les courbes expérimentales, les valeurs obtenues avec le modèle $M2$ sont généralement sous-estimées lorsqu'il y a présence d'un temps de latence. A l'inverse pour les courbes sans temps de latence observé, le modèle a tendance à donner une valeur positive. Le paramètre M_λ du modèle $M1$ a une moyenne *a posteriori* de -3.88h et un CI à 95% de [-7.08 ; 0.62]h (résultats non présentés). Avec ce modèle les bactéries ne subissent pas dans la plupart des cas de temps de latence, ou alors celui-ci est très court.

L'inférence bayésienne a aussi permis d'actualiser les valeurs *a priori* des paramètres μ_{opt} du modèle $M1$ et $M_{\mu_{opt}}$ des modèles $M2$ et $M2_{\lambda_{obs}}$. Les valeurs *a posteriori* obtenues pour ce paramètre sont différentes d'un modèle à l'autre. Pour le modèle $M1$, le taux de croissance optimal μ_{opt} est égal *a posteriori* en moyenne à 0.13h^{-1} et son intervalle de variation à 95% est $[0.09\,;\,0.19]\text{h}^{-1}$ (Tab. 3.9). Les valeurs moyennes *a posteriori* des paramètres $M_{\mu_{opt}}$ et $S_{\mu_{opt}}$ déterminées avec le modèle $M2$ sont respectivement 0.33h^{-1} et 0.14h^{-1} et celles déterminées avec le modèle $M2_{\lambda_{obs}}$ sont respectivement 0.40h^{-1} et 0.16h^{-1}. Le modèle $M2$ produit pour le taux de croissance optimal une valeur moyenne *a posteriori* de 0.33h^{-1} et un CI à 95% de $[0.03\,;\,0.68]\text{h}^{-1}$. Le modèle $M2_{\lambda_{obs}}$ propose une valeur moyenne *a posteriori* pour le taux de croisance optimal de 0.40h^{-1} et un CI de $[0.06\,;\,0.80]\text{h}^{-1}$.

3.7.2 Résultats de la validation

Représentation graphique des valeurs observées et des valeurs prédites

Les graphiques de la figure 3.5 représentent pour chaque modèle le tracé des valeurs prédites en fonction des valeurs observées pour chaque temps $(t_k^{(l)})_{k=1,...,K_l}$, de chaque courbe $l = 1, ..., L$. Pour les trois modèles, les points sont répartis autour de la bissectrice ce qui signifie que les modèles sont en adéquation avec les données. Les nuages de points sont

TAB. 3.9: Statistiques descriptives de la distribution *a posteriori* des paramètres du taux de croissance optimal de *L. monocytogenes* dans les salades (unité : h^{-1})

Modèle	Paramètre	Moyenne	Ecart-type	$P_{0.025}$	$P_{0.50}$	$P_{0.975}$
$M1$	μ_{opt}	0.13	0.03	0.09	0.13	0.19
$M2$	$M_{\mu_{opt}}$	0.33	0.08	0.21	0.32	0.52
	$S_{\mu_{opt}}$	0.14	0.03	0.09	0.14	0.22
	μ_{opt}	0.33	0.16	0.03	0.32	0.68
$M2_{\lambda_{obs}}$	$M_{\mu_{opt}}$	0.40	0.08	0.26	0.40	0.59
	$S_{\mu_{opt}}$	0.16	0.04	0.10	0.16	0.25
	μ_{opt}	0.40	0.19	0.06	0.39	0.80

toutefois assez larges et les fortes valeurs de concentrations prédites par les modèles sont en général plus faibles que celles observées.

Validation à l'aide d'un critère d'adéquation

Les tableaux 3.10 et 3.11 présentent la croissance observée et la croissance répliquée par les différents modèles pendant une durée donnée à différentes températures. La croissance est l'augmentation en log ufc de *L. monocytogenes*/g pour une durée donnée. La croissance moyenne observée est reportée dans les tableaux 3.10 et 3.11 ainsi que la croissance moyenne et son intervalle d'incertitude (UI) de niveau 95% obtenues en effectuant B =10 000 réplications pour les différents modèles. Les concentrations moyennes observées sont calculées à partir des différentes courbes disponibles pour chaque température. Les températures pour lesquelles le nombre de courbes disponibles pour un temps donné était inférieur à 4 n'ont pas été retenues. La durée a été choisie par un compromis entre un temps court qui permet un meilleur ajustement et un nombre de courbes suffisant. Pour information, pour chaque température la croissance minimale et la croissance maximale observées parmi les différentes courbes sont présentées dans les tableaux 3.10 et 3.11. La croissance minimale est considérée comme nulle (non croissance) pour les températures de 3°C à 10°C. A 15°C, la croissance minimale observée est de 1.14 log ufc/g en 4 jours.

A 3°C et pour une durée de 168h, la croissance moyenne en *L. monocytogenes* observée à partir de 22 courbes est de 0.43 log ufc/g. A cette température et pour une durée équivalente le modèle $M1$ prédit une croissance moyenne de 1.03 log ufc/g avec un UI à 95% de [0.61 ; 1.51] log ufc/g. Le modèle $M2$ produit des résultats à peu près équivalents : 0.94 log ufc/g avec un UI à 95% de [0.50 ; 1.45] log ufc/g. Le modèle $M2_{\lambda_{obs}}$ prédit une croissance moyenne

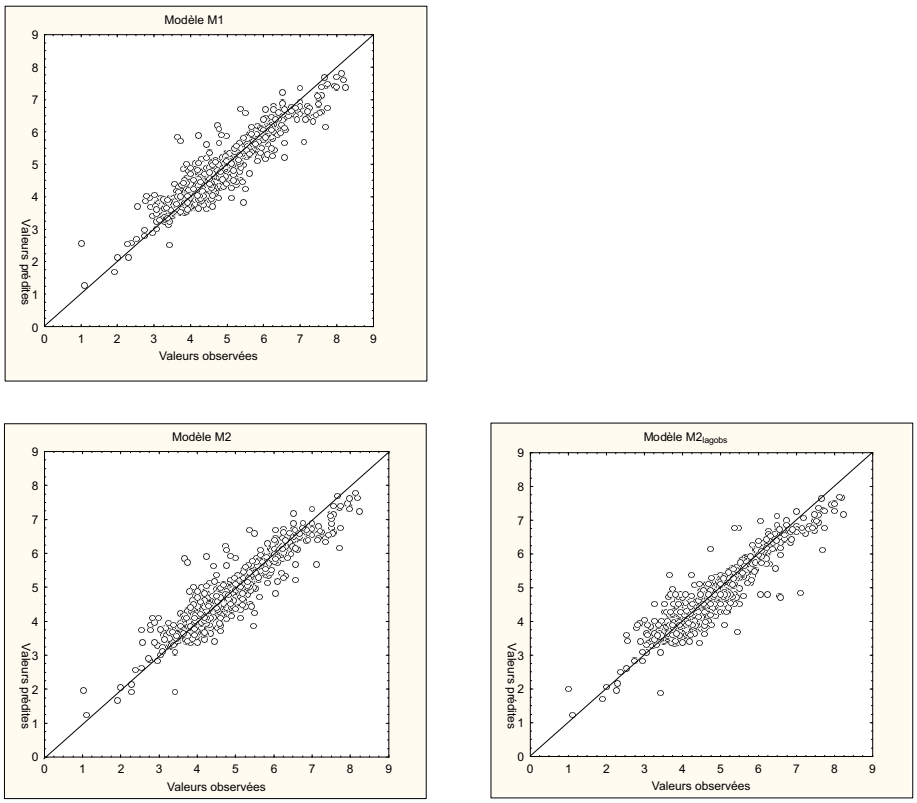

FIG. 3.5: Représentation graphique des valeurs observées et des valeurs prédites pour les trois modèles de croissance

TAB. 3.10: Résultats de la validation à l'aide d'un critère d'adéquation, des trois modèles d'estimation des paramètres de croissance (Observations et modèle $M1$)

			Observations			Modèle $M1$			
Temp.	Durée	Nb de	C. Moy.	C. Min.	C. Max.	C. Moy.	UI 95%		p_{valeur}
°C	h	courbes	log ufc/g	log ufc/g	log ufc/g	log ufc/g	log ufc/g		%
3.00	168.00	22.00	0.43	-∞	1.27	1.03	0.61	1.51	99.90
5.00	96.00	11.00	0.03	-∞	0.61	0.77	0.31	1.32	99.90
8.00	120.00	19.00	0.73	-∞	1.36	0.85	0.43	1.23	71.70
10.00	96.00	52.00	1.25	-∞	2.64	1.06	0.78	1.35	10.80
15.00	96.00	5.00	1.77	1.14	2.27	1.42	0.51	2.35	24.00

p_{valeur} : une valeur proche de 50% signifie que le modèle est en adéquation avec les données, une valeur supérieure (inférieure) à 50% que le modèle surestime (sous-estime) la croissance

TAB. 3.11: Résultats de la validation à l'aide d'un critère d'adéquation, des trois modèles d'estimation des paramètres de croissance (Modèle $M2$ et $M2_{\lambda_{obs}}$)

			Observations			Modèle $M2$				Modèle $M2_{\lambda_{obs}}$			
Temp. °C	Durée h	Nb de courbes	C. Moy. log ufc/g	C. Min. log ufc/g	C. Max. log ufc/g	C. Moy. log ufc/g	UI 95% log ufc/g		p*valeur* %	C. Moy. log ufc/g	UI 95% log ufc/g		p*valeur* %
3.00	168.00	22.00	0.43	-∞	1.27	0.94	0.50	1.45	99.30	0.82	0.43	1.23	97.70
5.00	96.00	11.00	0.03	-∞	0.61	0.72	0.21	1.30	99.80	0.32	0.02	0.68	96.90
8.00	120.00	19.00	0.73	-∞	1.36	0.75	0.32	1.24	52.30	0.83	0.37	1.29	68.50
10.00	96.00	52.00	1.25	-∞	2.64	1.12	0.77	1.50	23.80	1.23	0.93	1.55	44.20
15.00	96.00	5.00	1.77	1.14	2.27	1.52	0.57	2.52	28.50	1.69	0.75	2.61	43.70

p*valeur* : une valeur proche de 50% signifie que le modèle est en adéquation avec les données, une valeur supérieure (inférieure) à 50% que le modèle surestime (sous-estime) la croissance

plus faible que les deux modèles précédents : 0.82 log ufc/g avec un UI à 95% de [0.43 ; 1.23] log ufc/g. Notons que les intervalles de confiance des modèles $M1$ et $M2$ n'encadrent pas la valeur observée et se situent au-dessus. La croissance maximale observée pour une des cinétiques de croissance est de 1.27 log ufc/g. Les p-valeurs des trois modèles sont très élevés (sup. à 99%), ce qui signifie que dans plus de 99% des simulations les trois modèles produisent des valeurs moyennes supérieures à celle observée. A 5°C et pour une durée de 96h, la croissance moyenne en *L. monocytogenes* calculée à partir de 11 courbes est quasi nulle 0.03 log ufc/g. Cette valeur moyenne très faible est due à l'existence d'une décroissance du nombre de bactéries pour certaines courbes. Les modèles prédisent une croissance relativement faible mais non nulle, environ 0.70 log ufc/g pour les modèles $M1$ et $M2$ et 0.32 log ufc/g pour le modèle $M2_{\lambda_{obs}}$. L'intervalle de confiance calculé avec le modèle $M2_{\lambda_{obs}}$ est le seul intervalle à inclure la valeur moyenne observée pour cette température. De même qu'à 3°C, les p-valeurs calculées avec les trois modèles sont très élevées : 99.90% pour le modèle $M1$ et 96.90% pour le modèle $M2_{\lambda_{obs}}$. Notons cependant en observant les données qu'une cinétique réalisée à 5°C (Koseki & Isobe, 2005b) et qui n'a pas été retenue pour la validation car le temps d'incubation était de 94h, a une croissance de 0.8 log ufc/g, ce qui correspond à la croissance répliquée par les modèles.

Pour une durée de 120h à 8°C, la croissance moyenne en *L. monocytogenes* observée à partir de 19 courbes est 0.73 log ufc/g. Les trois modèles prédisent des résultats proches de cette valeur : une croissance moyenne de 0.85 log ufc/g avec un UI à 95% de [0.43 ; 1.23] log ufc/g pour le modèle $M1$, une croissance moyenne de 0.75 log ufc/g avec un UI à 95% de [0.32 ; 1.24] log ufc/g pour le modèle $M2$ et respectivement 0.83 log ufc/g et [0.37 ; 1.29]

log ufc/g pour le modèle $M2_{\lambda_{obs}}$. Les p-valeurs sont respectivement égales à 71.70%, 52.30% et 68.50% pour les modèles $M1$, $M2$ et $M2_{\lambda_{obs}}$. Les trois modèles prédisent correctement la croissance moyenne en *L. monocytogenes* à cette température, en particulier, le modèle $M2$ surestime moins souvent la croissance moyenne par rapport aux deux autres modèles.

A 10°C et pour une durée de 96h, la croissance moyenne en *L. monocytogenes* observée à partir de 52 cinétiques de croissance est 1.25 log ufc/g. Le modèle $M1$ prédit une croissance moyenne de 1.06 log ufc/g et un UI à 95% de [0.78 ; 1.35] log ufc/g. Le modèle $M2$ produit une croissance moyenne de 1.12 log ufc/g et un UI à 95% de [0.77 ; 1.50] log ufc/g. Le modèle $M2_{\lambda_{obs}}$ obtient une croissance moyenne similaire à celle observée : 1.25 log ufc/g de croissance moyenne observée v.s 1.23 log ufc/g et un UI à 95% de [0.93 ; 1.55] log ufc/g de croissance moyenne répliquée. Les p-valeurs des trois modèles sont respectivement 10.80%, 23.80% et 44.20%. Les modèles $M1$ et $M2$ prédisent respectivement dans 76.2% et 89.20% des cas des croissances moyennes en *L. monocytogenes* inférieures à celle observée à 10°C.

Pour une durée de 96h à 15°C, la croissance moyenne en *L. monocytogenes* obtenue à partir de 5 cinétiques est 1.77 log ufc/g. Notons que pour cette température le nombre de courbes est faible. Les croissances moyennes prédites sont respectivement 1.42 log ufc/g avec un UI à 95% de [0.51 ; 2.35] log ufc/g avec le modèle $M1$, 1.52 log ufc/g avec un UI à 95% de [0.57 ; 2.52] log ufc/g avec le modèle $M2$ et respectivement 1.69 log ufc/g avec un UI à 95% de [0.75 ; 2.61] log ufc/g avec le modèle $M2_{\lambda_{obs}}$. L'observation des p-valeurs permet de conclure que les modèles prédisent correctement la croissance moyenne à cette température en particulier le modèle $M2_{\lambda_{obs}}$ dont la p-valeur est 43.70%. Cependant, les modèles $M1$ et $M2$ prédisent plus souvent des croissances moyennes plus faibles que celle observée à 15°C.

3.7.3 Résultats de la sélection de modèles par le Critère d'Information de Déviance

Les Critères d'Information de Déviance obtenus pour les modèles $M1$, $M2$ et $M2_{\lambda_{obs}}$ sont respectivement 1438, 1344 et 1454. Le modèle qui minimise le DIC est donc le modèle $M2$ qui est le modèle comportant le plus de paramètres. Celui qui a le DIC le plus élevé est le modèle $M2_{\lambda_{obs}}$. Notons que ce critère ne va pas dans le même sens que le critère étudié précédemment selon lequel le modèle le plus performant est le modèle $M2_{\lambda_{obs}}$.

3.8 Discussion

Ce chapitre présente plusieurs modèles pour estimer les paramètres de croissance de *L. monocytogenes* dans les salades feuilles à partir de courbes de croissance de différentes origines. La validation proposée afin d'évaluer l'adéquation des modèles aux données montre que les trois modèles satisfont dans l'ensemble l'objectif principal qui est une prévision correcte de la croissance moyenne en *L. monocytogenes* pendant un temps fini et pour une température donnée. Si les modèles devaient être classés en fonction de ce critère, le meilleur modèle serait le modèle $M2_{\lambda_{obs}}$, suivi du modèle $M2$ et enfin le modèle $M1$. Selon le DIC, qui est un critère général le classement serait le modèle $M2$, puis le modèle $M1$ et enfin le modèle $M2_{\lambda_{obs}}$. Le modèle $M2$ est donc le modèle le mieux classé par les deux critères. Dans la suite du travail les distributions *a posteriori* des paramètres obtenues avec les modèles $M2$ et $M2_{\lambda_{obs}}$ seront utilisées pour simuler la croissance de *L. monocytogenes* dans la salade en sachet selon les conditions réelles de conservation de ce produit.

Le paramètre $T_{\min}$ correspond à la température minimale de croissance de la bactérie étudiée. Les trois modèles présentent des valeurs *a posteriori* particulièrement faibles (-5.29°C et -4.38°C en moyenne selon le modèle) pour ce paramètre en comparaison des études précédemment réalisées sur *L. monocytogenes.* En effet Augustin & Carlier (2000a) proposaient une valeur médiane de -2.7°C, Pouillot et al. (2003) obtenaient une valeur moyenne de -2.47°C et Delignette-Muller et al. (2006) trouvaient en moyenne -2.86°C. Aux valeurs de température obtenues la salade est gelée, ce paramètre n'aura donc pas ici une réelle signification biologique. Les faibles valeurs de ce paramètre peuvent être à l'origine de la surestimation de la croissance à de faibles températures. Aucune courbe de croissance de *L. monocytogenes* sur de la salade feuille n'a été réalisée à des températures élevées c'est-à-dire à plus de 30°C. Ainsi l'estimation des paramètres concernant les températures optimales et maximales de croissance ne semble pas réalisable avec les seules données disponibles. L'inférence bayésienne permet d'intégrer de l'information *a priori* sur ces paramètres. Leur attribuer une loi de probabilité comme *a priori* évite de les fixer à une valeur connue, et permet ainsi une certaine souplesse sur les valeurs de ces paramètres, ce qui a pour conséquence un meilleur ajustement des cinétiques de croissance par les modèles. Toutefois, pour vérifier la validité des résultats précédents, le modèle $M2$ a été testé en utilisant des valeurs fixes pour les paramètres $T_{\min}$, T_{opt} et $T_{\max}$. Les valeurs choisies sont les valeurs médianes obtenues par Augustin & Carlier (2000a), respectivement -2.7°C, 37°C et 45°C. Les résultats sont équivalents à ceux obtenus pour les modèles $M2$ et $M2_{\lambda_{obs}}$. Ainsi même si la valeur de $T_{\min}$ est plus haute que pour

les précédents modèles, la croissance moyenne répliquée à de faibles températures est quand même plus élevée que celle observée. On optera donc pour les modèles incluant une certaine incertitude sur les paramètres de température du modèle secondaire de croissance. Les valeurs *a posteriori* des paramètres représentant la concentration maximale peuvent parfois prendre des valeurs proches de 4 log ufc/g. Ce résultat est lié au fait que les valeurs de ces paramètres correspondent aux dernières valeurs ayant été relevées par les expérimentateurs et ne signifie pas pour autant que la phase stationnaire a été atteinte. Ce choix a été effectué pour privilégier la qualité de l'ajustement afin de déterminer le taux de croissance le plus adéquat pour chaque courbe. C'est aussi la raison pour laquelle la variance $\sigma^2_{y_{\max}}$ de ce paramètre a été considérée différente de la variance σ^2_y des autres points des cinétiques de croissance. Dans le chapitre suivant où la croissance de *Listeria monocytogenes* est simulée en conditions réelles de conservation, l'influence de la valeur de $y_{\max}$ sur les niveaux de concentration sera testée. A cette fin une distribution normale de moyenne 6.64 log ufc/g et d'écart-type 0.81 log ufc/g tronquée et définie sur l'intervalle [5, 8.5] log ufc/g est ajustée aux concentrations maximales des courbes pour lesquelles une phase stationnaire a été atteinte. Ces courbes sont au nombre de 59 et leur numéro est marqué par le symbole (*) dans le tableau 3.1. Notons enfin qu'en comparant avec les valeurs obtenues pour d'autres produits, le taux de croissance optimal de *L. monocytogenes* dans les salades feuilles (0.33h^{-1} ou 0.40h^{-1} en moyenne avec les modèles $M2$ et $M2_{\lambda_{obs}}$) est plus faible. En effet, dans le lait la valeur moyenne du taux de croissance optimal obtenue par Pouillot et al. (2003) est 0.69h^{-1}. Le logiciel Symprevius de simulation de la croissance propose des valeurs de taux de croissance optimal calculées à partir d'une ou plusieurs courbes de croissance par ses laboratoires. Par exemple, pour les rillettes le taux de croissance optimal de *L. monocytogenes* est en moyenne de 2.15h^{-1}, pour la viande de porc et le poulet cru celui-ci est respectivement en moyenne de 0.79h^{-1} et 1.24h^{-1}. La croissance dans le saumon fumé est la plus proche de celle dans la salade feuille avec un taux de croissance optimal de 0.42h^{-1}. La salade feuille est donc un aliment dans lequel *L. monocytogenes* se développe moins rapidement que dans les produits carnés, de la mer ou laitiers.

Peut-être serait-il intéressant de traiter différemment les cinétiques de croissance effectuées à de faibles températures de celles effectuées à des températures plus élevées. En effet, la phase stationnaire n'est jamais atteinte pour la plupart des cinétiques réalisées à des températures faibles. Dans ce cas, le modèle trilinéaire de Buchanan et al. (1996) serait peut-être mieux adapté que le modèle de Rosso (1995). De plus, si les données étaient disponibles, l'utilisation de modèles secondaires incluant l'effet des différents facteurs environnementaux par

système modulaire ainsi que l'interaction des facteurs entre eux (Adams et al., 1991; Wijtzes et al., 1993; Augustin & Carlier, 2000b) produirait peut-être de meilleures estimations.

Pour conclure, la méthodologie employée dans ce chapitre produit des résultats dans l'ensemble satisfaisants malgré la difficulté d'estimer des paramètres de croissance de *L. monocytogenes* dans la salade feuille. En effet, les résultats prédits par les modèles sont proches des valeurs observées et sont plutôt conservateurs, ce qui d'un point de vue du risque alimentaire est important. De plus, un des objectifs principaux de ce travail qui était de traiter un grand nombre de courbes de croissance de *L. monocytogenes* sur la salade feuille afin de capter l'hétérogénéité supposée de ce produit, a été atteint. Ainsi l'hétérogénéité du développement de *L. monocytogenes* dans ce produit a été confirmée. Un autre facteur a également été intégré : ce facteur est l'incertitude des concentrations relevées due aux différentes méthodes expérimentales mises en oeuvre par les laboratoires afin d'étudier la croissance. Ce travail est plutôt convaincant étant donné le peu de points expérimentaux disponibles par courbe de croissance. La méthodologie employée permet une fois l'implémentation informatique assez complexe effectuée de traiter rapidement et de manière uniforme, un grand nombre de courbes pour tout type de bactéries et d'estimer dans un même processus les paramètres de croissance des modèles primaire et secondaire.

Chapitre 4

Simulations de la contamination par *Listeria monocytogenes* des salades de IV^{ème} gamme à l'instant de la consommation

4.1 Introduction

Ce chapitre présente le modèle de simulation développé afin d'évaluer la contamination par *Listeria monocytogenes* des salades de IV^{ème} gamme de leur fabrication jusqu'à leur consommation. L'unité commercialisée étant le sachet de salade, les concentrations des salades contenues dans les sachets contaminés sont simulées tout au long de la chaîne alimentaire et le pourcentage de sachets contaminés est calculé. La contamination microbiologique des aliments est rarement relevée à chaque étape de la chaîne alimentaire. L'élaboration d'un tel modèle permet de disposer d'une estimation de la contamination aux différentes étapes de la fabrication du produit et du circuit logistique. De plus, ce modèle permet de tester des scénarios afin d'évaluer l'impact sur la contamination de la modification d'une étape du procédé de fabrication par exemple ou encore d'une amélioration des températures de la chaîne du froid.

Les étapes du procédé de fabrication et du circuit logistique suivi par les produits et qui ont un effet sur la prévalence et/ou sur les concentrations en micro-organismes sont identifiées et prise en compte dans le modèle. Le modèle de simulation nécessite de connaître

certaines quantités associées aux différentes étapes du procédé de fabrication et du circuit logistique. Ces quantités seront utilisées à des niveaux différents de la chaîne alimentaire comme paramètres d'entrée du modèle. Ces quantités sont la contamination de la matière première entrant dans l'usine, la réduction en micro-organismes lors de la fabrication du produit et les paramètres de croissance des bactéries au cours de la conservation du produit. Le modèle développé dans le second chapitre est appliqué aux données appropriées, afin d'estimer la distribution de la contamination de la matière première. Les distributions *a posteriori* des paramètres de croissance (taux de croissance optimal, température minimale de croissance, etc...) obtenues à partir des deux modèles sélectionnés dans le troisième chapitre sont utilisées pour simuler la croissance de *L. monocytogenes* sur les feuilles de salade des sachets contaminés lors de leur conservation de la sortie de la chaîne de fabrication jusqu'à leur consommation. L'effet du choix de l'un des deux modèles sur les concentrations à l'instant de la consommation peut ainsi être évalué. De même des scénarios portant sur la distribution du paramètre de croissance $y_{\max}$, qui est la concentration maximale atteignable par la population de micro-organismes, sont testés dans ce chapitre. En plus des paramètres de croissance, il est nécessaire de connaître les facteurs ayant une influence sur la croissance. Ces facteurs ont été identifiés dans le troisième chapitre et sont la température et la durée de conservation des produits. Les durées et températures sont variables d'une étape à une autre ainsi qu'au sein d'une même étape. Afin de prendre en compte la variabilité existante au sein d'une même étape pour ces facteurs, des distributions de probabilité provenant de sources variées (études nationales, dires d'experts, enquêtes auprès de consommateurs) sont utilisées pour simuler des valeurs de températures et de durées. La réduction du nombre de micro-organismes lors du passage des salades dans le bain de lavage est modélisée en ajustant une loi sur les réductions observées à partir de données provenant de la littérature. Afin d'évaluer l'impact du retrait du chlore du procédé de fabrication sur le niveau et le taux de contamination des salades en *L. monocytogenes*, deux scénarios sont testés : l'eau de lavage est traitée avec du chlore et l'eau de lavage ne contient aucune trace de chlore. Un test de vieillissement est simulé avec notre modèle. Ce test réalisé par les industriels afin de déterminer la date limite de consommation (DLC) des produits, consiste à relever les contaminations microbiologiques après conservation des produits à une température et une durée fixées (3 jours à 4°C et 6 jours à 8°C). Ces températures et durées ont été sélectionnées par les industriels afin de reproduire les conditions réelles de conservation des sachets de salade de la fin de leur fabrication jusqu'à leur consommation. Afin d'évaluer l'efficacité du test de vieillissement et ainsi l'effet sur l'estimation de la contamination de la prise en compte de

la variabilité de la température et de la durée de conservation, les concentrations obtenues par simulation du test de vieillissement sont comparées avec celles obtenues en utilisant des distributions de probabilité pour la température et la durée de conservation.

Ce chapitre est structuré comme suit. Une première section décrit le procédé de fabrication et le circuit logistique suivi par les salades en sachet. Une seconde section porte sur le modèle de simulation. Ainsi les étapes composant le modèle sont définies. Les modèles d'estimation de la contamination de la matière première, de la réduction en micro-organismes après lavage et de simulation de la croissance sont présentés. Après la présentation de l'ensemble des données disponibles dans une troisième section, les distributions qui sont utilisées comme paramètres d'entrée dans le modèle sont décrites dans une quatrième partie. La réalisation des simulations fait l'objet d'une cinquième partie. Les résultats sont présentés et commentés dans une sixième partie. Les hypothèses, les données nécessaires à l'élaboration d'un tel modèle ainsi que les résultats obtenus sont discutés dans la septième et dernière partie.

4.2 Procédé de fabrication et circuit logistique des salades de IV^{ème} gamme

Cette partie présente le procédé de fabrication des salades en sachet ainsi que les étapes constituant le circuit logistique de ces produits, de la fin de leur fabrication jusqu'à l'instant de leur consommation. Les trois principales étapes sont :

- la fabrication des salades en sachet,
- le transport et le stockage entre la fabrication et la vente des sachets chez le distributeur,
- la conservation du produit chez le consommateur incluant le trajet du point de vente jusqu'au domicile.

Procédé de fabrication des salades en sachet

Le procédé de fabrication présenté ici est une description générale des procédés des industries de IV^{ème} gamme qui diffèrent légèrement d'une entreprise à l'autre. Cette description est illustrée par la figure 4.1 disponible sur le site internet de la société Bonduelle (www.bonduelle.com).

Les entreprises françaises de fabrication de produits de IV^{ème} gamme privilégient les approvisionnements de proximité en matière première. La matière première est constituée de légumes frais qui sont pour la majorité des salades feuilles. En contre-saison, les produc-

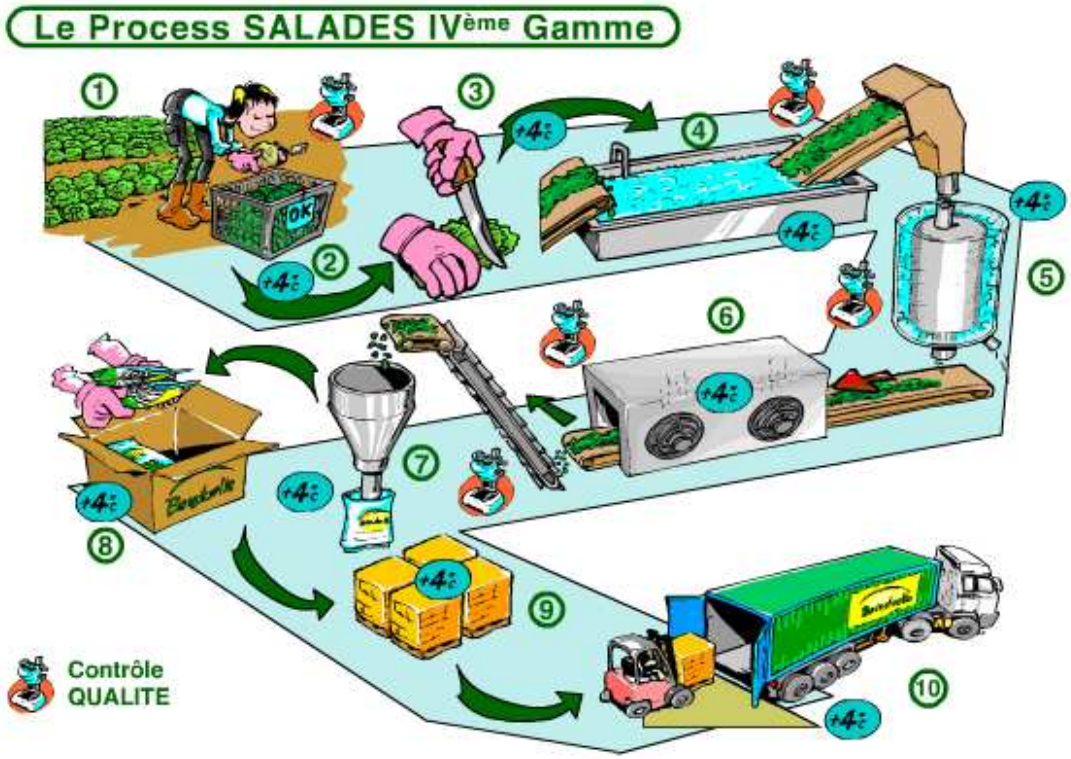

FIG. 4.1: Représentation shématique du procédé de fabrication des salades de IV^{ème} gamme (www.bonduelle.com). (1). Champ de salade (2). Réception /Agréage (3). Parage / Coupe (4). Lavage (5). Essorage (6). Séchage (7). Pesage / Ensachage (8). Conditionnement (9). Stockage (10). Chargement / Livraison.

teurs des pays limitrophes de la France comme l'Espagne, le Portugal et l'Italie fournissent une partie de la matière première (étape (1) de la figure 4.1). Le système de production des entreprises de IV^{ème} gamme est dit à "flux tendu", c'est-à-dire que les produits sont destinés à être vendu avec un temps de stockage réduit au minimum après leur fabrication. Ainsi les durées du procédé de fabrication et de livraison aux distributeurs (hypermarché, supermarché, supérette, etc...) sont courtes. Le temps nécessaire à la fabrication d'un sachet est de 20 minutes en moyenne et le temps de livraison varie de quelques heures à 4 jours et demi au maximum. Après réception des cagettes, les salades sont observées minutieusement (fraîcheur, état des feuilles, présence d'insectes, de terre, etc...) et analysées sur de nombreux critères (vérification du poids, taux de nitrate, de matière utile et de vert) c'est l'étape d'agréage (2). Si la matière première est conforme aux exigences, celle-ci est stockée dans une chambre froide (entre 4 et 6°C) constamment humidifiée. Le temps de stockage dans la

chambre froide est au minimum de 1h et au maximum de 1 jour et demi lorsque la fabrication est interrompue par le week-end. Les salades sont ensuite disposées sur les lignes de production. Les lignes de production sont spécifiques de la variété de salades à traiter et du produit à fabriquer (sachet composé d'une seule variété de salade ou de plusieurs variétés). Dans une première salle dans laquelle la température fluctue entre 4 et 6°C, les salades sont parées manuellement pour éliminer les premières couronnes de feuilles vertes et les feuilles abîmées puis découpées manuellement ou mécaniquement selon la fragilité de la variété (3). Ensuite, dans une deuxième salle maintenue entre 4 et 7°C (4), les feuilles sont convoyées sur un tapis roulant jusqu'au bain de lavage composé d'eau chlorée (entre 45 et 60 ppm en moyenne, avec une limite maximale à 80 ppm). Puis, elles sont rincées au moyen d'eau glacée (environ 1°C), essorées (5,6) et pesées pour être ensuite ensachées mécaniquement sous air ambiant ou parfois sous atmosphère modifiée, dans une troisième salle dont la température est de 4°C (7).

Transport et stockage entre la fabrication et la vente des sachets

Après vérification de la qualité des sachets (volume d'air, solidité de la soudure), ceux-ci sont disposés dans des cartons et stockés (8,9) en chambre froide (4°C) pour une durée d'environ 12h dans l'attente du chargement des camions frigorifiques (10). La livraison aux distributeurs est directe ou nécessite le déchargement des produits sur une plateforme de distribution. Les magasins livrés peuvent se situer jusqu'à 400 km de l'usine. La méthode de gestion de stockage FIFO (First In First Out) est pratiquée dans les usines de fabrication de produits de IV^{ème} gamme. Cette méthode consiste à expédier en priorité les produits dont la matière première est entrée en premier dans l'usine. Le temps de stockage du produit chez le distributeur dépend du débit du magasin et du temps défini par celui-ci pour garder le produit en vente avant sa date limite de consommation (DLC).

Transport après achat et conservation avant consommation

Les durées des trajets après achat sont variables en fonction du lieu d'habitation du consommateur. Selon certaines enquêtes les produits peuvent rester plus de 2 heures dans les véhicules des acheteurs (Stahl et al., 2006; Afchain, 2005). L'étude réalisée par l'Actia (Stahl et al., 2006) montre aussi qu'à son domicile le consommateur conserve la salade en sachet au réfrigérateur ou à température ambiante. La totalité du sachet est consommée en une seule prise alimentaire ou en plusieurs selon la quantité de salades contenues dans le sachet et le nombre de personnes dans le ménage. Il s'avère aussi que certaines personnes consomment la salade en sachet après la DLC. Certains résultats numériques de l'enquête Actia sont présentés dans la section 4.5.4.

4.3 Modèle de simulation de la contamination des sachets à l'instant de la consommation

4.3.1 Etapes du modèle

Les étapes retenues dans le modèle, durant lesquelles les concentrations en micro-organismes fluctuent, sont présentées dans la figure 4.2.

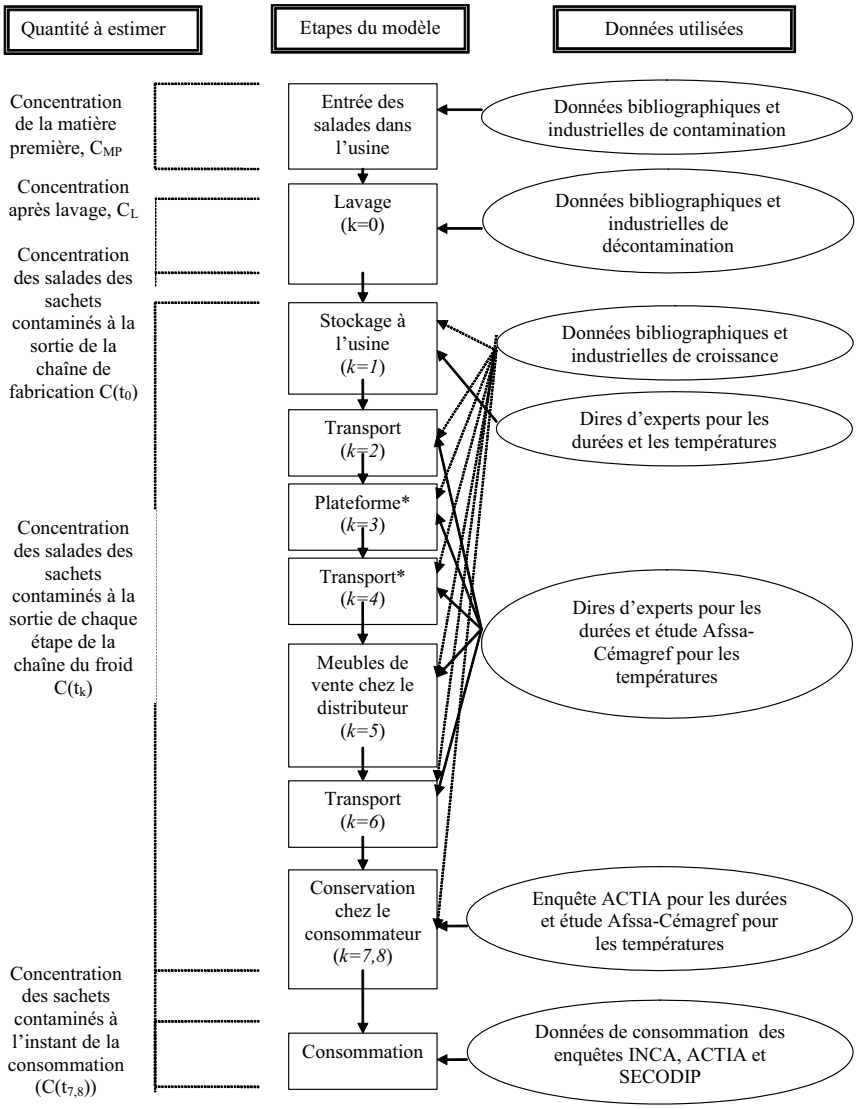

FIG. 4.2: Etapes du modèle d'estimation de la contamination des salades en sachet à l'instant de la consommation, données utilisées et quantité à estimer

La première étape déterminante pour la suite est l'estimation de la contamination de la matière première entrant dans l'entreprise. Le temps de fabrication d'un sachet étant très court (20 minutes) et les températures faibles, la croissance au sein de l'usine est considérée comme nulle. Ainsi au niveau du procédé de fabrication, seule l'étape de lavage de la matière première pouvant entraîner une réduction en micro-organisme, est retenue. La contamination des salades contenues dans les sachets est donc évaluée directement à partir de la distribution de contamination obtenue après le passage de la matière première dans l'eau de lavage. La croissance potentielle des bactéries lors du stockage des sachets avant chargement à bord des camions frigorifiques est prise en compte dans la modélisation ainsi que la croissance potentielle lors de chaque étape du circuit logistique. En accord avec les membres du comité de pilotage, huit étapes composant le circuit logistique subi par les salades en sachet de leur sortie de la chaîne de fabrication jusqu'à leur consommation sont déterminées. Ces 8 étapes sont :

1. le stockage dans la chambre froide de l'usine
2. le transport entre l'usine et la plateforme pour 80% des livraisons ou directement jusqu'au point de vente pour 20% des livraisons
3. le stockage sur la plateforme (pour 80% des livraisons)
4. le transport jusqu'au point de vente (pour 80% des livraisons)
5. le stockage chez le distributeur comprenant le stockage en chambre froide du magasin et dans les meubles de vente
6. le trajet jusqu'au domicile du consommateur
7. la conservation chez le consommateur avant la première consommation
8. la conservation du sachet entamé avant la deuxième et dernière consommation

En résumé, la contamination des salades en sachet par *Listeria monocytogenes* dépend de la concentration de la matière première, de la réduction en micro-organismes au cours du lavage et de la croissance des bactéries lors du stockage des produits de leur sortie de la chaîne de fabrication jusqu'à leur consommation.

4.3.2 Modèle d'estimation de la contamination de la matière première

Le modèle de mélange de deux lois normales présenté dans le chapitre 2 est appliqué ici pour établir la distribution de la contamination de la matière première entrant dans les

usines de fabrication de produit de IV^{ème} gamme. La loi jointe *a posteriori* des paramètres composant le modèle de mélange sera utilisée afin de simuler des valeurs de concentration de la matière première entrant dans l'usine.

4.3.3 Méthode d'estimation de la réduction en micro-organismes au cours du lavage à l'eau chlorée et du lavage à l'eau non chlorée

Des études publiées dans la littérature scientifique montrent les effets du lavage à l'eau chlorée et du lavage à l'eau non chlorée sur les concentrations en micro-organismes des salades. Le tableau 4.1 recense les études portant sur la réduction en *L. monocytogenes* (LM) et en flore aérobie mésophile (FAM) (flore de contamination, naturellement présente sur les produits et capable de se multiplier à l'air et à 30-37°C). Dans ces études, différentes concentrations en chlore ainsi que différents temps de contact entre les salades contaminées et l'eau de lavage ont été testés. D'autres facteurs ont été relevés comme le pH et la température du bain de lavage. L'effet du lavage sur les concentrations initiales est traduit par le nombre en log ufc/g de micro-organismes ayant été éliminés au cours d'un temps de contact t donné. Ce nombre est appelé par la suite "réduction en micro-organismes" et est calculé par : $y_0 - y(t)$ où y_0 est la concentration initiale en log ufc/g et $y(t)$ est la concentration en log ufc/g relevée après un temps de contact t. Notons que la réduction correspond à une division lorsque les concentrations sont exprimées en ufc/g. En étudiant la réduction en micro-organismes en fonction du temps de contact, il s'avère que l'effet du lavage (à l'eau non chlorée ou à l'eau chlorée) est quasiment instantané et se stabilise très rapidement. En effet, après 30 secondes ou 1 minute selon les expériences la réduction est quasiment maximale et les concentrations sont constantes lorsque le temps de contact augmente (Fig. 4.3). La comparaison de la distribution de la réduction obtenue au cours du lavage à l'eau chlorée et celle obtenue au cours du lavage à l'eau non chlorée à l'aide du test non-paramétrique de Wilcoxon-Mann-Whitney montre qu'il existe une différence significative entre ces deux distributions avec un risque de 5%.

La méthode proposée consiste à ajuster une première distribution sur les réductions obtenues au cours du lavage à l'eau non chlorée et une seconde distribution sur les réductions obtenues au cours du lavage à l'eau chlorée, quelles que soient les valeurs du pH, de la température du bain de lavage et le temps de contact. Toutefois, étant donné que les salades effectuent un court séjour dans le bain de lavage, les réductions ayant des temps de contact

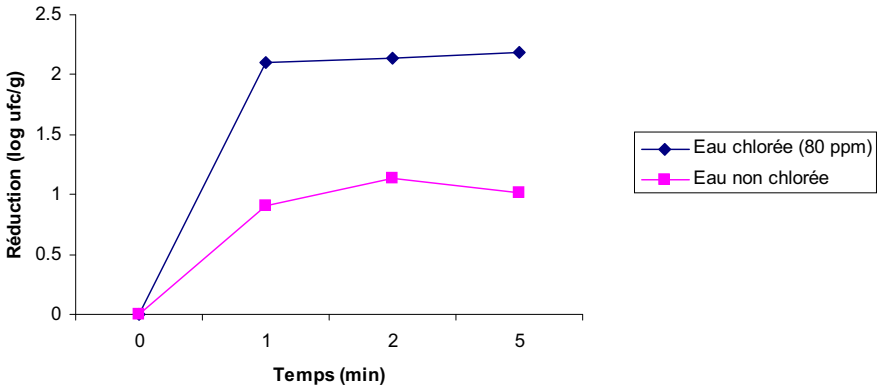

FIG. 4.3: Exemple de la réduction en micro-organismes en fonction du temps de contact avec l'eau de lavage chlorée ou non chlorée (Szabo et al. 2003)

supérieurs à 10 minutes n'ont pas été prises en compte. De plus, les concentrations en chlore des bains de lavage industriels ne dépassant pas les 80 ppm, les réductions correspondant à des concentrations en chlore supérieures à cette valeur n'ont pas été utilisées.

4.3.4 Modèle de simulation de la croissance de *L. monocytogenes* dans chaque étape du circuit logistique

L'objectif est de simuler la croissance de *L. monocytogenes* sur les feuilles de salades composant les sachets de leur sortie de la chaîne de fabrication jusqu'à leur consommation. L'évolution de la population bactérienne dépend de la température à laquelle est conservé le produit ainsi que la durée de la conservation. Les lois jointes *a posteriori* des paramètres de croissance obtenues dans le chapitre 3 à partir des modèles $M2$ et $M2_{\lambda_{obs}}$ sont utilisées ici pour simuler la croissance de *L. monocytogenes* en conditions réelles. Le temps de latence ne sera pas utilisé dans les simulations pour deux raisons. Premièrement, le temps de latence dépend non seulement de la température à laquelle les courbes de croissance ont été réalisées mais aussi des conditions de pré-culture des souches dont la température. Or les températures de pré-culture des courbes analysées sont supérieures à celles des *L. monocytogenes* natives (c'est-à-dire naturellement présentes sur les salades). Ainsi, les temps de latence calculés à partir de salades artificiellement contaminées ne peuvent être utilisés pour simuler la croissance des *Listeria monocytogenes* natives. Deuxièmement, le temps de latence dépend de l'adaptation de la bactérie à son substrat et du stress qu'elle peut subir. On peut

penser que les *L. monocytogenes* natives sont adaptées au substrat salade ce qui signifie que le temps de latence s'il existe doit être réduit. De plus, le passage dans l'eau de lavage n'a pas été identifié comme stressant pour *L. monocytogenes*.

Le circuit logistique des salades en sachet est défini par 8 étapes correspondant aux transports, à la vente et à la conservation chez le distributeur ou le consommateur. Soit $k = 1, ..., K = 8$, l'indice de l'étape dans laquelle se trouve le sachet de salade, le temps passé par le produit dans chaque étape k à la température T_k, est défini par $\triangle t_k = t_k - t_{k-1}$, $\triangle t_k \in R_+$. Une période de temps égale à zéro signifie que le produit ne passe pas par cette étape. Les variables aléatoires (v.a) $(\triangle t_k)_{k\geq 1}$ sont supposées indépendantes et chaque v.a. est distribuée selon une loi de probabilité G_k. De même, les v.a $(T_k)_{k\geq 1}$ sont supposées indépendantes et chaque v.a est distribuée selon une distribution H_k. Les paramètres propres aux modèles de croissance primaire et secondaire sont rassemblés dans le vecteur $\theta = \left\{T_{\min}, T_{opt}, T_{\max}, y_{\max}, \mu_{opt}, \sigma_y^{-2}, \sigma_{\mu_{\max}}^{-2}\right\}$. Les valeurs de ces paramètres sont indépendantes des étapes du circuit logistique. En effet, pour une simulation donnée, les paramètres de croissance des bactéries présentes dans un sachet contaminé sont constants d'une étape à l'autre. Ainsi, pour une simulation donnée, une unique valeur pour chaque paramètre est utilisée dans toutes les étapes du circuit logistique. La loi du vecteur θ ou encore la loi jointe *a posteriori* des paramètres composant ce vecteur a été établie dans le chapitre 3. Dans le cas de modèles hiérarchiques comme les modèles $M2$ et $M2_{\lambda_{obs}}$, la loi jointe *a posteriori* des hyperparamètres composant le vecteur $\psi = \{M_{T_{\min}}, S_{T_{\min}}^{-2}, ..., M_{\mu_{opt}}, S_{\mu_{opt}}^{-2}\}$ est utilisée. Cependant, pour simplifier la présentation, le vecteur des hyperparamètres ψ n'est pas mentionné dans cette partie.

Sachant l'étape k dans laquelle se trouve le sachet, la durée de cette étape $\triangle t_k$, la température T_k et le vecteur θ des paramètres de croissance des bactéries présentes dans le sachet, la concentration en log ufc/g à l'instant t_k est supposée distribuée selon une loi normale centrée autour de la concentration calculée avec les modèles primaire et secondaire présentés dans le chapitre 3

$$C(t_k) \sim N(f\left(T_k, \mu_{\max}, y_{\max}, \triangle t_k, C(t_{k-1})\right), \sigma_y^2). \tag{4.1}$$

Par convention, $t_0 = 0$ est choisi comme temps d'origine et $C_0 = C(t_0)$ est la contamination du sachet à la sortie de la chaîne de fabrication.

La fonction f du modèle logistique avec délai et rupture (Rosso (1995) et équation (3.1))

se réécrit avec les notations précédentes comme suivant

$$f\left(T_k, \mu_{\max}, y_{\max}, \triangle t_k, C(t_{k-1})\right) = y_{\max} - \log_{10}(1 + (10^{(y_{\max} - C(t_{k-1}))} - 1)\exp(-\mu_{\max}(T_k)(\triangle t_k))),$$

où la variable $\mu_{\max}$ est distribuée selon une loi normale

$$\sqrt{\mu_{\max}(T_k)} \sim N(\sqrt{g(T_k)}, \sigma^2_{\mu_{\max}}),$$

où la fonction g est définie par le modèle secondaire à valeurs cardinales (Rosso et al., 1993)

$$g(T_k) = \begin{cases} 0 & T_k \notin [T_{\min}, T_{\max}] \\ \frac{\mu_{opt}(T_k - T_{\max})(T_k - T_{\min})^2}{(T_{opt} - T_{\min})[(T_{opt} - T_{\min})(T_k - T_{opt}) - (T_{opt} - T_{\max})(T_{opt} + T_{\min} - 2T_k)]} & T_k \in [T_{\min}, T_{\max}]. \end{cases}$$

Le nombre de bactéries en log ufc/g s'étant développées au cours de la période $\triangle t_k$ se calcule par $C(t_k) - C(t_{k-1})$.

4.4 Nature et origine des données

4.4.1 Contamination de la matière première

Les données utilisées pour estimer la contamination des salades entrant dans l'usine sont les données portant sur la contamination des produits non transformés présentées dans le chapitre 2. Par souci d'exhaustivité, les données récemment recueillies par les centres techniques auprès des industries françaises de IV^{ème} gamme sont aussi intégrées pour estimer la contamination de la matière première. La limitation aux données de contamination par *L. monocytogenes* des salades non-transformées, françaises et des pays fournisseurs n'a pas pu être possible étant donné leur faible nombre et l'absence de concentration. La sélection a donc été élargie aux données de contamination de toutes les variétés de légumes et de germes non-transformés d'origine européenne et nord américaine (Etats-Unis et Canada). Au total 74 études de prévalence ont été utilisées. Ces prévalences varient entre 0% et 40%. La prévalence globale de la contamination, c'est à dire le nombre total de produits contaminés calculé à partir des 74 études, divisé par le nombre total de produits analysés est 2.2%. Pour quatre de ces études, les concentrations des produits contaminés ont été relevées. Les valeurs des concentrations fluctuent dans l'intervalle [-1.4 ; 4] log ufc/g. Quarante et une de ces études ont été menées après les années 2000. Seize d'entre elles (soit environ 22%) ont été réalisées

par les industriels partenaires du projet (Stahl et al., 2002, 2005).

4.4.2 Réduction en micro-organismes au cours du lavage à l'eau chlorée et du lavage à l'eau non chlorée

Les données concernant la réduction en micro-organismes au cours du lavage à l'eau non chlorée et du lavage à l'eau chlorée des salades, ont été extraites de 12 publications issues de la littérature scientifique et sont présentées dans le tableau 4.1.

Au total 103 observations ont été recueillies dont 60 sur la réduction en micro-organismes au cours du lavage à l'eau non chlorée et 43 sur la réduction en micro-organismes au cours du lavage à l'eau chlorée. Plusieurs critères de sélection, outre la pertinence des méthodes et l'absence d'ambiguïté dans la description des données, ont été respectés

1. la matrice alimentaire devait être les légumes feuilles (salades, romaine, laitue, scarole, etc...),
2. les micro-organismes sont *L. monocytogenes* (bactérie cible) ou la flore aérobie mésophile (flore installée, à laquelle appartiennent les *L. monocytogenes*). Une hypothèse ici est que *L. monocytogenes* ne présente aucune résistance particulière au lavage et que son comportement peut-être assimilé à celui de la flore aérobie mésophile,
3. l'eau de lavage est soit de l'eau n'ayant subi aucun traitement au sein de l'usine, soit de l'eau contenant de l'hypochlorite de sodium (donnant du HclO, la forme active), la forme de chlore la plus courante utilisée dans les entreprises. La quantité d'hypochlorite de sodium devait être supérieure ou égale à 0 et inférieure ou égale à 80 ppm. Les données sur le dioxyde de chlore (molécule différente) et autres molécules chlorées ont été exclues de l'analyse.

TAB. 4.1: Nombre d'observations de la réduction en micro-organismes et temps de contact par publication, type de micro-organisme et niveau de concentration en chlore

Publication	Micro-organisme	Nombre d'observations	Concentration en chlore (ppm)	Temps de contact (min)
Adams et al. (1989)	FAM	2	0	5, 10
		2	0.2	5
		1	10	5
		1	40	5
		1	80	5
Baur et al. (2005)	FAM	1	0	1
Beuchat et al. (2004)	LM	9	0	0.5
Burnett et al. (2004)	LM	20	0	5
Hellström et al. (2006)	FAM	2	0	1
	LM	1	0	1
Koseki et al. (2001)	FAM	4	0	1, 5, 10
Li et al. (2001)	FAM	1	0	1.5
		1	20	1.5
Li et al. (2002)	LM	1	0	1.5
		1	20	1.5
Park et al. (2001)	LM	4	45	1, 3
Pirovani et al. (2004)	FAM	1	26	8
		1	32	5
		1	41	5
		1	42	2
		1	53	8
		1	75	8
		2	76	2, 5
		1	77	5
		1	78	5
Szabo et al. (2003)	LM	3	0	1, 2, 5
		3	40	1, 2, 5
		3	80	1, 2, 5
Zhang & Farber (1996)	LM	16	0	1, 2, 5, 10
		8	25	1, 2, 5, 10
		8	50	1, 2, 5, 10

FAM : Flore Aérobie Mésophile
LM : *Listeria monocytogenes*

4.4.3 Durées et températures des différentes étapes du circuit logistique des salades en sachet

Aucune étude réalisée à ce jour n'a relevé les durées et les fluctuations des températures au cours des étapes du circuit logistique de la salade en sachet. Cependant des données provenant d'études variées peuvent être assimilées à celles-ci. Trois principales sources de données ont été utilisées pour reconstituer les couples durées-températures des huit étapes composant le circuit logistique de la salade en sachet :

1. une étude réalisée par l'Afssa avec la collaboration du Cemagref (Afchain, 2005; Afchain et al., 2005) qui a consisté à placer des capteurs thermiques dans des plaques de saumon fumé pour suivre l'évolution des températures au cours de leur circuit logistique. Les durées des différentes étapes ont également été déterminées. Cette étude a été effectuée dans le cadre d'un projet plus large (Afssa, 2006) incluant une analyse quantitative du risque lié à la présence de *Listeria monocytogenes* dans les plaques de saumon fumé (Pouillot, 2006; Pouillot et al., 2007). Etant donnée la similitude du saumon fumé et de la salade en sachet au niveau des recommandations de conservation (4°C), de l'emplacement des produits dans les meubles de vente et en considérant que la salade en sachet a la même réaction au changement de température que le saumon fumé, les distributions de probabilité des températures établies dans l'étude Afssa-Cemagref peuvent être assimilées à celles de la salade en sachet pour les étapes similaires,
2. des dires d'experts interrogés sur la durée et la température de chaque étape subie par les salades en sachet de la sortie de la chaine de fabrication jusqu'à la vente,
3. une enquête réalisée par l'Actia (Stahl et al., 2006) relevant les habitudes d'achat, de conservation et de consommation de 900 personnes.

Suivi thermique des plaques de saumon fumé au cours de la chaîne du froid, l'étude Afssa-Cemagref

Protocole Deux campagnes de mesure ont été menées afin de collecter les profils thermiques subis par les plaques de saumon au cours de leur circuit logistique et afin d'ajuster des lois de probabilité sur ces données. La première campagne a consisté à suivre des produits dès la fin de leur fabrication jusqu'à la plateforme de distribution ou dans quelques cas jusqu'au meuble de vente. La deuxième campagne a porté sur la partie suivante du circuit logistique, c'est à dire de la plateforme jusqu'à la consommation. Les températures ont été

enregistrées à l'aide de capteurs placés dans les produits. L'intervalle de temps entre deux enregistrements était constant pour tous les capteurs et toutes les étapes. Les agents de l'Afssa sont intervenus directement sur le terrain auprès des professionnels et des consommateurs. Un biais non-sécuritaire (c'est-à-dire minimisant les températures) sur les mesures est donc possible dû à un changement de comportement des individus soumis à une étude. A l'inverse un biais sécuritaire est possible puisque les températures des meubles de vente ont été relevées pendant les heures d'ouverture du magasin, heures auxquelles la température est la plus élevée. La chaîne du froid a été séquencée en huit étapes, une étape pouvant se retrouver plusieurs fois dans un même circuit logistique : transport frigorifique, entrepôt (temps entre deux chargement de camion), plateforme, chambre froide du magasin, meuble frigorifique de vente, transport après achat, réfrigérateurs domestiques et conservation domestique à température ambiante.

Modélisation statistique : modèle DMS (Afchain et al., 2005) Dans l'article de Afchain et al. (2005), il est montré que le profil durée-température d'un produit nécessaire à la modélisation de la croissance des bactéries peut être caractérisé par les trois variables suivantes : la durée de l'étape Δt, la température moyenne m et la variance des températures s^2. Des distributions de probabilité telles que la distribution exponentielle pour Δt, la distribution normale pour m et la distribution gamma pour s^2 ont été ajustées aux données récoltées pour chaque étape (cf. Afchain et al. (2005) pour les résultats complets). Les paramètres de ces distributions ont été obtenus par maximisation de la vraisemblance. Notons que pour quelques étapes, les températures relevées par une partie des capteurs, se sont avérées constantes ainsi $s = 0$. Cependant, le pourcentage de températures constantes étant en majorité faible, ce cas ne sera pas considéré ici. Les coefficients de corrélation (intra-étapes) des rangs de Spearman entre les variables $(\Delta t, m, s^2)$ d'une même étape ont été calculés par Pouillot (2006) et sont présentés dans le tableau 4.2. On remarque que les corrélations ne sont en général par très élevées à part entre la moyenne et l'écart-type des températures de l'entrepôt et du transport jusqu'au domicile du consommateur ainsi qu'entre la durée et l'écart-type de la température de la plateforme.

Procédure d'utilisation des distributions temps-températures dans l'étude Afssa-Cemagref Plusieurs profils de circuit logistique, chacun comprenant un nombre variable d'étapes ont été identifiés. Certains profils incluent jusqu'à quatre transports frigorifiques. Les proportions des différents profils de circuit logistique ont été calculées. La reproduction

TAB. 4.2: Corrélation des rangs de Spearman intra-étape calculée à partir de l'étude Afssa-Cemagref (Pouillot 2006)

Etapes	$Cor(\Delta t_k, m_k)$	$Cor(\Delta t_k, s_k)$	$Cor(m_k, s_k)$
Transport frigorifique	0.03	0.42	0.21
Entrepôt	0.15	0.40	0.63
Plateforme	0.30	0.67	0.10
Chambre froide	−0.22	0.40	−0.08
Meuble de vente	−0.16	0.13	0.36
Transport après achat	0.14	0.30	0.68
Conservation domestique à température ambiante	−0.12	−0.03	−0.30
Réfrigérateur domestique	0.30	−0.08	0.00

Δt_k : durée de l'étape k, m_k et s_k : moyenne et écart-type de la température de l'étape k

du circuit logistique des plaques de saumon fumé a été réalisée en utilisant des simulations de Monte-Carlo (Pouillot, 2006), c'est à dire en effectuant des tirages aléatoires dans les distributions respectives des trois variables $(\Delta t, m, s^2)$ ainsi que dans les profils. Les corrélations entre ces trois variables ont été reconstruites par la méthode d'Iman & Conover (1982). Cette procédure a été validée par comparaison graphique avec la procédure consistant à intégrer les températures observées sur tous les pas de temps pour lesquels les enregistrements ont été effectués (Afchain, 2005; Pouillot et al., 2007). Cette validation permet de montrer que la température moyenne et la variance suffisent à reconstituer les profils thermiques. Elle montre aussi qu'il n'est pas nécessaire de prendre en compte les corrélations (inter-étapes) entre les variables $(\Delta t, m, s^2)$ des différentes étapes, difficiles à intégrer à cause de la grande taille de la matrice des coefficients de corrélation des rangs de Spearman.

Dires d'experts

Les experts interrogés sont les responsables qualité des entreprises partenaires du projet réunis lors d'un des comités de pilotage. Leur expertise professionnelle et leurs connaissances des durées et des températures de stockage avant embarquement des produits par les transporteurs ainsi que celles des livraisons peuvent servir comme données ou comme validation de données existantes. Pour les étapes 1 à 5, il a été demandé aux différents experts de s'accorder sur la durée minimale, la plus probable et maximale de chacune des étapes ainsi que sur la température minimale, la plus probable et maximale.

Enquête ACTIA

Un questionnaire sur les habitudes d'achat, de conservation et de consommation des salades en sachets a été élaboré avec les membres du comité de pilotage ainsi que des personnes de la section analyse sensorielle de l'ACTIA (Stahl et al., 2006). Au total 900 personnes ont renseigné le questionnaire présenté en annexe. Certains résultats de cette enquête, utilisés dans ces travaux, sont donnés dans la section 4.5.4.

4.5 Distributions des paramètres du modèle

4.5.1 Distribution des paramètres du modèle de contamination de la matière première

Le modèle de mélange construit dans le chapitre 2 est appliqué aux données de contamination des produits végétaux non-transformés d'origine européenne et nord américaine. La loi jointe *a posteriori* des paramètres ainsi obtenue est utilisée afin de simuler des valeurs de concentration pour la matière première entrant dans l'usine. Les deux chaines convergent pour tous les paramètres après 5 000 itérations. Il est nécessaire d'effectuer 40 000 itérations supplémentaires afin que pour chaque paramètre soit respectée la régle sur l'erreur de Monte-Carlo (cf. 2.5.2). Des statistiques descriptives sont calculées pour chaque paramètre à partir des valeurs *a posteriori* ainsi obtenues et sont présentées dans le tableau 4.3. Rappelons que le paramètre β représente la proportion de matière première dont les concentrations sont supérieures à la limite de détection. La moyenne et l'écart-type des concentrations supérieures à la limite de détection sont données par les paramètres μ_1 et σ_1. Les paramètres μ_2 et σ_2 représentent respectivement la moyenne et l'écart-type des concentrations inférieures à la limite de détection et le paramètre σ_3 l'écart-type intra-étude.

4.5.2 Distribution de la réduction en micro-organismes au cours du lavage à l'eau chlorée et du lavage à l'eau non chlorée

La réduction en micro-organismes ne peut pas prendre de valeur négative. Ainsi les distributions testées pour modéliser la réduction sont défines sur $\mathbb{R}^+$. Une loi normale tronquée de moyenne 2.09 log ufc/g et d'écart-type 0.70 log ufc/g est utilisée pour modéliser la réduction en micro-organisme au cours du lavage à l'eau chlorée (Fig. 4.4).

TAB. 4.3: Statistiques descriptives des distributions *a posteriori* des paramètres de contamination par *L. monocytogenes* de la matière première entrant dans l'usine (unité : log ufc/g excepté β)

Paramètre	Moyenne	Ecart-type	$P_{0.025}$	$P_{0.50}$	$P_{0.975}$
β	0.022	0.002	0.019	0.022	0.026
μ_1	1.87	0.60	0.68	1.87	3.09
μ_2	-3.75	0.21	-4.21	-3.73	-3.40
σ_1	0.94	0.88	0.24	0.63	3.46
σ_2	0.96	0.17	0.68	0.94	1.35
σ_3	0.79	0.07	0.66	0.78	0.95

Cette hypothèse est validée par le test de Kolmogorov-Smirnov d'ajustement à une loi continue avec une p-valeur de 0.56. Le deuxième histogramme de la figure 4.4 montre que la distribution de la réduction après lavage à l'eau non chlorée serait tri-modale et qu'il serait nécessaire d'utiliser un mélange de lois. Or, selon les membres du comité de pilotage, l'utilisation d'une distribution tri-modale n'a aucune justification biologique et ce phénomène serait uniquement dû à un manque de données. Il a donc été décidé d'utiliser une loi unimodale. Le test de Kolmogorov-Smirnov rejette l'hypothèse selon laquelle la réduction en micro-organismes est distribuée selon une loi normale tronquée de moyenne 0.99 log ufc/g et d'écart-type 0.73 log ufc/g (p-valeur$<$ 0.01) mais valide celle d'une répartition selon une loi gamma de paramètre de forme 1.49 et de paramètre d'échelle 0.67 avec une p-valeur de 0.42. Ainsi, la loi $Gamma(1.49, 0.67)$ de moyenne 1 log ufc/g et d'écart-type 0.82 log ufc/g est utilisée pour simuler la réduction en *L. monocytogenes* au cours du lavage des salades à l'eau non chlorée. L'ajustement des distributions a été réalisé par maximum de vraisemblance à l'aide du logiciel R (Team, 2007).

4.5.3 Distribution de la quantité de salades contenues dans les sachets

Les sachets de salade vendus sur le marché ont des poids différents. Les chiffres de la production des sachets en fonction de la quantité de salades qu'ils contiennent, ont été fournis par les industriels partenaires du projet. Afin de prendre en compte les différentes tailles des sachets dans la simulation, une distribution discrète[1] (Vose, 2000) ayant comme paramètres les couples poids-proportions présentés dans la figure 4.5 a été utilisée.

[1] $X \sim Discrete(\{x_i\}, \{p_i\})$, $i = 1$ à $n : \mathbb{P}(X = x_i) = p_i$

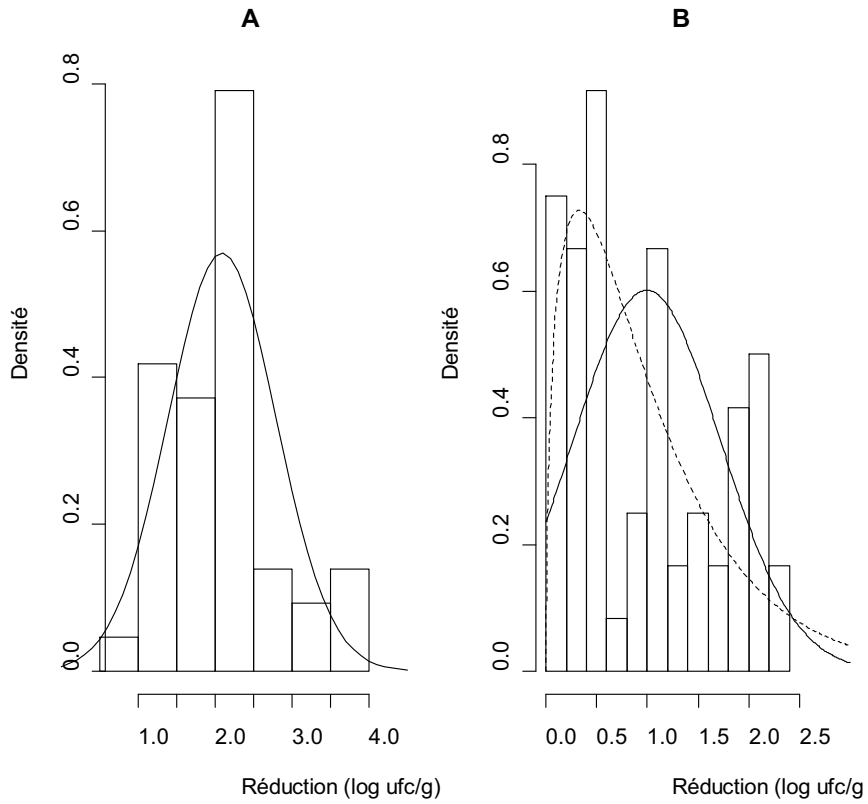

FIG. 4.4: Histogramme de la réduction en micro-organismes au cours du lavage à l'eau chlorée (A) et à l'eau non chlorée (B) et ajustement de distributions : A. Loi $Normale(2.09, 0.70)I(0, +\infty)$. B. Loi $Normale(0.99, 0.73)I(0, +\infty)$ (trait plein) et loi $Gamma(0.67, 1.49)$ (pointillé)

4.5.4 Distributions des durées et des températures de chaque étape du circuit logistique

Données mesurées ou dires d'experts ?

Utiliser les distributions établies à partir de données réellement observées plutôt que celles déterminées à partir des connaissances des professionnelles de la filière présente l'avantage de pouvoir inclure dans la modélisation des valeurs "accidentelles" ou "inhabituelles" c'est-à-dire des valeurs plus élevées ou plus basses que les valeurs réglementaires mais réellement observées lors d'accidents techniques. Ainsi, en comparant les températures moyennes déterminées par les experts (résultats non présentés) et celles estimées à partir de l'étude

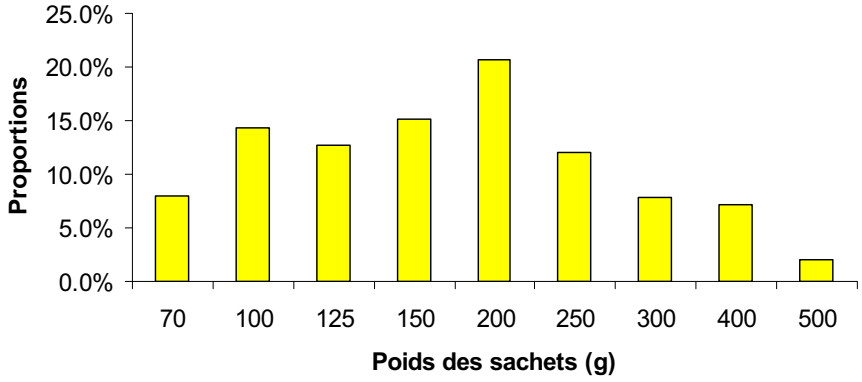

FIG. 4.5: Proportions des poids des sachets de salade vendus dans le commerce au détail

Afssa-Cemagref, il s'avère que la différence entre les deux est quasi nulle : une moyenne commune de 4°C pour les transports frigorifiques, ou une légère différence d'environ 1°C pour les températures des plateformes et des meubles de vente. Par contre, la dispersion des valeurs des températures est beaucoup plus grande dans l'étude Afssa-Cemagref. Par exemple, la valeur au $97.5^{ème}$ percentile des températures des transports frigorifiques obtenues par simulation à partir des distributions de m_2 et s_2^2 est de 10.8°C tandis que la valeur maximale déterminée par les experts est 8°C. De même, 2.5% des températures des plateformes sont supérieures à 9°C tandis que la valeur maximale déterminée par les experts est 6°C. En effet, un des avantages de l'étude Afssa-Cemagref est que les distributions ont été établies pour la moyenne m des températures et pour l'écart-type s des températures. On peut donc considérer ces distributions comme des distributions d'incertitude des estimations des paramètres de la distribution normale des températures. Ainsi la variabilité des températures des étapes et l'incertitude des estimations sont intégrées.

Par ailleurs, au niveau des durées des étapes de nombreuses différences sont constatées entre l'étude Afssa-Cemagref et les valeurs données par les experts. Un exemple est la durée des transports en camions frigorifiques qui est beaucoup plus courte, 1.8h en moyenne, que celle annoncée par les professionnels de la filière $IV^{ème}$ gamme qui est de 6h en moyenne. Cette différence s'explique par le fait que les transports sont plus nombreux dans le circuit logistique du saumon. De même, étant donné que la durée de conservation du saumon fumé est plus grande que celle de la salade en sachet, ces produits restent plus longtemps dans les meubles de vente et sont consommés moins rapidement ($>$ à 17 jours pour 2.5% des plaques

de saumon). Il semble donc plus adapté d'utiliser les distributions obtenues à partir des dires d'experts et de l'étude ACTIA pour les étapes des transports frigorifiques, de la plateforme, de la vente et de la conservation au domicile. La distribution triangulaire (Vose, 2000), généralement employée pour la facilité avec laquelle ses paramètres peuvent être déterminés (par exemple dans les travaux de Lindqvist et al., 2002; Shorten et al., 2006), a été utilisée pour traduire les dires d'experts sur les durées des étapes.

En conclusion, les températures des étapes du circuit logistique sont simulées à partir des distributions de la température moyenne m et de la variance des températures s^2, déterminées dans l'étude Afssa-Cemagref. Les distributions de ces paramètres sont décrites dans le tableau 4.4 pour chaque étape $k = 2, ..., 8$. Pour chaque étape k, la température T_k est considérée distribuée selon une loi normale de paramètres m_k et s_k^2. Les distributions des durées sont simulées à l'aide des distributions établies à partir des dires d'experts pour les étapes constituant le circuit logistique de la sortie de la chaîne de fabrication jusqu'à la vente (4.4).

TAB. 4.4: Distributions des durées et des températures des étapes du circuit logistique de la salade en sachet

Etapes du circuit logistique	Durées $\triangle t_k$ (h)	Température T_k ($^\circ$C)	
Chambre froide de l'usine k=1	Triang(0,12,48)	N(4,0.5)	
		Moyenne m_k	Variance s^2_k
Premier transport frigorifique k=2	Triang(3,6,12)	N(3.6,2.8^2)	G(2.2,0.7)
Plateforme k=3	Ber(0.8)×T(1,9,36)	N(2.9,2.9^2)	G(0.5,0.8)
Second transport frigorifique k=4	Ber(0.8)×T(3,6,12)	N(3.6,2.8^2)	G(2.2,0.7)
Meuble de vente k=5	Triang(4,84,233)	N(5.6,2.2^2)	G(1.0,0.9)
Trajet domicile k=6	Exp(1.10)	N(13,3.6^2)	G(0.6,0.1)
Conservation au réfrigérateur k=7,8	Ber(0.992)×[Triang(0,85,150,1)+Ber(0.567)×Triang(0,24,132)]	N(7.0,3.0^2)	G(1.4,0.6)
Conservation à température ambiante k=7	Ber(0.008)×U[0,24]	N(17.9,5.1^2)	G(0.6,0.1)

$X \sim Ber(p)$, distribution de Bernoulli de paramètre p. $X \sim N(m, s^2)$, distribution normale de moyenne m et d'écart-type s

$X \sim Triang(a, b, c)$, distribution triangulaire où a est la valeur minimale, b la valeur la plus probable et c la valeur maximale.

$X \sim G(a, b)$, distribution gamma de paramètre de forme a, de paramètre d'échelle b, d'espérance $E(X) = a \times b$ et de variance $V(X) = a \times b^2$.

$X \sim Exp(a)$, distribution exponentielle de paramètre $1/a$ et d'espérance $E(X) = a$.

Les statistiques des simulations de la durée et de la température sont présentées dans les tableaux 4.5 et 4.6. Chez le consommateur, les distributions des durées de conservation sont celles obtenues à partir de l'enquête menée par l'Actia. Chaque étape est détaillée ci-dessous et les distributions adoptées sont comparées avec celles de la bibliographie quand

elles existent.

TAB. 4.5: Durées des grandes étapes (livraison, vente, consommation) du circuit logistique de la salade en sachet. Les durées de chaque étape sont obtenues par tirage aléatoire dans les distributions correspondantes du tableau 1.4 (unité : h)

Etapes	Moyenne	Ecart-type	$P_{0.025}$	$P_{0.50}$	$P_{0.975}$
Sortie usine - Distributeur $k = 1$ à 4	44.0	14.3	16.6	44.1	71.0
Vente $k = 5$	72.8	34.8	18.2	68.5	144.0
Vente - Consommation $k = 6$ à 8	86.5	35.2	21.3	85.3	153.7

TAB. 4.6: Températures des étapes du circuit logistique de la salade en sachet. Les températures sont obtenues pour chaque étape par tirage aléatoire dans les distributions correspondantes du tableau 1.4 (unité : °C)

Etapes	Moyenne	Ecart-type	$P_{0.025}$	$P_{0.50}$	$P_{0.975}$
Chambre froide de l'usine $k = 1$	4.0	0.5	3.0	4.0	5.0
Transport frigorifique $k = 2$ ou $k = 4$	3.6	3.4	-2.3	3.6	10.6
Plateforme $k = 3$	3.1	2.9	-2.3	3.2	9.1
Meubles de vente $k = 5$	5.7	2.3	1.9	5.6	11.0
Trajet $k = 6$	13.2	3.5	6.1	13.2	19.6
Réfrigérateur $k = 7, 8$	7.1	3.4	0.2	7.1	13.4
Air ambiant $k = 7$	17.9	5.3	7.8	18.0	28.5

Stockage dans la chambre froide de l'usine

Les distributions utilisées pour la durée et la température de stockage des produits dans la chambre froide de l'usine lors de leur sortie de la chaîne de fabrication, sont celles déterminées à partir de dires d'experts. Le temps d'attente le plus courant avant chargement des camions est de 12h à une température moyenne de 4°C. Les distributions utilisées sont respectivement pour la durée (en h) et la température (en °C) la distribution $Triang(0, 12, 48)$ et la distribution $Normale(4, 0.5)$ (Tab. 4.4).

Stockage et transport de la sortie de la chaîne de fabrication jusqu'à la mise en rayon des sachets chez le distributeur

Les distributions des paramètres m et s nécessaires pour simuler les températures des étapes 2 et 3 sont présentées dans le tableau 4.4. La température moyenne des transports

frigoriques est 3.6°C avec un écart-type moyen de 4°C et la température moyenne de la plateforme est 2.9°C avec un écart-type d'environ 3°C. La distribution établie à partir des dires d'experts pour la durée (en h) des trajets entre l'usine et la plateforme et entre la plateforme et le point de vente est la distribution $Triang(3, 6, 12)$. La distribution déterminée pour le temps (en h) de stockage dans la plateforme est la distribution $Triang(1, 9, 36)$.

Chez le distributeur : la chambre froide et les meubles de vente

La température moyenne des meubles de vente est 5.6°C avec un écart-type moyen de 2.5°C (Tab. 4.4 et 4.6). Ces températures, établies dans l'étude Afssa-Cemagref, sont légèrement plus élevées que celles observées dans l'étude menée par la DGCCRF (Pierre, 1996) sur les températures des rayons « libre service » des grandes et moyennes surfaces et sur lesquelles a été ajustée une distribution $Normale(4, 2.17)$ utilisée par Nauta et al. (2003b).

La distribution du temps (en h) d'exposition des sachets de salade dans les meubles de vente est la distribution $Triang(4, 84, 233)$ établie à partir des dires d'experts. Cette durée comprend le temps de stockage dans la chambre froide de l'usine. Ainsi aucune distinction au niveau des températures n'est faite entre la chambre froide et les meubles de vente. Cela peut entraîner une légère surestimation des températures car selon l'étude Afssa-Cemagref les températures sont plus faibles de 1°C dans la chambre froide. Cependant le temps passé par les produits en chambre froide est très faible comparé à celui observé dans les meubles de vente. Ces deux étapes sont donc confondues dans la modélisation.

Trajet jusqu'au domicile du consommateur

La moyenne des températures du trajet jusqu'au domicile du consommateur, déterminée dans l'étude Afssa-Cemagref est 13°C avec un écart-type de 3.5°C (Tab. 4.4 et 4.6). Ces températures correspondent à peu près à celles recueillies auprès d'experts dans le cadre d'un projet européen (Evans, 1998; Nauta et al., 2003b) où était proposée une distribution $BetaPert(4, 10, 25)$[2] de moyenne 11.5°C et dont les paramètres sont la température minimale de 4°C, la température maximale de 25°C et la température la plus probable de 10°C. La distribution de la durée des trajets utilisée ici est la loi exponentielle de moyenne 1.10h issue

[2]La distribution $BetaPert(a, b, c)$ est utilisée, comme la distribution triangulaire, lorsque l'on connait la valeur minimale a, maximale c et la plus probable b, de la variable étudiée :
$BetaPert(a, b, c) = Beta(\alpha_1, \alpha_2) \times (c - a) + a$ où $Beta(\alpha_1, \alpha_2)$, Densité : $f(x) = \frac{x^{\alpha_1 - 1}(1-x)^{\alpha_2 - 1}}{\int_0^1 t^{\alpha_1 - 1}(1-t)^{\alpha_2 - 1} dt}$,
$\alpha_1 = \frac{(\mu - a) \times (2b - a - c)}{(b - \mu) \times (c - a)}$, $\alpha_2 = \frac{\alpha_1 \times (c - \mu)}{(\mu - a)}$, $\mu = \frac{a + 4 \times b + c}{6}$.

de l'étude Afssa. La moyenne des trajets est alors d'environ 1h, 16% des trajets ont une durée supérieure à 2h et 2.5% des trajets ont une durée supérieure à 4h. Notons qu'avec cette distribution le pourcentage de personnes ayant un trajet de plus de 2h est beaucoup plus élevé que celui obtenu dans l'enquête Actia (4%), ou celui issu d'une étude anglaise (Evans et al., 1991) (1%). Dans cette dernière, une distribution $Gamma(5.24, 8.17)$ a été ajustée aux données et ainsi le trajet moyen est de 42.8 minutes avec un écart-type de 18.7 minutes.

Conservation au domicile du consommateur

Lieu de conservation et température L'enquête Actia montre que 97.6% des personnes interrogées conservent leur sachet de salade au réfrigérateur, 1.6% dans un endroit frais (cave, balcon) et 0.8% ne prennent aucune précaution particulière et le laissent à température ambiante. Ainsi deux modes de conservation seront envisagés le premier étant la conservation au réfrigérateur dans 99.2% des cas, incluant la conservation dans un endroit frais et le second la conservation à température ambiante dans 0.8% des cas mais avec une durée plus courte que dans le cas précédent. Les distributions de la température des réfrigérateurs et de la température ambiante ont été déterminées dans l'étude Afssa-Cemagref (Tab. 4.4). Les simulations de ces températures produisent une moyenne de 7°C pour les réfrigérateurs et 17.9°C pour l'air ambiant et des écart-types respectifs de 3.4°C et 5.3°C (Tab. 4.6).

Plusieurs données européennes sur la température (en °C) des réfrigérateurs ont été rassemblées dans l'article de Nauta et al. (2003b) et une loi $Normale(6.64, 2.35)$ a été ajustée à ces données. En France, Laguerre et al. (2002) réalisent une étude en 1999 portant sur la température des réfrigérateurs de 119 ménages. Ils obtiennent des résultats similaires et ajustent une loi $Normale(6.6, 2.3)$. Les températures moyennes sont à peu près égales à celle de l'étude Afssa-Cemagref. Par contre, l'écart-type obtenu dans l'étude Afssa-Cemagref est plus grand, ainsi des températures plus élevées ont été observées. La différence de température entre les compartiments des réfrigérateurs et les pratiques des consommateurs ont également été étudiées par Laguerre et al. (2002) ainsi que par les anglais James & Evans (1992). Les deux enquêtes concluent qu'il n'existe pas de différence significative entre la température globale du réfrigérateur et celle du haut, du milieu ou du bas du réfrigérateur. Un tel degré de raffinement au niveau des températures des réfrigérateurs n'est donc pas pris en compte.

Durée de conservation L'enquête effectuée par l'Actia comporte également des questions sur la durée de conservation du produit avant consommation. D'après le questionnaire, 37% des personnes entament leur sachet au maximum 2 jours après son achat, 56% entre 3 et 5 jours et 7% plus de 5 jours après. La durée (en h) de conservation du produit avant ouverture pour une consommation totale ou partielle du sachet est donc modélisée par une distribution $Triang(0, 85, 150)$. La consommation du sachet en plusieurs prises alimentaires (deux actes de consommation au maximum) étant possible, des questions portaient sur cette pratique. Ainsi 42% des personnes interrogées répondent "oui souvent" et 49% des personnes "parfois" à la question portant sur la consommation en plusieurs prises alimentaires. Parmi ces personnes, 71% consomment le sachet entamé après 2 jours, 28% entre 3 et 5 jours et 0.7% après plus de 5 jours. En considérant que la réponse "parfois" signifie dans 30% des cas, 56.7% des sachets seront considérés consommés lors de deux actes de consommation séparés par une reconservation du produit. La distribution $Triang(0, 24, 132)$ est utilisée pour modéliser la durée (en h) de reconservation du produit entamé. N'ayant aucune information sur le mode de consommation (durée, nombre de consommation) des personnes conservant le sachet à température ambiante, une loi uniforme sur l'intervalle de temps [0, 24]h a été utilisée.

Date limite de consommation (DLC) La date limite de consommation est souvent dépassée par 4% des personnes interrogées, parfois par 53% et ne l'est jamais par 43%. Dans un but sécuritaire, la réponse "parfois" est considérée représenter 30% des fois, ainsi 20% des consommations pourront être effectuées jusqu'à 48h après la date limite de consommation.

4.6 Réalisation des simulations

Un nombre $n_v = 10\ 000 \times n_u = 1000$ de simulations sont effectuées suivant la procédure de Monte-Carlo de second ordre (cf. chapitre 1) décrite ci-dessous.

1. Une valeur d'incertitude pour chaque paramètre et hyperparamètre du modèle de mélange $(\mu_1, \mu_2, \sigma_1, \sigma_2, \sigma_3, \beta)$ est tirée parmi leur loi jointe *a posteriori*. Puis n_v concentrations en log ufc/g notées $(C_{MP,v})_{1\leq v\leq n_v}$ sont simulées à partir du mélange des deux lois normales ayant comme paramètres les valeurs tirées précédemment dans la loi jointe *a posteriori*.
2. De ces valeurs est retranchée la concentration en micro-organismes en log ufc/g éliminée au cours du lavage, $(R_{Cl,v})_{1\leq v\leq n_v}$ dans le cas d'eau chlorée ou $(R_{NCl,v})_{1\leq v\leq n_v}$ dans le cas

d'eau non chlorée. Des valeurs de concentration de la matière première après lavage, notées $(C_{L,v})_{1\leq v\leq n_v}$sont ainsi obtenues.

3. Puis n_v poids $(W_v)_{1\leq v\leq n_v}$ sont tirés dans la distribution de la quantité de salades contenues dans les sachets vendus sur le marché. Ainsi n_v concentrations de *L. monocytogenes* en log ufc par sachet, $(C_{S,v})_{1\leq v\leq n_v}$, sont calculées en croisant les concentrations obtenues à l'étape 2, $(C_{L,v})_{1\leq v\leq n_v}$ et les poids $(W_v)_{1\leq v\leq n_v}$. Le pourcentage p_c, de sachets contenant des salades contaminées parmi les n_v simulés est calculé : $p_c = \frac{1}{n_v}\sum_{v=1}^{n_v} 1_{C_{s,v}\geq 0}$.
4. Pour les $n_c = n_v \times p_c$ sachets contenant des salades contaminées, la croissance de *L. monocytogenes* est simulée comme suit :
   - pour la première étape, n_c durées et températures sont tirées aléatoirement dans leur distribution respective.
   - n_u valeurs des variables $\triangle t_k$, m_k et s_k sont tirées aléatoirement de leur distribution respective pour chaque étape $k = 2, ..., 8$. La corrélation entre ces trois varaibles est reconstruite par la méthode d'Iman & Conover (1982) en utilisant les coefficients des rangs de Spearman observés dans l'étude Afssa-Cemagref (Tab. 4.2). De nouvelles valeurs de $\triangle t_k$, m_k et s_k avec $k = 2, ..., 8$, respectant les corrélations observées sont obtenues. Cette étape est effectuée une seule fois pour l'ensemble des simulations.
   - Puis, pour chaque étape $k = 2, ..., 8$, une valeur d'incertitude est tirée parmi les nouvelles valeurs de s_k et m_k. Ensuite n_c valeurs de températures sont obtenues à partir de la loi normale ayant comme paramètre les valeurs de m_k et s_k tirées précédemment, ainsi que n_c valeurs de durées tirées parmi les nouvelles valeurs de $\triangle t_k$. Pour les étapes $k = 3$ et $k = 4$ un tirage à partir d'une distribution de Bernoulli de paramètre $p = 0.8$ est réalisé afin que dans 20% des simulations la durée de ces étapes soit nulle.
   - Les simulations qui correspondent à des scénarios aberrants sont éliminées. Ces scénarios sont :
     - la DLC est dépassée avant la vente
     - la consommation s'effectue deux jours ou plus après la DLC
     - plus de 20% des produits sont consommés après la DLC.
   - Pour les modèles hiérarchiques $M2$ et $M2_{\lambda_{obs}}$ de simulation de la croissance, une valeur d'incertitude pour chaque hyperparamètre du vecteur ψ est tirée aléatoirement dans les réalisations de leur loi jointe *a posteriori*. n_c valeurs de variabilité sont ensuite tirées dans les distributions des paramètres composant le vecteur θ. Les valeurs des

paramètres de ces distributions sont les valeurs tirées précédemment.
- Pour chaque étape $k = 1, ...8$, n_c valeurs de concentration sont calculées à partir des équations (3.1) et (3.2) et de la distribution définie en (4.1) tronquée à 8.5 log ufc/g.

5. Des statistiques des concentrations à chaque étape ainsi que le nombre de sachets contaminés, sont calculés à partir des n_c valeurs.
6. Les étapes 1 à 5 exceptée l'étape de reconstruction des corrélations, sont effectuées $n_u = 1\ 000$ fois. Les $50^{ème}$, $2.5^{ème}$, et $97.5^{ème}$ percentiles des n_u statistiques obtenues sont calculés afin d'obtenir un estimateur de chaque statistique ainsi qu'un intervalle d'incertitude (UI).

La programmation a été réalisée avec le logiciel R (Team, 2007). Les nombres $n_v = 10\ 000$ et $n_u = 1\ 000$ ont été choisis par un compromis entre le nombre de simulations réalisables sous le logiciel R et le nombre de simulations nécessaires à la stabilité des résultats. Le nombre de simulations réalisées permet de calculer les percentiles désirés sans approximation linéaire, c'est-à-dire que les valeurs de ces percentiles sont des valeurs simulées et non extrapolées linéairement. De plus, le pourcentage de sachets contaminés étant faible, il est nécessaire d'effectuer un grand nombre de simulations pour pouvoir calculer des statistiques sur un nombre de sachets contaminés conséquent. Le nombre de simulations de la variabilité a donc été privilégié par rapport à celui de l'incertitude, $n_v = n_u = 10\ 000$ étant trop grand.

4.7 Résultats

4.7.1 Contamination de la matière première entrant dans l'usine

Les statistiques descriptives de la distribution prédictive de la contamination des salades entrant dans l'usine, obtenue avec le modèle de mélange développé dans le chapitre 2 sont présentées dans le tableau 4.7. Pour reprendre l'unité employée dans le chapitre 2, les résultats de ce paragraphe sont exprimés en logarithme du nombre de *L. monocytogenes* viables par gramme (log vlm/g). La contamination moyenne de la matière première est -3.62 log vlm/g avec un intervalle d'incertitude (UI) à 95% de [-4.08 ; -3.29] log vlm/g. En d'autres termes, la contamination moyenne varie entre 1 et 5 *L. monocytogenes* dans 10 kg de matière première. La contamination médiane est similaire à la contamination moyenne. Le $97.5^{ème}$ percentile de la contamination de la matière première par *L. monocytogenes* est -0.40 log vlm/g avec un UI à 95% de [-1.07 ; 0.42] log vlm/g. Cela signifie que dans 2.5% des cas, la concentration

de la matière première est supérieure à une valeur située entre 1 *L. monocytogenes* dans 10g et environ 3 *L. monocytogenes* dans 1g.

TAB. 4.7: Contamination par *L. monocytogenes* de la matière première entrant dans l'usine (unité : log ufc/g)

C_{MP}	Moyenne	Ecart-type	$P_{0.025}$	$P_{0.5}$	$P_{0.975}$
Estimation	-3.62	1.48	-6.15	-3.71	-0.40
UI_{inf}	-4.08	1.27	-7.10	-4.16	-1.07
UI_{sup}	-3.29	1.78	-5.49	-3.38	0.42

UI_{inf} et UI_{sup} borne inférieure et borne supérieure de l'intervalle d'incertitude à 95%

4.7.2 Contamination de la matière première après lavage à l'eau chlorée et à l'eau non chlorée

La concentration moyenne de la matière première après son passage dans l'eau de lavage contenant du chlore est -5.74 log ufc/g avec un UI à 95% de [-6.21 ; -5.41] log ufc/g (Tab. 4.8). La contamination de la matière première dans le cas où celle-ci serait lavée uniquement à l'eau non chlorée est plus élevée, avec une concentration moyenne de -4.60 log ufc/g et un UI à 95% de [-5.04 ; -4.29] log ufc/g. Notons que les intervalles d'incertitude de la concentration moyenne obtenues après passage dans l'eau chlorée et dans l'eau non chlorée ne se chevauchent pas. Les $2.5^{ème}$ et $97.5^{ème}$ percentiles de la contamination de la matière première se situent à 95% respectivement dans les intervalles de concentration [-9.58 ; -8.00] log ufc/g et [-2.77 ; -1.62] log ufc/g après lavage à l'eau chlorée et [-8.65 ; -7.25] log ufc/g et [-1.67 ; -0.55] log ufc/g après lavage à l'eau non chlorée.

TAB. 4.8: Contamination par *L. monocytogenes* de la matière première après lavage à l'eau chlorée et à l'eau non chlorée (unité : log ufc/g)

Eau de lavage	C_L	Moyenne	Ecart-type	$P_{0.025}$	$P_{0.5}$	$P_{0.975}$
Chlorée	Estimation	-5.74	1.62	-8.61	-5.82	-2.23
	UI_{inf}	-6.21	1.43	-9.58	-6.29	-2.77
	UI_{sup}	-5.41	1.90	-8.00	-5.48	-1.62
Non chlorée	Estimation	-4.60	1.69	-7.78	-4.62	-1.14
	UI_{inf}	-5.04	1.50	-8.65	-5.10	-1.67
	UI_{sup}	-4.29	1.97	-7.25	-4.30	-0.55

UI_{inf} et UI_{sup} borne inférieure et borne supérieure de l'intervalle d'incertitude à 95%

TAB. 4.9: Contamination par *L. monocytogenes* des salades des sachets contaminés après lavage à l'eau chlorée et à l'eau non chlorée (unité : log ufc/g)

Eau de lavage	$C(t_0)$	Moyenne	Ecart-type	$P_{0.025}$	$P_{0.5}$	$P_{0.975}$
Chlorée	Estimation	-0.49	1.18	-2.34	-0.54	1.90
	UI_{inf}	-0.98	0.96	-2.46	-1.27	1.11
	UI_{sup}	0.67	2.17	-2.15	0.52	5.38
Non chlorée	Estimation	-0.82	1.51	-2.45	-1.48	2.59
	UI_{inf}	-1.17	1.18	-2.52	-1.68	1.81
	UI_{sup}	-0.26	2.21	-2.38	-0.81	5.44

UI_{inf} et UI_{sup} borne inférieure et borne supérieure de l'intervalle d'incertitude à 95%

4.7.3 Contamination à la sortie de la chaîne de fabrication

Le pourcentage de sachets contenant des salades contaminées après lavage de la matière première à l'eau chlorée est 2.55% avec un UI à 95% de [1.75 ; 3.42]%. Dans le cas où l'eau de lavage ne contiendrait pas de chlore, ce pourcentage augmenterait et serait de 6.1% avec un UI à 95% de [4.21 ; 8.92]%. Après lavage à l'eau chlorée, les concentrations des salades contenues dans les sachets contaminés sont en moyenne de -0.49 log ufc/g avec un UI à 95% de [-0.98 ; 0.67] log ufc/g (Tab. 4.9). La contamination au $97.5^{ème}$ percentile est de 1.90 log ufc/g avec un UI à 95% de [1.11 ; 5.38] log ufc/g. Après lavage à l'eau non chlorée, les concentrations des salades contenues dans les sachets contaminés sont en moyenne de -0.82 log ufc/g avec un UI à 95% de [-1.17 ; -0.26] log ufc/g. La contamination au $97.5^{ème}$ percentile est de 2.59 log ufc/g avec un UI à 95% de [1.81 ; 5.44] log ufc/g.

Remarquons que la moyenne, les valeurs aux faibles percentiles et la médiane des concentrations après lavage à l'eau non chlorée sont légèrement inférieures à celles obtenues après lavage à l'eau chlorée. Ce phénomène peut être expliqué par l'étude de la figure 4.6 qui représente un exemple des distributions de la contamination de l'ensemble des sachets selon le type de lavage. La distribution de la contamination des sachets dont la matière première a été lavée sans chlore se situe à droite de la distribution de l'ensemble des sachets lavés à l'eau chlorée. Cependant, lorsque l'on s'intéresse uniquement à la distribution des sachets contaminés, la droite verticale marquant la présence d'au moins une bactérie se situe plus à gauche de la distribution des sachets dont les salades ont été lavées à l'eau non chlorée. Ainsi de plus faibles valeurs de concentration sont prises en compte ce qui produit par exemple une valeur médiane de contamination plus faible que celle obtenue pour les sachets dont les salades ont été lavées à l'eau chlorée. Toutefois on remarque qu'il y a aussi plus de valeurs

élevées qu'avec un lavage à l'eau chlorée ce qui entraîne des valeurs plus élevées aux forts percentiles. Ainsi après lavage à l'eau chlorée, le pourcentage de sachets simulés contenant des salades dont la concentration en *L. monocytogenes* est supérieure à 1 log ufc/g et à 2 log ufc/g est respectivement 0.28% avec un UI à 95% de [0.08 ; 0.89]% et 0.05% avec un UI à 95% de [0.00 ; 0.58]%. Dans le cas où le lavage s'effectue avec de l'eau non chlorée, le pourcentage de sachet simulés contenant des salades dont la concentration en *L. monocytogenes* est supérieure à 1 log ufc/g et à 2 log ufc/g serait respectivement 1.01% avec un UI à 95% de [0.51 ; 1.61]% et 0.38% avec un UI à 95% de [0.12 ; 1.02]%.

4.7.4 Contamination à chaque étape du circuit logistique

Les statistiques des concentrations des salades contenues dans les sachets contaminés calculées à chaque étape de la sortie de la chaîne de fabrication à la consommation sont présentées dans le tableau 4.10. Les simulations ont été effectuées en utilisant le modèle $M2$ et en considérant que la matière première a été lavée à l'eau chlorée. La distribution utilisée en entrée du modèle de simulation est donc celle décrite dans les trois premières lignes du tableau 4.9. Entre la sortie de la chaîne de fabrication des sachets de salade et l'arrivée chez le distributeur, le temps moyen est 44h (Tab. 4.5), la croissance moyenne 0.6 log ufc/g et la croissance au $97.5^{ème}$ percentile 1.3 log ufc/g. Pendant la vente qui dure en moyenne 72.8h à une température moyenne de 5.7°C (Tab. 4.6), la croissance est en moyenne de 1.29 log ufc/g. Au niveau du $97.5^{ème}$ percentile, la croissance pendant la vente est de 2.64 log ufc/g. Même si les températures durant le trajet jusqu'au domicile du consommateur sont élevées avec une moyenne de 13.2°C et une valeur au $97.5^{ème}$ percentile de 19.6°C, sa durée étant très courte (1h en moyenne), la croissance pendant cette étape est quasi nulle. Le temps moyen entre l'achat et la consommation entière du sachet, c'est-à-dire en incluant la deuxième consommation, est de 86.5h. La croissance durant cette période est de 1.58 log ufc/g en moyenne.

Les résultats du tableau 4.10 peuvent être présentés par rapport à l'ensemble des produits simulés et non plus par rapport aux produits contaminés. Par exemple à l'instant de la première consommation, le pourcentage de produits simulés dépassant 1 log ufc/g est de 1.97% avec un UI à 95% de [1.26 ; 2.62]%. De même, les pourcentages de sachets simulés contenant des salades ayant des concentrations supérieures à 2 log ufc/g et 3 log ufc/g sont respectivement 1.63% avec un UI à 95% de [1.01 ; 2.25]% et 1.27% avec un UI à 95% de [0.72 ; 1.89]%.

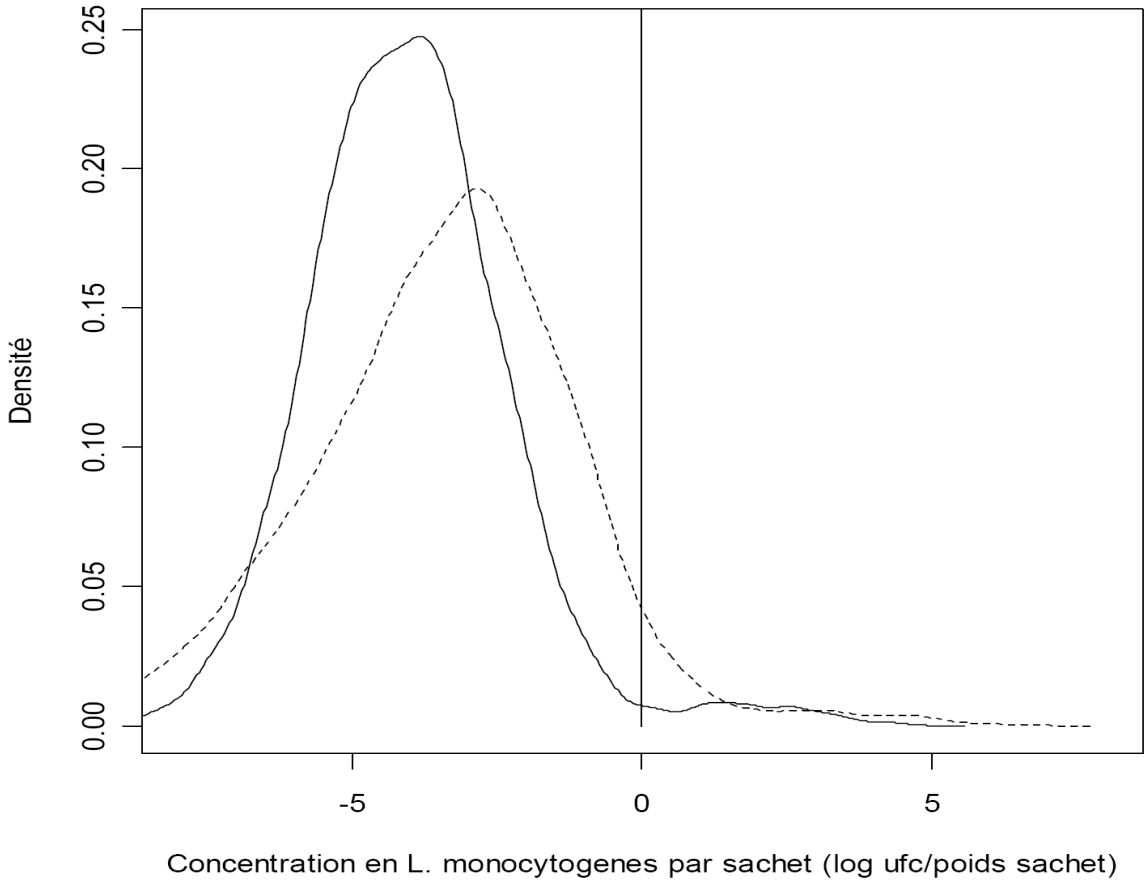

FIG. 4.6: Distributions de la concentration des salades contenues dans les sachets simulés selon la teneur en chlore de l'eau de lavage : chlorée (trait plein), non chlorée (pointillés). Les distributions sont tracées pour une valeur d'incertitude fixée. La droite verticale délimite les sachets contenant au moins une bactérie de ceux non contaminés.

Le tableau 4.11 présente les concentrations à l'instant de la première consommation des salades contenues dans les sachets contaminés et ayant été lavées à l'eau chlorée ou à l'eau non chlorée. Les simulations ont été effectuées pour chacune des eaux de lavage avec les modèles de croissance $M2$ et $M2_{\lambda_{obs}}$. Comme cela avait été observé dans le chapitre 3, la croissance est légèrement plus rapide avec le modèle $M2_{\lambda_{obs}}$. La concentration à l'instant de la première consommation calculée avec le modèle $M2_{\lambda_{obs}}$ est supérieure d'environ 0.56 log ufc/g pour la moyenne et d'environ 0.25 log ufc/g pour le $97.5^{ème}$ percentile par rapport à celle obtenue avec le modèle $M2$. L'eau de lavage employée (chlorée ou non chlorée) n'a pas d'impact sur les niveaux de concentration. Par exemple, les concentrations obtenues avec le modèle $M2$ sont à peu près égales : 2.61 log ufc/g en moyenne pour l'eau chlorée contre 2.65 log ufc/g pour l'eau non chlorée. De même, au $97.5^{ème}$ percentile les concentrations sont respectivement de 6.74 log ufc/g et 7.01 log ufc/g. Par contre, comme cela a déjà été précisé précédemment le nombre de sachets contaminés est plus élevé lorsque l'eau de lavage ne contient pas de chlore.

4.7.5 Test de vieillissement

Le test de vieillissement est un test réalisé par les fabricants afin de déterminer et valider la date limite de consommation (DLC) des produits. La DLC doit être déterminée pour qu'à cette date le produit commercialisé soit conforme aux critères microbiologiques réglementaires. Le test de vieillissement consiste à effectuer des analyses microbiologiques après conservation des sachets de salade 3 jours à 4°C et 6 jours à 8°C. Nous avons simulé la croissance de *L. monocytogenes* dans les conditions du test de vieillissement. Les résultats obtenus avec le modèle $M2$ et lorsque la matière première a été lavée à l'eau chlorée (concentrations des trois premières lignes du tableau 4.9), sont présentés dans le tableau 4.12. La concentration moyenne est de 3.21 log ufc/g avec un UI à 95% de [2.33 ; 4.2] log ufc/g. La concentration au $97.5^{ème}$ percentile est égale à 7.12 log ufc/g avec un UI à 95% de [6.54 ; 7.68] log ufc/g. Le test de vieillissement produit des résultats similaires (légèrement supérieurs) à ceux obtenus par tirage aléatoire des températures et des durées dans leur distribution respective (Tab. 4.1). Ainsi, en considérant que notre modèle est valide, le test de vieillissement, 3 jours à 4°C et 6 jours à 8°C, est un outil adapté pour prévoir les concentrations des produits à l'instant de la consommation.

4.7.6 Scénarios pour le paramètre y_{max}

Dans le chapitre 3, la distribution du paramètre $y_{\max}$ a été déterminée à partir du traitement de 206 courbes de croissance. Nous avons mentionné que ce choix pouvait entraîner une erreur puisque la phase stationnaire n'a pas été atteinte pour l'ensemble des courbes. Un ajustement sur les concentrations maximales atteintes a été effectué en utilisant uniquement les 59 courbes pour lesquelles la phase stationnaire avait été clairement identifiée. Ainsi la croissance de *L. monocytogenes* a été de nouveau simulée en considérant que les $y_{\max}$ étaient distribués selon la loi $Normale(6.64; 0.81^2)I(5, 8.5)$. Les calculs présentés dans le tableau 4.13 ont été effectués avec le modèle $M2$ et à partir de la matière première ayant été lavée à l'eau chlorée (concentrations dans les trois premières lignes du tableau 4.9). Les concentrations moyennes et aux différents percentiles sont équivalentes à celles obtenues avec la loi *a posteriori* de $y_{\max}$ estimée à partir de l'ensemble des courbes. Si l'on compare la distribution de $y_{\max}$ ajustée uniquement à partir des courbes pour lesquelles la phase stationnaire a été atteinte avec la distribution *a posteriori* obtenue pour le modèle $M2$ avec l'ensemble des courbes (Tab. 3.7, la moyenne est 6.03 log ufc/g et l'écart-type 1.05 log ufc/g), celles-ci ne sont pas très différentes. Ce résultat montre qu'il n'est pas nécessaire dans notre cas d'utiliser pour $y_{\max}$ une distribution ajustée uniquement à partir des courbes présentant une phase stationnaire. Précisons aussi qu'il est préférable, afin de conserver les corrélations pouvant exister entre les paramètres et hyperparamètres, d'utiliser la loi jointe *a posteriori* obtenue dans le chapitre 3.

L'article de Gnanou Besse et al. (2006) montre qu'il existe un effet significatif de la taille de l'inoculum y_0 sur la concentration maximale atteinte $y_{\max}$, c'est-à-dire que jusqu'à une certaine valeur, plus la concentration initiale est élevée, plus la taille de la population bactérienne atteinte à la phase stationnaire est élevée. L'expérience a été menée sur *L. monocytogenes* en milieu de culture et dans du saumon fumé à l'air ambiant ou sous vide. Il a aussi été précisé que l'effet de y_0 sur $y_{\max}$ interagit avec l'aliment, le milieu de culture et l'emballage. Nous n'avons pas pris en compte dans les calculs réalisés précédemment l'effet de la concentration initiale sur le niveau de concentration atteint à la phase sationaire car les courbes de croissance de *L. monocytogenes* dans les salades feuilles présentaient uniquement des tailles d'inoculum élevées : entre 2 et 5.57 log ufc/g avec une moyenne de 4 log ufc/g. Une seule courbe de croissance avait une taille d'inoculum plus faible (1 log ufc/g). Il ne nous a donc pas été possible d'estimer la relation pouvant exister entre y_0 et $y_{\max}$ dans la salade. Toutefois, nous avons quand même voulu étudier l'effet que cette relation pourrait avoir sur la concentration des salades à l'instant de la consommation. Plusieurs scénarios ont

été testés. Le premier scénario est de considérer que $y_{\max} = y_0 + 6$ log ufc/g. Cette relation est établie à partir du graphique de la publication de Gnanou Besse et al. (2006) pour le saumon fumé de type A. Deux autres scénarios ont été testés : $y_{\max} = y_0 + 4$ log ufc/g et $y_{\max} = y_0 + 2$ log ufc/g. Les niveaux de contamination à la première consommation selon les différents scénarios testés sur $y_{\max}$ sont présentés dans le tableau 4.13. Les calculs ont été effectués avec le modèle $M2$ et à partir de la matière première ayant été lavée à l'eau chlorée (concentrations dans les trois premières lignes du tableau 4.9). Le premier scénario pour $y_{\max}$ n'a pas d'impact sur le niveau de concentration à l'instant de la consommation. Pour préciser, la moyenne obtenue avec le scénario 1 est 2.54 log ufc/g et les valeurs aux $2.5^{ème}$ et $97.5^{ème}$ percentiles sont respectivement -1.67 log ufc/g et 6.85 log ufc/g contre 2.65 log ufc/g en moyenne et -1.65 log ufc/g pour le $2.5^{ème}$ percentile et 6.74 log ufc/g pour le $97.5^{ème}$ percentile en utilisant la distribution *a posteriori* de $y_{\max}$ obtenue pour le modèle $M2$, à partir de l'ensemble des courbes dans le chapitre 3 (Tab. 4.1). Avec le scénario 2 ($y_{\max} = y_0 + 4$ log ufc/g), la différence de concentration par rapport à celle obtenue avec la distribution *a posteriori* de $y_{\max}$ est de -0.52 pour la moyenne et 0 log ufc/g et -1.34 log ufc/g pour le $2.5^{ème}$ et $97.5^{ème}$ percentile. Ce scénario a donc peu d'effet sur les concentrations faibles ou moyennes, par contre au $97.5^{ème}$ percentile, la différence avec les concentrations obtenues pour ce percentile avec la distribution *a posteriori* de $y_{\max}$ est importante. L'impact sur le niveau de contamination est plus important avec le scénario 3, ($y_{\max} = y_0 + 2$ log ufc/g), pour lequel la concentration moyenne est de 1.09 log ufc/g et le $97.5^{ème}$ percentile de 3.82 log ufc/g. Par contre, la concentration au $2.5^{ème}$ percentile est égale à celle obtenue avec la loi *a posteriori* de $y_{\max}$. L'ajustement de la valeur de $y_{\max}$ en fonction de y_0 a un impact important sur la moyenne et les percentiles élevés des concentrations à l'instant de la consommation lorsque l'écart entre ces deux paramètres est de 4 log ufc/g ou de 2 log ufc/g.

4.8 Discussion

Ce chapitre présente le modèle de simulation de la contamination par *L. monocytogenes* des salades contenues dans les sachets de leur fabrication jusqu'à leur consommation. Les résultats attendus sont le pourcentage de sachets contenant des salades contaminées et leur concentration à chaque étape du modèle. Après une description du procédé de fabrication et du circuit logistique suivi par les salades en sachets de leur sortie de la chaîne de fabrication jusqu'à leur consommation, les étapes composant le modèle ont été définies. Ce chapitre intègre les résultats méthodologiques (élaboration d'un modèle de mélange de deux lois

normales afin d'estimer la distribution des contaminations) ou numériques (distributions *a posteriori* des paramètres de croissance) obtenus dans les deux chapitres précédents. Une méthode d'estimation de la réduction en micro-organismes a été développée afin d'évaluer la concentration en *L. monocytogenes* de la matière première après lavage. Les résultats numériques des taux et niveaux de contamination sont intéressants cependant il a semblé pertinent de les compléter par des scénarios concernant le modèle de simulation ou le procédé de fabrication. La simulation de la croissance a été réalisée en utilisant les deux modèles produisant le meilleur ajustement des données parmi les trois présentés dans le chapitre 3. Les résultats obtenus à partir de ces modèles de croissance ont été comparés afin d'évaluer l'impact du choix du modèle sur la concentration des produits lors de leur consommation. Un test de vieillissement qui consiste à conserver les produits pendant 3 jours à 4°C et 6 jours à 8°C a été simulé. Ainsi les concentrations obtenues en utilisant des distributions ajustées sur les valeurs observées des températures et des durées de chaque étape ont pu être comparées avec les concentrations obtenues lors d'un test pratiqué à des températures fixées par les industriels. Au niveau du procédé de fabrication, un scénario correspondant à la suppression du chlore du bain de lavage, pratique récemment mise en place dans certaines usines européennes, a été testé. Différents scénarios sur la valeur du paramètre $y_{\max}$ ont aussi été mis en place. Les données utilisées dans ce travail sont des valeurs numériques ou des distributions extraites principalement d'articles ou de rapports scientifiques, ou encore des dires d'experts à partir desquels des distributions de probabilité ont été élaborées.

4.8.1 Difficultés de la reproduction d'un procédé de fabrication et d'un circuit logistique

Une des difficulté de ce travail est la reproduction des étapes constituant la fabrication et le circuit logistique des salades de IV^{ème} gamme. Nous nous sommes efforcés de les reproduire avec précision en effectuant un travail considérable de tri et de recueil de données d'origines diverses. Cependant il faut souligner que même si la compilation de données de nature et de source différentes est couramment utilisée dans l'analyse des risques microbiologiques, il est préférable afin d'intégrer les dépendances entre les différents facteurs (par exemple les températures et les durées de conservation), de disposer de données ciblées sur un produit et issues d'une même étude effectuée sur l'ensemble du procédé de fabrication et de conservation du produit, comme par exemple l'étude Afssa-Cemagref pour le saumon fumé. Le manque de données numériques pour certains paramètres nous a contraint à utili-

ser des dires d'experts et certaines améliorations pourraient être effectuées, notamment au niveau des distributions utilisées pour traduire ces dires d'experts. Par exemple, pour simuler les durées de certaines étapes, des distributions triangulaires ont été établies à partir des dires d'experts ou de l'enquête de consommation Actia. Ce choix a été effectué d'un point de vue pratique. Les questions aux experts n'ont pas été posées dans un cadre formel d'élicitation (Morris, 1977; Kadane & Wolfson, 1998; O'Hagan & Oakley, 2007) mais à l'occasion d'un comité de pilotage. Ainsi les questions étaient simples et portaient principalement sur la durée maximale, minimale et la plus probable. Avec ces informations, il n'était pas possible par exemple, de déterminer les paramètres de la loi exponentielle qui est généralement utilisée pour modéliser la durée d'un événement (Vose, 2000). Par contre, à partir des réponses données par les experts les paramètres de la distribution triangulaire étaient facilement identifiables. Un autre avantage de la distribution triangulaire est qu'il est possible de gérer ses bornes, contrairement à la loi exponentielle. Toutefois, ce type de distribution implique que les bornes inférieures et supérieures utilisées soient des minimums et maximums absolus, il n'est pas possible d'obtenir des valeurs inférieures ou supérieures à ces bornes ce qui ne semble pas réaliste. Il aurait été pertinent, si l'information avait été disponible, d'utiliser la distribution triangulaire proposée par Vose (2000) qui permet d'intégrer un pourcentage de valeurs au-dessus et en dessous des bornes. Une autre distribution ayant les mêmes facilités d'utilisation que la triangulaire et souvent employée dans l'analyse des risques microbiologiques (particulièrement par les utilisateurs du logiciel @risk) pour modéliser les dires d'experts est la distribution BetaPert. Cette distribution est plus sensible que la loi triangulaire à la valeur la plus probable et contrairement à cette dernière ne souffre pas de biais pour l'estimation de la moyenne lorsque la valeur de la borne supérieure est grande. Il est donc possible d'améliorer le choix des distributions concernant principalement les durées des étapes soit par la mise en place d'une véritable procédure d'élicitation soit par l'utilisation de lois plus souples.

Nous nous sommes attachés à utiliser des distributions de probabilité pour modéliser les différents paramètres et hyperparamètres d'entrée du modèle de simulation afin de tenir compte des différentes sources de variabilité et d'incertitude. Cependant, au niveau des températures et de la durée des étapes du circuit logistique, la similitude des résultats obtenus en utilisant des températures fixées (test de vieillissement) ou variables laisseraient à penser qu'intégrer leur variabilité et leur incertitude n'est pas indispensable. Mais ce phénomène pourrait aussi être dû au modèle de croissance qui serait peu sensible aux variations des températures et des durées. Ainsi dans le cas d'un modèle plus sensible, l'utilisation de dis-

tributions pour les températures et les durées des étapes pourrait s'avérer utile. C'est pour cela que cette source de variabilité et d'incertitude n'a pas été négligée ici.

Afin de prendre en compte tous les facteurs pouvant avoir un impact sur la contamination, le nombre d'étapes constituant la fabrication et le circuit logistique de la salade en sachet qui ont été sélectionnées est élevé. Cependant par manque d'information, de données et pour simplifier le modèle, certaines étapes n'ont pas été retenues. Par exemple, la recontamination et la contamination croisée pouvant exister lors du procédé de fabrication n'ont pas été modélisées. En effet, il existe des "niches" comme les perforations dans les tapis de transport par exemple, dans lesquelles les bactéries sont susceptibles de se déposer, de se développer et de contaminer de nouvelles salades. Les lames de coupe peuvent aussi transmettre des bactéries d'une salade à l'autre (Lainé & Michard, 1988). La contamination peut également être d'origine humaine. Par exemple, le personnel préposé à la coupe peut contaminer la matière première par contact manuel ou par des outils de coupe souillés. De plus, les feuilles de salade étant parfois en contact les unes avec les autres, les bactéries peuvent se déplacer d'une feuille à l'autre. Les sources de contamination sont donc potentiellement nombreuses à l'intérieur de l'usine. Cependant, un nettoyage pluriquotidien, l'utilisation d'un procédé où il existe peu de contact entre les employés et la matière première et l'utilisation d'eau chlorée limitent les recontaminations et les contaminations croisées, ce qui justifie la non prise en compte de ces sources de contamination dans le modèle.

4.8.2 De l'influence du volume de données sur le niveau d'incertitude : exemple des données de contamination

L'incertitude autour des estimations de la contamination de la matière première d'origine européenne ou nord américaine (Tab. 4.7) est plus élevée que l'incertitude autour des estimations de la contamination des végétaux du monde entier obtenue dans le second chapitre (Tab. 2.4). En effet, lorsque l'on compare les distributions des paramètres μ_1, μ_2, $\sigma_{1,}$, σ_2 et σ_3 reflétant l'incertitude, il s'avère que les valeurs de ces paramètres fluctuent dans des intervalles plus grands pour la matière première (Tab. 4.3) que pour l'ensemble des végétaux (Tab. 2.2). Ce phénomène n'est pas étonnant étant donné que la contamination de la matière première a été estimée à partir d'un nombre de données moins important que la contamination de l'ensemble des végétaux. Il faut donc tenir compte de ce niveau élevé d'incertitude lors de l'analyse des résultats. Par exemple, le fait que les concentrations supérieures à la limite de détection (paramètre μ_1) soient plus élevées pour la matière première

des pays européens et nord américains que pour les végétaux provenant du monde entier est surprenant. En effet, il semble contradictoire que les concentrations en bactéries pathogènes soient plus élevées dans des pays où les moyens déployés pour contrecarrer les risques microbiologiques sont très nombreux. Ce résultat doit donc être plutôt attribué au fait que peu d'analyses ont été effectuées et que les seuls produits analysés contenaient des concentrations particulièrement élevées. Cet exemple confirme qu'il est intéressant de séparer la variabilité de l'incertitude afin d'interpréter correctement les résultats et de déterminer s'il est possible de réduire l'incertitude totale (variabilité et incertitude) par l'enrichissement en données pertinentes.

4.8.3 Effet du chlore sur la contamination des salades de IV^{ème} gamme

Le retrait du chlore de l'eau de lavage augmente le pourcentage de sachets contaminés. En effet, le nombre de sachets contenant des salades contaminées serait deux fois et demi plus élevé en cas de retrait du chlore de l'eau de lavage. Il semble donc que le lavage à l'eau chlorée soit efficace au niveau de la prévalence de *L. monocytogenes*. Dans le chapitre suivant l'impact de l'utilisation de l'eau chlorée ou non chlorée sur le risque sera évalué.

4.8.4 Peut-on valider le modèle ?

La validation de ce type de modèle est un exercice difficile pour plusieurs raisons. Une des raisons, déjà mentionnée dans le second chapitre, est que les données relevant les niveaux de concentration microbiologique sont rares. De plus, la comparaison entre les résultats obtenus avec notre modèle et des valeurs observées de concentration n'est pas directement réalisable étant donné qu'à notre connaissance aucune étude sur les concentrations en *L. monocytogenes* des salades en sachet avant consommation n'a été effectuée auprès de consommateurs (comme par exemple les études d'alimentation dupliquée réalisées pour des analyses de risques toxicologiques Leblanc et al. (2000) et Leblanc et al. (2005)). Il existe cependant des données sur la contamination des salades de IV^{ème} gamme mais les articles d'où sont tirées ces données ne mentionnent pas en général la date exacte à laquelle les produits ont été prélevés du circuit logistique, ce qui rend difficile la comparaison avec nos résultats. Une autre raison est que les quantités prélevées pour les analyses sont en général faibles par rapport à la contenance des sachets vendus, ce qui implique que les concentrations observées ne sont que partielles.

Toutefois, nous proposons de valider notre modèle en :

- comparant les prévalences en *L. monocytogenes* observées à différentes étapes du circuit logistique des salades en sachets et les prévalences simulées avec notre modèle. Notons que puisque c'est la prévalence qui est étudiée, et que celle-ci est constante au cours du temps il n'est pas nécessaire de connaître la date d'analyse du produit.
- en comparant les concentrations simulées avec notre modèle à celles observées par Odumeru et al. (1997) dans leur étude sur les concentrations de salades en sachet provenant d'usines de IV^{ème} gamme.

Les données utilisées pour comparer les prévalences simulées et observées sont les données de contamination des salades transformées (lavées, coupées, mises en sachet, etc...) d'origine européenne et nord américaine (Etats-Unis et Canada) ayant servi à l'élaboration du modèle du chapitre 2 (Bind, 1992; Tiwari & Aldenrath, 1990; Gombas et al., 2003; Levre et al., 1995; Little et al., 1997; Lainé & Michard, 1988; Breer & Schopfer, 1989; Breer & Baumgartner, 1992; Beaufort et al., 1992; Francis & O'Beirne, 2006; DGCCRF, 1993-1996 and 1997-2001) ainsi que des données recueillies auprès des industriels participant au projet (Stahl et al., 2002, 2005). Les prévalences étant observées pour des quantités de 25g, les simulations sont réalisées également en utilisant des quantités de 25g. La prévalence en *L. monocytogenes* calculée sur l'ensemble de ces données est 1.97%. Le pourcentage simulé de quantités de 25g contaminée est 1.85% avec un UI à 95% de [0.98 ; 2.32]%. Le modèle prévoit donc une prévalence en *L. monocytogenes* dans les salades en sachet similaire à celle observée, ce qui est satisfaisant.

Le protocole expérimental de l'article d'Odumeru et al. (1997) est le suivant. Les salades de IV^{ème} gamme sont conservées plusieurs jours à deux températures différentes : 4°C et à 10°C. Les concentrations sont relevées à différentes dates : 1, 4, 7 et 11 jours et les quantités analysées sont en moyenne de 1 800g. L'étude d'Odumeru et al. (1997) est intéressante car elle permet d'une part de connaître les concentrations des salades de IV^{ème} gamme en *L. monocytogenes* et d'autre part d'observer la croissance de *L. monocytogenes* natives (et non artificiellement inoculées) en fonction du temps. De plus, les quantités analysées sont plus grandes que celles généralement prélevées pour des analyses microbiologiques. Les résultats de l'étude sont présentés dans la colonne "Concen. observée" du tableau 4.14. Une première analyse des données montre que les concentrations observées sont très variables d'un produit à l'autre et qu'il est difficile d'admettre un processus continu de croissance, étant donné que les concentrations les plus élevées ne correspondent pas aux temps de conservation les plus longs. Par exemple, après 4 jours de conservation à 4°C, une valeur de 1.81 log ufc/g a été observée tandis qu'après 7 jours de conservation à la même température la concentration

est de 0.96 log ufc/g. Ou encore après 7 jours de conservation à 10°C, les concentrations observées sont 1.36 log ufc/g et 3.04 log ufc/g, soit une différence d'environ 2 log ufc/g entre ces deux analyses.

L'objectif est de déterminer à quel niveau se situent les concentrations observées par rapport aux concentrations simulées avec notre modèle dans les mêmes conditions (quantité, durée et température), c'est-à-dire si les observations correspondent plutôt à de faibles percentiles de la distribution de concentration simulée ou plutôt à des percentiles élevés. Le modèle est considéré performant lorsque les observations sont équitablement réparties entre les faibles et les forts percentiles de la distribution de concentration simulée avec une majorité de points autour du $50^{ème}$ percentile de la distribution. Les percentiles de la distribution simulée correspondant aux valeurs observées sont donc déterminés par le pourcentage de valeurs simulées inférieures ou égales aux valeurs observées. Les intervalles d'incertitude autour de chaque percentile sont aussi calculés. Le modèle de croissance utilisé est le modèle M_2 avec plusieurs scénarios pour $y_{\max}$ (distribution *a posteriori* obtenue dans le chapitre 3 à partir des 206 courbes de croissance, $y_{\max} = y_0 + 4$ log ufc/g et $y_{\max} = y_0 + 2$ log ufc/g). Les résultats sont présentés dans le tableau 4.14 et dans la figure 4.7. Une première remarque importante est que les concentrations simulées encadrent bien les valeurs observées. En effet, les valeurs observées se situent selon la température et le scénario, entre le $4^{ème}$ et $96^{ème}$ percentile des valeurs simulées. La tendance générale pour les deux premiers scénarios (distribution *a posteriori* et $y_{\max} = y_0 + 4$ log ufc/g) est que les concentrations observées après 1 et 4 jours se situent dans les percentiles élevés de la distribution de concentration simulée par le modèle (entre le $66^{ème}$ et le $86^{ème}$). *A contrario* après des temps de conservation plus longs (7 et 11 jours), les concentrations observées se situent plutôt dans les faibles percentiles de la distribution de concentration simulée. Pour le dernier scénario ($y_{\max} = y_0 + 2$ log ufc/g), les valeurs observées se situent toujours dans les percentiles élevés de la distribution de concentration simulée (entre le $72^{ème}$ et le $96^{ème}$). Ainsi par rapport à ces observations, le modèle a tendance à produire des concentrations inférieures à celles observées lorsque le temps de conservation est compris entre 1 et 4 jours et supérieures à celles observées lorsque la conservation est plus longue, excepté pour le scénario ($y_{\max} = y_0 + 2$ log ufc/g) , pour lequel les concentrations simulées sont constamment inférieures à celles observées. Notons que le scénario ($y_{\max} = y_0 + 4$ log ufc/g), basé sur certaines conclusions des travaux de Gnanou Besse et al. (2006), fournit les concentrations les plus en adéquation avec les valeurs observées par Odumeru et al. (1997). En effet, sur la figure 4.7 on remarque que pour ce scénario les points sont répartis au dessus et au-dessous du $50^{ème}$ percentile.

Dans ce travail, nous montrons donc que même si les données de contamination sont rares et souvent incomplètes car manquant d'informations indispensables à la validation, il existe toutefois des études permettant de valider notre modèle. En conclusion, le modèle est satisfaisant au niveau de la prévalence de la contamination. Au niveau des concentrations, pour chaque scénario testé, les simulations obtenues encadrent bien les valeurs observées par Odumeru et al. (1997), et le scénario ($y_{\max} = y_0 + 4$ log ufc/g) semble être celui à privilégier. De plus, nous confirmons avec cette validation l'importance de développer des modèles permettant de prendre en compte et de propager tout au long de la chaîne alimentaire la variabilité des contaminations microbiologiques.

TAB. 4.10: Contamination par *L. monocytogenes* des salades des sachets contaminés après chaque étape du circuit logistique. Ces valeurs sont obtenues à partir de la contamination de la matière première lavée à l'eau chlorée et en utilisant le modèle de croissance $M2$ (unité : log ufc/g)

Etape	$C(t_k)$	Moyenne	Ecart-type	$P_{0.025}$	$P_{0.5}$	$P_{0.975}$
Sortie chambre froide de	Estimation	-0.23	1.29	-2.4	-0.29	2.37
l'usine $k = 1$	UI_{inf}	-0.74	1.08	-2.67	-0.95	1.6
	UI_{sup}	0.98	2.37	-2.06	0.81	5.71
Après premier transport	Estimation	-0.13	1.36	-2.51	-0.2	2.61
frigorifique $k = 2$	UI_{inf}	-0.66	1.16	-2.83	-0.82	1.8
	UI_{sup}	1.07	2.37	-2.1	0.9	5.84
Sortie plateforme $k = 3$	Estimation	0.02	1.45	-2.56	-0.05	2.95
	UI_{inf}	-0.52	1.24	-2.98	-0.66	2.14
	UI_{sup}	1.2	2.42	-2.06	1.04	5.99
Après second transport	Estimation	0.11	1.52	-2.65	0.04	3.17
frigorifique $k = 4$	UI_{inf}	-0.45	1.31	-3.1	-0.57	2.29
	UI_{sup}	1.29	2.46	-2.1	1.16	6.09
A l'achat $k = 5$	Estimation	1.4	2.07	-2.16	1.2	5.81
	UI_{inf}	0.63	1.8	-2.76	0.4	4.67
	UI_{sup}	2.53	2.56	-1.51	2.49	6.89
Après le trajet jusqu'au	Estimation	1.44	2.11	-2.22	1.26	5.88
domicile $k = 6$	UI_{inf}	0.67	1.85	-2.79	0.43	4.74
	UI_{sup}	2.59	2.58	-1.56	2.55	6.98
A l'instant de la $1^{ère}$	Estimation	2.61	2.31	-1.65	2.53	6.74
consommation $k = 7$	UI_{inf}	1.69	2.08	-2.37	1.49	5.93
	UI_{sup}	3.7	2.56	-0.86	4.09	7.4
A l'instant de la consommation	Estimation	2.98	2.34	-1.49	3	6.98
($1^{ère}$ et $2^{ème}$) $k = 7, 8$	UI_{inf}	2.04	2.12	-2.3	1.87	6.28
	UI_{sup}	4	2.56	-0.58	4.55	7.61

UI_{inf} et UI_{sup} borne inférieure et borne supérieure de l'intervalle d'incertitude à 95%

TAB. 4.11: Contamination par *L. monocytogenes* des salades des sachets contaminés à l'instant de la première consommation selon l'eau de lavage (chlorée ou non chlorée) et le modèle de croissance ($M2$ et $M2_{\lambda_{obs}}$) utilisés (unité : log ufc/g)

Modèle de croissance	Eau de lavage	$C(t_7)$	Moyenne	Ecart-type	$P_{0.025}$	$P_{0.5}$	$P_{0.975}$
$M2$	Chlorée	Estimation	2.61	2.31	-1.65	2.53	6.74
		UI_{inf}	1.69	2.08	-2.37	1.49	5.93
		UI_{sup}	3.7	2.56	-0.86	4.09	7.4
$M2_{\lambda_{obs}}$	Chlorée	Estimation	3.17	2.34	-1.36	3.25	6.99
		UI_{inf}	2.26	2.11	-2.18	2.14	6.42
		UI_{sup}	4.23	2.58	-0.42	4.79	7.53
$M2$	Non chlorée	Estimation	2.65	2.51	-1.97	2.6	7.01
		UI_{inf}	1.83	2.34	-2.59	1.58	6.42
		UI_{sup}	3.6	2.68	-1.23	4.00	7.53
$M2_{\lambda_{obs}}$	Non chlorée	Estimation	3.29	2.49	-1.63	3.49	7.19
		UI_{inf}	2.4	2.3	-2.3	2.25	6.71
		UI_{sup}	4.11	2.66	-0.91	4.78	7.6

UI_{inf} et UI_{sup} borne inférieure et borne supérieure de l'intervalle d'incertitude à 95%

TAB. 4.12: Contamination par *L. monocytogenes* des salades des sachets contaminés après un test de vieillissement : 3 jours à 4 °C et 6 jours à 8 °C. Ces valeurs sont obtenues à partir de la contamination de la matière première lavée à l'eau chlorée et en utilisant le modèle de croissance $M2$ (unité : log ufc/g)

	Moyenne	Ecart-type	$P_{0.025}$	$P_{0.5}$	$P_{0.975}$
Estimation	3.21	2.35	-1.17	3.24	7.12
UI_{inf}	2.33	2.13	-1.79	2.11	6.54
UI_{sup}	4.2	2.55	-0.38	4.87	7.68

UI_{inf} et UI_{sup} borne inférieure et borne supérieure de l'intervalle d'incertitude à 95%

TAB. 4.13: Contamination par *L. monocytogenes* des salades des sachets contaminés à la première consommation selon les différents scénarios pour $y_{\max}$. Ces valeurs sont obtenues à partir de la contamination de la matière première lavée à l'eau chlorée et en utilisant le modèle de croissance $M2$ (unité : log ufc/g)

Scénarios pour $y_{\max}$	$C(t_7)$	Moyenne	Ecart-type	$P_{0.025}$	$P_{0.5}$	$P_{0.975}$
$y_{\max} \sim N(6.64; 0.81)I(5; 8.5)$	Estimation	2.61	2.31	-2.22	2.53	6.74
	UI_{inf}	1.69	2.08	-3.04	1.49	5.93
	UI_{sup}	3.7	2.56	-1.41	4.09	7.4
$y_{\max} = y_0 + 6$	Estimation	2.54	2.27	-1.67	2.5	6.85
	UI_{inf}	1.76	2.04	-2.38	1.59	5.78
	UI_{sup}	3.86	2.8	-0.78	3.89	8.31
$y_{\max} = y_0 + 4$	Estimation	2.09	1.86	-1.65	2.15	5.4
	UI_{inf}	1.38	1.64	-2.43	1.38	4.59
	UI_{sup}	3.22	2.61	-0.8	3.24	8.12
$y_{\max} = y_0 + 2$	Estimation	1.09	1.44	-1.68	1.04	3.82
	UI_{inf}	0.53	1.21	-2.41	0.45	3.04
	UI_{sup}	2.27	2.44	-0.98	2.13	7.23

UI_{inf} et UI_{sup} borne inférieure et borne supérieure de l'intervalle d'incertitude à 95%

TAB. 4.14: Comparaison entre les concentrations observées dans l'article de Odumeru et al. (1997) en fonction du nombre de jours et de la température de conservation et les concentrations simulées par le modèle $M2$ en utilisant différents scénarios pour le paramètre $y_{\max}$

Température $°C$	Nb jours	Concentrations observées log ufc/g	Percentiles de la distribution simulée correspondant aux concentrations observées selon différents scénarios pour $y_{\max}$		
			Dist. *a post.*[a]	$y_{\max} = y_0 + 4$	$y_{\max} = y_0 + 2$
4	1	0.56	86 [73 ; 94][b]	86 [73 ; 94]	86 [73 ; 94]
	4	1.81	82 [69 ; 91]	85 [71 ; 92]	90 [77 ; 96]
	7	0.96	40 [27 ; 55]	48 [35 ; 61]	74 [59 ; 85]
	11	1.04	17 [9 ; 29]	34 [25 ; 44]	72 [57 ; 84]
		2.38	28 [16 ; 43]	67 [52 ; 79]	89 [75 ; 95]
10	4	2.66	66 [46 ; 81]	81 [67 ; 90]	93 [81 ; 97]
		3.04	70 [51 ; 84]	84 [71 ; 92]	96 [84 ; 99]
	7	1.36	15 [5 ; 27]	43 [32 ; 54]	76 [62 ; 87]
		3.04	27 [12 ; 44]	74 [60 ; 86]	95 [82 ; 98]
	11	1.63	4 [1 ; 10]	47 [35 ; 57]	79 [64 ; 89]
		2.66	7 [2 ; 15]	67 [52 ; 79]	91 [78 ; 96]

[a] Distribution *a posteriori* établie à partir du traitement des 206 courbes de croissance dans le chapitre 2

[b] Est [UI_{inf}, UI_{sup}]: estimation, borne inférieure et borne supérieure de l'intervalle d'incertitude à 95%

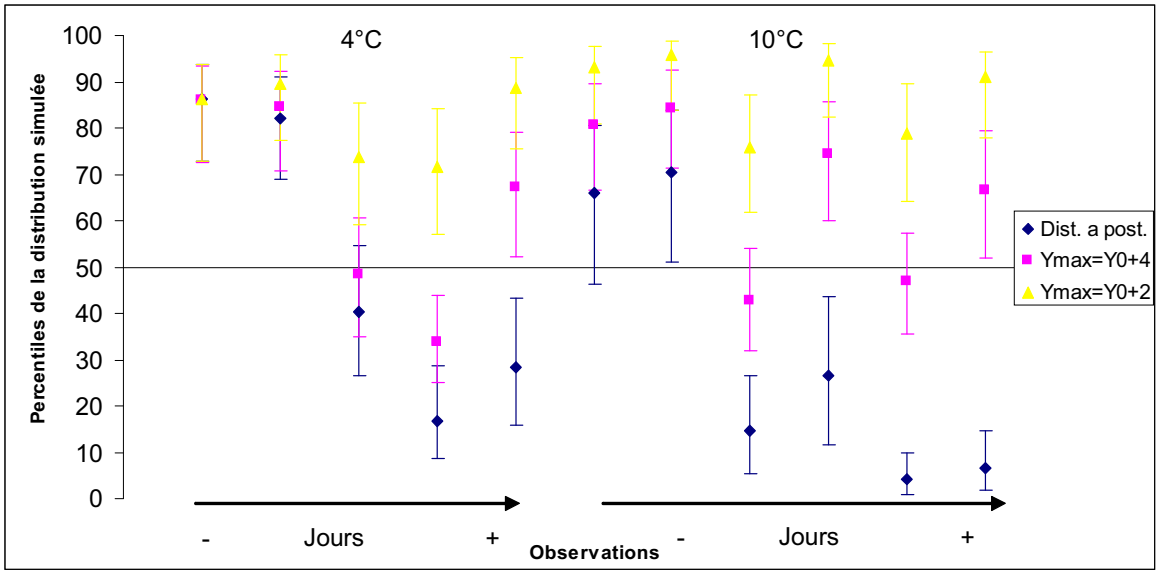

FIG. 4.7: Comparaison entre les concentrations observées dans l'article de Odumeru et al. (1997) (présentées dans le même ordre que dans le tableau 4.14) et les concentrations simulées par le modèle $M2$ en utilisant différents scénarios pour le paramètre $y_{\max}$. L'intervalle d'incertitude (UI) autour de chaque percentile est représenté par une barre.

Chapitre 5

Evaluation de l'exposition et caractérisation des risques liés à la présence de *Listeria monocytogenes* dans les salades de IV^{ème} gamme

5.1 Introduction

Ce chapitre est consacré à l'évaluation de l'exposition à *L. monocytogenes* des consommateurs de salades en sachet et à la caractérisation des risques encourus. L'évaluation de l'exposition nécessite de connaître les quantités de salade en sachet consommées ainsi que leurs fréquences de consommation sur une période donnée et les concentrations en *L. monocytogenes* de ce produit à l'instant de la consommation.

Les habitudes de consommation alimentaire des français sont très variables d'un individu à l'autre. Il s'agit alors d'intégrer cette variabilité dans l'évaluation de l'exposition et du risque en utilisant des données individuelles de consommation. Pour cela, nous proposons d'utiliser les données individuelles de consommation de l'enquête INCA. Cette enquête relève la consommation individuelle des français sur une semaine. Cependant, dans sa nomenclature, la consommation de salades non transformées n'est pas distinguée de la consommation de salades de IV^{ème} gamme. Afin d'intégrer la proportion consommée de salade de IV^{ème} gamme dans les calculs, nous utilisons le panel des ménages Sécodip. Cette étude répertorie les achats de ménages français sur une année en distinguant la salade transformée et la salade

non-transformée. L'utilisation des données individuelles de consommation permet alors de prendre en compte la relation de dépendance pouvant exister entre quantités consommées par un individu et la fréquence de consommation de cet individu mais aussi entre la sensibilité à L. monocytogenes de l'individu et sa consommation. En effet, certaines sous-populations comme les personnes âgées, les femmes enceintes ou les personnes immunodéprimées sont plus sensibles à l'ingestion de L. monocytogenes. La sensibilité de ces personnes est généralement prise en compte dans la relation dose-réponse par l'utilisation d'une probabilité de listériose plus élevée que celle de la population dite normale. Or il est aussi intéressant de prendre en compte les habitudes de consommation de ces sous-populations qui diffèrent peut-être de celles de la population normale. C'est ce que nous nous emploierons à faire dans cette partie. Notons que les données de consommations individuelles sont très peu employées en analyse des risques microbiologiques mais que pour les raisons précédemment énoncées, il nous semble nécessaire des les utiliser.

La concentration en *L. monocytogenes* des quantités consommées (exposition) est estimée dans ce travail selon deux méthodes. Dans la première, le nombre de bactéries par quantité consommée est déterminé à partir de la contamination du sachet dans lequel la quantité a été prélevée. Or le mode de transfert des bactéries du sachet à la quantité doit tenir compte du fait que la répartition des bactéries n'est pas homogène dans le sachet mais que celles-ci sont plutôt groupées en amas. La méthode de "partitioning" de Nauta (2005) est utilisée ici pour prendre en compte ce phénomène. Le problème de cette méthode est que la valeur du coefficient de groupement pour la salade en sachet et plus généralement pour les aliments, est inconnu. De ce fait, la plupart des analyses de risques microbiologiques sont réalisées en considérant que la répartition des bactéries est homogène dans les aliments. Nous proposerons toutefois une estimation de ce coefficient. Cette méthode sera comparée à celle employée par Sanaa et al. (2004) pour des fromages qui consiste à simuler la croissance des bactéries directement dans les quantités consommées jusqu'à l'instant de la consommation. Ainsi l'exposition est directement obtenue sans passer par le calcul de la contamination du sachet.

L'exposition établie, le risque de listériose est étudié pour chacune des sous-populations ainsi que le risque de listériose de la population entière. Or pour que le risque de listériose estimé à partir de la population INCA soit extrapolable à celui de la population française il est nécessaire que la représentation de chaque sous-population soit la même que celle de la population française. L'enquête INCA présente un nombre d'enfants supérieur à celui de la population française. Pour redresser cet échantillon nous affecterons à chaque individu des

différentes sous-populations un poids de redressement calculé par rapport aux statistiques de la population française INSEE, Institut National de la Statistique et des Etudes Economiques (2002).

Les deux premières parties de ce chapitre sont consacrées à la consommation : les données disponibles et leur utilisation. Les parties 3 à 5 présentent les méthodes utilisées ici pour calculer l'exposition, le modèle dose-réponse employé ainsi que les définitions générales du risque individuel par acte de consommation, du risque individuel annuel et du nombre de cas de listériose. Les résultats sont ensuite présentés dans une sixième partie et une discussion clôt ce chapitre.

5.2 Les données de consommation disponibles en France

En France, la consommation alimentaire est évaluée selon quatre types de données présentés ci-dessous

- les données de production qui permettent d'avoir une idée des quantités moyennes consommées mais qui tendent à surestimer les consommations individuelles réelles.
- les enquêtes auprès des ménages qui sont de deux sortes. Les premières ont pour cible la dépense des ménages et recueillent par exemple les tickets de caisses des supermarchés. Les secondes recueillent en plus les achats des ménages (comme les données du panel français Sécodip). Serra-Majem et al. (2003) ont montré pour le Canada et certains pays européens, que ce type d'enquête peut donner une bonne approximation des quantités consommées. Cependant ces données ont tendance à surévaluer la consommation de certains produits de longue conservation comme les céréales ou le sucre et à sous-estimer la consommation de produits comme le poisson, la viande ou les légumes frais, notamment par l'absence de prise en compte de la consommation hors domicile. De plus, afin d'obtenir des consommations individuelles l'évaluateur de risque a recours à une simple division de la consommation du ménage par le nombre d'individus qui le composent, ce qui entraîne de nombreux biais.
- les enquêtes individuelles qui utilisent deux méthodes différentes pour récolter les données. Soit il est demandé aux participants de noter ses consommations dans un carnet pendant un ou plusieurs jours (7 jours pour les données INCA), soit les individus sont interrogés sur leurs consommations antérieures de 24h (dites "rappels de 24h") ou bien plus globalement sur leurs habitudes de consommation (dites "de questionnaire de fréquence").

– les repas dupliqués qui consistent à ce que le consommateur prépare en double les quantités d'aliments qu'ils consomment. La part dupliquée est analysée afin d'évaluer son niveau de concentration pour le(s) contaminant(s) recherché(s). Ce type d'enquête est principalement utilisé dans les analyses de risque toxicologique (Leblanc et al., 2000, 2005).

L'enquête individuelle de consommation alimentaire (INCA 1999) et les données d'achat des ménages du panel Sécodip utilisées dans ces travaux pour évaluer la consommation de salade en sachet en France sont détaillées ci-dessous.

5.2.1 L'enquête individuelle nationale sur les consommations alimentaires (INCA)

L'enquête INCA a été réalisée par le centre de recherche pour l'étude et l'observation des conditions de vie (CREDOC) en 1998-1999, pour le compte de clients institutionnels et privés. Les commanditaires sont l'Afssa, les groupes Danone et Nestlé, ainsi que les offices et les interprofessions de plusieurs secteurs agro-alimentaires (CREDOC-AFSSA-DGAL, 1999).

Cette enquête vise quatre objectifs

– connaître les consommations individuelles réelles ainsi que leurs déterminants, par occasion (petit déjeuner, déjeuner, goûter, dîner, en-cas) et par lieu de consommation (au domicile et hors foyer),
– suivre l'évolution des pratiques et des connaissances dans le domaine de l'alimentation et de la nutrition,
– identifier les apports nutritionnels à partir des consommations déclarées et en déduire la situation nutritionnelle des consommateurs en fonction de leurs besoins,
– analyser les opinions et attitudes des consommateurs, notamment dans le domaine de la nutrition et de la sécurité alimentaire.

Le relevé des consommations alimentaires est effectué à l'aide d'un carnet de consommation, sur une période de 7 jours consécutifs, l'identification des aliments et des portions étant facilitée par l'utilisation d'un carnet photographique (carnet réalisé pour l'enquête SUI.VI.MAX, Hercberg et al. (2004)). Les 3003 carnets de consommation recueillis correspondent à 75889 prises alimentaires et environ 900 références alimentaires formant 48 groupes d'aliments.

L'enquête INCA repose sur un échantillon constitué de 1985 personnes de 15 ans et plus et de 1018 enfants et jeunes adolescents de 3 à 14 ans, soit au total 3003 individus. Pour

éviter les effets de grappe, tous les individus ne sont pas systématiquement interrogés : sur un total de 1865 foyers enquêtés, tous les membres du foyer de 3 ans ou plus sont interviewés dans 812 ménages, tandis que dans les 1053 autres ménages, la personne interrogée est tirée au sort. Cette méthodologie permet à la fois de disposer de résultats au niveau ménage et au niveau individuel, pour les adultes et les enfants. La représentativité de l'échantillon est assurée par stratification sur les variables région géographique et taille d'agglomération et par la méthode des quotas sur les variables âge, sexe, profession et catégorie socioprofessionnelle, taille du ménage. Pour intégrer les effets de saisonnalité, la réalisation des enquêtes est programmée sur une période de 11 mois (août 1998 à juin 1999), segmentée en quatre vagues. La méthodologie retenue vise donc à éviter certains biais bien connus : non-représentativité nationale, saisonnalité, effet de lassitude en fin d'enquête. La surreprésentation des enfants, qui constituent plus d'un tiers de l'échantillon (34%), est volontaire de la part des commanditaires, afin de cibler sur leur comportement alimentaire. Ceci implique qu'il faut considérer les deux échantillons, celui des enfants et celui des adultes séparément. Ainsi l'étude entière de la population en terme de risque nécessite le redressement de l'échantillon afin de prendre en compte la surreprésentation des plus jeunes (cf. 5.3.3).

Certaines personnes qui sous-déclarent leur consommation peuvent être identifiées par un apport énergétique trop faible pour être vraisemblable. La sous déclaration est un biais bien connu, en particulier pour ce qui concerne les boissons alcoolisées dont la valeur énergétique est élevée. Ces personnes sont généralement écartées des analyses. Cependant, le retrait de ces personnes (511 adultes sur 1985) fausse la représentativité de l'échantillon. Nous n'éliminerons donc pas ces personnes d'autant que la salade ne semble pas faire partie des produits sous-déclarés. La durée de l'enquête (7 jours), bien que plutôt longue pour une enquête individuelle, les enquêtes de deux ou trois jours étant les plus fréquentes en Europe, engendre un biais d'infréquence d'achat. En effet, les produits consommés que très rarement peuvent ne pas être captés par ce type d'enquête malgré les différentes vagues de l'enquête. Un autre biais peut se produire dû aux objectifs nutritionnels affichés et la courte durée de l'enquête. En effet, il est probable que le comportement alimentaire se trouve modifié durant la période de l'enquête, en faveur des aliments à connotation nutritionnelle positive.

L'enquête INCA fournit les données sur la consommation nécessaire à l'élaboration d'une analyse de risque pour un produit régulièrement consommé. Cependant, la distinction entre la salade en sachet et la salade non transformée n'est pas mentionnée dans la nomenclature de la base INCA. Le recours au panel de ménage Sécodip est nécessaire afin d'obtenir cette information.

5.2.2 Le panel Sécodip

La société privée Sécodip (Société d'Etudes de la Consommation, de la Distribution et de la Publicité, qui s'appelle dorénavant TNS Secodip, http ://www.secodip.fr) répertorie les achats des ménages français sur une année depuis1989. L'INRA achète ces données dans un délai de 2 ans suivant leur recueil (voir la présentation détaillée de Boizot, 2005). Les achats étaient initialement relevés sur papier de manière hebdomadaire, la liste des produits augmentant progressivement, puis, en 1996 a été introduite la "scannette" permettant la lecture optique des codes-barres et depuis 2005 le "palm" est mis en place. Pour limiter le temps de recueil et favoriser la participation au panel, chaque ménage ne répertorie pas l'ensemble de ses achats : de 1989 à 1995, deux panels disjoints de ménages relevaient leurs achats de produits (types de produits différents selon les panels) ; depuis 1996, le panel général comprend deux sous-panels, l'un relevant les "Viandes et Poissons frais et Vins" et l'autre les "Fruits et Légumes frais" en plus des autres produits. On ne dispose donc pas pour un même ménage de l'ensemble de ses achats et ainsi selon le produit étudié il faut utiliser l'un ou l'autre des deux panels. Par exemple, pour l'étude des achats de salade transformée et non transformée, il faut utiliser uniquement les personnes appartenant au panel "Fruit et Légumes frais". Pour chacun des achats, sont fournis : la quantité (en kilogrammes, litres ou nombres d'unités) le prix d'achat, le lieu d'achat (type de magasin), la marque du produit, etc. Concernant les ménages, la composition du foyer en termes d'âge, de sexe, de catégorie socioprofessionnelle, de niveau d'étude est détaillée et des informations globales de type région, commune, classe économique et sociale, possession d'appareils électroménagers, présence d'animaux domestiques, existence d'un jardin, etc. sont aussi fournies. Les panels sont initialement constitués par un sondage aléatoire stratifié selon la région de résidence et le type d'habitat, puis renouvelés tous les quatre ans. Certaines populations sensibles (personnes âgées) sont recrutées directement pour assurer une certaine représentativité nationale des ménages. Par ailleurs, TNS Sécodip fournit pour chaque panel des poids d'activité, calculés par la méthode du calage sur marges, pour chaque ménage.

Les relevés des achats alimentaires des ménages permettent d'obtenir une évaluation de la consommation alimentaire à domicile. Ces données présentent l'avantage de couvrir des périodes longues (un ménage est présent dans la base en moyenne 4 ans) et sont très utiles pour l'analyse de risque chronique comme le méthyle mercure dans les produits de la mer. La durée de l'enquête permet aussi d'évaluer les consommations occasionnelles, ne pouvant être recueillies par une enquête de 7 jours comme l'enquête INCA. Le fait que les achats soient recensés sur une longue période et dans un but commercial permet probablement d'éviter

le biais dû à un changement de comportement du consommateur en faveur des produits à connotation nutritionnelle positive. Toutefois, cette enquête présente deux inconvénients majeurs pour une analyse de risque microbiologique : il s'agit d'achats et non de consommations, ce n'est donc qu'un "proxy" de la consommation à domicile, les achats sont faits par le ménage et les quantités sont recueillies au niveau des ménages et non des individus.

5.3 Utilisation de l'enquête INCA et du panel Sécodip

Dans le cadre d'une évaluation quantitative des risques, l'idéal est bien sûr de disposer de données individuelles précises sur une longue période. Ce type de données n'existe pas en France et certaines méthodes se développent pour palier ce manque. Ainsi Allais & Tressou (2006) proposent une méthode d'estimation non-paramétrique inspirée de Chesher (1997) afin de décomposer les consommations des ménages du panel Sécodip en consommations individuelles. Crépet et al. (2007b) développent une méthode utilisant la vraisemblance empirique pour combiner les données de consommations INCA et les données des achats des ménages du panel Sécodip, afin d'évaluer la probabilité de dépasser une certaine dose de méthyle mercure par consommation de produits de la mer. Les auteurs répondent également au problème de surreprésentation des enfants dans l'enquête INCA en utilisant des contraintes marginales et proposent un "facteur de correction" permettant de prendre en compte les consommations hors domicile. De telles techniques nécessitant une formalisation supplémentaire ainsi que des algorithmes complexes ne seront pas utilisées ici.

Nous proposons alors d'utiliser les données individuelles de consommation INCA pour évaluer les quantités de salades en sachet consommées ainsi que les fréquences auxquelles s'effectuent les consommations (nombre d'actes de consommation par semaine). Les données d'achat du panel Sécodip sont utilisées pour estimer la proportion de salades en sachet consommées par rapport aux salades non transformées. De plus, afin que les résultats des simulations du risque soient extrapolables à la population française, un redressement de la population INCA est effectué par pondération. Des poids de redressement pour les différentes sous-populations sont alors calculés.

5.3.1 Statistiques de la consommation de salade

L'enquête INCA compte 1979 consommateurs de salade parmi $L = 3003$ individus interrogés. Pour chaque individu $l = 1, ..., L$, la quantité, $q_{l,a}$, de salade consommée par acte de

consommation a est disponible. Le nombre d'actes de consommation de salade effectué par l'individu l sur une semaine (ou fréquence de consommation) est noté n_l. Les statistiques descriptives de la consommation de salade des différentes sous-populations de l'enquête INCA (les adultes entre 15 et 59 ans, les enfants, les personnes âgées de 60 ans et plus et les femmes enceintes) sont données dans les tableaux 5.1 et 5.2.

TAB. 5.1: Quantités en gramme de salades consommées par acte de consommation par les différentes sous-populations de l'enquête INCA

Sous-population	Nb d'individus	Quantités ingérées par acte de consommation (g)			
		Moyenne	Ecart-type	Minimale	Maximale
Adultes 15-59 ans (exceptées femmes enceintes)	1089	41.86	17.93	3	200
Enfants 3-14 ans	523	31.37	17.79	3	150
Personnes âgées 60 ans et plus	329	41.20	20.72	2	250
Femmes enceintes	38	37.39	14.89	6	70

TAB. 5.2: Nombre d'actes de consommation de salades effectués sur une semaine par les différentes sous-populations de l'enquête INCA

Sous-population	Nb d'individus	Nombre d'actes de consommation / semaine			
		Moyenne	Ecart-type	Minimale	Maximale
Adultes 15-59 ans (exceptées femmes enceintes)	1089	3.06	2.10	1	13
Enfants 3-14 ans	523	2.24	1.52	1	9
Personnes âgées 60 ans et plus	329	3.40	2.16	1	12
Femmes enceintes	38	2.95	1.80	1	8

TAB. 5.3: Proportions des différentes sous-populations dans INCA et dans la population française et poids de redressement

Sous-population	Nb d'individus dans INCA	Proportion dans INCA (%)	Nb de consommateurs de salade dans INCA	Proportion[a] dans la population française (%)	Poids p_r de redressement
Adultes 15-59 ans (exceptées femmes enceintes)	1523	50.72	1089	59.44	1.17[b]
Enfants 3-14 ans	1018	33.90	523	18.71	0.55
Personnes âgées 60 ans et plus	420	13.99	329	20.57	1.47
Femmes enceintes	42	1.40	38	1.28	0.92
Total	3003	100	1979	100	-

[a]INSEE (2002) site web, [b]1.17=(59.44/50.72)

La quantité moyenne de salade consommée varie entre 31g et 42g par acte de consommation selon la sous-population étudiée. L'écart-type des quantités consommées est assez élevé et varie entre 15g et 21g. Ainsi les quantités consommées sont très variables d'un individu à l'autre. La plus petite quantité consommée est 2g et la quantité maximale est 250g. Les petites quantités (2-3g) correspondent à l'utilisation de la salade comme décoration des assiettes. Le nombre moyen d'acte de consommation de salade par semaine pour les adultes de 15-59 ans, les femmes enceintes et les personnes âgées se situe autour de 3 actes, avec un nombre légèrement plus élevé (3.4) pour les personnes âgées. Le nombre d'actes de consommation varie de 1 à 13 par semaine. Les enfants consomment moins souvent de la salade et en moins grande quantité que les autres sous-populations.

5.3.2 Proportion de salades en sachet consommées par rapport aux salades non transformées

La proportion p_s de salades en sachet consommées par rapport à la salade non transformée est approchée par la proportion p'_s, de salades en sachet achetées par les ménages du panel Sécodip de l'année 2004. Le marché de la salade en sachet s'étant beaucoup développé depuis ces dix dernières années, l'année 2004, qui est l'année la plus récente pour laquelle nous disposons des données Sécodip, est choisie. Ainsi, la proportion p'_s annuelle est calculée pour chacun des 2805 ménages consommant de la salade transformée et/ou non transformée. Nous choisissons de classer les ménages en trois catégories : les ménages consommant majoritairement de la salade en sachet c'est-à-dire les ménages pour lesquels la proportion p'_s est supérieure ou égale à 0.75, les ménages consommant majoritairement des salades non transformées c'est-à-dire les ménages pour lesquels la proportion p'_s est inférieure ou égale à 0.25 et les ménages consommant les deux types de salades c'est-à-dire les ménages pour lesquels la proportion p'_s est supérieure à 0.25 et inférieure à 0.75. Les pourcentages de ménage correspondant à ces trois catégories sont respectivement 21.6%, 58.6% et 19.8%.

Ainsi, la proportion p_s prendra les valeurs suivantes

$$p_s = \begin{cases} = 1 \text{ avec une probabilité de } 0.216 \\ = 0 \text{ avec une probabilité de } 0.586 \\ = 0.5 \text{ avec une probabilité de } 0.198. \end{cases} \tag{5.1}$$

La proportion moyenne de salade en sachet consommée par les 19.8% de personnes qui consomment les deux types de salades est 0.49, c'est pourquoi dans ce cas p_s est choisie égal

à 0.5. Le nombre d'actes de consommation de salade en sachet par individu par semaine s'écrit alors $p_s \times n_l$.

5.3.3 Poids de redressement de la population INCA

Pour que le risque calculé à partir de l'échantillon de l'enquête INCA, soit extrapolable à celui de la population française, il est nécessaire de vérifier si la structure de l'échantillon par rapport à l'âge, le sexe, la catégorie socioprofessionnelle, etc... est représentative de celle de la population française. Si certaines sous-populations sont sous ou sur représentées, nous utiliserons des poids de redressement pour pondérer leurs poids dans les calculs. Les sous-populations ont été définies par rapport à leur sensibilité à *L. monocytogenes* (personnes âgées de 60 ans et plus et les femmes enceintes) ou par rapport à leur représentation dans l'enquête INCA (adultes et enfants). Les poids de redressement sont calculés afin que, pour chaque sous-population, la proportion pondérée de personnes (consommatrices et non consommatrices de salade) de l'enquête INCA soit la même que dans la population française. Les proportions dans la population française proviennent des données de recensement de l'INSEE (INSEE, Institut National de la Statistique et des Etudes Economiques, 2002). Le tableau 5.3 présente les quatre sous-populations étudiées (les adultes entre 15 et 59 ans, les enfants, les personnes âgées de 60 ans et plus et les femmes enceintes), les proportions de ces sous-populations présentes dans l'enquête INCA et dans la population française ainsi que les poids de redressement utilisés. On remarque que les personnes âgées ainsi que les adultes sont sous-représentés dans l'enquête INCA ainsi un poids de redressement supérieur à 1 leur est attribué. Pour les enfants, surreprésentés, une pondération de 0.55 sera utilisée dans les calculs du risque pour diminuer leur influence. Le poids de redressement proche de 1 montre que les femmes enceintes sont correctement représentées dans l'enquête INCA. Les personnes immunodéprimées qui sont plus sensibles à *L. monocytogenes* ne sont pas ici étudiées spécifiquement car l'état immunologique des individus interrogés dans INCA n'est pas précisé. De plus, ces personnes présentent des maladies très différentes, il est donc difficile d'obtenir le pourcentage de personnes immunodéprimées dans la population française.

5.4 Evaluation de l'exposition

5.4.1 Définition de l'exposition

Soit C_Q le nombre de bactéries présent dans une quantité Q en gramme de produit consommée par un individu. L'exposition au contaminant étudié de cet individu est donc C_Q pour une quantité Q consommée. Pour un nombre $a = 1, ..., A$ d'acte de consommation relevés dans une population, chaque acte a étant représenté par la quantité q_a observée de produit consommé, l'estimateur de la distribution de l'exposition par acte de consommation est la fonction de répartition empirique définie par

$$F_A(x) = \frac{1}{A} \sum_{a=1}^{A} \mathbb{1}(c_{q_a} \leq x),$$

où $\mathbb{1}b$ est égal à 1 si b est vrai et à 0 sinon.

Précisons que comme le pourcentage de quantités consommées contaminées est faible, la distribution de l'exposition est composée principalement de 0. Afin d'avoir une meilleure représentation des expositions non nulles, seule l'exposition par consommation de quantités contaminées est mentionnée dans les résultats. Le pourcentage de quantités contaminées est également donné dans les résultats.

5.4.2 Evaluation de l'exposition à partir des sachets de salade

Soit Q la quantité de salade en gramme extraite d'un sachet de poids W ($Q \leq W$) contaminé à l'instant de la consommation par un nombre de bactéries C_W (ufc/W g). Dans le cas où le sachet ne contient pas de bactéries, C_W est égal à 0 et ainsi le nombre C_Q de bactéries dans la quantité Q issue du sachet est nul. Lorsque le sachet est contaminé, C_W prend ses valeurs dans la distribution de $C(t_7)$ ou celle de $C(t_8)$ exprimées en ufc/W g (Tab. 4.10). Dans ce cas, si la répartition des bactéries sur toutes les feuilles de salade d'un même sachet est considérée homogène, le nombre C_Q de bactéries présentes dans la quantité Q est distribué selon la loi binomiale

$$C_Q \sim Binomiale(C_W, Q/W).$$

Or pour un aliment tel que la salade l'hypothèse d'homogénéité, légitime dans le cas d'un liquide, n'est pas réaliste. En effet, lorsqu'un aliment solide est contaminé, les bactéries se

développent à partir d'une ou plusieurs colonies se situant dans différentes zones de l'aliment et forment alors des amas. Ainsi dans un sachet de salade il existe des zones "saines" (sans bactéries) et des zones "contaminées" (contenant des bactéries). C'est pourquoi, lorsqu'une quantité de salade est extraite d'un sachet, celle-ci peut soit provenir d'une zone "saine", soit d'une zone "contaminée", ou encore être issue d'un mélange de feuilles se situant dans les deux zones. Nauta (2005) propose une méthode permettant de prendre en compte ce phénomène de groupement ("clustering"). Dans ce cas, la probabilité qu'une quantité Q soit contaminée n'est plus Q/W mais possède une distribution de probabilité qui est la distribution $Beta(b, b(W/Q-1))$ de moyenne Q/W et où $b > 0$ est le coefficient de groupement. Ainsi le nombre de bactéries présent dans une quantité Q suit une distribution Beta-Binomiale[1]

$$C_Q \sim Beta - Binomiale(C_W, b, b(W/Q-1)), \tag{5.2}$$

de moyenne $C_W \times Q/W$ et de variance $C_W(W/Q-1)(bW/Q+C_W)/((W/Q)^2(bW/Q+1))$.

Lorsque le coefficient de groupement b tend vers l'infini, la probabilité qu'une quantité Q soit contaminée est Q/W et la répartition des bactéries est alors homogène. Si b tend vers 0, le phénomène de regroupement est maximal et dans ce cas, soit la quantité Q contient toutes les bactéries présentes dans la quantité W soit la quantité Q n'est pas contaminée.

Détermination de b :

Les données permettant d'estimer le paramètre b seraient, par exemple, la prévalence en *L. monocytogenes* calculée à partir de plusieurs prises d'essai provenant des différentes feuilles de salade composant un même sachet. Or l'origine des prises d'essai et en particulier le fait que les prises d'essai proviennent du même sachet ou non, n'est pas mentionnée dans les publications. La solution que nous proposons est de déterminer b par simulation. Pour une valeur de b fixée le pourcentage de portions de 25g contaminées obtenues à partir des concentrations des sachets est simulé. Cette opération est effectuée n_u fois afin de prendre en compte l'incertitude sur ce pourcentage. Lorsque la valeur médiane des pourcentages obtenus à partir des simulations est égale au pourcentage observé de portions de 25g contaminées, la valeur de b ayant produit les simulations est retenue. Les données utilisées pour calculer le pourcentage observé sont celles mentionnées dans la discussion du chapitre précédent (Bind, 1992; Tiwari & Aldenrath, 1990; Gombas et al., 2003; Levre et al., 1995; Little et al., 1997; Lainé & Michard, 1988; Breer & Schopfer, 1989; Breer & Baumgartner, 1992; Beaufort et al., 1992; Francis & O'Beirne, 2006; DGCCRF, 1993-1996 and 1997-2001; Stahl et al.,

[1] $Beta - Binomiale(n, \alpha, \beta)$, Densité : $f(x) = \binom{n}{x} \frac{(\alpha+x-1)!(n+\beta-x-1)!(\alpha+\beta-1)!}{(\alpha+\beta+n-1)!(\alpha-1)!(\beta-1)!}$

2002, 2005). Rappelons que ces données portent sur la contamination des salades de IV^{ème} gamme des pays européens et d'Amérique du Nord et que le pourcentage de portions de 25g contaminées est 1.97%. La valeur de b produisant ce pourcentage est 3.5. Si seules les données françaises de prévalence de *L. monocytogenes* sont utilisées, le pourcentage observé de portions de 25g contaminées est 0.97% et la valeur de b correspondante est 0.085. Notons que le cas $b = 3.5$ est presque équivalent au cas homogène. En effet, sous l'hypothèse d'homogénéité, 2.01% des portions de 25g produites à partir des sachets sont contaminées. L'intervalle d'incertitude à 95% associé à ce pourcentage est [1.27 ; 2.55]%.

5.4.3 Evaluation de l'exposition par simulation de la croissance des bactéries directement dans les quantités consommées

Sanaa et al. (2004), dans le cadre d'une évaluation du risque de listériose lié à la consommation de deux fromages, proposent de calculer directement le nombre de *Listeria monocytogenes* présent dans une portion consommée à partir de la contamination moyenne du lait ayant servi à la fabrication du fromage, sans au préalable déterminer la concentration du fromage à partir duquel la portion est prélevée. La croissance de chaque ufc de *Listeria monocytogenes* est simulée indépendamment. Ainsi la valeur utilisée pour $y_{\max}$ représente la concentration maximale pouvant être atteinte par une bactérie. Ensuite afin de calculer l'exposition le nombre de bactéries produites à partir d'une ufc est multiplié par le nombre de bactéries présentes dans la portion. La méthode qui est proposée ici est inspirée de celle de Sanaa et al. (2004). La différence est que comme la taille maximale de la population pouvant être atteinte dans la salade à partir d'une ufc de *Listeria monocytogenes* est inconnue, la croissance est simulée à partir de la concentration en *Listeria monocytogenes* de la quantité consommée. Cette concentration est calculée en croisant la concentration C_L (log ufc/g) estimée dans le chapitre 4, après le passage dans le bain de lavage des feuilles de salade et le logarithme en base 10 de la quantité consommée Q. La simulation de la croissance s'effectue comme dans le chapitre précédent à la différence que l'unité dans laquelle la croissance est étudiée n'est plus le sachet mais la quantité consommée.

5.5 Caractérisation des dangers

Le modèle dose-réponse utilisé est le modèle exponentiel élaboré dans le rapport FAO/WHO (2004) pour l'évaluation des risques liés à *Listeria monocytogenes* dans les aliments prêts à

consommer. La relation dose-réponse y est modélisée d'une manière similaire à celle décrite dans Buchanan & Bagi (1997) et dans Lindqvist & Westöö (2000). Ces auteurs utilisent conjointement des données d'enquêtes alimentaires et des statistiques annuelles sur la maladie pour élaborer un modèle dose-réponse exponentiel. Le modèle exponentiel a été choisi par les experts de la FAO en raison de son adéquation reconnue pour modéliser la listériose invasive, de sa simplicité en tant que modèle à un seul paramètre et de sa nature linéaire lorsqu'il est extrapolé à la fourchette des faibles doses. Il s'écrit[2]

$$R = 1 - \exp(-r \times D), \tag{5.3}$$

où R est la probabilité de tomber malade, D est la dose ingérée (le nombre de *L. monocytogenes* ingérées), et r est la probabilité qu'une seule cellule cause la maladie. En effet, ce modèle implique qu'une seule *L. monocytogenes* a une probabilité très faible mais bien définie de causer la maladie. Des valeurs de r spécifiques pour les populations moins sensibles (saines) et les populations plus sensibles ont été établies en considérant que la quantité de *L. monocytogenes* ingérée était la même pour les deux sous-populations. Le rapport FAO mentionne que la variabilité observée dans la virulence des différentes souches de *L. monocytogenes* existantes est considérée puisque les données, et donc les valeurs de r, reflètent les caractéristiques moyennes de nombreuses souches, y compris la fréquence des cas et l'importance de la virulence. De même, il est aussi tenu compte de la variabilité parmi les individus exposés au pathogène étant donné que l'incidence annuelle de la listériose comprend toute la population désignée. Toutefois, comme le précise Pouillot (2006) on peut se demander dans quelle mesure cette variabilité est prise en compte étant donné que le mécanisme de la loi dose-réponse exponentielle suppose une probabilité d'infection par une ufc de *L. monocytogenes* constante quel que soit l'individu ou la souche considérée. L'incertitude est prise en compte dans ces travaux en considérant plusieurs paramètres d'entrée comme incertains. La concentration maximale pouvant être atteinte dans une quantité consommée est fixée à cinq niveaux différents variant de 7.5 à 10.5 log ufc. Une distribution uniforme ayant comme paramètres 1888 et 3 148 correspondant à $\mp$ 25% du nombre total de cas de listériose aux États-Unis d'Amérique (2518 cas) est utilisée pour modéliser le nombre de cas de listériose. Les valeurs médianes et au $5^{ème}$ et $95^{ème}$ percentiles de r ainsi obtenues et utilisées dans ce chapitre, pour les deux types de populations, saines et sensibles sont présentées dans le tableau 5.4. Les différents groupes de population sensibles sont les personnes âgés (plus de

[2]Le modèle dose-réponse exponentiel est issu de l'approximation de la loi $Binomiale(n, p)$ par la loi de $Poisson(np)$ lorsque $n \to \infty$, $p \to 0$ et np est fini.

TAB. 5.4: Valeurs du paramètre r du modèle dose-réponse exponentiel utilisées pour la population sensible et saine (FAO/WHO, 2004)

sous-population	$P_{0.5}$	$P_{0.25}$	$P_{0.975}$
Sensible[a]	1.06×10^{-12}	2.47×10^{-13}	9.32×10^{-12}
Saine	2.37×10^{-14}	3.55×10^{-15}	2.70×10^{-13}

[a] personnes âgées et immunodéprimées, femmes enceintes et nourissons

60 ans), les nourrissons, les femmes enceintes et les personnes immunodéprimées. Les données individuelles de consommation de l'enquête INCA permettent de prendre en compte les consommations des différentes populations (sensible ou saine). Ainsi une valeur de r spécifique est attribuée à chaque personne selon si cette personne est enceinte, âgée ou fait partie de la population saine. Rappelons que pour les raisons mentionnées précédemment nous ne considérerons pas les personnes immunodéprimées. Notons aussi que les nourrissons font partis de la population sensible mais qu'étant donné que l'aliment vecteur (la salade de IV^{ème} gamme) de *L. monocytogenes* ne fait pas partie de leur régime alimentaire, cette population n'est pas étudiée.

5.6 Caractérisation des risques

5.6.1 Risque individuel par acte de consommation

En utilisant le modèle dose-réponse précédent (équation (5.3)), le risque de listériose invasive de l'individu l appartenant à la sous-population i par acte de consommation a s'écrit

$$R_{l,a}^{(i)} = 1 - \exp(-r_i \times c_{q_{l,a}}),$$

où $c_{q_{l,a}}$ est l'exposition de l'individu l ayant consommée la quantité $q_{l,a}$ contaminée à un niveau de concentration fixé. Soit la quantité consommée est issue d'un sachet non contaminé et $c_{q_{l,a}}$ prend la valeur 0, soit le sachet est contaminé et $c_{q_{l,a}}$ est une réalisation de la distribution (5.2) ou alors $c_{q_{l,a}}$ est estimée par simulation de la croissance dans la quantité $q_{l,a}$ consommée. Précisons qu'à un niveau de contamination fixé, la probabilité de ne pas tomber malade de l'individu l est $1 - R_{l,a}^{(i)} = \exp(-r_i c_{q_{l,a}})$.

Afin d'évaluer le risque individuel par acte de consommation de la population française, chaque risque dû à un acte de consommation effectué par l'individu l est multiplié par le poids de redressement $p_r^{(i)}$ de la sous-population i à laquelle cet individu appartient.

5.6.2 Risque individuel annuel

Le risque individuel annuel dépend du nombre d'actes de consommation effectués par l'individu l sur une année. Le nombre d'actes de consommation n_l étant uniquement observé sur une semaine de l'année, il est considéré que l'individu l effectue un nombre d'actes de consommation constant sur chaque semaine j de l'année. Ainsi l'individu l aura consommé sur un an $(n_l \times 52)$ fois de la salade dont le nombre de salade en sachet est $p_s \times n_l \times 52$. Les quantités consommées d'une semaine à l'autre sont aussi considérées constantes, ainsi pour chaque semaine $j = 1, ..., 52$, les quantités sont $(q_{l,a})_{a=1,...,p_s \times n_l}$. L'exposition de l'individu l à chaque acte a de consommation réalisé sur la semaine j est noté $c_{q_{l,a,j}}$.

Le risque individuel annuel de l'individu l appartenant à la sous-population i s'écrit

$$\begin{aligned} \pi_l^{(i)} &= 1 - \prod_{j=1}^{52} \prod_{a=1}^{[p_s \times n_l]} \exp(-r_i \times c_{q_{l,a,j}}) \\ &= 1 - \exp(-r_i \sum_{j=1}^{52} \sum_{a=1}^{[ps \times n_l]} c_{q_{l,a,j}}). \end{aligned} \tag{5.4}$$

Le risque de chaque individu est multiplié par le poids de redressement $p_r^{(i)}$ de la sous-population i à laquelle il appartient afin de calculer des statistiques sur le risque individuel annuel pour la population française.

5.6.3 Nombre de cas annuel de listériose du à la consommation de salade en sachet

Chaque individu l de la sous-population i a un risque annuel de tomber malade $\pi_l^{(i)} \times p_r^{(i)}$, l'état de l'individu l (malade ou non malade) $Y_l^{(i)}$ suit donc une distribution de Bernoulli de paramètre $\pi_l^{(i)} \times p_r^{(i)}$, et a pour espérance $\pi_l^{(i)} \times p_r^{(i)}$.

A l'ensemble des $(c_{q_{l,a,j}})_{\substack{a=1,...,p_s \times n_l \\ j=1,...,52}}$ fixés, le nombre de cas attendu de listériose invasive par an $K^{(i)}$ dans la sous-population i composé d'un nombre $L_{i,c}$ de consommateurs de salade, est égal à

$$K^{(i)} = E(\sum_{l=1}^{L_{i,c}} Y_l^{(i)}) = \sum_{l=1}^{L_{i,c}} E(Y_l^{(i)}) = \sum_{l=1}^{L_{i,c}} \pi_l^{(i)} \times p_r^{(i)}.$$

Le nombre annuel de cas de listériose du à la consommation de salade en sachet sur un an dans la population française composée de $N_{pop} = 60\ 350\ 000$ (1^{er} décembre 2002, INSEE) d'individus divisés en quatre sous-populations ($i = 1, ..., 4$), est

$$\frac{K^{(1)} + K^{(2)} + K^{(3)} + K^{(4)}}{L} \times N_{pop},$$

où L est le nombre total d'individus composant la population pour laquelle les quantités consommées sont disponibles, ici 3003.

5.7 Résultats

Les simulations sont réalisées par modélisation de Monte-Carlo à deux dimensions selon les mêmes principes que ceux énoncés dans l'introduction et le chapitre 4. Notons toutefois que dans cette partie la dimension de la variabilité, n_v, n'est plus donnée arbitrairement mais elle est déterminée soit par le nombre d'actes de consommation soit par le nombre d'individus présents dans l'enquête INCA, selon si le risque est calculé par acte de consommation ou par individu.

Les tableaux 5.5, 5.6 et 5.7 rassemblent les statistiques des distributions de l'exposition, du risque par acte de consommation et du risque individuel annuel pour l'ensemble de la population selon les différents scénarios utilisés pour :

- le modèle de croissance ($M2$ ou $M2_{\lambda_{obs}}$),
- le traitement de l'eau de lavage (chlorée ou non chlorée),
- le lieu de croissance (sachets ou quantités) et le mode de répartition des bactéries des sachets aux quantités consommées (coefficient de regroupement b),
- la distribution de la concentration maximale atteignable $y_{\max}$ (distribution *a posteriori* calculée à partir des 206 courbes dans le chapitre 3, la valeur de $y_{\max}$ dépend de la concentration initiale y_0 : $y_{\max} = y_0 + 4$ log ufc/g et $y_{\max} = y_0 + 2$ log ufc/g)
- la valeur du coefficient r de la dose réponse (médiane ou $2.5^{ème}$ percentile).

L'intervalle d'incertitude à 95% autour de l'estimation de chaque statistique est également mentionné dans les tableaux de résultats. Pour ne pas alourdir la présentation des résultats, le risque pour les différentes sous-populations est présenté uniquement pour un scénario dans le tableau 5.8. Le nombre de cas de listériose est également décrit selon les différents scénarios dans le tableau 5.9. La description détaillée des résultats est effectuée dans les parties suivantes.

5.7.1 Exposition estimée à *L. monocytogenes* par consommation de quantités de salades de IV^{ème} gamme contaminées

Le tableau 5.5 présente l'exposition à *L. monocytogenes* par consommation de quantités de salades de IV^{ème} gamme contaminées. Le pourcentage de quantités contaminées est 2.44% avec un UI à 95% de [1.69 ; 3.51]% lorsque la contamination des quantités consommées est calculée à partir de la contamination des sachets dont les feuilles ont été lavées à l'eau chlorée et lorsque le coefficient de regroupement est de 3.5. On remarque que ce pourcentage est très proche du pourcentage de sachets contaminés ce qui confirme que ce scénario est quasiment équivalent au cas d'une répartition homogène des bactéries. Pour ce scénario, la concentration moyenne des quantités contaminées à l'instant de la consommation (exposition) est 2.81 log ufc/g avec un UI à 95% de [2.18 ; 3.41] log ufc/g. L'exposition peut atteindre des valeurs élevées, par exemple la valeur au 97.5^{ème} percentile est de 6.48 log ufc/g avec un UI à 95% de [5.99 ; 6.85] log ufc/g. L'utilisation du modèle $M2_{\lambda_{obs}}$ dont le taux de croissance est plus élevé que celui du modèle $M2$, produit des valeurs d'exposition plus élevées. Par exemple, la moyenne est de 3.28 log ufc/g avec un UI à 95% de [2.58 ; 4.01] log ufc/g. *A contrario*, pour les scénarios intégrant une dépendance entre la valeur de la concentration maximale atteignable et celle de la concentration initiale, l'exposition est plus faible. On observe par exemple une différence sur l'exposition moyenne de -0.6 log ufc/g et -1.64 log ufc/g respectivement pour les scénarios $y_{\max} = y_0 + 4$ log ufc/g et $y_{\max} = y_0 + 2$ log ufc/g par rapport à celle obtenue avec la distribution *a posteriori* de $y_{\max}$. De plus les distributions sont moins étalées.

Lorsque le phénomène de regroupement en amas des bactéries est plus important ($b = 0.085$), le pourcentage de quantités contaminées est alors plus faible : 1.74% avec un UI à 95% [1.13 ; 2.43]%. Par contre le niveau d'exposition est à peu près le même, bien que légèrement plus faible, que pour le scénario supposant la répartition des bactéries homogène. La moyenne avec ce scénario est de 2.32 log ufc/g avec un UI à 95% de [1.75 ; 2.88] log ufc/g et l'exposition au 97.5^{ème} percentile est d'environ 6.5 log ufc/g.

Lorsque l'eau de lavage est non traitée avec du chlore, le pourcentage de quantités contaminées est plus élevé (5.78% avec un UI à 95% de [3.97 ; 8.54]%) que lorsque l'eau est traitée avec du chlore. L'exposition est à peu près équivalente entre ces deux scénarios.

Lorsque l'exposition est estimée par simulation de la croissance des bactéries directement dans les quantités consommées, il s'avère que le pourcentage de quantités contaminées est plus faible (0.91% avec un UI à 95% de [0.39 ; 1.56]%). Par contre les concentrations des quantités contaminées sont plus élevées, leur moyenne est estimée à environ 6 log ufc/g

et 95% des valeurs se situent entre [2.43 ; 8.89] log ufc/g. Le fait que les concentrations des quantités contaminées soient dans ce cas plus élevées que celles obtenues à partir des concentrations des sachets est dû au mode de calcul. Par exemple un sachet de 200g est contaminé lorsqu'il contient au moins 0.005 ufc/g (soit 1 ufc/200g) par contre une quantité de 40g est contaminée lorsqu'elle contient au moins 0.025 ufc/g (soit 1 ufc/40g). Ainsi pour qu'une quantité soit contaminée il faut que la concentration par gramme soit plus élevée, ce qui est plus rare. Le modèle produit alors un pourcentage de quantités contaminées moins élevé que celui de sachets contaminés mais contenant des concentrations supérieures. Ce lien, paradoxal, entre la taille de la quantité étudiée et le niveau de concentration a déjà était montré dans la section (Fig. 4.6) et sera discuté dans la dernier chapitre de cette thèse.

De manière générale, la taille des intervalles d'incertitude, montre que la variabilité de l'exposition prédomine sur l'incertitude des estimations.

5.7.2 Risque estimé de listériose par acte de consommation de salade de IV^{ème} gamme

Le tableau 5.6 présente le risque de listériose par acte de consommation de salade de IV^{ème} gamme. Etant donné que le pourcentage estimé de quantités contaminées varie entre 1.74% et 5.78% selon le scénario utilisé, seules les valeurs des percentiles les plus élevés de la distribution du risque par acte de consommation sont différents de zéro. Ainsi, la valeur du risque moyen par acte de consommation est principalement due aux valeurs extrêmes des percentiles élevés. L'estimation du risque moyen par acte de consommation varie entre 3.E-10 et 2.E-07 selon le scénario envisagé. Les risques les plus importants sont obtenus avec le scénario pour lequel l'eau de lavage est non chlorée et à celui utilisant le modèle de croissance $M2_{\lambda_{obs}}$. Les risques les moins importants sont obtenus avec les scénarios $y_{\max} = y_0 + 4$ log ufc/g et $y_{\max} = y_0 + 2$ log ufc/g, pour lesquels l'exposition est plus faible. Dans le cas où la croissance est simulée directement à partir des quantités consommées, la valeur estimée au 99ème percentile est à la différence des autres scénarios, nulle (avec un UI à 95% de [0 ; 6.E-08]), puisque le pourcentage estimé de quantités contaminées est inférieur à 1%. Cependant, comme les concentrations des quantités contaminées sont très élevées en moyenne, le risque moyen est à peu près équivalent à ceux obtenus avec les scénarios précédents. L'utilisant d'une valeur de r (probabilité qu'une bactérie cause la maladie) plus faible (valeur au 2.5^{ème} percentile) pour la dose réponse a relativement peu d'influence sur l'estimation du risque.

5.7.3 Risque estimé de listériose annuel

Le tableau 5.7 présente le risque individuel annuel de listériose pour l'ensemble de la population du à la consommation de salade de IV^{ème} gamme. L'estimation du risque annuel moyen varie entre 0 et 1.E-05 selon le scénario utilisé. Pour l'ensemble des scénarios, la valeur estimée de la médiane ainsi que l'intervalle d'incertitude autour de cette valeur sont de 0, excepté pour le scénario pour lequel la croissance est simulée directement dans les quantités consommées. Pour ce scénario, l'estimation du risque annuel est égale à 0, avec cependant une valeur de 1.E-4 pour la borne supérieure de l'intervalle d'incertitude. Le premier graphique de la figure 5.1 représente la fonction de répartition du risque individuel annuel pour le premier scénario (modèle de croissance $M2$, lavage à l'eau chlorée, coefficient de groupement de 3.5 et r médian) avec son intervalle d'incertitude à 95%. Sur le deuxième graphique les fonctions de répartition du risque pour différents scénarios sont représentées.

Le tableau 5.8 permet de comparer le risque annuel de listériose pour les différentes sous-populations étudiées. Le scénario utilisé est le modèle de croissance $M2$, lavage à l'eau chlorée, croissance simulée dans les sachets, coefficient de regroupement égal à 3.5, $y_{\max}$ a pour distribution celle la loi *a posteriori* établie dans le chapitre 3 et la valeur r de la dose réponse est la valeur médiane. Les sous-populations ayant le risque annuel le plus élevé sont les personnes âgées et les femmes enceintes avec un risque annuel moyen respectivement de 2.E-05 et 1.E-05. Cela n'est pas étonnant puisque ces populations sont plus sensibles à l'exposition à *L. monocytogenes*. Les enfants ont un risque annuel légèrement moins important que les adultes qui est dû à une consommation moins importante de salade en sachet. Notons que pour les femmes enceintes, à la différence des autres sous-populations, la borne supérieure de l'intervalle d'incertitude autour de la médiane n'est pas de 0 mais est égale à 6.E-10.

TAB. 5.5: Estimation de l'exposition à *L. monocytogenes* par consommation de quantités contaminées de salades de IV^{ème} gamme selon différents scénarios

Modèle de croissance	Eau de lavage	Lieu de croissance	Scénarios y_{max}		Proportions de quantités contaminées (%)	Exposition par consommation de quantités contaminées				
						Moyenne	Ecart-type	$P_{0.025}$	$P_{0.5}$	$P_{0.975}$
M2	Chlorée	Sachet, b=3.5	Dist. a post.[a]	Estimation	2.44	2.81	2.15	-1.16	2.78	6.48
				UI_{inf}	1.69	2.18	1.95	-1.60	1.93	5.99
				UI_{sup}	3.51	3.41	2.33	-0.56	3.68	6.85
M2	Chlorée	Sachet, b=0.085	Dist. a post.[a]	Estimation	1.74	2.32	2.31	-1.46	2.11	6.58
				UI_{inf}	1.13	1.75	2.09	-1.60	1.41	5.97
				UI_{sup}	2.43	2.88	2.50	-1.12	2.96	7.13
M2	Chlorée	Sachet, b=3.5	$y_{max}=y_0+4$	Estimation	2.44	2.20	1.71	-1.03	2.15	5.47
				UI_{inf}	1.69	1.76	1.50	-1.37	1.68	4.78
				UI_{sup}	3.51	2.71	1.98	-0.58	2.73	6.52
M2	Chlorée	Sachet, b=3.5	$y_{max}=y_0+2$	Estimation	2.44	1.17	1.39	-1.30	1.08	3.98
				UI_{inf}	1.69	0.79	1.19	-1.50	0.65	3.28
				UI_{sup}	3.51	1.77	1.91	-0.85	1.62	5.91
M2	Non chlorée	Sachet, b=3.5	Dist. a post.[a]	Estimation	5.78	2.60	2.26	-1.30	2.52	6.56
				UI_{inf}	3.97	2.08	2.13	-1.60	1.79	6.26
				UI_{sup}	8.54	3.11	2.40	-1.00	3.26	6.82
$M2\lambda_{obs}$	Chlorée	Sachet, b=3.5	Dist. a post.[a]	Estimation	2.44	3.28	2.18	-1.00	3.44	6.65
				UI_{inf}	1.69	2.58	2.00	-1.30	2.45	6.29
				UI_{sup}	3.51	4.01	2.37	-0.37	4.56	6.91
M2	Chlorée	Quantité	Dist. a post.[a]	Estimation	0.91	5.98	1.85	2.43	6.23	8.89
				UI_{inf}	0.39	5.12	1.47	1.16	5.10	7.94
				UI_{sup}	1.56	7.04	2.20	3.80	7.34	9.69

[a]Distribution *a posteriori* établie à partir du traitement des 206 courbes de croissance dans le chapitre 3

UI_{inf} et UI_{sup} borne inférieure et borne supérieure de l'intervalle d'incertitude à 95%

TAB. 5.6: Estimation du risque de listériose par acte de consommation de salades de IV^{ème} gamme selon différents scénarios

Modèle de croissance	Eau de lavage	Lieu de croissance	Scénarios y_{max}	Valeur de r		Risque par acte de consommation					
						Moyenne	Ecart-type	$P_{0.025}$	$P_{0.5}$	$P_{0.975}$	$P_{0.99}$
M2	Chlorée	Sachet,	Dist. a post.[a]	Médiane	Estimation	8.E-08	4.E-06	0	0	0	7.E-09
		b=3.5			UI_{inf}	9.E-09	3.E-07	0	0	0	2.E-10
					UI_{sup}	4.E-07	2.E-05	0	0	8.E-11	1.E-07
M2	Chlorée	Sachet,	Dist. a post.[a]	Médiane	Estimation	6.E-08	3.E-06	0	0	0	1.E-10
		b=0.085			UI_{inf}	2.E-09	8.E-08	0	0	0	2.E-13
					UI_{sup}	6.E-07	3.E-05	0	0	0	6.E-09
M2	Chlorée	Sachet,	$y_{max}=y_0+4$	Médiane	Estimation	8.E-09	4.E-07	0	0	0	1.E-09
		b=3.5			UI_{inf}	7.E-10	2.E-08	0	0	0	5.E-11
					UI_{sup}	2.E-07	1.E-05	0	0	3.E-11	1.E-08
M2	Chlorée	Sachet,	$y_{max}=y_0+2$	Médiane	Estimation	3.E-10	1.E-08	0	0	0	6.E-11
		b=3.5			UI_{inf}	2.E-11	7.E-10	0	0	0	4.E-12
					UI_{sup}	6.E-08	3.E-06	0	0	2.E-12	9.E-10
M2	Non chlorée	Sachet,	Dist. a post.[a]	Médiane	Estimation	2.E-07	8.E-06	0	0	2.E-09	3.E-07
		b=3.5			UI_{inf}	5.E-08	1.E-06	0	0	7.E-11	3.E-08
					UI_{sup}	7.E-07	3.E-05	0	0	7.E-08	2.E-06
M2	Chlorée	Sachet,	Dist. a post.[a]	$P_{0.025}$	Estimation	2.E-08	9.E-07	0	0	0	1.E-09
		b=3.5			UI_{inf}	1.E-09	4.E-08	0	0	0	2.E-11
					UI_{sup}	1.E-07	4.E-06	0	0	7.E-12	3.E-08
M2λ_{obs}	Chlorée	Sachet,	Dist. a post.[a]	Médiane	Estimation	1.E-07	5.E-06	0	0	0	3.E-08
		b=3.5			UI_{inf}	2.E-08	7.E-07	0	0	0	3.E-10
					UI_{sup}	5.E-07	2.E-05	0	0	2.E-10	4.E-07
M2	Chlorée	Quantité	Dist. a post.[a]	Médiane	Estimation	7.E-08	3.E-06	0	0	0	0
					UI_{inf}	5.E-09	2.E-07	0	0	0	0
					UI_{sup}	2.E-06	1.E-04	0	0	0	6.E-08

[a]Distribution *a posteriori* établie à partir du traitement des 206 courbes de croissance dans le chapitre 3

UI_{inf} et UI_{sup} borne inférieure et borne supérieure de l'intervalle d'incertitude à 95%

TAB. 5.7: Estimation du risque pour l'ensemble de la population de listériose lié à la consommation annuelle de salades de IV^{ème} gamme selon différents scénarios

Modèle de croissance	Eau de lavage	Lieu de croissance	Scénarios y_{max}	Valeur de r		Risque annuel Moyenne	Ecart-type	$P_{0.025}$	$P_{0.5}$	$P_{0.975}$	$P_{0.99}$
M2	Chlorée	Sachet, b=3.5	Dist. a post.[a]	Médiane	Estimation	6.E-06	5.E-05	0	0	3.E-05	2.E-04
					UI_{inf}	2.E-06	2.E-05	0	0	6.E-06	4.E-05
					UI_{sup}	1.E-05	1.E-04	0	0	1.E-04	4.E-04
M2	Chlorée	Sachet, b=0.085	Dist. a post.[a]	Médiane	Estimation	6.E-06	7.E-05	0	0	2.E-05	1.E-04
					UI_{inf}	2.E-06	2.E-05	0	0	3.E-06	2.E-05
					UI_{sup}	1.E-05	1.E-04	0	0	6.E-05	4.E-04
M2	Chlorée	Sachet, b=3.5	$y_{max}=y_0+4$	Médiane	Estimation	8.E-07	1.E-05	0	0	3.E-06	1.E-05
					UI_{inf}	1.E-07	1.E-06	0	0	5.E-07	3.E-06
					UI_{sup}	6.E-06	6.E-05	0	0	2.E-05	1.E-04
M2	Chlorée	Sachet, b=3.5	$y_{max}=y_0+2$	Médiane	Estimation	2.E-08	3.E-07	0	0	9.E-08	4.E-07
					UI_{inf}	3.E-09	3.E-08	0	0	2.E-08	6.E-08
					UI_{sup}	3.E-06	4.E-05	0	0	7.E-06	6.E-05
M2	Non chlorée	Sachet, b=3.5	Dist. a post.[a]	Médiane	Estimation	1.E-05	9.E-05	0	0	1.E-04	4.E-04
					UI_{inf}	6.E-06	4.E-05	0	0	4.E-05	2.E-04
					UI_{sup}	2.E-05	2.E-04	0	0	3.E-04	7.E-04
M2	Chlorée	Sachet, b=3.5	Dist. a post.[a]	$P_{0.025}$	Estimation	1.E-06	1.E-05	0	0	7.E-06	4.E-05
					UI_{inf}	4.E-07	4.E-06	0	0	1.E-06	8.E-06
					UI_{sup}	3.E-06	3.E-05	0	0	3.E-05	1.E-04
$M2_{\lambda_{obs}}$	Chlorée	Sachet, b=3.5	Dist. a post.[a]	Médiane	Estimation	8.E-06	7.E-05	0	0	7.E-05	3.E-04
					UI_{inf}	3.E-06	3.E-05	0	0	2.E-05	9.E-05
					UI_{sup}	2.E-05	1.E-04	0	0	2.E-04	5.E-04
M2	Chlorée	Quantité	Dist. a post.[a]	Médiane	Estimation	0	0	0	0	0	0
					UI_{inf}	0	0	0	0	0	0
					UI_{sup}	9.E-03	3.E-02	0	1.E-04	9.E-02	1.E-01

[a]Distribution *a posteriori* établie à partir du traitement des 206 courbes de croissance dans le chapitre 3

UI_{inf} et UI_{sup} borne inférieure et borne supérieure de l'intervalle d'incertitude à 95%

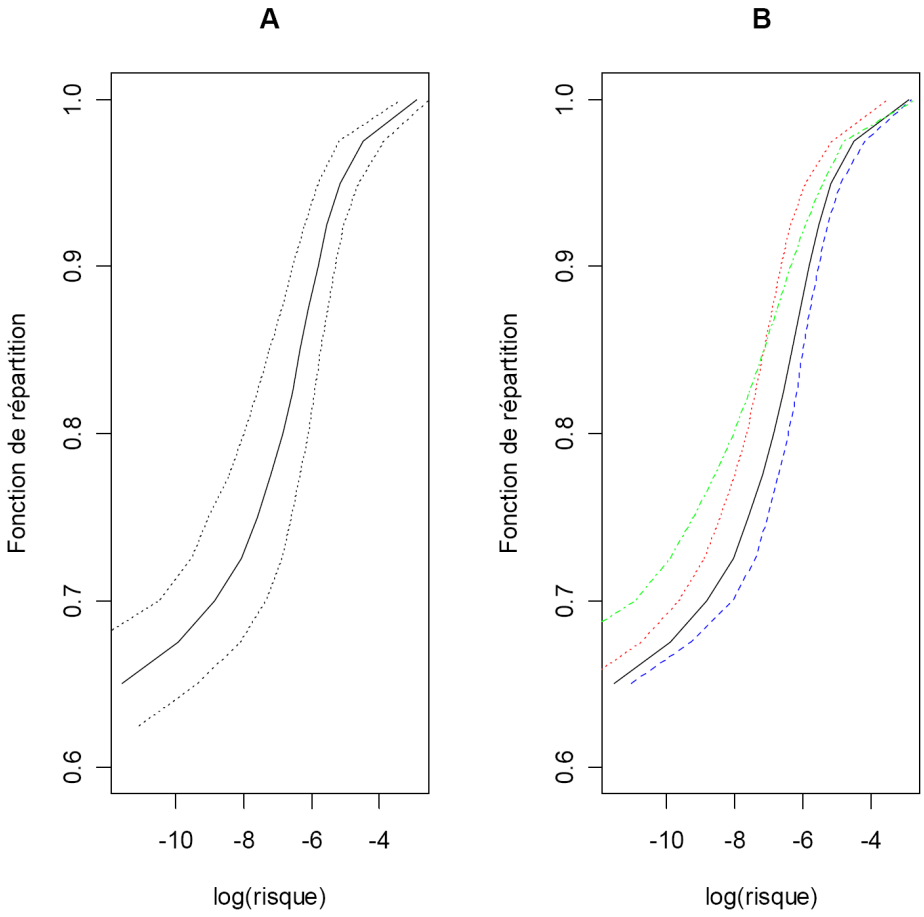

FIG. 5.1: A. Fonction de répartition du risque annuel pour le scénario (M2, chlore, $b = 3.5$ et r médian). Les pointillés représente l'intervalle d'incertitude à 95%. B. Fonction de répartition du risque annuel pour différents scénarios : -.-. : M2, chlore, $b = 0.085$, r médian ; .... : M2, chlore, $b = 3.5$, $P_{0.025}$ pour r; — : M2, chlore, $b = 3.5$, r médian ; - - - : $M2_{\lambda_{obs}}$, chlore, $b = 3.5$, r médian.

TAB. 5.8: Estimation du risque de listériose pour les différentes sous-populations lié à la consommation annuelle de salades de IV^{ème} gamme pour le scénario utilisant le modèle de croissance M2, le lavage à l'eau chlorée, le paramètre de groupement b=3.5 et la valeur médiane pour r

Sous-population		Moyenne	Ecart-type	$P_{0.025}$	$P_{0.5}$	$P_{0.975}$	$P_{0.99}$
Adultes 15-59 ans	Estimation	5.E-07	2.E-06	0	0	6.E-06	1.E-05
(exceptées femmes enceintes)	UI_{inf}	1.E-07	8.E-07	0	0	2.E-06	4.E-06
	UI_{sup}	1.E-06	4.E-06	0	0	1.E-05	2.E-05
Enfants 3-14 ans	Estimation	3.E-07	9.E-07	0	0	1.E-06	3.E-06
	UI_{inf}	8.E-08	3.E-07	0	0	4.E-07	1.E-06
	UI_{sup}	6.E-07	2.E-06	0	0	3.E-06	7.E-06
Personnes âgées 60 ans et plus	Estimation	2.E-05	1.E-04	0	0	3.E-04	6.E-04
	UI_{inf}	6.E-06	3.E-05	0	0	9.E-05	2.E-04
	UI_{sup}	5.E-05	2.E-04	0	0	8.E-04	1.E-03
Femmes enceintes	Estimation	1.E-05	4.E-05	0	0	1.E-04	2.E-04
	UI_{inf}	1.E-06	4.E-06	0	0	7.E-06	2.E-05
	UI_{sup}	6.E-05	2.E-04	0	6.E-10	5.E-04	1.E-03

UI_{inf} et UI_{sup} borne inférieure et borne supérieure de l'intervalle d'incertitude à 95%

5.7.4 Nombre estimé de cas de listériose en France par consommation de salade de IV^{ème} gamme

Le nombre estimé de cas de listériose dans la population française prévu par les différents scénarios sont présentés dans le tableau 5.9. Le nombre estimé de cas est très variable selon le scénario utilisé, entre 0 et 531 cas. Le scénario considérant que le phénomène de regroupement en amas des bactéries est plus important ($b = 0.085$) a peu d'influence sur le nombre de cas : 239 cas avec un UI à 95% de [67 ; 551] cas vs 233 cas avec un UI à 95% de [71 ; 556]. Cette faible différence s'explique par le fait que l'utilisation de la valeur 0.085 pour le coefficient de groupement entraîne une baisse de moins de 1% de quantités contaminées par rapport à la valeur 3.5. Par contre, les scénarios introduisant dans la modélisation de la croissance une dépendance entre $y_{\max}$ et y_0 ont une grande influence sur le nombre de cas estimé : 32 cas avec un UI à 95% de [6 ; 237] cas pour le scénario $y_{\max} = y_0 + 4$ log ufc/g et 1 cas avec un UI à 95% de [0 ; 123] pour le scénario $y_{\max} = y_0 + 2$ log ufc/g. Notons qu'avec ce dernier scénario le nombre de cas peut être nul. L'utilisation du 2.5^{ème} percentile pour la valeur de r a également un effet important sur le nombre de cas : 54 cas avec un UI à 95% de [15 ; 129] cas par an. L'utilisation du modèle de croissance $M2_{\lambda_{obs}}$ entraîne un nombre de cas de listériose important : 351 cas avec un UI à 95% de [130 ; 684] cas. De même le scénario n'intégrant pas le chlore dans le procédé de fabrication prévoit un nombre de cas deux fois plus élevé que le scénario avec chlore : 531 avec un UI à 95% de [233 ; 1031] cas. Le nombre de cas en simulant la croissance directement dans les quantités consommées est estimé à 0 avec un UI à 95% de [0 ; 357270]. L'incertitude est alors très grande autour de ce nombre de cas. Ce chiffre élevé est dû au fait que la borne supérieure de l'intervalle d'incertitude du pourcentage de quantités contaminées est égale à 1.56% et que les concentrations des quantités contaminées peuvent être très élevées pour ce scénario.

5.8 Discussion

Ce chapitre présente les données et la méthodologie utilisée afin d'évaluer l'exposition à *Listeria monocytogenes* par consommation de salade de IV^{ème} gamme et les risques encourus par la population française. Deux points sont discutés dans cette partie : l'intérêt d'utiliser les données individuelles de consommation ainsi que les caractéristiques socio-démographiques des individus d'une part et les résultats obtenus d'autre part.

TAB. 5.9: Nombre de cas de listériose estimé dans la population française par consommation de salade de IV^{ème} gamme selon différents scénarios

Modèle de croissance	Lieu de croissance	Eau de lavage	Scénarios y_{max}	Valeur de r	Nbre cas de listériose Estimation	UI_{inf}	UI_{sup}
M2	Sachet, b=3.5	Chlorée	Dist. a post.[a]	Médiane	233	71	556
M2	Sachet, b=0.085	Chlorée	Dist. a post.[a]	Médiane	239	67	551
M2	Sachet, b=3.5	Chlorée	$y_{max}=y_0+4$	Médiane	32	6	237
M2	Sachet, b=3.5	Chlorée	$y_{max}=y_0+2$	Médiane	1	0	123
M2	Sachet, b=3.5	Non chlorée	Dist. a post.[a]	Médiane	531	233	1031
M2	Sachet, b=3.5	Chlorée	Dist. a post.[a]	$P_{0.025}$	54	15	129
$M2_{\lambda_{obs}}$	Sachet, b=3.5	Chlorée	Dist. a post.[a]	Médiane	351	130	684
M2	Quantité	Chlorée	Dist. a post.[a]	Médiane	0	0	357270

[a] Distribution *a posteriori* établie à partir du traitement des 206 courbes de croissance dans le chapitre 3

UI_{inf} et UI_{sup} borne inférieure et borne supérieure de l'intervalle d'incertitude à 95%

5.8.1 Intérêt d'utiliser les données individuelles de consommation et les caractéristiques socio-démographiques des individus

L'étude de la consommation montre que certaines populations ont des comportements alimentaires particuliers voir extrêmes. Il est donc important dans l'évaluation du risque de prendre en compte la variabilité de la consommation selon les individus et de ne pas uniquement étudier le comportement d'un individu moyen. La plupart des évaluations des risques microbiologiques (Sanaa et al., 2004; Bemrah et al., 1998; Farber et al., 1996), considère que tous les individus consomme une quantité moyenne (ou parfois maximale) du produit étudié un nombre moyen de fois dans l'année, ainsi les comportements à risque ne sont pas pris en compte. Pouillot (2006) prend en compte la variabilité de la consommation de saumon par l'utilisation des données de consommation individuelle INCA mais ne tient pas compte de la dépendance entre la consommation d'une personne et sa sensibilité à *L. monocytogenes*. Or étant donné que la sensibilité à *L. monocytogenes* n'est pas la même pour certaines sous-populations il est nécessaire de pouvoir caractériser le comportement alimentaire de ces personnes afin de prendre en compte la dépendance entre la sensibilité de l'individu et son régime alimentaire. Ce travail permet ainsi de caractériser les populations les plus à risque non plus uniquement par leur sensibilité à la bactérie mais aussi par rapport à leur mode de consommation du produit. Dans notre étude, il s'avère que les populations les plus à risque sont également les populations les plus sensibles à *L. monocytogenes*. Cela n'est pas étonnant puisque la salade de IV^{ème} gamme est consommée à peu près de la même manière par

les différentes sous-populations. Par contre, pour un produit moins couramment consommé que la salade par l'ensemble des sous-populations, la corrélation entre population à risque et population sensible pourrait ne pas être aussi forte. Nous avons aussi mis en évidence que les enfants ont un risque plus faible que les autres sous-populations étant donné qu'ils ont une consommation moins élevée que le reste de la population.

Pour ne pas compliquer l'algorithme de simulation déjà très complexe, nous avons considéré ici que les contaminations des quantités consommées étaient indépendantes. Or, il serait peut-être intéressant dans un travail futur de prendre en compte la dépendance entre la contamination des quantités consommées par différentes personnes appartenant à un même ménage. En effet, le sachet de salade étant consommé en général par plusieurs personnes d'un même foyer, la somme des concentrations bactériennes $(C_{Q,j})_{j=1,\ldots,N}$ des N quantités d'un même sachet doit être égale à la concentration C_W du sachet : $\sum_{j=1}^{N} C_{Q,j} = C_W$. Il faut alors utiliser une loi multinomiale et non plus une loi binomiale lors de la répartition des bactéries dans les N quantités d'un même sachet.

Dans ce travail, afin que les résultats soient extrapolables à la population française, nous avons utilisé des poids de redressement calculés par rapport aux différentes classes d'âge. Cette méthode pourrait être généralisée en prenant en compte dans le calcul des poids l'ensemble des caractéristiques socio-démographiques des individus.

5.8.2 Une estimation du risque de listériose par consommation de salade de IV^{ème} gamme relativement élevé

Le risque de listériose moyen par acte de consommation est d'environ 8.E-8 avec le scénario de base (modèle de croissance $M2$, lavage à l'eau chlorée, coefficient de groupement de 3.5 et r médian). La valeur moyenne du risque est très influencée par les valeurs extrêmes se situant dans la queue de la distribution, il est alors préférable de s'intéresser à la médiane ou directement aux percentiles élevés. Ici la médiane du risque par acte de consommation est égale à 0 et les valeurs au 97.5^{ème} percentile et au 99ème percentile sont respectivement pour le scénario de base : 0 et 7.E-09. La médiane du risque de listériose annuel est égale à 0 pour le scénario de base, les valeurs au 97.5^{ème}percentile et au 99èmepercentile sont respectivement de.3E-05 et de 2.E-04. La fréquence de consommation du produit a une grande influence dans le calcul du risque annuel. La salade de IV^{ème} gamme étant un produit fréquemment consommé, le risque annuel est d'autant plus élevé.

Dans cette modélisation, deux facteurs pouvant avoir une forte influence sur le niveau

de risque n'ont pas été, ou seulement partiellement, pris en considération par manque de données permettant de les modéliser : les interactions entre *Listeria monocytogenes* et la flore totale, et la variabilité de virulence des souches.

Dans l'aliment, la concentration maximale peut être affectée par la présence d'autres micro-organismes (sans que le taux de croissance ne soit modifié) (Carlin et al., 1996c; Buchanan & Bagi, 1997; Campo et al., 2001). Les données extraites de la littérature scientifique que nous avons utilisées pour déterminer les paramètres de notre modèle de croissance ne rendent pas compte de ce phénomène. En effet, les inoculations artificielles sont réalisées avec des inoculums concentrés et donc la concentration maximale est rapidement atteinte. En conditions naturelles, en revanche, les niveaux de contamination sont peu élevés, et proportionnellement à des niveaux incommensurablement moins élevés que ceux de la flore aérobie mésophile. Bien que cela reste à démontrer formellement, on peut alors penser que *Listeria monocytogenes* aura du mal à atteindre la population maximale théoriquement atteignable, en raison de fortes interactions avec une abondante flore mésophile, ou de l'épuisement du milieu qui stoppe les populations bactériennes dès lors que la population majoritaire a atteint son maximum (effet Jameson, Gnanou Besse et al. (2006)). Ainsi les scénarios pour lesquels la valeur de la concentration maximale pouvant être atteinte dépend de la concentration initiale, reproduisent dans une certaine mesure l'influence de la compétition de la flore totale sur la croissance de *L. monocytogenes*. Nous avons montré que le risque est divisé par 10 ou 100 (selon le scénario envisagé) dès lors que l'on fait l'hypothèse d'un plafonnement de la concentration en *L. monocytogenes* lié à sa concentration initiale.

Par ailleurs, le risque est calculé en considérant que toutes les souches de *L. monocytogenes* présentes dans les quantités consommées ont la même probabilité de provoquer la maladie. Or il existe des souches de *L. monocytogenes* dont la virulence est plus ou moins élevée. Le risque obtenu dans ces travaux est donc calculé pour le scénario le pire pouvant être envisagé. En effet, pour l'ensemble des souches de *L. monocytogenes* présentes dans les quantités consommées, il est considéré que celles-ci ont la même probabilité que la souche ayant servi à établir la courbe dose-réponse, c'est à dire une souche particulièrement virulente puisqu'elle est à l'origine d'une toxi-infection.

Comment se situe le risque obtenu pour *L. monocytogenes* dans les salades de IV^{ème} gamme par rapport à d'autres travaux déjà réalisés pour *L. monocytogenes* ? En préambule toutefois, il est important de souligner que dans ces travaux, les produits ou le pays étudiés ne sont pas les mêmes que dans nos travaux, que les méthodes de calculs diffèrent considérablement les unes des autres ainsi que la forme sous laquelle sont donnés les résultats. L'étude

réalisée par l'FDA-USDA (2003) sur le risque de listériose par consommation de légumes prévoit une valeur médiane de 3.E-12 par acte de consommation, les valeurs aux $2.5^{ème}$ et $95^{ème}$ percentiles sont de 3.E-16 et 2.E-10. En comparaison, nous obtenons pour ces percentiles un risque égal à 0. Le risque moyen de listériose lié à la consommation d'une ration de saumon fumé est de 5.E-6 pour la population sensible et de 1.E-7 pour la population normale, les deux types de populations étant étudiées séparément (Pouillot, 2006). La médiane du risque est égal à 0 pour les deux populations. Les moyennes du risque par acte de consommation sont plus élevées que les notre. Bemrah et al. (1998) étudiant le risque par consommation d'une portion de fromage au lait cru, obtiennent une médiane égale à 1.86E-8 pour la population sensible et à 9.74E-13 pour la population normale. Une fois encore les résultats sont plus élevés que ceux obtenus dans notre étude. Précisons que le saumon fumé et les fromages au lait cru sont des produits pour lesquels *L. monocytogenes* est plus fréquemment détectée et avec des concentrations plus élevées que dans la salade.

5.8.3 Un nombre de cas de listériose plausible ?

Incidence de la Listériose invasive en France

L'InVS, chargé de surveiller l'incidence des maladies en France mentionne que l'incidence des bactériémies et des méningites à *Listeria monocytogenes*, relativement stable de 1987 à 1989, chute brutalement de 1989 à 1990. Malgré le pic épidémique de 1992, la tendance à la baisse se poursuit avec une réduction d'un facteur 3,5 du nombre annuel de bactériémies et de méningites pour la période 1987-1997. Sur la même période, à partir de plusieurs enquêtes nationales sur la listériose on observe une réduction du nombre de nouveaux cas annuels sporadiques par 3 (1987 : 661 cas, 1992 : 458 cas, 1997 : 225 cas). De 1987 à 1997, le nombre de cas de moins de 65 ans sans terrain prédisposant a été divisé par 8, le nombre de cas associés à la grossesse par 7, le nombre de sujets âgés de plus de 65 ans par 3 et le nombre de sujets immunodéprimés par 1.4. L'incidence de la listériose invasive en 2002 est de 3.66 pour un million d'habitants. Ainsi le nombre de cas déclarés de listériose en France en 2002 est 220.9, soit 3.66×60.35. Une étude plus récente de Vaillant et al. (2005) estime le nombre de cas de listériose par an à environ 300, tout aliment confondu. Précisons que bien que l'incidence de la listériose en France soit hautement surveillée, le nombre de cas fourni par l'InVS représente uniquement le nombre de cas déclaré, ce qui laisse supposer que le nombre de cas réel serait plus important.

L'estimation du nombre de cas de listériose obtenu avec notre modèle varie entre 0 et

531 selon le scénario envisagé. Etant donné que la consommation de salade de IV^{ème} gamme n'est pas la seule cause de listériose en France, il est clair que le nombre de cas est surestimé pour la plupart des scénarios envisagés. On peut attribuer la surestimation du nombre de cas aux hypothèses conservatrices utilisées dans la modélisation (absence de temps de latence, absence d'interactions entre *L. monocytogenes* et la flore totale) ou encore à la prise en compte des comportements extrêmes au niveau de la consommation, des durées et des températures de conservation des produits. Ces hypothèses sont fondées pour la plupart sur de réelles observations (comme par exemple l'observation de températures de conservation très élevées) ou pour certaines sur l'absence d'information les contredisant (par exemple pour l'interaction entre les différentes bactéries). Or, par exemple, prendre en compte une possible compétition entre la flore totale et *L. monocytogenes* (scénarios pour lesquels $y_{\max}$ dépend de y_0) qui aurait pour effet de limiter la croissance de cette dernière a pour conséquence une importante diminution du nombre estimé de cas de listériose. Ces résultats confirment ceux déjà obtenus dans le chapitre 4 sur l'importance de considérer une possible limitation de la croissance de *L. monocytogenes*. Cependant, les recherches doivent être encore approfondies afin de mieux définir ce mécanisme qui dépend en particulier de la matrice alimentaire.

Enfin, le nombre de cas obtenu avec le scénario intégrant le retrait du chlore de l'eau de lavage, est estimé deux fois supérieur à celui obtenu en considérant le lavage à l'eau chlorée. Cette élévation du risque dû à la suppression du chlore était attendue, mais sa quantification est une information intéressante et totalement nouvelle. La suppression du chlore de l'industrie de la IV^{ème} gamme est un enjeu industriel important et particulièrement d'actualité. C'est typiquement ce genre d'estimation de risque qui peuvent amener les industriels à choisir certaines options de gestion comme la réduction du niveau de contamination de la matière première en cas afin d'éliminer le chlore des usines.

Caractéristiques des produits incriminés dans les cas simulés de listériose

Nous avons effectué une première analyse afin de déterminer quelles sont les caractéristiques des produits (concentrations, conditions de conservation, quantité consommée, caractéristiques du consommateur) incriminés dans les cas de listériose. Cette analyse est effectuée pour le scénario de base : modèle de croissance $M2$, lavage à l'eau chlorée, coefficient de groupement de 3.5 et r médian. Il s'avère que les cas de listériose sont principalement dus à la consommation de quantités très fortement contaminées par des individus appartenant à la population sensible. Les concentrations moyennes des quantités consommées par ces personnes sont supérieures à 6 log ufc/g. Le temps total ainsi que la température moyenne de conservation des sachets produisant des quantités contaminées à des concentrations su-

périeures à 6 log ufc/g sont présentés dans la figure 5.2. Les températures moyennes de conservation de ces sachets tout au long du circuit logistique varient entre 2°C et 9°C, avec une moyenne à 5°C. Ces températures bien qu'en moyenne supérieure à la température de 4°C recommandée par les industriels ne sont pas très élevées par rapport à certaines valeurs pouvant être observées (cf. chapitre 4). Les durées de conservation des sachets correspondant à de fortes concentrations dans les quantités consommées sont par contre très élevées (entre 84h et 288h avec une moyenne de 220h). La distribution des taux de croissance produisant ces concentrations élevées n'est pas significativement différente de la distribution du taux de croissance utilisé pour l'ensemble des quantités consommées (résultats non montrés). Les concentrations supérieures à 6 log ufc/g seraient donc essentiellement dues à des durées de conservation du produit proche et même supérieure à la date limite de conservation (DLC). Nous avons considéré dans les simulations que la DLC pouvait être dépassée dans 20% des cas et que dans 56.7% des cas les sachets étaient consommés en deux temps séparés. Ces durées ont été calculées en considérant l'hypothèse la plus conservatrice pour le risque mais elles semblent toutefois surestimées. En effet, la salade en sachet est un produit dont l'altération par la flore microbienne est visible (nécroses, brunissements tendant à s'accroître avec la durée de conservation). On peut alors penser que dans la majorité des cas le sachet est jeté lorsque la DLC est dépassée. Notre étude est une des premières réalisées sur la question et les estimations qu'elles proposent doivent être corroborées par d'autres analyses des risques menées de manière indépendante. Cependant, dans la mesure où la plupart des cas de listériose sont liés à des durée de vie prolongées, l'effet d'un meilleur respect de la DLC sur l'estimation du risque de listériose mériterait d'être d'évalué. D'autres scénarios pourraient également être envisagés comme par exemple l'utilisation d'une date limite de conservation moins longue, ou également d'une simulation de la conservation avec 3 jours à 4°C et 6 jours à 8°C, classiquement appliquée dans les protocoles de détermination de durée de vie et l'estimation du nombre de produits dépassant la norme de 100 *L. monocytogenes*/g fixée par le règlement européen 2073 de l'année 2005.

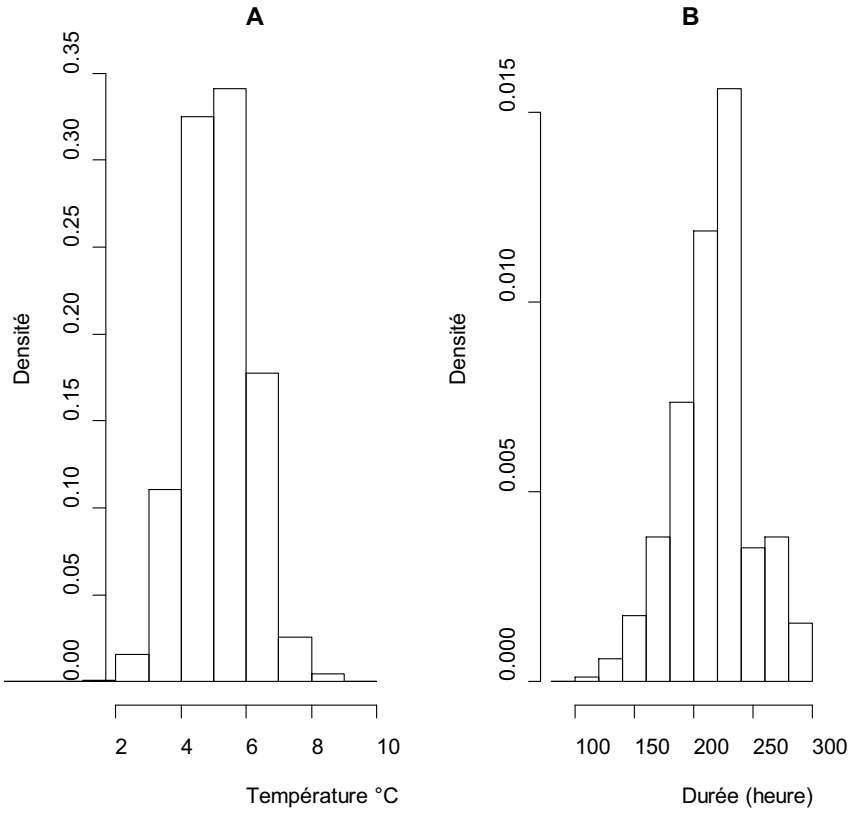

FIG. 5.2: Histogrammes des températures moyennes (A) et des durées (B) du circuit logistique des quantités consommées dont la concentration est supérieure à 6 log ufc/g.

Chapitre 6

Discussion générale

Ce travail de thèse présente un modèle d'évaluation des risques lié à la présence de *L. monocytogenes* dans les salades de IV^{ème} gamme. L'objectif était de représenter au mieux les phénomènes biologiques (contamination, croissance, décroissance), les procédés technologiques (procédé de fabrication, circuit logistique) et le comportement humain (mode de conservation, habitudes de consommation) ayant une influence sur les concentrations bactériennes afin d'évaluer avec précision l'exposition et le risque des consommateurs de salade de IV^{ème} gamme. On doit néanmoins insister sur les limites des méthodes utilisées et sur la spécificité d'une telle recherche dans le cadre d'une évaluation des risques.

6.1 Limites des travaux et pistes de recherche

La modélisation est un mécanisme intellectuel consistant à conceptualiser des phénomènes réels dans le but de les analyser et de les prévoir. Malgré toute l'inventivité et l'ingéniosité dont le modélisateur peut faire preuve, les modèles restent néanmoins une simplification de la réalité et sont construits à partir de nombreuses hypothèses qui peuvent toujours être discutées et remises en cause.

En particulier, une hypothèse souvent posée en analyse des risques microbiologiques est celle d'homogénéité de la contamination dans les aliments. Cette hypothèse est utilisée d'un point de vue pratique mais semble peu réaliste et conduit à des situations contradictoires. Dans ces travaux, des outils ont été mis en place afin de prendre en compte le groupement des bactéries en amas. Ainsi la loi Beta-Binomiale a été utilisée pour établir la contamination des quantités consommées à partir de la contamination des sachets. Toutefois, cette méthode est à améliorer notamment par une estimation du coefficient de groupement à partir de réelles

observations. De même, des améliorations sont nécessaires au niveau de la transmission des bactéries de la matière première au sachet, puisque ici le phénomène de groupement n'a pas été pris en considération.

Nous avons voulu testé l'hypothèse considérant une limitation de la croissance de *L. monocytogenes* dans la salade de IV^{ème} gamme. Les résultats obtenus montrent que la prise en compte de ce phénomène a une forte influence sur le risque et que ne pas considérer ce phénomène peut être source d'une surestimation importante du nombre de cas. En attendant la validation de ce phénomène par des recherches en microbiologie des aliments, il semble nécessaire de l'intégrer dans l'évaluation des risques microbiologiques.

Les résultats obtenus par simulations de Monte-Carlo de second ordre ont montré qu'il était nécessaire de prendre en compte et de propager séparément la variabilité et l'incertitude. De plus, la combinaison de méthodes de simulation de Monte-Carlo et de méthodes d'inférence bayésienne s'avère très utile. En effet, l'inférence bayésienne intégrant des modèles hiérarchiques en produisant des distributions *a posteriori* pour les paramètres et les hyperparamètres facilite la détermination des paramètres reflétant la variabilité des paramètres de ceux reflétant l'incertitude. Toutefois ces deux techniques nécessitent des temps de simulations très longs (plus de 12h pour certains).

Une des difficultés majeures en modélisation est de valider les modèles proposés. En particulier, il est très difficile de valider les modèles d'évaluation du risque puisqu'ils sont souvent très complexes et que l'observation des nombreux paramètres étudiés (contamination, risque, nombre de cas) est limitée aux seules enquêtes effectuées, celles-ci ne pouvant être exhaustives. Leurs résultats sont déjà intégrés dans le modèle et ne peuvent alors servir pour valider. Il est toutefois possible et nécessaire de situer les résultats proposés par rapport à des études indépendantes de celles qui ont servi à l'estimation ou de les confronter avec des analyses de risques effectuées pour d'autres produits. C'est ce que nous nous sommes attachés à faire tout au long de ces travaux.

6.2 Statistique et évaluation des risques : un dispositif complexe

La spécificité de ce travail en statistique est de s'intégrer dans le processus de l'évaluation des risques. Ceci nécessite des collaborations pluridisciplinaires, des échanges avec les industriels ainsi qu'avec les gestionnaires du risque.

6.2.1 Pluridisciplinarité scientifique

L'évaluation des risques microbiologiques dans les aliments propose un domaine d'application des statistiques riche au niveau des problématiques abordées, des modèles envisagés et des collaborations possibles. Ce champ d'application des statistiques fait également intervenir d'autres domaines de recherche. En effet, l'évaluation des risques microbiologiques dans les aliments nécessite de rassembler un grand nombre d'informations, de techniques et de données pertinentes issues de domaines divers. Ces domaines sont la microbiologie des aliments, la microbiologie prévisionnelle, l'épidémiologie, la vieille sanitaire, l'étude de la consommation et du comportement du consommateur, le génie des procédés industriels et la technologie alimentaire. Ainsi l'évaluation des risques nécessite que l'évaluateur possède, en plus de la maîtrise de la statistique, une bonne connaissance des techniques développées et des ressources disponibles dans chacun de ces domaines. De même, la collaboration étroite entre l'évaluateur et les personnes spécialistes des différents domaines auxquels l'évaluation des risques fait appel est indispensable afin de produire des évaluations de qualité. Ces collaborations sont aussi nécessaires afin d'améliorer non pas la qualité des données produites dans les différents domaines mais leur pertinence pour l'évaluateur et le modélisateur. En effet, l'évaluateur peut orienter les analyses effectuées en amont de l'évaluation des risques afin que les résultats produits soient plus pertinents et plus facilement intégrables. Prenons l'exemple des données de contamination microbiologique. Ces données sont pour la plupart issues d'un grand nombre d'analyses, parfois près de 4 000 échantillons analysés dans une seule étude, cependant, très souvent, seule la présence du contaminant a été relevée et aucune information n'est fournie sur le nombre de bactéries présentes. Or, bien que la prévalence de la contamination soit un paramètre important dans l'évaluation des risques, la quantification de la dose de contaminant présent dans l'aliment est un paramètre non moins important puisque la sévérité de la réponse en dépend. L'analyse d'un nombre plus réduit de produits mais comportant systématiquement la numération des bactéries semble plus pertinente. Ainsi, l'évaluateur peut aider le microbiologiste à définir ses plans d'expériences afin que celui-ci produise des résultats à la fois intéressants dans son domaine mais qui peuvent aussi être réutilisés dans d'autres domaines comme celui de l'évaluation des risques. De même, le dialogue constant entre les personnes produisant les données et l'évaluateur, souvent de formation statistique, permet que celui-ci soit moins exigeant au niveau des données et réticent à l'application de ces compétences dans ce domaine.

6.2.2 Collaboration avec l'industrie

De plus, la collaboration entre l'évaluateur et les entreprises, même si celle-ci peut être controversée puisqu'il peut y avoir conflits d'intérêts, est très enrichissante pour les deux parties lorsque celle-ci est bien menée. Les industriels peuvent en effet fournir des éléments importants dans l'évaluation des risques comme les données de contamination, des informations sur le procédé de fabrication. L'expertise terrain des responsables qualités des entreprises peut être très utile pour pallier le manque de données numériques, on l'a vu par exemple pour les durées des étapes du circuit logistique, ou pour accentuer les recherches vers les facteurs importants de contamination. Ils peuvent aussi aider l'évaluateur à envisager des scénarios de réduction des risques techniquement réalisables.

6.2.3 Communication avec les gestionnaires

Les résultats de ces travaux sont destinés à appuyer les discussions des gestionnaires du risque. C'est pourquoi l'évaluateur du risque pose ses hypothèses afin que celles-ci soient conservatrices, c'est-à-dire qu'il sélectionne parmi les hypothèses possibles celles qui produiront le risque le plus élevé. De cette manière, si le risque est faible et considéré comme "acceptable", il n'est pas nécessaire de mettre en place des mesures de gestion. Dans le cas contraire, les hypothèses sont tout d'abord vérifiées. Puis le modèle est utilisé pour évaluer l'impact sur le risque de différents scénarios de mesure de gestion. Ainsi, dans le cas d'un risque avéré comme la listériose, les modèles d'évaluation des risques sont conçus afin d'identifier les paramètres les plus influents et de tester différentes mesures de gestion du risque. Nous avons déjà évalué l'impact de plusieurs scénarios sur les niveaux de concentration et sur le risque. Il serait intéressant, dans un travail futur de mettre en oeuvre, d'autres techniques d'analyse de sensibilité plus poussées.

Les travaux développés dans cette thèse sont des travaux de recherche intégrés dans le processus d'analyse du risque. L'objectif était alors de développer des méthodes afin d'intégrer l'ensemble des événements pouvant se produire ainsi que l'incertitude autour des estimations. De ce fait, les résultats obtenus sont nombreux et leur forme complexe (distributions, intervalle d'incertitude, etc...), ce qui peut entraîner des difficultés d'interprétation pour les gestionnaires du risque ou encore pour les industriels. Cependant, nous avons organisé tout au long de ce travail, des discussions avec les personnes intéressées afin d'expliquer et de simplifier les résultats.

Bibliographie

Adams, M., Hartley, A. & Cox, L. (1989). Factors affecting the efficacy of washing procedures used in the production of prepared salads. *Food Microbiol.* **6**, 69–77.

Adams, M. R., Little, C. L. & Easter, M. C. (1991). Modelling the effect of ph, acidulant and temperature on the growth rate of *Yersinia enterocolitica*. *J. Appl. Bacteriol.* **71**, 65–71.

Afchain, A. L. (2005). Couplage de modèles thermique et microbiologique en vue de l'évaluation de l'exposition à un risque sanitaire. Rapport interne. Cemagref. Tech. rep.

Afchain, A. L., Derens, E., Guilpart, J. & Cornu, M. (2005). Statistical modelling of cold-smoked salmon thermal profiles for risk assessment of *Listeria monocytogenes*. *Acta Horticulturae* **674**, 383–388.

Afssa (2006). Agence française de sécurité sanitaire des aliments. Construction d'une démarche interdisciplinaire de description du processus sanitaire modulant l'exposition au danger *Listeria monocytogenes* dans les produits réfrigérés. Tech. rep.

Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle*. Intl. Symp. on Information Theory. Akadémiai Kiado, Budapest, B. Petrov and F. Csaki ed.

Allais, O. & Tressou, J. (2006). Using decomposed household food aquisitions as inputs of a kinetic dietary exposure model. Document de travail.

Arnold, G. & Coble, J. (1995). Incidence of *Listeria species* in foods in NSW. *Food Aust.* **47**, 71–75.

Arumugaswamy, R., Gulam, R. & Siti, N. (1994). Prevalence of *Listeria monocytogenes* in foods in malaysia. *Int. J. Food Microbiol.* **23**, 117–121.

Augustin, J. C. & Carlier, V. (2000a). Mathematical modelling of the growth rate and lag time for *Listeria monocytogenes*. *Int. J. Food Microbiol.* **56**, 29–51.

Augustin, J. C. & Carlier, V. (2000b). Modelling the growth rate of listeria monocytogenes with a multiplicative type model including interactions between environmental factors. *Int. J. Food Microbiol.* **56**, 53–70.

Aureli, P., Fiorucci, G. C., Caroli, D., Marchiaro, B., Novara, O., Leone, L. & Salmoso, S. (2000). An outbreak of febrile gastroenteritis associated with corn contaminated by listeria monocytogenes. *New England Journal of Medicine* **342**, 1236–1241.

Baranyi, J. & Roberts, T. A. (1994). A dynamic approach to predicting bacterial-growth in food. *Int. J. Food Microbiol.* **23**, 277–294.

Baranyi, J., Roberts, T. A. & McClure, P. J. (1993). A non-autonomous differential equation to model bacterial growth. *Food microbiol.* **10**, 43–59.

Barker, G., Malakar, P. & Peck, M. (2005). Germination and growth from spores : variability and uncertainty in the assessment of food borne hazards. *Int. J. Food Microbiol.* **100**, 67–76.

Barlow, S. M., Greig, J. B., Bridges, J. W., Carere, A., Carpy, A. J. M., Galli, C. L., Kleiner, J., Knudsen, I., Koëter, H. B. W. M. & Levy, L. S. (2002). Hazard identification by methods of animal-based toxicology. *Food and Chemical Toxicology* **40**, 145–191.

Baur, S., Klaiber, R., Wei, H., Hammes, W. P. & Carle, R. (2005). Effect of temperature and chlorination of pre-washing water on shelf-life and physiological properties of ready-to-use iceberg lettuce. *Inn. Food Sci. Emerg. Technol.* **6**, 171–182.

Beaufort, A., Poumeyrol, G. & Rudelle, S. (1992). Fréquence de contamination par *Listeria* et *Yersinia* d'une gamme de produits de $4^{ème}$ gamme. *Rev. Gen. Froid* **27**, 28–31.

Beckers, H., In't-Veld, P., Soentoro, P. & Delfgou van Asch, E. (1988). The occurrence of *Listeria* in food. Foodborne listeriosis. In *Proceedings of a symposium on September 7, 1988 in Wiesbaden, FRG. Behr's Verlag, Hambourg.*

Bemrah, N., Bergis, H., Colmin, C., Beaufort, A., Millemann, Y., Dufour, B., Benet, J. J., Cerf, O. & Sanaa, M. (2003). Quantitative risk assessment of human salmonellosis from the consumption of a turkey product in collective catering establishments. *Int. J. Food Microbiol.* **80**, 17–30.

Bemrah, N., Sanaa, M., Cassin, M. H., Griffiths, M. W. & Cerf, O. (1998). Quantitative risk assessment of human listeriosis from consumption of soft cheese made from raw milk. *Prev. Vet. Med.* **37**, 129–45.

Beuchat, L. (1996). *Listeria monocytogenes* : incidence of vegetables. *Food Cont.* **7**, 223–228.

Beuchat, L. & Brackett, R. (1990). Survival and growth of *Listeria monocytogenes* on lettuce as influenced by shredding, chlorine treatment, modified atmosphere packaging and temperature. *J. Food Sci.* **55**, 755–758, 870.

Beuchat, L. R., Adler, B. B. & Lang, M. M. (2004). Efficacy of chlorine and a peroxyacetic acid sanitizer in killing *Listeria monocytogenes* on iceberg and romaine lettuce using simulated commercial processing conditions. *J. Food Prot.* **67**, 1238–1242.

Bill, J. (1990). *Epidemiology of listeriosis in Europe, with special reference to the Swiss outbreak. Foodborne Listeriosis.* New York : Miller, A. J. and Smith, J. L. and Somkuti, G. A.

Bind, J. (1992). La listériose. *Bull. Group. Tech. Vet.* **89(5)**, 59–77.

Bloch, J. (1998). *Modèlisation de mesures répétées de titres obtenues par la mtéthode des dilutions. Application au suivi des virémies VIH.* Thèse de doctorat, Faculté de médecine Paris-Sud, ParisXI.

Bloch, J. & Chavance, M. (1998). A mixed model for repeated dilution assays. *Biometrics* **54**, 482–92.

Blodgett, R. J. & Garthright, W. E. (1998). Several MPN models for serial dilutions with suppressed growth at low dilutions. *Food. Microbiol.* **15**, 91–99.

Boardman, A., Greenberg, D., Aidan, R. & Weimer, D. (1996). *Cost-Benefit : Concepts and Practice.* New Jersey : Prentice Hall. Upper Saddle River.

Boizot, C. (2005). Présentation du panel de données SECODIP. Tech. rep., INRA-CORELA.

Borraz, O., Besançon, J. & Clergeau, C. (2006). *It is just about trust ? The Partial Reform of French Food Safety Regulation. Whatś the Beef ? The Contested Governance of European Food Safety.* Ansell, C. and Vogel, D. MIT Press.

Breer, C. & Baumgartner, A. (1992). Vorkommen un verhalten von *Listeria monocytogenes* auf Salaten und Gemüsen sowie in frischgeprebten Gemüsesäften. *Arch. Lebensmittelhyg.* **43**, 108–110.

Breer, C. & Schopfer, K. (1989). Listerien in nahrungsmitteln. *Schweiz Med. Wochenschr.* **119**, 306–311.

Brooks, P. & Gelman, A. (1998). General method for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.* **7**, 434–455.

Broome, C. V., Gellin, B. & Schwartz, B. (1990). *Epidemiology of listeriosis in the United States. Foodborne Listeriosis.* New York : Miller, A. J. and Smith, J. L. and Somkuti, G. A.

Buchanan, R. & Cygnarowicz, M. (1990). A mathematical approach toward defining and calculating on the duration of the lag phase. *Food Microbiol.* **7**, 237–240.

Buchanan, R., Whiting, R. & Damert, W. (1996). When is simple good enougth. comparison of the baranyi, the gompertz and the three linear models for fitting bacterial growth curves. In *International Conference on Predictive Microbiology.* Hobart.

Buchanan, R. L. & Bagi, L. K. (1997). Microbial competition : effect of culture conditions on the suppression of *Listeria monocytogenes Scott A* by *Carnobacterium piscicola. J. Food Prot.* **60**, 254–261.

Burnett, A. B., Iturriaga, M. H., Escartin, E. F., Pettigrew, C. A. & Beuchat, L. R. (2004). Influence of variations in methodology on populations of *Listeria monocytogenes* recovered from lettuce treated with sanitizers. *J. Food Prot.* **67**, 742–750.

Campo, J. D., Carlin, F. & Nguyen-The, C. (2001). Effects of epiphytic *Enterobacteriaceae* and *pseudomonas* on the growth of *Listeria monocytogenes* in model media. *J. Food Prot.* **64**, 721–724.

Carlin, F., Broussolle, V., Perelle, S., Litman, S. & Fach, P. (2004). Prevalence of *Clostridium botulinum* in food raw materials used in REPFEDs manufactured in france. *Int. J. Food. Microbiol.* **91**, 141–145.

Carlin, F., Nguyen-the, C. & Abreu da Silva, A. (1995). Factors affecting the growth of *Listeria monocytogenes* on minimally processed fresh endive. *J. Appl. Bacteriol.* **78**, 636–646.

Carlin, F., Nguyen-the, C., Abreu da Silva, A. & Cochet, C. (1996a). Effects of carbon dioxide on the fate of *Listeria monocytogenes*, of aerobic bacteria and on the development of spoilage in minimally processed fresh endive. *Int. J. Food Microbiol.* **32**, 159–172.

Carlin, F., Nguyen-the, C. & Morris, C. (1996b). Influence of background microflora on *Listeria monocytogenes* on minimally processed fresh endive (cichorium endivia var. latifolia). *J. Food Prot.* **59**, 698–703.

Carlin, F., Nguyen The, C. & Morris, C. E. (1996c). Influence of background microflora on *Listeria monocytogenes* on minimally processed fresh broad-leaved endive (*Cichorium endivia var. latifolia*). *J. Food Prot.* **59**, 698–703.

Carriquiry, A. L., Jensen, H. H. & Nusser, S. M. (1990). Modeling chronic versus acute human risk from contaminants in food. Tech. Rep. 90-WP 69, Center for Agricultural and Rural Development.

Cassin, M. H., Lammerding, A. M., Todd, E. C., Ross, W. & McColl, R. S. (1998). Quantitative risk assessment for *Escherichia coli* O157 :H7 in ground beef hamburgers. *Int. J. Food Microbiol.* **41**, 21–44.

Chesher, A. (1997). Diet revealed ? : Semiparametric estimation of nutrient intake-age relationships. *Journal of the Royal Statistical Society A* **160**, 389–428.

Cho, S. Y., Park, B. K., Moon, K. D. & Oh, D. H. (2004). Prevalence of *Listeria monocytogenes* and related species in minimally processed vegetables. *J. Microbiol. Biotechnol.* **14**, 515–519.

Clough, H. E., Clancy, D., O'Neill, P. D., Robinson, S. E. & French, N. P. (2005). Quantifying uncertainty associated with microbial count data : A bayesian approach. *Biometrics* **61**, 610–616.

Cochran, W. G. (1950). Estimation of bacterial densities by means of the "most probable number". *Biometrics* **6**, 105–16.

Codex Alimentarius Commission (2003). *Principles and guidelines for the conduct of microbiological risk assesment.* Roma : FAO.

Crainiceanu, C., Stedinger, J. R., Ruppert, D. & Behr, C. T. (2003). Modeling the U.S. national distribution of waterborne pathogen concentrations with application to *Cryptosporidium parvum*. *Water resources research* **39(9)**, 1235–1950.

CREDOC-AFSSA-DGAL (1999). *Enquête INCA (individuelle et nationale sur les consommations alimentaires)*. Lavoisier, Paris, TEC&DOC ed. (Coordinateur : J.L. Volatier).

Crépet, A., Albert, I., Dervin, C. & Carlin, F. (2007a). Estimation of microbial contamination of food from prevalence and concentration data : Application to *Listeria monocytogenes* in fresh vegetables. *Appl. Environ. Microbiol.* **73**, 250–258.

Crépet, A., Harari-Kermadec, H. & Tressou, J. (2007b). Combining data by empirical likelihood. Document de travail.

Curtis, M., de Franceschi, O. & de Castro, N. (2002). *Listeria monocytogenes* in vegetables minimally processed ready-to-use. *Arch. Latinoam. Nutr.* **52**, 282–288.

Dalton, C., Austin, C. C., Sobel, J., Hayes, P. S., Bibb, W. F., Graves, L. M. & Swaminathan, B. (1997). An outbreak of gastroenteritis and fever due to *Listeria monocytogenes* in milk. *New England Journal of Medicine* **336**, 100–105.

de Simon, M., Tarrago, C. & Ferrer, M. D. (1992). Incidence of *Listeria monocytogenes* in fresh foods in Barcelona (Spain). *Int. J. Food Microbiol.* **16**, 153–156.

Delaquis, P., Stewart, S., Cazaux, S. & Toivonen, P. (2002). Survival and growth of *Listeria monocytogenes* and *Escherichia coli O157 : H7* in ready-to-eat iceberg lettuce washed in warm chlorinated water. *J. Food Prot.* **65**, 459–464.

Delignette-Muller, M. L. (1998). Relation between the generation time and the lag time of bacterial growth kinetics. *Int. J. Food Microbiol.* **43**, 97–104.

Delignette-Muller, M. L., Cornu, M., Pouillot, R. & Denis, J. B. (2006). Use of bayesian modelling in risk assessment : application to growth of *Listeria monocytogenes* and food flora in cold-smoked salmon. *Int. J. Food Microbiol.* **106**, 195–208.

Delignette-Muller, M. L. & Rosso, L. (2000). Biological variability and exposure assessment. *Int. J. Food Microbiol.* **58**, 203–12.

DGCCRF (1993-1996 and 1997-2001). Plan de surveillance de la contamination des aliments par *Listeria monocytogenes*, internal report.

Dhanashree, B., Otta, S., Karunasagar, I., Goebel, W. & Karunasagar, I. (2003). Incidence of *Listeria* spp. in clinical and food samples in Mangalore, India. *Food Microbiol.* **20**, 447–453.

Dodds, K. L. (1993). *Clostridium botulinum. Ecology and control in foods.* New York : Marcel Dekker.

Dybing, E., Doe, J., Groten, J., Kleiner, J., O'Brien, J., Renwick, A. G., Schlatter, J., Steinberg, P., Tritscher, A., Walker, R. & Younes, M. (2002). Hazard characterisation of chemicals in food and diet : dose response, mechanisms and extrapolation issues. *Food and Chemical Toxicology* **40**, 237–282.

Efron, B. & Tibshirani, J. T. (1993). *An introduction to the bootstrap.* Chapman & Hall.

Evans, J. (1998). *Consumer perceptions and practice in the handling of chilled foods. Sous vide and Cook Chill Processing for the Food Industry.* Gaithersburg, Maryland, USA : Aspen Publishers, Ghazala, S. ed.

Evans, J., Stanton, J. I., Russel, S. L. & James, S. J. (1991). *Consumer handling of chilled foods : A survey of time and temperature conditions.* London : MAFF.

FAO/WHO (2004). Risk assessment of *Listeria monocytogenes* in ready-to-eat foods. microbiological risk assessment series, No. 5.

Farber, J., Ross, W. & Harwig, J. (1996). Health risk assessment of *Listeria monocytogenes* in canada. *Int. J. Food Microbiol.* **30**, 145–156.

Farber, J., Sanders, G. & Johnston, M. (1989). A survey of various foods for the presence of *Listeria species. J. Food Prot.* **52**, 456–458.

Farber, J. M., Wang, S. L., Cai, Y. & Zhang, S. (1998). Changes in populations of *Listeria monocytogenes* inoculated on packaged fresh-cut vegetables. *J. Food Prot.* **61**, 192–195.

FDA-USDA (2003). Quantitative assessment of relative risk to public health from foodborne *Listeria monocytogenes* among selected categories of ready-to-eat foods. FDA-Center for Food Safety and Applied Nutrition and USDA-Food Safety and Inspection Service and Centers for Disease Control and Prevention.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London,* **Ser. A, 222**, 309–368.

Francis, G. & O'Beirne, D. (1997). Effects of gas atmosphere, antimicrobial dip and temperature on the fate of listeria innocua and *Listeria monocytogenes* on minimally processed lettuce. *Int. J. Food Sci. Technol.* **32**, 141–151.

Francis, G. & O'Beirne, D. (2001). Effects of acid adaptation on the survival of *Listeria monocytogenes* on modified atmosphere packaged vegetables. *Int. J. Food Sci. Technol.* **36**, 477–487.

Francis, G. A. & O'Beirne, D. (2005). Variation among strains of *Listeria monocytogenes* : differences in survival on packaged vegetables and in response to heat and acid conditions. *Food Cont.* **16**, 687–694.

Francis, G. A. & O'Beirne, D. (2006). Isolation and pulsed-field gel electrophoresis typing of *Listeria monocytogenes* from modified atmosphere packaged fresh-cut vegetables collected in ireland. *J. Food Prot.* **69**, 2524–2528.

Garcia-Gimeno, R., Zurera-Cosano, G. & Amaro-Lopez, M. (1996a). Incidence, survival and growth of *Listeria monocytogenes* in ready-to-use mixed vegetable salads in Spain. *J. Food Saf.* **16**, 75–86.

Garcia-Gimeno, R., Zurera-Cosano, G. & Amaro-Lopez, M. (1996b). Incidence, survival and growth of *Listeria monocytogenes* in ready-to-use mixed vegetable salads in Spain. *J. Food Saf.* **16**, 75–86.

Garriga, M., Aymerich, M. T., Costa, S., Monfort, J. M. & Hugas, M. (2002). Bactericidal synergism through bacteriocins and high pressure in a meat model system during storage. *Food Microbiol.* **19**, 509–518.

Gelfand, A. & Smith, A. (1990). Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* **85**, 398–409.

Gelman, A., Carlin, J., Stern, H. & Rubin, D. (2004a). *Bayesian Data Analysis.* Boca Raton : Chapman & Hall, 2nd ed.

Gelman, A., Carlin, J., Stern, H. & Rubin, D. (2004b). *Bayesian Data Analysis.* Boca Raton : Chapman and Hall, 2nd ed.

Geman, S. & Geman, D. (1984). Stochastic relaxation, gibbs distribution and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-6**, 721–741.

Gill, C. O. (1996). The hygienic condition of manufacturing beef destined for the manufacture of hamburger patties. *Food Microbiol.* **13**, 391–396.

Gill, C. O., Rahn, K., Sloan, K. & McMullen, L. M. (1997). Assessment of the hygienic performances of hamburger patty production processes. *Int. J. Food. Microbiol.* **36**, 171–8.

Gnanou Besse, N., Audinet, N., Barre, L., Cauquil, A., Cornu, M. & Colin, P. (2006). Effect of the inoculum size on *Listeria monocytogenes* growth in structured media. *Int. J. Food Microbiol.* **110**, 43–51.

Gohil, V., Ahmed, M., Davies, R. & Robinson, R. (1995). Incidence of *Listeria spp.* in retail foods in the united arab emirates. *J. Food Prot.* **58**, 102–104.

Gombas, D. E., Chen, Y., Clavero, R. S. & Scott, V. N. (2003). Survey of *Listeria monocytogenes* in ready-to-eat foods. *J. Food Prot.* **66**, 559–569.

Guerra, M. M., McLauchlin, J. & Bernardo, F. (2001). *Listeria* in ready-to-eat and unprocessed foods produced in portugal. *Food Microbiol.* **18**, 423–429.

Gunasena, D., Kodikara, C., Kumuda, G. & Widanapathirana, S. (1995). Occurrence of *Listeria monocytogenes* in food in Sri Lanka. *J. Natn. Sci. Coun. Sri Lanka* **23**, 107–114.

Halvorson, H. O. & Ziegler, N. R. (1933). Applications of statistics to problems in bacteriology. i. a means of determining bacterial population by the dilution method. *J. Bacteriol.* **25**, 101–121.

Harvey, J. & Gilmour, A. (1993). Occurrence and characteristics of *Listeria* in foods produced in Northern Ireland. *Int. J. Food Microbiol.* **19**, 193–205.

Hastings, W. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika* **57**, 97–109.

Heisick, J., Harrel, F., Peterson, E., McLaughlin, S., Wagner, D., Wesley, I. & Bryner, J. (1989a). Comparison of four procedures to detect *Listeria* spp. in foods. *J. Food Prot.* **52**, 154–157.

Heisick, J. E., Wagner, D. E., Nierman, M. L. & Peeler, J. T. (1989b). *Listeria spp.* found on fresh market produce. *Appl. Environ. Microbiol.* **55**, 1925–1927.

Hellström, S., Kervinen, R., Lyly, M., Ahvenainen-Rantala, R. & Korkeala, H. (2006). Efficacy of disinfectants to reduce *Listeria monocytogenes* on precut iceberg lettuce. *J. Food Prot.* **69**, 1565–1570.

Helsel, D. (2004). *Nondetects and Data Analysis : Statistics for Censored Environmental Data. Statistics in Practice.* Wiley ed.

Hercberg, S., Galan, P., Preziosi, P., Bertrais, S., Mennen, L., Malvy, D., Roussel, A.-M., Favier, A. & Briançon, S. (2004). The SU.VI.MAX study : a randomised placebo-controlled trial of the health effects of antioxidant vitamins and minerals. *Archives Internal Medecine* **164**, 2335–2342.

Iman, R. L. & Conover, W. J. (1982). A distribution-free approach to inducing rank correlation among input variables. *Commun. Statist.-Simula. Computa.* **11**, 311–334.

Inoue, S., Nakama, A., Arai, Y., Kokubo, Y., Maruyama, T., Saito, A., Yoshida, T., Terao, M., Yamamoto, S. & Kumagai, S. (2000). Prevalence and contamination levels of *Listeria monocytogenes* in retail foods in Japan. *Int. J. Food Microbiol.* **59**, 73–77.

INSEE, Institut National de la Statistique et des Etudes Economiques (2002). La situation démographique en 2002 - Mouvements de la population et enquête emploi de janvier 2002. Tech. rep.

Jacxsens, L., Devlieghere, F. & Debevere, J. (2002). Predictive modelling for packaging design : equilibrium modified atmosphere packages of fresh-cut vegetables subjected to a simulated distribution chain. *Int. J. Food Microbiol.* **73**, 331–341.

Jacxsens, L., Devlieghere, F., Falcato, P. & Debevere, J. (1999). Behavior of *Listeria monocytogenes* and *Aeromonas* spp. on fresh-cut produce packaged under equilibrium-modified atmosphere. *J. Food Prot.* **62**, 1128–1135.

James, S. J. & Evans, J. (1992). Consumer handling of chilled foods : Temperature performance. *Int. J. Refrig.* **15**, 299–306.

Jarvis, B. (2000). Sampling for microbiological analysis. In *The Microbiological Quality and Safety of Food. Volume II*, B. Lund, A. Baird-Parker & G. Gould, eds. Gaithersburg, : Aspen Publishers, pp. 1691–1733.

Jaykus, L. A. (1996). The application of quantitative risk assessment to microbial food safety risks. *Critical Reviews in Microbiology* .

Johannessen, G. S., Loncarevic, S. & Kruse, H. (2002). Bacteriological analysis of fresh produce in norway. *Int. J. Food Microbiol.* **77**, 199–204.

Kadane, J. B. & Wolfson, L. J. (1998). Experiences in elicitation. *The Statisticians* **47**, 3–19.

Kakiomenou, K., Tassou, C. & Nychas, G.-J. (1998). Survival of salomonella enteritidis and *Listeria monocytogenes* on salad vegetables. *World J. Microbiol. Biotechnol.* **14**, 383–387.

Kaneko, K., Hayashidani, H., Ohtomo, Y., Kosuge, J., Kato, M., Takahashi, K., Shiraki, Y. & Ogawa, M. (1999). Bacterial contamination of ready-to-eat foods and fresh products in retail shops and food factories. *J. Food Prot.* **62**, 644–649.

Kaplan, E. & Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **53**, 457–481.

Kono, T. (1968). Kinetics of microbial cell growth. *Biotech. Bioeng.* **10**, 105–131.

Koseki, S. & Isobe, S. (2005a). Growth of *Listeria monocytogenes* on iceberg lettuce and solid media. *International Journal of Food Microbiology* **101**, 217–25.

Koseki, S. & Isobe, S. (2005b). Growth of *Listeria monocytogenes* on iceberg lettuce and solid media. *Int. J. Food Microbiol.* **101**, 217–25.

Koseki, S., Yoshida, K., Isobe, S. & Itoh, K. (2001). Decontamination of lettuce using acidic electrolyzed water. *J. Food Prot.* **64**, 652–658.

Kroes, R., Muller, D., Lambe, J., Lowick, M. R. H., v. Klaveren, J., Kleiner, J., Massey, R., Mayer, S., Urieta, I., Verger, P. & Visconti, A. (2002). Assessment of intake from the diet. *Food and Chemical Toxicology* **40**, 327–385.

Laguerre, O., Derens, E. & Palagos, B. (2002). Study of domestic refrigerator temperature and analysis of factors affecting temperature : a French survey. *Int. J. Refrig.* **25**, 653–659.

Lainé, K. & Michard, J. (1988). Fréquence et abondance des *Listeria* dans des légumes frais découpés prêts à l'emploi. *Microbiol. Alim. Nutr.* **6**, 329.

Leblanc, J., Malmauret, L., Guérin, T., Bordet, F., Boursier, B. & Verger, P. (2000). Estimation of the dietary intake of pesticides residues, lead, cadmium, arsenic and radionucléides in France. *Food Additives and Contaminants* **17**, 925–932.

Leblanc, J. C., Guérin, T., Noöl, L., C.-T. G., Volatier, J. L. & Verger, P. (2005). Dietary exposure estimates of 18 elements from the 1st French total diet study. *Food Additives and Contaminants* **22**, 624–641.

Levre, E., Valentini, P. & Caroli, G. (1995). Prodotti orticoli e contaminazione da *Listeria*. *Ig. Mod.* **103**, 475–486.

Li, Y., Brackett, R., Chen, J. & Beuchat, L. (2002). Mild heat treatment of lettuce enhances growth of *Listeria monocytogenes* during subsequent storage at 5°c or 15°c. *J. of Appl. Microbiol.* **92**, 269–275.

Li, Y., Brackett, R., Shewfelt, R. & Beuchat, L. (2001). Changes in appearance and natural microflora on iceberg lettuce treated in warm, chlorinated water and then stored at refrigeration temperature. *Food Microbiol.* **18**, 299–308.

Lindqvist, R., Sylvén, S. & Vagsholm, I. (2002). Quantitative microbial risk assessment exemplified by *Staphylococcus aureus* in unripened cheese made from raw milk. *Int. J. Food. Microbiol.* **78**, 155–170.

Lindqvist, R. & Westöö, A. (2000). Quantitative risk assessment for *Listeria monocytogenes* in smoked or gravad salmon/rainbow trout in Sweden. *International Journal of Food Microbiology* **58**, 181–196.

Little, C., Monsey, H., Nichols, G. & de Louvais, J. (1997). The microbiological quality of refrigerated salads and crudites. an analysis of the results from the 1995 European community coordinated food control programme for England and Wales. *PHLS Microbiol. Dig.* **14**, 142–146.

Lorimer, M. F. & Kiermeier, A. (2007). Analysing microbiological data : Tobit or not tobit ? *Int. J. Food. Microbiol.* **116**, 313–8.

MacGowan, A. P., Bowker, K., McLauchlin, J., Bennett, P. M. & Reeves, D. S. (1994). The occurrence and seasonal changes in the isolation of *Listeria* spp. in shop bought food stuffs, human faeces, sewage and soil from urban sources. *Int. J. Food Microbiol.* **21**, 325–334.

McCrady (1915). The numerical interpretation of fermentation-tube results. *J. Infect. Dis.* **17**, 183–212.

McLauchlin, J. (1997). The pathogenicity of *Listeria monocytogenes* : A public health perspective. *Reviews in Medical Microbiology* **8**, 1–14.

McLauchlin, J. & Gilbert, R. (1990). *Listeria* in food. *PHLS Microbiol. Dig.* **7**, 54–55.

McMeekin, T., Olley, J., Ross, T. & Ratkowsky, D. (1993). *Predictive Microbioloby : theory and application.* Research Studies Press. LTD, Taunton.

Messi, P., Casolari, C., Fabio, A., Fabio, G., Gibertoni, C., Menziani, G. & Quaglio, P. (2000). Occurrence of *Listeria* in food matrices. *Ind. Aliment.* **39**, 151–157.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. & Teller, E. (1953). Equation of state calculations by fast computing machine. *Journal of Chemical Physics* **21**, 1087–1091.

Monge, R. & Arias, M. (1996). Presencia de microorganismos patogenos en las hortalizas de cosumo crudo en Costa Rica. *Archiv. Latinoam. Nutr.* **46**, 293–294.

Morris, P. A. (1977). Combining expert judgments : a bayesian approach. *Management Science* **23**, 679–693.

Murase, M., Miyata, T., Kimata, H. & Kurokawa, M. (2002). Investigation of the actual conditions of pathogenic bacteria among imported fresh vegetables and fruits. *Jap. J. Food Microbiol.* **19**, 71–75.

Nauta, M. (2000). Separation of uncertainty and variability in quantitative microbial risk assessment models. *Int. J. Food Microbiol.* **57**, 9–18.

Nauta, M. (2005). Microbiological risk assessment models for partioning and mixing during food handling. *Int. J. Food Microbiol.* **100**, 311–322.

Nauta, M. J., Litman, S., Barker, G. C. & Carlin, F. (2003a). A retail and consumer phase model for exposure assessment of *Bacillus cereus*. *Int. J. Food Microbiol.* **83**, 205–218.

Nauta, M. J., Litman, S., Barker, G. C. & Carlin, F. (2003b). A retail and consumer phase model for exposure assessment of *Bacillus cereus*. *Int. J. Food Microbiol.* **83**, 205–18.

Nguyen-the, C. & Carlin, F. (2000). Fresh and processed vegetables. In *The Microbiological Quality and Safety of Food*, B.-P. A. Lund, B.M. & G. Gould, eds. Gaithersburg : Aspen Publishers, pp. 620–684.

Nguyen-The, C., Halna-du Fretay, F. & Abreu da Silva, A. (1996). The microbiology of mixed salad containing raw and cooked ingredients without dressing. *Int. J. Food Sci. and Technol.* **31**, 481–487. 27.

Norrung, B., Andersen, J. & Schlundt, J. (1999). Incidence and control of *Listeria monocytogenes* in foods in Denmark. *Int. J. Food Microbiol.* **53**, 195–203.

Odumeru, J. A., Mitchell, S. J., Alves, D. M., Lynch, J. A., Yee, A. J., Wang, S. L., Styliadis, S. & Farber, J. M. (1997). Assessment of the microbiological quality of ready-to-use vegetables for health-care food services. *J. Food Prot.* **60**, 954–960. 0362-028X.

O'Hagan, A. & Oakley, J. E. (2007). Uncertainty in prior elicitations : a nonparametric approach. *Biometrika* **94**, 427–441.

Ohsone, M., Kaneko, K., Hayashidani, H., Takahashi, T. & Ogawa, M. (1999). Growth and inhibition by acids of five species of pathogenic bacteria inoculated in salad vegetables. *J. Food Hyg. Soc. Japan* **40**, 297–303.

Pacini, R., Panizzi, L., Quagli, E., Galassi, R., Malloggi, L. & Morganti, R. (1993). Diffusione di *Listeria monocytogenes*. *Ind. Alim.* **32**, 1086–1089.

Park, C., Hung, Y., Doyle, M., Ezeike, G. & Kim, C. (2001). Pathogen reduction and quality of lettuce treated with electrolyzed oxidizing and acidified chlorinated water. *J. Food Sci.* **66**, 1368–1372.

Pasanisi, A. (2004). *Aide à la décision dans la gestion des parcs de compteurs d'eau potable.* Thèse de doctorat, Ecole Nationale du Génie Rural, des Eaux, et des Forêts.

Paulo, M., van der Voet, H., Jansen, M.J.W, t. B. C. & van Klaveren J.D. (2005). Risk assessment of dietary exposure to pesticides using a bayesian method. *Pest Manag. Sci.* **61**, 759–766.

Pierre, O. (1996). Température de conservation de certaines denrées alimentaires très périssables dans les rayons libre service des grandes et moyennes surfaces. *Option Qualité* **138**, 12–18.

Pirovani, M., Piagentini, A., Guemes, D. & Arkwright, S. (2004). Reduction of chlorine concentration and microbial load during washing-disinfection of shredded lettuce. *Int. J. Food Sci. Technol.* **39**, 341–347.

Porto, E. & Uboldi Eiroa, M. N. (2001). Occurrence of *Listeria monocytogenes* in vegetables. *Dairy Food Environ. Sanit.* **2**, 282–286.

Pouillot, R. (2006). *Appréciation quantitative des risques en hygiène des aliments : développements et mise en oeuvre pour la prise en compte des recommmandations internationales.* Thèse de doctorat, Faculté de médecine Paris-Sud, ParisXI.

Pouillot, R., Albert, I., Cornu, M. & Denis, J. B. (2003). Estimation of uncertainty and variability in bacterial growth using bayesian inference. Application to *Listeria monocytogenes*. *Int. J. Food Microbiol.* **81**, 87–104.

Pouillot, R., Miconnet, N., Afchain, A. L., Delignette-Muller, M. L., Beaufort, A., Rosso, L., Denis, J. B. & Cornu, M. (2007). Quantitative risk assessment of *Listeria monocytogenes* in french cold-smoked salmon : I. quantitative exposure assessment. *Risk Anal.* **27(3)**, 683–700.

Prazak, A. M., Murano, E. A., Mercado, I. & Acuff, G. R. (2002). Prevalence of *Listeria monocytogenes* during production and postharvest processing of cabbage. *J. Food Prot.* **65**, 1728–1734.

Previdi, M., Tomasoni, L. & Bondi, B. (2000). Minimally processed and frozen vegetables : microbiological quality and incidence of pathogens in commercial products. *Ind. Conserv.* **75**, 383–391.

Ranta, J., Tuominen, P. & Maijala, R. (2005). Estimation of true *Salmonella* prevalence jointly in cattle herd and animal populations using bayesian hierarchical modeling. *Risk Anal.* **25(1)**, 23–37.

Ratkowsky, D. A., Olley, J., McMeekin, T. A. & Ball, A. (1982). Relationship between temperature and growth rate of bacterial cultures. *J. Bacteriol.* **149**, 1–5.

Reinders, R. D., De Jonge, R. & Evers, E. G. (2003). A statistical method to determine whether micro-organisms are randomly distributed in a food matrix, applied to coliforms and *Escherichia coli* O157 in minced beef. *Food Microbiol.* **20**, 297–2003.

Renwick, A. G., Barlow, S. M., Hertz-Picciotto, I., Boobis, A. R., Dybing, E., Edler, L., Eisenbrand, G., Greig, J. B., Kleiner, J., Lambe, J., Müller, D. J., Smith, M. R., Tritscher, A., Tuijtelaars, S., van den Brandt, P. A., Walker, R. & Kroes, R. (2003). Risk characterisation of chemicals in food and diet. *Food and Chemical Toxicology* **41**, 1211–1271.

Richardson, S. & Gilks, W. (1993). A bayesian approach to measurement error problems in epidemiology using conditional independence models. *American Journal of Epidemiology* **138(6)**, 430–442.

Riedo, F. X., Pinner, R. W., De Lourdes Tosca, M., Cartter, M. L., Graves, L. M., Reeves, M. W., Weaver, R. E., Plikaytis, B. D. & Broome, C. V. (1994). A point-source foodborne outbreak : Documented incubation period and possible mild illness. *Journal of Infectious Diseases* **170**, 693–696.

Ringlé, P., Vincent, J.-P. & Catteau, M. (1991). Evolution de listeria dans les produits de 4ème gamme. In *Les microorganismes contaminants dans les industries alimentaires : colonisation, détection, maîtrise. 7ème colloque de la SFM, Institut Pasteur, Paris, 13-14 mars 1991*, C. Lahellec, ed. Paris : Société Française de Microbiologie, pp. 324–329.

Rivot, E., Prévost, E., Parent, E. & Baglinière, J. (2004). A bayesian state-space modelling framework for fitting a salmon stage-structured population dynamic model to multiple time series of field data. *Ecological modelling* **179**, 463–485.

ROBERT, C. (2001). *The Bayesian Choice.* New York : Springer, 2nd ed.

ROBERT, C. P. (1996). *Méthodes de Monte Carlo par Chaînes de Markov.* Economica ed.

ROCOURT, J. (1996). Risk factor for listeriosis. *Food Cont.* **7**, 192–222.

ROSSO, L. (1995). *Modélisation et microbiologie prévisionnelle : élaboration d'un nouvel outil pour l'agro-alimentaire.* Thèse de doctorat, Claude Bernard Lyon I.

ROSSO, L., LOBRY, J. R. & FLANDROIS, J. P. (1993). An unexpected correlation between cardinal temperatures of microbial growth highlighted by a new model. *J. Theor. Biol.* **162**, 447–63.

SAGOO, S., LITTLE, C. & MITCHELL, R. (2001). The microbiological examination of ready-to-eat organic vegetables from retail establishments in the united kingdom. *Lett. Appl. Microbiol.* **33**, 434–439.

SAGOO, S. K., LITTLE, C. L. & MITCHELL, R. T. (2003a). Microbiological quality of open ready-to-eat salad vegetables : effectiveness of food hygiene training of management. *J. Food Prot.* **66**, 1581–1586.

SAGOO, S. K., LITTLE, C. L., WARD, L., GILLESPIE, I. A. & MITCHELL, R. T. (2003b). Microbiological study of ready-to-eat salad vegetables from retail establishments uncovers a national outbreak of salmonellosis. *J. Food Prot.* **66**, 403–409.

SALAMINA, G., DALLE DONNE, E., NICCOLINI, A., PODA, G., CESARONI, D., BUCCI, M., FINI, R., MALDIN, M., SCHUCHAT, A., SWAMINATHAN, B., BIBB, W., ROCOURT, J., BINKIN, N. & SALMASOL, S. (1996). A foodborne outbreak of gastroenteritis involving *Listeria monocytogenes. Epidemiology Infection* **117**, 429–436.

SANAA, M., COROLLER, L. & CERF, O. (2004). Risk assessment of listeriosis linked to the consumption of two soft cheeses made from raw milk : Camembert of Normandy and Brie of Meaux. *Risk Anal.* **24**, 389–399.

SCHAFFNER, D. W. (1994). Application of a statistical bootstrapping technique to calculate growth rate variance for modelling psychrotrophic pathogen growth. *Int. J. Food Microbiol.* **24**, 309–14.

Schlech, W., Lavigne, P., Bortolussi, R., Allen, A., Haldane, E., Wort, A., Hightower, A., Johnson, S., King, S., Nicholls, E. & Broome, C. (1983). Epidemic listeriosis-evidence for transmission by food. *The New England Journal of Medicine* **30**, 203–206.

Schwartz, D. (1994). *Le jeu de la science et du hasard. La statistique et le vivant.* Flammarion.

Serra-Majem, L., MacLean, D., Ribas, L., Brule, D., Sekula, W., Prattala, R., Garcia-Closas, R., Yngve, A. & Petrasovits, M. L. A. (2003). Comparative analysis of nutrition data from national, household, and individual levels : results from a WHO-CINDI collaborative project in Canada, Finland, Poland, and Spain. *Journal of Epidemiology and Community Health* **57**, 74–80.

Shorten, P. R., Pleasants, A. B. & Soboleva, T. K. (2006). Estimation of microbial growth using population measurements subject to a detection limit. *Int. J. Food. Microbiol.* **108**, 369–75.

Sizmur, K. & Walker, C. W. (1988). *Listeria* in prepacked salads. *The Lancet* **331**, 1167.

Smith, T. C., Spiegelhalter, D. J. & Thomas, A. (1995). Bayesian approaches to random-effects meta-analysis : a comparative study. *Stat. Med.* **14**, 2685–99.

Soriano, J., Rico, H., Molto, J. & Manes, J. (2001). *Listeria species* in raw and ready-to-eat foods from restaurants. *J. Food Prot.* **64**, 551–553.

Spiegelhalter, D., Thomas, A., Best, N. & Lunn, D. (2005). OpenBUGS Version 2.10, User Manual. MRC Biostatistics Unit, Cambridge, UK, Dept of Epidemiology& Public Health, Imperial College, UK and Dept of Mathematics& Statistics, University of Helsinki, Finland.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *J. R. Statist. Soc.* **64**, 583–639.

Stahl, V., Carlin, F., Denis, C., Dervin, C. & Huchet, V. (2002). Sécurité des produits végétaux commercialisés par l'industrie agro-alimentaire par le développement d'un outil permettant l'étude du comportement des flores pathogènes et l'étude de l'incidence

de choix technologiques. AERIAL Illkirch, ADRIA Normandie et ADRIA Quimper, INRA (France), Rapport Interne ACTIA 99.15. Tech. rep.

Stahl, V., Denis, C. & Huchet, V. (2005). Survey on the contamination of minimally processed Fresh salads with major foodborne pathogens. AERIAL Illkirch, ADRIA Normandie et ADRIA Quimper (France), ACTIA 03.13. Tech. rep.

Stahl, V., Denis, C. & Huchet, V. (2006). Estimation quantitative du danger microbiologique, étudiée pour la filière ivème gamme. AERIAL Illkirch, ADRIA Normandie, ADRIA Quimper, INRA (France), ACTIA 03.13. Tech. rep.

Steinbruegge, E., Maxcy, R. & Liewen, M. (1988). Fate of *Listeria monocytogenes* on ready to serve lettuce. *J. Food Prot.* **51**, 596–599.

Swaminathan, B., Cabanes, D., Zhang, W. & Cossart, P. (2007). *Listeria monocytogenes. In Food Microbiology : Fundamentals and Frontiers.* Washington D.C. : ASM Press : Doyle, M.P. and Beuchat, L.R.

Szabo, E., Scurrah, K. & Burrows, J. (2000). Survey for psychrotrophic bacterial pathogens in minimally processed lettuce. *Lett. Appl. Microbiol.* **30**, 456–460.

Szabo, E. A., Simons, L., Coventry, M. J. & Cole, M. B. (2003). Assessment of control measures to achieve a food safety objective of less than 100 CFU of listeria monocytogenes per gram at the point of consumption for fresh precut iceberg lettuce. *J. Food Prot.* **66**, 256–264.

Team, R. D. C. (2007). R : A Language and Environment for Statistical Computing (Version R.2.5.1). Vienna : R Foudation for Statistical Computing.

Thunberg, R. L., Tran, T. T., Bennett, R. W., Matthews, R. N. & Belay, N. (2002). Microbial evaluation of selected fresh produce obtained at retail markets. *J. Food Prot.* **65**, 677–682.

Tiwari, N. & Aldenrath, S. (1990). Occurrence of *Listeria species* in food and environmental samples in Alberta. *Can. Inst. Food Sci. Technol.* **23**, 109–113.

Tressou, J. (2006). Non parametric modelling of the left censorship of analytical data in food risk exposure assessment. *J. Am. Stat. Assoc.* **101 (476)**, 1377–1386.

Tressou, J., Leblanc, J., Feinberg, M. & Bertail, P. (2004). Statistical methodology to evaluate food exposure to a contaminant and influence of sanitary limits : application to ochratoxin A. *Regul. Toxicol. Pharmacol.* **40**, 252–63.

Vaillant, V., de Valk, H., Baron, E., Ancelle, T., Colin, P., Delmas, M. C., Dufour, B., Pouillot, R., Le Strat, Y., Weinbreck, P., Jougla, E. & Desenclos, J. C. (2005). Foodborne infections in france. *Foodborne Pathog. Dis.* **2**, 221–232.

Villari, P., Greco, G. & Torre, I. (1997). *Listeria* spp. in ready-to-eat foods. *Ig. Mod.* **107**, 403–410.

Vose, D. (1998). The application of quantitative risk assessment to microbial food safety. *J. Food Prot.* **61**, 640–648.

Vose, D. (2000). *Risk Analysis : A Quantitative Guide.* Chichester : John Wiley and Sons, 2nd ed.

Walsh, D., Duffy, G., Sheridan, J., Blair, I. & McDowell, D. (1998). Comparison of selective and nonselective enrichment media for the isolation of *Listeria species* from retail foods. *J. Food Saf.* **18**, 85–99.

Wijtzes, T., McClure, P. J., Zwietering, M. H. & Roberts, T. A. (1993). Modelling bacterial growth of *Listeria monocytogenes* as a function of water activity, ph and temperature. *Int. J. Food. Microbiol.* **18**, 139–49.

Wong, H. C., Chao, W. L. & Lee, S. J. (1990). Incidence and characterization of *Listeria monocytogenes* in foods available in Taiwan. *Appl. Environ. Microbiol.* **56**, 3101–3104.

Zhang, S. & Farber, J. (1996). The effects of various disinfectants against *Listeria monocytogenes* on fresh-cut vegetables. *Food Microbiol.* **13**, 311–321.

Zwietering, M. H., Cuppers, H. G., de Wit, J. C. & van 't Riet, K. (1994). Evaluation of data transformations and validation of a model for the effect of temperature on bacterial growth. *Appl. Environ. Microbiol.* **60**, 195–203.

Zwietering, M. H., Jongenburger, I., Rombouts, F. M. & van 't Riet, K. (1990). Modeling of the bacterial growth curve. *Appl. Environ. Microbiol.* **56**, 1875–1881.

Annexes

Communications orales et posters

Les travaux de cette thèse ont fait l'objet de communications orales/poster lors de congrès internationaux avec comités de lecture.

- *Microbial Quantitative Exposure Assessment Using Hierarchical Bayesian Modelling and Second-Order Monte Carlo Simulations.* Communication orale, 5th International Conference Predictive Modelling in Foods, Sept 2007, Athènes.
- *Estimation de la contamination des denrées alimentaires à partir des données de prévalence et de concentration. Application à L. monocytogenes dans les légumes frais.* Poster, 7ème Congrès de la Société Française de Microbiologie, Juin 2007, Nantes.
- *Bayesian method combining information from prevalence and concentration data to estimate microbial contamination.* Communication orale, 25th European Meeting of Statisticians, Juillet 2005, Oslo.
- *Non parametric modelling of contamination data. Application to ready-to-eat salads.* Poster, COMPSTAT 2004, 16th Symposium of IASC, Août 2004, Prague.

Annexes suivantes

Dans les pages suivantes, sont fournis :

- l'article intitulé Estimation of microbial contamination of food from prevalence and concentration data : Application to Listeria monocytogenes in fresh vegetables, co-écrit avec Albert, I., Dervin, C. and F. Carlin, paru dans *Applied and Environmental Microbiology*, Volume 73, No. 1, Pages 250-258.
- le questionnaire de l'enquête ACTIA.

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Jan. 2007, p. 250–258 Vol. 73, No. 1
0099-2240/07/$08.00+0 doi:10.1128/AEM.00351-06
Copyright © 2007, American Society for Microbiology. All Rights Reserved.

Estimation of Microbial Contamination of Food from Prevalence and Concentration Data: Application to *Listeria monocytogenes* in Fresh Vegetables▽

Amélie Crépet,[1]* Isabelle Albert,[1] Catherine Dervin,[1] and Frédéric Carlin[2]

INRA-Mét@risk, UR 1204, Méthodologies d'Analyse de Risque Alimentaire, F-75231 Paris, France,[1] and INRA, UMR 408, Sécurité et Qualité des Produits d'Origine Végétale, F-84914 Avignon, and Université d'Avignon, F-84029 Avignon, France[2]

Received 13 February 2006/Accepted 29 October 2006

A normal distribution and a mixture model of two normal distributions in a Bayesian approach using prevalence and concentration data were used to establish the distribution of contamination of the food-borne pathogenic bacteria *Listeria monocytogenes* in unprocessed and minimally processed fresh vegetables. A total of 165 prevalence studies, including 15 studies with concentration data, were taken from the scientific literature and from technical reports and used for statistical analysis. The predicted mean of the normal distribution of the logarithms of viable *L. monocytogenes* per gram of fresh vegetables was −2.63 log viable *L. monocytogenes* organisms/g, and its standard deviation was 1.48 log viable *L. monocytogenes* organisms/g. These values were determined by considering one contaminated sample in prevalence studies in which samples are in fact negative. This deliberate overestimation is necessary to complete calculations. With the mixture model, the predicted mean of the distribution of the logarithm of viable *L. monocytogenes* per gram of fresh vegetables was −3.38 log viable *L. monocytogenes* organisms/g and its standard deviation was 1.46 log viable *L. monocytogenes* organisms/g. The probabilities of fresh unprocessed and minimally processed vegetables being contaminated with concentrations higher than 1, 2, and 3 log viable *L. monocytogenes* organisms/g were 1.44, 0.63, and 0.17%, respectively. Introducing a sensitivity rate of 80 or 95% in the mixture model had a small effect on the estimation of the contamination. In contrast, introducing a low sensitivity rate (40%) resulted in marked differences, especially for high percentiles. There was a significantly lower estimation of contamination in the papers and reports of 2000 to 2005 than in those of 1988 to 1999 and a lower estimation of contamination of leafy salads than that of sprouts and other vegetables. The interest of the mixture model for the estimation of microbial contamination is discussed.

Quantitative microbial risk assessment (QMRA) is in rapid development in the area of food safety. To obtain a quantitative exposure assessment or a quantitative risk characterization for a given food and a given food-borne pathogen, statistical distributions of microbial concentrations are used as input values (such as bacterial concentrations in raw food materials subjected to further process) and/or as output values (predicted contamination in foods after processing and/or storage) (35, 75). Microbial contaminations in foods are expressed in two forms: (i) prevalence data (percentage of positive samples in a given study, i.e., growth/no growth of a target pathogen after enrichment in an appropriate broth of an aliquot of a food sample) and (ii) concentration data expressed as CFU per gram or CFU per milliliter. Data on the concentration of pathogenic bacteria in foods, while scarce, are essential to QMRA.

Previous works on QMRA did not consider low levels of concentrations, i.e., concentrations below the threshold of detection of microbiological methods (6, 15, 44, 53). However, low concentration estimates in raw materials subjected to further processing and storage are necessary for proper QMRA. Finding low initial numbers of bacteria does not mean that the bacteria will not increase to reach numbers critical for human health.

The objective of this work is to estimate bacterial concentration levels from prevalence and concentration data using two statistical methods: (i) a determination of the parameters of a normal distribution on the logarithm base 10 (log) of microbial concentrations and (ii) a mixture model of two normal distributions on the logarithm base 10 of microbial concentrations using a Bayesian approach (37). The methods are both based on the most-probable-number (MPN) theory to estimate bacterial concentrations from prevalence data (10, 11, 28) and allow modeling of both prevalence and concentration data in a single frame. In the method for determining the parameters of the normal distribution, the mean of the distribution was estimated from the MPN estimates and its standard deviation was estimated from the concentration data. The assumption of the mixture model is to consider that the concentrations are distributed by two normal distributions, one representing low concentrations (below the threshold of detection) and one representing high concentrations (higher than the threshold of detection). Using a probabilistic approach of the MPN theory on prevalence data, the low concentrations and the variability of concentrations between studies were estimated. High concentrations and variability of concentrations between samples representing the variability within studies were directly estimated from the con-

* Corresponding author. Mailing address: INRA-Mét@risk, Méthodologies d'Analyse de Risque Alimentaire, INA-PG, 16 rue Claude Bernard, 75231 Paris Cedex 5, France. Phone: 33-1-44-08-86-56. Fax: 33-1-44-08-72-76. E-mail: crepet@inapg.inra.fr.

▽ Published ahead of print on 10 November 2006.

TABLE 1. Number of studies extracted from the 51 papers and reports on the contamination of fresh and minimally processed vegetables with *L. monocytogenes* per type of product with total, minimum, and maximum numbers of samples analyzed

Product	No. of studies			No. of samples[d]			References
	Total	Prevalence only	Prevalence and concn	Total	Min	Max	
Minimally processed vegetables[a,c]	48	42	6	8,511	1	2,934	1, 3–5, 9, 16, 17, 19, 26, 29, 39, 42, 46–50, 55, 60, 62, 63, 69
Minimally processed leafy salad vegetables[b,c]	15	10	5	7,511	3	3,849	4, 16, 19, 22, 25, 42, 46, 62, 64, 69, 70, 72
Sprouts and germinated seed, mung bean, alfalfa, and bean sprouts	12	10	2	785	6	276	3, 16, 19, 51, 71
Unprocessed vegetables[a]	67	66	1	3,934	1	425	1, 3, 8, 9, 13, 14, 18, 20, 21, 24, 26, 27, 31, 36, 38, 39, 42, 47–49, 51, 56, 57, 59, 66, 67, 69, 71, 72, 74, 78, 81
Unprocessed leafy salad vegetables[b]	23	22	1	4,337	10	2,966	9, 13, 21, 25, 27, 32, 38, 41, 42, 45, 57, 67, 69, 71, 78
Total	165	150	15	25,078	1	3,849	

[a] Vegetables included beets, carrots, cabbage, cauliflower, celery, cucumber, roots, bulbs, tomato, spinach, broccoli, leek, various herbs, and mixed vegetables.
[b] Vegetables included lettuce, rocket, lamb's lettuce, and endive.
[c] Vegetables were trimmed, cut or shredded and washed.
[d] Min, minimum; max, maximum.

centration data. The estimations were carried out as in a variance component model and by using a hierarchical Bayesian approach (12).

In a Bayesian framework (37), model parameters are random variables and are firstly described by a prior distribution. When information about parameters is absent or poor, uninformative or vague prior distributions are used. In contrast, available expert knowledge may contribute to more-informative prior distributions. Posterior distributions over all parameters are computed using Bayes' theorem, which combines prior distributions and information available in the data set. From these posterior distributions, distributions of predicted quantities of interest can be computed. Hierarchical Bayesian models use an additional level on parameters, i.e., observations are modeled conditionally on certain parameters of interest, which themselves are given distributions that conditionally lead to further parameters, known as hyperparameters. In these models, hyperprior distributions are assigned to hyperparameters. Through the hyperprior distributions, hierarchical Bayesian models account for the variability and the uncertainty of the parameters of interest (23, 58, 61).

The two statistical methods were applied to the estimation of the contamination of vegetables with the food-borne pathogenic bacterium *Listeria monocytogenes*. *L. monocytogenes* has caused several recognized outbreaks of food-borne infections in which vegetables were implicated; it is regularly isolated from fresh vegetables and can survive and grow even at low temperatures (7, 8, 34, 54, 65). *L. monocytogenes* is therefore a major hazard for fresh and minimally processed vegetables. Its prevalence in vegetables is rather well documented and monitored by the food industry, facilitating the collection of a significant amount of data from international literature or from food companies, along with a comparison of concentration levels of different categories of vegetables, periods of publications, degrees of processing, and geographical origins of vegetables. In addition, the methods of detection of food-borne pathogens show various sensitivities (78, 80), and their influence was tested by taking advantage of the flexibility of the Bayesian approach. However, the prevalence of *L. monocytogenes* in vegetables is generally below 5% and concentration data are scarce and relevant for the highest microbial concentrations. Establishing the distribution of contamination from such structured data requires proper methods, and these are tested in the present work.

MATERIALS AND METHODS

Data related to *L. monocytogenes* prevalence and concentration in vegetables. Data related to *L. monocytogenes* contamination of fresh unprocessed or minimally processed vegetables were extracted from the scientific literature, the FSTA database (Food Science Technology Abstracts; IFIS Publishing, Reading, United Kingdom), reports from the French food inspection service DGCCRF (Direction Générale de la Concurrence, de la Consommation et de la Répression des Fraudes, Paris, France) (19), and food companies (69). Papers and reports not using proper methods for *L. monocytogenes* detection or confirmation or those containing ambiguous information about the number of samples analyzed or *L. monocytogenes* prevalence or concentrations were excluded. Only fresh unprocessed (undressed and uncooked) or minimally processed (trimmed, cut or shredded, or washed) vegetables were considered. Data were kept in the format of the bibliographic source. Consequently, for each paper or report, several studies of different vegetables may have been analyzed. From 51 selected papers and reports, J (=165) prevalence studies (absence/presence of *L. monocytogenes*) were identified (Table 1). For each study, the number of analyzed samples (n), the number of positive samples (r), and the analyzed quantity (q) were recorded. A number J_0 (=15) out of the J (=165) studies and reports gave concentrations of *L. monocytogenes* in class intervals (counts in CFU *L. monocytogenes* per gram), which were also recorded. For instance, in the paper of Sagoo et al. (64), *L. monocytogenes* prevalence data are n, r, and q equal 3,849, 90, and 25, respectively, and, out of 90 positive samples, the concentrations are lower than 20 CFU/g for 88 samples, between 20 and 99 CFU/g for 1 sample, and between 100 and 999 CFU/g for 1 sample.

To generate concentration values from the class interval of concentration, we used the following procedure. For each study j of the J_0, $k = 1, \ldots, r_j$ concentration values, noted as $c_{j,k}$, were randomly selected in each class interval by a triangular distribution: triangular($\min_{j,k}$, $\mathrm{mod}_{j,k}$, $\max_{j,k}$), where the $\min_{j,k}$ and the $\mathrm{mod}_{j,k}$ were assumed to be the lower bound of the interval containing the concentration of the sample k of the study j, and the $\max_{j,k}$ was its upper bound. Afterwards, the randomly selected concentration values $c_{j,k}$ were considered to be observed concentrations for the mixture model. All data were stored in a Microsoft Excel spreadsheet.

Determination of the parameters of a normal distribution to establish the distribution of the contamination. (i) MPN calculations. The MPN estimation, based on the assumption of a Poisson distribution of the bacterial numbers in a

volume, was used to determine the concentrations from prevalence studies (10, 11, 28). Let n_j denote the number of analyzed samples for the study j, r_j the number of positive samples, and q_j the quantity in grams of analyzed samples. Let m_j denote the mean concentration expressed in MPN per gram of the samples for the study j. With the Poisson assumption, the probability that a sample of the study j is contaminated is

$$p_j = 1 - \exp(-q_j m_j) \quad (1)$$

This probability can be estimated as r_j/n_j, and therefore, the mean concentration of the samples of the study j is

$$m_j = \frac{\ln[n_j/(n_j - r_j)]}{(q_j)} \quad (2)$$

Further calculations will use *L. monocytogenes* counts expressed as CFU per gram or MPN per gram. These counts will be equally designated as viable *L. monocytogenes* bacteria per gram.

(ii) Determination of the parameters of normal distribution. *L. moncytogenes* mean log concentrations were estimated for the J prevalence studies by using equation 2. The logarithms base 10 of these concentrations were assumed to be independently and identically normally distributed as proposed by Jarvis (37). No positive sample ($r_j = 0$) was detected in 100 out of the 165 analyzed studies. For these studies, m_j was 0 and consequently an overestimated value of one positive sample ($r_j = 1$) was attributed to make possible the determination of a $\log(m_j)$ value. The mean M of the J $\log(m_j)$ values gave the mean of the normal distribution. Concentrations given by the J_0 concentration studies were higher than MPN estimates and therefore were assumed to be part of the high percentiles of the contamination distribution. The standard deviation of the normal distribution (S) was derived from the proportion α of concentration values higher than a fixed value $\log(d)$, corresponding to the high percentiles of the contamination distribution. The $\log(d)$ values of 1 and 2 log viable *L. monocytogenes* organisms per gram, which are the highest and the most common limits of detection of *L. monocytogenes* counts in studies reviewed in this work, were chosen to determine the standard deviation.

Estimation of the distribution of the contamination using a mixture model. The mixture model is based on the assumption that the distribution of the concentrations noted $\log(c)$ is a mixture of two normal distributions and can be represented by

$$\log(c)|\mu_l, \mu_h, \sigma_l, \sigma_h \sim (1 - \beta)\ \text{normal}(\mu_l, \sigma_l) + \beta\ \text{normal}(\mu_h, \sigma_h) \quad (3)$$

where μ_l and μ_h are the means of the low and the high log concentrations, respectively, and σ_l and σ_h are the standard deviations of the low and the high log concentrations, respectively, and β denotes the proportion of high concentrations from the normal(μ_h, σ_h). Assuming that all of the positive samples are contaminated at levels comprised in the distribution of the observed concentrations, an estimation of the proportion β can be expressed as the equation $\hat{\beta} = \sum_{j=1}^{J} r_j \Big/ \sum_{j=1}^{J} n_j$.

The parameters of the distribution of low log concentrations were estimated in the following way. Consider J independent prevalence studies, with study j estimating the random parameter p_j from the number n_j of analyzed samples and from the number r_j of positive samples following a binomial distribution:

$$r_j|m_j, q_j \sim \text{binomial}(n_j, p_j), \text{ for } j = 1, \ldots, J \quad (4)$$

The parameter p_j, which is the probability for a sample of study j being positive, is expressed as a function of the concentration m_j of study j, and q_j is expressed as the quantity in grams of analyzed samples by equation 1.

A hierarchy is included in the model to account for the variability between studies. Let mp_j be $\log(m_j)$; the mp_j values are viewed as a sample from a common normal distribution with hyperparameters (μ_p, σ_p)

$$mp_j|\mu_p, \sigma_p \sim \text{normal}(\mu_p, \sigma_p) \quad (5)$$

where μ_p and σ_p are the common mean and the standard deviation, respectively, of the log concentrations of the studies, with σ_p denoting the variability between studies. Then the mean and the variance of the low log concentrations are given by the equations $\mu_l = \mu_p$ and $\sigma_l^2 = \sigma_p^2 + s^2$, respectively, where s is the standard deviation of the sample log concentrations estimated from observed concentrations (see below).

The parameters of the distribution of high log concentrations were estimated as follows. Let a number J_0 of studies indexed by $j = 1, \ldots, J_0$ out of the J studies, with observed concentrations. The logarithms base 10 of these concentrations, $\log(c_{j,k})$, are then supposed to be normally distributed as follows:

$$\log(c_{j,k}) \mid mc_j, s \sim \text{normal}(mc_j, s), \text{ for } j = 1, \ldots, J_0; k = 1, \ldots, r_j \quad (6)$$

where s is the standard deviation of the sample log concentrations denoting the variability within studies and the parameter mc_j is the log concentration of the study j. As for low log concentrations, a hierarchical level is considered and therefore the mc_j values are distributed by a common normal distribution with hyperparameters (μ_c andσ_c)

$$mc_j|\mu_c, \sigma_c \sim \text{normal}(\mu_c, \sigma_c) \quad (7)$$

where μ_c and σ_c are the common mean and the standard deviation, respectively, of log concentrations of the studies with observed concentrations, with σ_c denoting the variability between studies. Then the mean and the variance of the high log concentrations are given by the equations $\mu_h = \mu_c$ and $\sigma_h^2 = \sigma_c^2 + s^2$, respectively.

A graphical representation of the model and of its conditional dependences is shown in Fig. 1.

Posterior distributions of the hyperparameters and the parameter. Vague distributions were used to assign hyperprior distributions for the hyperparameters μ_p, σ_p, μ_c, and σ_c and the parameter s: a normal(0, 10^6) for each mean μ_p, μ_c, and a uniform(0, 10^3) for each standard deviation, σ_p, σ_c, or s.

The analytical calculations of posterior distributions are often difficult, especially for models including many parameters, but computer-intensive techniques, such as those for the Markov chain Monte Carlo method, are powerful for the generation of chains of simulated values for parameters converging to the posterior distribution of interest. With these techniques for each hyperparameter, μ_p, σ_p, μ_c, or σ_c, and the parameter, s, a vector of a number of B simulated posterior values was computed.

Distribution of predicted concentrations. The distribution of predicted (pred) concentrations given the observed data (c, r, q, and n) is obtained by simulating B log concentrations from the mixture of the two normal distributions

$$\log(c_b^{\text{pred}})|c, r, q, n \sim (1 - \hat{\beta})\ \text{normal}(mp_b^{\text{pred}}, s_b) + \hat{\beta}\ \text{normal}(mc_b^{\text{pred}}, s_b), \quad \text{for } b = 1, \ldots, B \quad (8)$$

Their means, mp_b^{pred} and mp_c^{pred}, are the realizations obtained by

$$mp_b^{\text{pred}}|c, r, q, n \sim \text{normal}(\mu_{p,b}, \sigma_{p,b}), \text{ for } b = 1, \ldots, B \quad (9)$$

$$mc_b^{\text{pred}}|c, r, q, n \sim \text{normal}(\mu_{c,b}, \sigma_{c,b}), \text{ for } b = 1, \ldots, B \quad (10)$$

Introduction in the mixture model of a rate of sensitivity for the detection methods. A rate of sensitivity of the method was introduced in the mixture model to account for the inability of any method to detect all contaminated samples. Let t_j denote the updated number of positive samples of the study j. t_j was introduced in the mixture model by $r_j \mid v_j \sim \text{binomial}(t_j, v_j)$, for $j = 1, \ldots, J$, where r_j is the observed number of positive samples of study j, and v_j is the rate of sensitivity attributed to study j. This sensitivity rate was assumed either to be fixed or to be variable and distributed as a triangular distribution to account for the unpredictability of the sensitivity rate expressed in the equation $v_j \sim \text{triangular}(\min_v, \text{mod}_v, \max_v)$.

The parameters $\min_v$ and $\max_v$ of the triangular distribution represent the lowest and the highest value of v, respectively, and the parameter mod_v represents the most probable value of v.

Then expression 4 is replaced with $t_j|m_j, p_j \sim \text{binomial}(n_j, p_j)$ and the parameter β is estimated as the equation $\hat{\beta} = \sum_{j=1}^{J} t_j \Big/ \sum_{j=1}^{J} n_j$.

Integration of qualitative variables in the mixture model. Let (X_j) with $j = 1, \ldots, J$ denote a qualitative variable representing the product category, publication period, degree of processing, or geographical origin of the studies. Given a vector of one input variable with M modalities ($X_j = X_{j,0}, \ldots, X_{j,m}, \ldots, X_{j,M-1}$), the output log concentrations mp_j and mc_j are estimated by replacing expressions 5 and 7 with $mp_j|\alpha_{p,m}, \sigma_p \sim \text{normal}\left(\sum_{m=0}^{M-1} X_{j,m}\alpha_{p,m}, \sigma_p\right)$ and $mc_j \mid \alpha_{c,m}, \sigma_c \sim \text{normal}\left(\sum_{m=0}^{M-1} X_{j,m}\alpha_{c,m}, \sigma_c\right)$, where $\alpha_{p,m}$ and $\alpha_{c,m}$ are the coefficients of the modality m and $X_{j,m}$ takes the value 1 if the study j has the modality m and 0 otherwise. For $m = 0$, $X_{j,m}$ is always equal to 1. For the coefficients $\alpha_{p,m}$ and $\alpha_{c,m}$, the vague hyperprior distribution normal(0, 10^6) was attributed.

Computing tools. Bayesian calculations were performed using OpenBUGS software (MRC Biostatistics Unit, Cambridge, United Kingdom, Department of Epidemiology and Public Health, Imperial College, London, United Kingdom, and Department of Mathematics and Statistics, University of Helsinki, Helsinki,

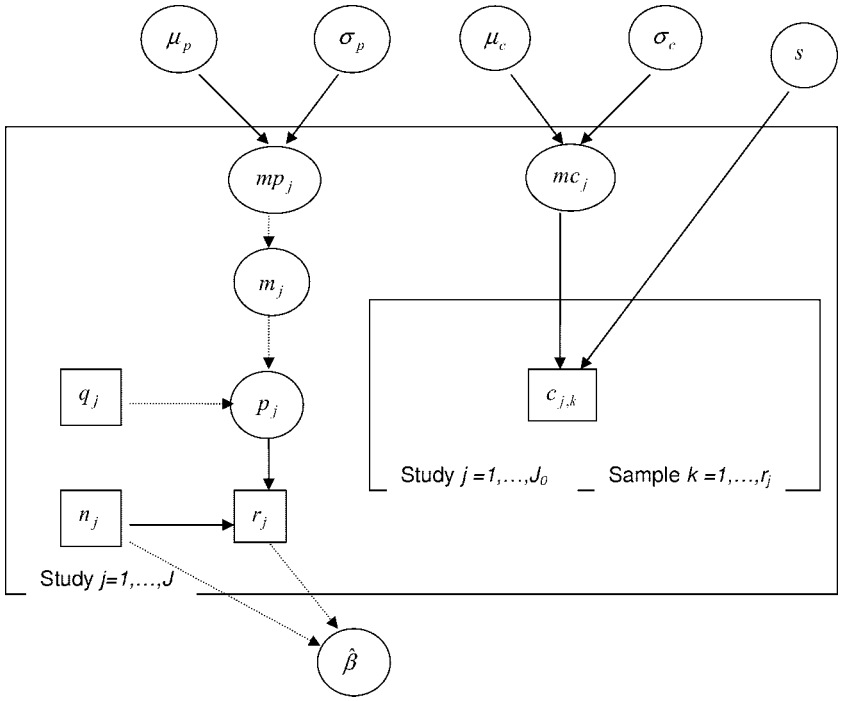

FIG. 1. Directed acyclic graph of the model estimating the mixture component parameters. All model quantities are presented as nodes. Data are denoted by rectangles, and parameters and hyperparameters are denoted by ellipses. Arrows run between nodes from their direct influence ("parents") to the "descendants," indicating the conditional dependence assumptions of the model: given its parent nodes, each node is independent of all other nodes in the graph except its "descendants." A solid arrow indicates probabilistic dependences while dashed arrows indicate logical functions. Probabilistic and logical links are described in Materials and Methods. The mixture process to estimate the distribution of predicted concentrations is described in the "Distribution of predicted concentrations" section of Materials and Methods.

Finland) (68). A triangular distribution function was developed to complete the proposed distribution functions. The convergence of the Markov chain Monte Carlo algorithm was checked by visual analysis of three independent chains and was obtained after 2,000 iterations. Other B (=10,000) iterations were carried out to obtain posterior values of the hyperparameters μ_p, σ_p, μ_c, and σ_c and the parameter s and predicted concentrations.

RESULTS

Distribution of the contamination using a normal distribution. The predicted mean of the normal distribution of the logarithms of viable *L. monocytogenes* organisms per gram in fresh vegetables was −2.63 log viable *L. monocytogenes* organisms/g. The percentage of concentration data higher than log(d) = 2 log viable *L. monocytogenes* organisms/g was α = 0.09% (11 samples out of 12,451 samples tested). The value of the standard deviation of the normal distribution, S, was estimated as 1.48 log viable *L. monocytogenes* organisms/g (Table 2 and Fig. 2). The percentage of concentration data higher

TABLE 2. Descriptive statistics of the log concentrations of vegetables with *L. monocytogenes* determined by a normal distribution and by the mixture model[a]

Method	Mean	SD	2.5th percentile	Median	97.5th percentile
Normal distribution	−2.63	1.48	−5.53	−2.63	0.28
Mixture model	−3.38	1.46	−5.97	−3.45	0.07

[a] Results are expressed as log viable *L. monocytogenes* organisms per gram.

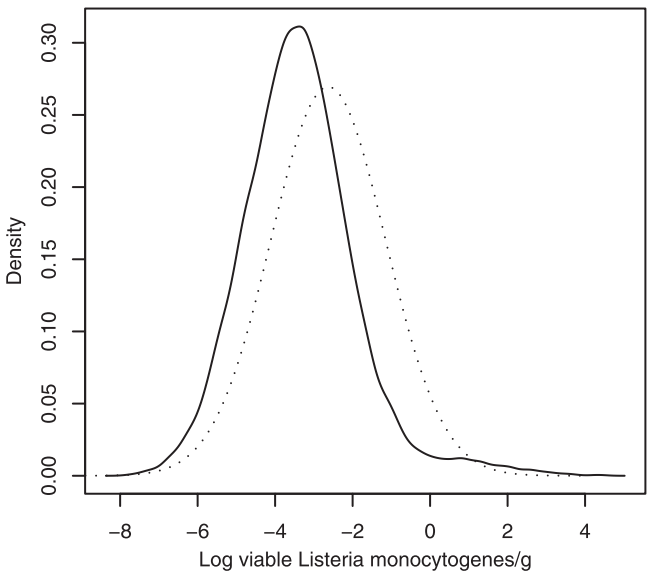

FIG. 2. Density of log concentrations (expressed as log viable *L. monocytogenes* organisms per gram) of fresh unprocessed and minimally processed vegetables with *L. monocytogenes* determined by a normal distribution (dotted line) with an S value of 1.48 log viable *L. monocytogenes* organisms/g [corresponding to an α value of 0.09% of samples with log(d) counts higher than 2 log viable *L. monocytogenes* organisms/g], or predicted by a mixture model (solid line).

TABLE 3. Descriptive statistics of the empirical posterior distributions of hyperparameters and parameter of the mixture model[a]

Hyperparameter and parameter of the mixture model	Value	Mean	SD	2.5th percentile	Median	97.5th percentile
μ_p	Mean	−3.51	0.13	−3.79	−3.50	−3.27
μ_c		1.01	0.28	0.46	1.02	1.57
σ_p	SD	1.10	0.11	0.90	1.09	1.34
σ_c		1.03	0.23	0.69	1.00	1.58
s		0.57	0.0001	0.54	0.57	0.61

[a] Results are expressed as log viable *L. monocytogenes* organisms per gram.

than log(d) = 1 log viable *L. monocytogenes* organisms/g was α = 1.09% (104 samples out of 9,517 samples tested), and the derived standard deviation S was 1.56 log viable *L. monocytogenes* organisms/g. In conclusion, selecting log(d) values of 1 or 2 log viable *L. monocytogenes* organisms/g (and consequently different α values) had almost no influence on the estimation of the contamination distribution using a normal distribution.

Distribution of the contamination using a mixture model. The percentage of positive samples calculated for all studies is expressed as $\hat{\beta}$ = 3.01% (754 positive samples out of 25,078 tested samples of fresh unprocessed and minimally processed vegetables). Table 3 shows statistics of the posterior distributions of the hyperparameters. The posterior means of the mean, μ_p, and the standard deviation, σ_p, of the log concentrations, estimated from prevalence data, were −3.51 log viable *L. monocytogenes* organisms/g and 1.10 log viable *L. monocytogenes* organisms/g, respectively. Therefore, the future log concentrations mp^{pred} had a mean of −3.51 log viable *L. monocytogenes* organisms/g and a 95% credible interval of [−5.70; −1.33] log viable *L. monocytogenes* organisms/g, i.e., the log concentrations had a probability of 95% to belong to this interval (Table 4). The posterior means of the hyperparameters estimated with concentration data were 1.01 log viable *L. monocytogenes* organisms/g for the mean μ_c and 1.03 log viable *L. monocytogenes* organisms/g for the standard deviation σ_c of the log concentrations estimated from studies with observed concentrations. Therefore the predicted log concentrations mc^{pred} had a mean of 1.03 log viable *L. monocytogenes* organisms/g and a 95% credible interval of [−1.12; 3.22] log viable *L. monocytogenes* organisms/g (Table 4). The posterior mean of the standard deviation of the sample concentrations was 0.57 with a very low standard deviation of 10^{-4} (Table 3). The variability between studies of prevalence and studies with observed concentrations was quite similar and was higher than the variability within studies.

Table 2 and Fig. 2 show that the predicted log concentrations of vegetables with *L. monocytogenes* had a mean of −3.38 log viable *L. monocytogenes* organisms/g and a 95% credible interval of [−5.97; 0.07] log viable *L. monocytogenes* organisms/g. The model predicts probabilities of 1.44, 0.63, and 0.17% of samples of fresh unprocessed and minimally processed vegetables to be contaminated with concentrations higher than 1, 2, and 3 log viable *L. monocytogenes* organisms/g, respectively.

TABLE 4. Descriptive statistics of the predicted mean log concentrations of vegetables with *L. monocytogenes* estimated by the mixture model[a]

Mean log concn type	Mean	SD	2.5th percentile	Median	97.5th percentile
mp^{pred}	−3.51	1.11	−5.70	−3.50	−1.33
mc^{pred}	1.03	1.10	−1.12	1.02	3.22

[a] Results are expressed as log viable *L. monocytogenes* organisms per gram.

Influence of the sensitivity of the detection methods. Considering previous works performed on naturally contaminated or artificially spiked samples (2, 30, 40, 43, 73, 77–80), the percentage of sensitivity of the methods (samples testing positive/samples known to be contaminated with *L. monocytogenes*) ranges from 40 to 100%, being commonly close to 80%. In the absence of specific tests, these values were assumed to be the same for the detection of *L. monocytogenes* in vegetables and the triangular(40, 80, 100) was used for a variable sensitivity rate.

Introducing a fixed sensitivity rate of 80 or 95% in the mixture model had a small effect on the estimation of the contamination (Table 5). In contrast, introducing a low fixed sensitivity rate (40%) resulted in marked differences in the distribution, especially for high percentiles. The predicted mean log concentration was −2.74 log viable *L. monocytogenes* organisms/g and the 97.5th percentile was 1.56 log viable *L. monocytogenes* organisms/g (Table 5), while these values were −3.38 log viable *L. monocytogenes* organisms/g and 0.07 log viable *L. monocytogenes* organisms/g, respectively, when no sensitivity rate was used (which corresponds to a sensitivity rate of 100%) (Table 2). With a variable sensitivity rate distributed by a triangular distribution, the predicted mean log concentration and the 97.5th percentile were between those obtained with a sensitivity rate of 40% and those obtained without introducing a rate of sensitivity. Because of an increase in the proportion $\hat{\beta}$ of concentrations from the distribution of high concentrations, most differences appeared on the estimate of the 97.5th percentile, but not on the 2.5th percentile, the median, or the mean. For instance, a marked increase from $\hat{\beta}$ at 3.01% to $\hat{\beta}$ at 7.52% resulted from the introduction of a level of sensitivity of 40% (Table 5).

Influence of qualitative variables. Table 6 shows the descriptive statistics of the concentrations of vegetables with *L. monocytogenes* estimated with the mixture model for the different qualitative variables which are the type of product (three levels, leafy salads, sprouts, and other vegetables), the process degree (two levels, unprocessed vegetables, minimally processed vegetables), the period of publication (two levels, 1988 to 1999 and 2000 to 2005), and the geographical origin of the studies (three levels, North America, Europe, and other regions). The coefficients of the variables "type of product" and "period of publication" were significantly different from zero (with a credible interval of 95%). Leafy salads, with a mean log concentration of −3.36 log viable *L. monocytogenes* organisms/g and a 95% credible interval of [−5.99; −0.17] log viable *L. monocytogenes* organisms/g, showed lower predicted concentrations than those of sprouts and other vegetables, which had means of log concentrations of −3.09 and −3.43 log viable

TABLE 5. Descriptive statistics of the log concentrations of vegetables with *L. monocytogenes* estimated for different values of the sensitivity rate of the methods of detection[a]

Level of sensitivity[b] (%)	Proportion $\hat{\beta}$ of high concn (%)[c]	Mean	SD	2.5th percentile	Median	97.5th percentile
40	7.52	−2.74	1.61	−5.39	−2.91	1.56
80	3.76	−3.22	1.46	−5.73	−3.32	0.55
95	3.16	−3.33	1.41	−5.77	−3.41	0.10
Triangular (40; 80; 100)	4.12	−3.21	1.56	−5.95	−3.30	0.69

[a] Results are expressed as log viable *L. monocytogenes* organisms per gram unless otherwise indicated.
[b] A rate of sensitivity of 100% corresponds to estimations made without the introduction of a sensitivity level in the calculation and is found in Table 2.
[c] See Materials and Methods for details.

L. monocytogenes organisms/g, respectively, and 95% credible intervals of [−5.88; 0.74] and [−6.10; 0.69] log viable *L. monocytogenes* organisms/g, respectively. Concentrations of *L. monocytogenes* in fresh vegetables recorded in publications issued between 1988 and 1999, with a mean log concentration of −2.79 log viable *L. monocytogenes* organisms/g and a 95% credible interval of [−5.30; 1.49] log viable *L. monocytogenes* organisms/g, were higher than those from publications issued between 2000 and 2005 with a mean log concentration of −3.94 log viable *L. monocytogenes* organisms/g and a 95% credible interval of [−6.34; −1.23] log viable *L. monocytogenes* organisms/g.

DISCUSSION

Previous works on quantitative microbial risk assessment stuck strictly to prevalence and concentration data but treated them separately. Such an approach results in a nonestimation of low concentration levels. The method that we proposed to determine the parameters of a normal distribution using prevalence and concentration data can give an easy-to-apprehend estimate of a distribution of contamination. However, it leads to an intrinsic overestimation of the mean concentration because of the artificial consideration of one contaminated sample in studies in which all samples are in fact negative. In contrast, the mixture model offers a framework simultaneously accounting for both prevalence and concentration data. The mixture model using a probabilistic approach of the MPN theory offers an appropriate treatment of studies with all negative samples which are, however, worth considering. It allows an estimation of concentrations that are below the threshold of detection of microbiological methods, but that nevertheless have a high significance for QMRA. The overestimation of the normal distribution relative to the mixture model distribution can be seen on Fig. 2, showing the densities of the two distributions. The mixture model satisfactorily accounts for high concentrations. There are slight differences between the values predicted by the model and those generated by the random selection in the triangular distribution. For instance, the model predicts 1.44% of samples with more than 1 log viable *L. monocytogenes* organism/g and 0.63% of samples with more than 2 log viable *L. monocytogenes* organisms/g against 1.87 and 0.61%, respectively, for generated data. However, the differences between concentration data in class intervals (1.09 and 0.09% of observed samples with more than 1 and 2 log viable *L. monocytogenes* organisms/g, respectively) and predicted high concentrations are due to the constraint of data generation using a triangular distribution, which is nevertheless the most appropriate procedure. Parameters of a triangular distribution can be determined unambiguously from the format of the data, and the distribution accounts for a decrease in probability of high concentrations. The estimation process could be improved using methods adapted to class interval data, such as those for censored data (but with an increased complexity of the model) (33), or with a better availability of defined concentration data. Anyway, in the current situation, the estimation process is likely slightly conservative for high percentiles and therefore keeps its interest for risk assessors or risk managers.

Integrating variability and uncertainty presents great interest when output distributions have to be used as input values for

TABLE 6. Descriptive statistics of the log concentrations of the vegetables with *L. monocytogenes* by the type of product, the process degree, the period of publication, and the geographical origin of the studies[a]

Qualitative variable	Modality	No. of studies	Mean	SD	2.5th percentile	Median	97.5th percentile
Type of product	Leafy salads	38	−3.36 A	1.42 A	−5.99 A	−3.41 A	−0.17 A
	Sprouts	12	−3.09 B	1.55 B	−5.88 B	−3.17 B	0.74 B
	Other vegetables	115	−3.43 B	1.55 B	−6.10 B	−3.53 B	0.69 B
Process degree	Unprocessed	91	−3.40	1.47	−6.01	−3.48	0.11
	Minimally processed	74	−3.37	1.53	−6.03	−3.47	0.57
Period of publication	1988–1999	80	−2.79 C	1.53 C	−5.30 C	−2.93 C	1.49 C
	2000–2005	85	−3.94 D	1.31 D	−6.34 D	−3.97 D	−1.23 D
Geographical origin	North America	33	−3.73	1.39	−6.32	−3.78	−0.80
	Europe	95	−3.45	1.50	−6.05	−3.53	0.30
	Other countries	97	−3.07	1.38	−5.67	−3.12	−0.13

[a] For each variable, significant differences at the 95% credible interval between modalities are indicated by different letters. Results are expressed as log viable *L. monocytogenes* organisms per gram unless otherwise indicated.

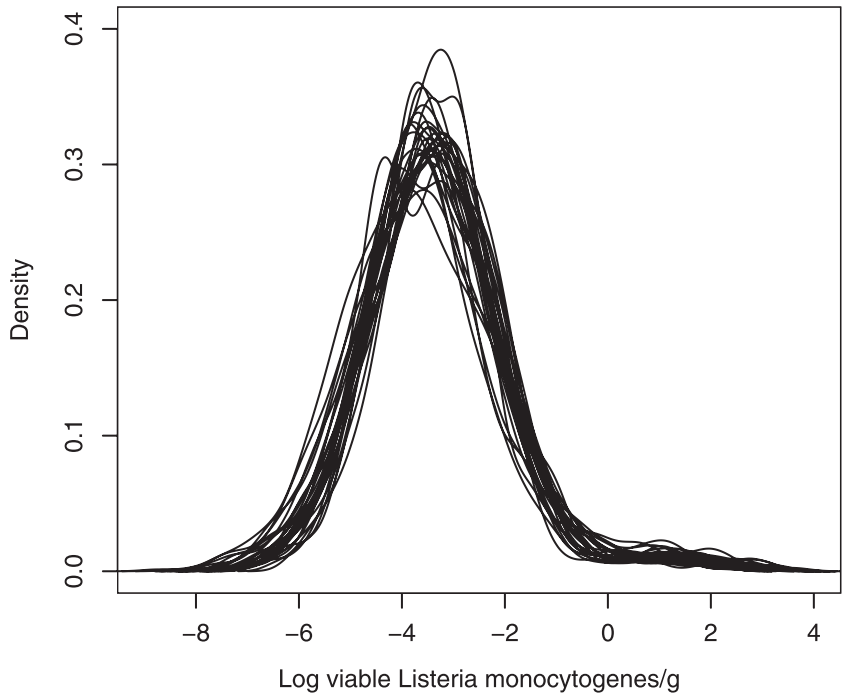

FIG. 3. Diversity of the distributions of predicted log concentrations of fresh vegetables with *L. monocytogenes* generated by variables μ_p, σ_p, μ_c, σ_c, and s in the mixture model. Fifty values were randomly selected among the B posterior values of each hyperparameter and parameter. Each curve represents one of the 50 derived density functions of $\log(c_b^{\text{pred}})$ $b = 1, \ldots 50$.

further modeling, such as for further QMRA (52, 76). The hierarchical structure of the mixture model makes possible the introduction in modeling of the "between" and the "within" variability and the uncertainty on the hyperparameters. The variability of the contaminations of fresh unprocessed and minimally processed vegetables between studies is higher than the variability within studies. To approach the overall variability of contaminations, a large number of studies from different contexts (time period, origin, and product) are necessary to reflect the variability of the contaminations, assuming that the variability due to the methods of detection has a minor contribution to the overall variability. The results of the mixture model shown in Table 3, i.e., the mean posterior values of the hyperparameters μ_p, σ_p, μ_c, and σ_c and the model parameter s can be used in expressions 8, 9, and 10 to simply establish a distribution of *L. monocytogenes* in fresh unprocessed and minimally processed vegetables. However this computation does not account for the uncertainty of the hyperparameters and parameter. The Bayesian approach makes possible the introduction of several distributions of contamination in simulation processes, including QMRA, determined from a random selection (accounting for the correlations between the hyperparameters and the parameter) of the values of the hyperparameters μ_p, σ_p, μ_c, and σ_c, and the parameter s in their posterior distributions. Figure 3 shows the diversity of distributions of contamination of fresh vegetables with *L. monocytogenes* generated by variables μ_p, σ_p, μ_c, σ_c, and s, reflecting the importance of considering variability and uncertainty. In contrast, Fig. 2 represents the distributions of contamination of fresh unprocessed and minimally processed vegetables with *L. monocytogenes*, when all variability and uncertainty are integrated, i.e., when all of the distributions of the hyperparameters and the parameter are used.

No statistically significant difference was observed between vegetables from different origins or between unprocessed and minimally processed vegetables. In contrast, studies of 2000 to 2005 showed lower levels of *L. monocytogenes* contamination than did the previous ones starting in 1988. This decrease could be attributed to improved management of risk as a consequence of the improved knowledge of the biology and ecology of *L. monocytogenes* since the first recognized food-borne outbreaks in the early 1980s. This suggests use of the most recent data for QMRA whenever sufficient data are available. In the same way, leafy salads also showed significantly lower numbers of predicted contaminations than sprouts and other vegetables and this suggests the preferential use of disaggregated data or of data relevant to the type of vegetable of interest.

Most studies report the presence-absence of *L. monocygenes* in series of food samples. As shown in this work, these data have a strong influence on the estimation of the mean concentration but are of limited interest in the estimation of the highest levels of concentration, precisely where major uncertainty lies. Studies combining prevalence and concentration data represent a much heavier and more tedious analytical effort, but have a tremendous impact on the quality of the estimation of the distribution of *L. monocytogenes* contamination. In addition, the sensitivity of the methods of detection must be high enough (80% showed to be quite satisfactory in this work) to limit high rates of false-negative samples, which may markedly alter the quality of the prediction. The transformation of data obtained by the food microbiologist into a statistical distribution is of fundamental interest to the risk

assessor who needs such structured data as input values for risk assessment (75). The mixture model developed in this work is a generic tool responding to the expectations of the risk assessors. It can be applied to contamination data in foods of other major food-borne pathogens, such as *Salmonella*, *Escherichia coli* O157:H7, and *Clostridium botulinum*, which have the same characteristics (majority of prevalence data with a low frequency of positive samples and scarce concentration data) as those of *L. monocytogenes* in vegetables.

ACKNOWLEDGMENTS

Amélie Crépet received a grant from INRA (Paris, France), the Ministry of Agriculture (Paris, France), and ACTIA (Paris, France) (Association de Coordination Technique pour l'Industrie Agro-alimentaire) under contract 03-13 through Aérial, Technology Resource Centre (Illkirch 67, France).

We thank Olivier Gouze for developing the triangular distribution and Valérie Stahl, Catherine Denis, Véronique Huchet, Olivier Cerf, Laurent Rosso, Jean-Jacques Daudin, and Jessica Tressou for stimulating discussions.

REFERENCES

1. **Arnold, G. J., and J. Coble.** 1995. Incidence of *Listeria* species in foods in NSW. Food Aust. **47:**71–75.
2. **Art, D., and P. Andre.** 1991. Comparison of three selective isolation media for the detection of *L. monocytogenes* in foods. Zentbl. Bakteriol. **275:**79–84.
3. **Arumugaswamy, R. K., R. R. A. Gulam, and N. B. A. H. Siti.** 1994. Prevalence of *Listeria monocytogenes* in foods in Malaysia. Int. J. Food Microbiol. **23:**117–121.
4. **Beaufort, A., G. Poumeyrol, and S. Rudelle.** 1992. Fréquence de contamination par *Listeria* et *Yersinia* d'une gamme de produits de 4e gamme. Rev. Gen. Froid. **March 1992:**28–31.
5. **Beckers, H. J., P. H. In't-Veld, P. S. S. Soentoro, and E. H. M. Delfgou van Asch.** The occurrence of *Listeria* in food. Foodborne listeriosis. Proceedings of a symposium on September 7, 1988 in Wiesbaden. Behr's Verlag, Hamburg, Germany.
6. **Bemrah, N., H. Bergis, C. Colmin, A. Beaufort, Y. Millemann, B. Dufour, J. J. Benet, O. Cerf, and M. Sanaa.** 2003. Quantitative risk assessment of human salmonellosis from the consumption of a turkey product in collective catering establishments. Int. J. Food Microbiol. **80:**17–30.
7. **Beuchat, L. R.** 2002. Ecological factors influencing survival and growth of human pathogens on raw fruits and vegetables. Microbes Infect. **4:**413–423.
8. **Beuchat, L. R.** 1996. *Listeria monocytogenes*: incidence of vegetables. Food Control **7:**223–228.
9. **Bind, J. L.** 1992. La listériose. Bull. Group Tech. Vet. **89:**59–77.
10. **Blodgett, R.** 2001. Bacteriological analytical manual online, appendix 2, most probable number from serial dilutions. http://www.cfsan.fda.gov/~ebam/bam-a2.html.
11. **Blodgett, R. J., and W. E. Garthright.** 1998. Several MPN models for serial dilutions with suppressed growth at low dilutions. Food Microbiol. **15:**91–99.
12. **Box, G. E. P., and G. C. Tiao.** 1973. Bayesian inference in statistical analysis. Addison-Wesley, Reading, MA.
13. **Breer, C., and A. Baumgartner.** 1992. Vorkommen un Verhalten von *Listeria monocytogenes* auf Salaten und Gemüsen sowie in frischgepreßten Gemüsesäften. Arch. Lebensmittelhyg. **43:**108–110.
14. **Breer, C., and K. Schopfer.** 1989. Listerien in Nahrungsmitteln. Schweiz. Med. Wochenschr. **119:**306–311.
15. **Cassin, M. H., A. M. Lammerding, E. C. Todd, W. Ross, and R. S. McColl.** 1998. Quantitative risk assessment for *Escherichia coli* O157:H7 in ground beef hamburgers. Int. J. Food Microbiol. **41:**21–44.
16. **Cho, S. Y., B. K. Park, K. D. Moon, and D. H. Oh.** 2004. Prevalence of *Listeria monocytogenes* and related species in minimally processed vegetables. J. Microbiol. Biotechnol. **14:**515–519.
17. **de Curtis, M. L., O. de Franceschi, and N. de Castro.** 2002. *Listeria monocytogenes* in vegetables minimally processed ready-to-use. Arch. Latinoam. Nutr. **52:**282–288.
18. **de Simón, M., C. Tarrago, and M. D. Ferrer.** 1992. Incidence of *Listeria monocytogenes* in fresh foods in Barcelona (Spain). Int. J. Food Microbiol. **16:**153–156.
19. **DGCCRF.** 2002. 1993–1996 and 1997–2001. Plan de surveillance de la contamination des aliments par *Listeria monocytogenes*. Internal report. DGCCRF, Paris, France.
20. **Dhanashree, B., S. K. Otta, I. Karunasagar, W. Goebel, and I. Karunasagar.** 2003. Incidence of *Listeria* spp. in clinical and food samples in Mangalore, India. Food Microbiol. **20:**447–453.
21. **Farber, J. M., G. W. Sanders, and M. A. Johnston.** 1989. A survey of various foods for the presence of *Listeria* species. J. Food Prot. **52:**456–458.
22. **Garcia-Gimeno, R. M., G. Zurera-Cosano, and M. Amaro-Lopez.** 1996. Incidence, survival and growth of *Listeria monocytogenes* in ready-to-use mixed vegetable salads in Spain. J. Food Saf. **16:**75–86.
23. **Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin.** 2004. Bayesian data analysis, 2nd ed. Chapman & Hall, Boca Raton, FL.
24. **Gohil, V. S., M. A. Ahmed, R. Davies, and R. K. Robinson.** 1995. Incidence of *Listeria* spp. in retail foods in the United Arab Emirates. J. Food Prot. **58:**102–104.
25. **Gombas, D. E., Y. Chen, R. S. Clavero, and V. N. Scott.** 2003. Survey of *Listeria monocytogenes* in ready-to-eat foods. J. Food Prot. **66:**559–569.
26. **Guerra, M. M., J. McLauchlin, and F. A. Bernardo.** 2001. *Listeria* in ready-to-eat and unprocessed foods produced in Portugal. Food Microbiol. **18:**423–429.
27. **Gunasena, D. K., C. P. Kodikara, G. Kumuda, and S. Widanapathirana.** 1995. Occurrence of *Listeria monocytogenes* in food in Sri Lanka. J. Natl. Sci. Counc. Sri Lanka **23:**107–114.
28. **Halvorson, H. O., and N. R. Ziegler.** 1933. Applications of statistics to problems in bacteriology. I. A means of determining bacterial population by the dilution method. J. Bacteriol. **25:**101–121.
29. **Harvey, J., and A. Gilmour.** 1993. Occurrence and characteristics of *Listeria* in foods produced in Northern Ireland. Int. J. Food Microbiol. **19:**193–205.
30. **Hayes, P. S., L. M. Graves, G. W. Ajello, B. Swaminathan, R. E. Weaver, J. D. Wenger, A. Schuchat, and C. V. Broome.** 1991. Comparison of cold enrichment and U.S. Department of Agriculture methods for isolating *Listeria monocytogenes* from naturally contaminated foods. Appl. Environ. Microbiol. **57:**2109–2113.
31. **Heisick, J. E., F. M. Harrel, E. H. Peterson, S. McLaughlin, D. E. Wagner, I. V. Wesley, and J. Bryner.** 1989. Comparison of four procedures to detect *Listeria* spp. in foods. J. Food Prot. **52:**154–157.
32. **Heisick, J. E., D. E. Wagner, M. L. Nierman, and J. T. Peeler.** 1989. *Listeria* spp. found on fresh market produce. Appl. Environ. Microbiol. **55:**1925–1927.
33. **Helsel, D. R.** 2004. Nondetects and data analysis: statistics for censored environmental data. Statistics in practice. Wiley, New York, NY.
34. **Ho, J. L., K. N. Shands, G. Friedland, P. Eckind, and D. W. Fraser.** 1986. An outbreak of type 4b *Listeria monocytogenes* involving patients from eight Boston hospitals. Arch. Intern. Med. **146:**520–524.
35. **Hoornstra, E., and S. Notermans.** 2001. Quantitative microbiological risk assessment. Int. J. Food Microbiol. **66:**21–29.
36. **Inoue, S., A. Nakama, Y. Arai, Y. Kokubo, T. Maruyama, A. Saito, T. Yoshida, M. Terao, S. Yamamoto, and S. Kumagai.** 2000. Prevalence and contamination levels of *Listeria monocytogenes* in retail foods in Japan. Int. J. Food Microbiol. **59:**73–77.
37. **Jarvis, B.** 2000. Sampling for microbiological analysis, p. 1691–1733. *In* B. M. Lund, A. C. Baird-Parker, and G. W. Gould, The microbiological quality and safety of food, volume 2. Aspen Publishers, Gaithersburg, MD.
38. **Johannessen, G. S., S. Loncarevic, and H. Kruse.** 2002. Bacteriological analysis of fresh produce in Norway. Int. J. Food Microbiol. **77:**199–204.
39. **Kaneko, K., H. Hayashidani, Y. Ohtomo, J. Kosuge, M. Kato, K. Takahashi, Y. Shiraki, and M. Ogawa.** 1999. Bacterial contamination of ready-to-eat foods and fresh products in retail shops and food factories. J. Food Prot. **62:**644–649.
40. **Kornacki, J. L., D. J. Evanson, W. Reid, K. Rowe, and R. S. Flowers.** 1993. Evaluation of the USDA protocol for detection of *Listeria monocytogenes*. J. Food Prot. **56:**441–443.
41. **Laine, K., and J. Michard.** 1988. Fréquence et abondance des *Listeria* dans des légumes frais découpés prêts à l'emploi. Microbiol. Alim. Nutr. **6:**329–335.
42. **Levre, E., P. Valentini, and G. Caroli.** 1995. Prodotti orticoli e contaminazione da *Listeria*. Ig. Mod. **103:**475–486.
43. **Liming, S. H., Y. Zhang, J. Meng, and A. A. Bhagwat.** 2004. Detection of *Listeria monocytogenes* in fresh produce using molecular beacon-real-time PCR technology. J. Food Sci. **69:**M240–M245.
44. **Lindqvist, R., and A. Westöö.** 2000. Quantitative risk assessment for *Listeria monocytogenes* in smoked or gravad salmon and rainbow trout in Sweden. Int. J. Food Microbiol. **58:**181–196.
45. **Little, C., D. Y. Roberts, E. Youngs, and J. DeLouvois.** 1999. Microbiological quality of retail imported unprepared whole lettuces: a PHLS food working group study. J. Food Prot. **62:**325–328.
46. **Little, C. L., H. A. Monsey, G. L. Nichols, and J. de Louvais.** 1997. The microbiological quality of refrigerated salads and crudites. An analysis of the results from the 1995 European Community Coordinated Food Control Programme for England and Wales. PHLS Microbiol. Dig. **14:**142–146.
47. **MacGowan, A. P., K. Bowker, J. McLauchlin, P. M. Bennett, and D. S. Reeves.** 1994. The occurrence and seasonal changes in the isolation of *Listeria* spp. in shop bought food stuffs, human faeces, sewage and soil from urban sources. Int. J. Food Microbiol. **21:**325–334.
48. **McLauchlin, J., and R. J. Gilbert.** 1990. *Listeria* in food. PHLS Microbiol. Dig. **7:**54–55.
49. **Messi, P., C. Casolari, A. Fabio, G. Fabio, C. Gibertoni, G. Menziani, and P. Quaglio.** 2000. Occurrence of *Listeria* in food matrices. Ind. Aliment. **39:**151–157.

50. **Monge, R., and M. L. Arias.** 1996. Presencia de microorganismos patogenos en las hortalizas de cosumo crudo en Costa Rica. Arch. Latinoam. Nutr. **46:**293–294.
51. **Murase, M., T. Miyata, H. Kimata, and M. Kurokawa.** 2002. Investigation of the actual conditions of pathogenic bacteria among imported fresh vegetables and fruits. Jpn. J. Food Microbiol. **19:**71–75.
52. **Nauta, M. J.** 2000. Separation of uncertainty and variability in quantitative microbial risk assessment models. Int. J. Food Microbiol. **57:**9–18.
53. **Nauta, M. J., S. Litman, G. C. Barker, and F. Carlin.** 2003. A retail and consumer phase model for exposure assessment of *Bacillus cereus*. Int. J. Food Microbiol. **83:**205–218.
54. **Nguyen-the, C., and F. Carlin.** 2000. Fresh and processed vegetables, p. 620–684. *In* B. M. Lund, A. C. Baird-Parker, and G. W. Gould, The microbiological quality and safety of food. Aspen Publishers, Gaithersburg, MD.
55. **Nørrung, B., J. K. Andersen, and J. Schlundt.** 1999. Incidence and control of *Listeria monocytogenes* in foods in Denmark. Int. J. Food Microbiol. **53:**195–203.
56. **Pacini, R., L. Panizzi, E. Quagli, R. Galassi, L. Malloggi, and R. Morganti.** 1993. Diffusione di *Listeria monocytogenes*. Ind. Aliment. **32:**1086–1089.
57. **Porto, E., and M. N. Uboldi Eiroa.** 2001. Occurrence of *Listeria monocytogenes* in vegetables. Dairy Food Environ. Sanit. **2:**282–286.
58. **Pouillot, R., I. Albert, M. Cornu, and J. B. Denis.** 2003. Estimation of uncertainty and variability in bacterial growth using Bayesian inference. Application to *Listeria monocytogenes*. Int. J. Food Microbiol. **81:**87–104.
59. **Prazak, A. M., E. A. Murano, I. Mercado, and G. R. Acuff.** 2002. Prevalence of *Listeria monocytogenes* during production and postharvest processing of cabbage. J. Food Prot. **65:**1728–1734.
60. **Previdi, M. P., L. Tomasoni, and B. Bondi.** 2000. Minimally processed and frozen vegetables: microbiological quality and incidence of pathogens in commercial products. Ind. Conserve **75:**383–391.
61. **Robert, C. P.** 2001. The Bayesian choice, 2nd ed. Springer, New York, NY.
62. **Sagoo, S. K., C. L. Little, and R. T. Mitchell.** 2001. The microbiological examination of ready-to-eat organic vegetables from retail establishments in the United Kingdom. Lett. Appl. Microbiol. **33:**434–439.
63. **Sagoo, S. K., C. L. Little, and R. T. Mitchell.** 2003. Microbiological quality of open ready-to-eat salad vegetables: effectiveness of food hygiene training of management. J. Food Prot. **66:**1581–1586.
64. **Sagoo, S. K., C. L. Little, L. Ward, I. A. Gillespie, and R. T. Mitchell.** 2003. Microbiological study of ready-to-eat salad vegetables from retail establishments uncovers a national outbreak of salmonellosis. J. Food Prot. **66:**403–409.
65. **Schlech, W. F., III, P. M. Lavigne, R. A. Bortolussi, A. C. Allen, E. V. Haldane, A. J. Wort, A. W. Hightower, S. E. Johnson, S. H. King, E. S. Nicholls, and C. V. Broome.** 1983. Epidemic listeriosis—evidence for transmission by food. N. Engl. J. Med. **308:**203–206.
66. **Sizmur, K., and C. W. Walker.** 1988. *Listeria* in prepacked salads. Lancet **i:**1167.
67. **Soriano, J. M., H. Rico, J. C. Molto, and J. Manes.** 2001. *Listeria* species in raw and ready-to-eat foods from restaurants. J. Food Prot. **64:**551–553.
68. **Spiegelhalter, D. J., A. Thomas, N. Best, and D. Lunn.** 2005. OpenBUGS version 2.10, user manual. MRC Biostatistics Unit, Cambridge, United Kingdom; Department of Epidemiology and Public Health, Imperial College, London, United Kingdom; and Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland.
69. **Stahl, V., C. Denis, and V. Huchet.** Survey on the contamination of minimally processed fresh salads with major foodborne pathogens. Internal report. ACTIA 03.13. AERIAL, Illkirch, France.
70. **Szabo, E. A., K. J. Scurrah, and J. M. Burrows.** 2000. Survey for psychrotrophic bacterial pathogens in minimally processed lettuce. Lett. Appl. Microbiol. **30:**456–460.
71. **Thunberg, R. L., T. T. Tran, R. W. Bennett, R. N. Matthews, and N. Belay.** 2002. Microbial evaluation of selected fresh produce obtained at retail markets. J. Food Prot. **65:**677–682.
72. **Tiwari, N. P., and S. G. Aldenrath.** 1990. Occurrence of *Listeria* species in food and environmental samples in Alberta. Can. Inst. Food Sci. Technol. J. **23:**109–113.
73. **Twedt, R. M., A. D. Hitchins, and G. A. Prentice.** 1994. Determination of the presence of *Listeria monocytogenes* in milk and dairy products: IDF collaborative study. J. AOAC Int. **77:**395–402.
74. **Villari, P., G. Greco, and I. Torre.** 1997. *Listeria* spp. in ready-to-eat foods. Ig. Mod. **107:**403–410.
75. **Vose, D.** 1998. The application of quantitative risk assessment to microbial food safety. J. Food Prot. **61:**640–648.
76. **Vose, D.** 2000. Risk analysis: a quantitative guide, 2nd ed. John Wiley and Sons, Chichester, United Kingdom.
77. **Waak, E., W. Tham, and M. L. Danielsson-Tham.** 1999. Comparison of the ISO and IDF methods for detection of *Listeria monocytogenes* in blue veined cheese. Int. Dairy J. **9:**149–155.
78. **Walsh, D., G. Duffy, J. J. Sheridan, I. S. Blair, and D. A. McDowell.** 1998. Comparison of selective and nonselective enrichment media for the isolation of *Listeria* species from retail foods. J. Food Saf. **18:**85–99.
79. **Warburton, D. W., J. M. Farber, A. Armstrong, R. Caldeira, T. Hunt, S. Messier, R. Plante, N. P. Tiwari, and J. Vinet.** 1991. A comparative study of the 'FDA' and 'USDA' methods for the detection of *Listeria monocytogenes* in foods. Int. J. Food Microbiol. **13:**105–117.
80. **Westöö, A., and M. Peterz.** 1992. Evaluation of methods for detection of *Listeria monocytogenes* in foods: NMKL collaborative study. J. AOAC Int. **75:**46–52.
81. **Wong, H. C., W. L. Chao, and S. J. Lee.** 1990. Incidence and characterization of *Listeria monocytogenes* in foods available in Taiwan. Appl. Environ. Microbiol. **56:**3101–3104.

C.Q. N.O. PAGE /

Enquête sur les habitudes d'achat et de consommation
de salades en sachets prêtes à consommer

Vous êtes

☐ Homme

☐ Femme

Votre âge

☐ 20 à 30 ans

☐ 31 à 40 ans

☐ 41 à 50 ans

☐ 51 à 60 ans

☐ + de 60 ans

Situation familiale

☐ célibataire

☐ marié(e)ou vie maritale

☐ veuf(ve)

☐ divorcé(e)

Catégorie socio professionnelle

☐ agriculteur

☐ artisan- commerçant

☐ cadre supérieur - Profession libérale

☐ cadre moyen - Agent de maîtrise

☐ employé

☐ étudiant

☐ femme ou homme au foyer

☐ ouvrier

☐ recherche d'emploi

☐ retraité

Lieu d'habitation

☐ Nord-est

☐ Nord-ouest

☐ Centre

☐ Sud-est

☐ Sud-ouest

☐ Ile de France

Type d'habitation

☐ zone rurale

☐ ville

Nombre de personnes au foyer (vous inclus) :_____ âge:_____

dont nombre d'enfants:_____ âges :_____

Qui dans le foyer fait les courses alimentaires?

☐ vous

☐ autre(s)

C.Q. N.O. PAGE /

Enquête sur les habitudes d'achat et de consommation
de salades en sachets prêtes à consommer

Fréquences d'achat de salades en sachet prêtes à consommer?

	en hiver	au printemps	en été	en automne
jamais	☐	☐	☐	☐
1 à 2 fois par saison	☐	☐	☐	☐
1 fois par mois	☐	☐	☐	☐
plusieurs fois par mois	☐	☐	☐	☐
1 fois par semaine	☐	☐	☐	☐
plusieurs fois par semaine	☐	☐	☐	☐

Quelle est la proportion de salades en sachet prêtes à consommer achetées par rapport à des salades entières (jardin- marché-grande surface)?

	en hiver	au printemps	en été	en automne
toujours (100% en sachets)	☐	☐	☐	☐
souvent (75% en sachets)	☐	☐	☐	☐
régulièrement (50% en sachets)	☐	☐	☐	☐
parfois (25% en sachets)	☐	☐	☐	☐
occasionnellement (entre 0 et 25% en sachets)	☐	☐	☐	☐
jamais (o% en sachets)	☐	☐	☐	☐

Quels types de salades en sachet prêtes à consommer achetez vous le plus souvent? (plusieurs réponses possibles)

☐ mâche ☐ laitue ☐ scarole ☐ jeunes pousses

☐ frisée ☐ mélange ☐ bio ☐ autres(précisez): _____

Achetez vous des salades prêtes à consommer plutôt quand vous êtes nombreux à table? oui ☐ pas spécialement ☐ non ☐

VOS HABITUDES D'ACHAT

Où achetez vous la salade en sachet prête à consommer le plus souvent?

magasin de proximité shopi,coop,...	au supermarché Atac champion,..	hard discount Lidl,Aldi,...	à l'hypermarché Carrefour,cora,...
☐	☐	☐	☐

en général, combien de temps s'écoule entre l'achat et l'arrivée à votre domicile? moins de 1 heure ☐ entre 1 et 2 heures ☐ plus de 2 heures ☐

C.Q. N.O. PAGE /

Enquête sur les habitudes d'achat et de consommation
de salades en sachets prêtes à consommer

Comment transportez vous les salades en sachets prêtes à consommer durant le trajet?

	sans précaution particulière	en sac isotherme
Hiver	☐	☐
Printemps	☐	☐
Eté	☐	☐
Automne	☐	☐

Dans les rayons choisissez vous

- ☐ le premier sachet qui se présente
- ☐ un des sachets derrière
- ☐ en fonction de l'aspect
- ☐ en fonction de la date limite de consommation
- ☐ indifféremment

VOS HABITUDES DE CONSOMMATION

Chez vous, comment conservez vous la salade en sachet prête à consommer

☐ à température ambiante ☐ au réfrigérateur ☐ dans un endroit frais (cave,balcon)

si vous la conservez au réfrigérateur

☐ bac à légumes ☐ milieu ☐ haut du réfrigérateur

Combien de temps conservez vous le sachet de salade non ouvert?

☐ 1 ou 2 jours maximum ☐ entre 3 et 5 jours ☐ plus de 5 jours

Lisez vous la mention "à consommer jusqu'au" indiquée sur le sachet au moment de l'achat?

☐ oui ☐ non

Lisez vous la mention "à consommer jusqu'au" indiquée sur le sachet au moment de la consommation?

☐ oui ☐ non

Si oui, vous arrive t-il de consommer le produit après la date indiquée?

☐ oui souvent ☐ parfois ☐ non jamais

C.Q. . . . . N.O. . . . . PAGE . . . . /. . . .

Enquête sur les habitudes d'achat et de consommation
de salades en sachets prêtes à consommer

Vous arrive t-il d'entamer le sachet et de le conserver à nouveau?

☐ oui souvent ☐ parfois ☐ non jamais

Si oui, combien de temps conservez vous le sachet entamé?

1 ou 2 jours maximum ☐ entre 3 et 5 jours ☐ plus de 5 jours ☐

Si oui,comment conservez vous la salade prête à consommer une fois le sachet entamé?

☐ dans son sachet d'origine ☐ dans une boîte plastique fermée ☐ dans une boîte plastique ouverte

☐ dans un autre sachet (sac congélateur, film plastique)

Comment consommez vous la salade en sachet le plus souvent?

☐ sans sauce ☐ avec sauce

Combien de temps avant le repas préparez vous la salade?

juste avant ☐ moins de 30 min avant ☐ + de 30 min avant ☐

Consommez vous la salade en sachet prête à consommer plutôt?

☐ le midi ☐ le soir

Comment préparez vous la salade prête à consommer?

(plusieurs réponses possibles)

vous l'utilisez telle quelle ☐ vous triez certaines feuilles ☐ vous lavez les feuilles ☐

Qui dans votre foyer consomment de la salade en sachet prête à consommer

(plusieurs réponses possibles)

vous ☐ votre conjoint ☐ âge_____ vos enfants ☐ âge(s)_____

En général, combien de parts considérez vous dans un paquet (250g)

moins de 4 parts ☐ 4 parts ☐ 6 parts ☐ 8 parts ☐

☐ ne sais pas

MERCI DE VOTRE PARTICIPATION

Résumé

Listeria monocytogenes par sa présence dans les végétaux et sa capacité à se développer à de faibles températures représente un danger pour les consommateurs de salades de IV^{ème} gamme. L'objectif de ces travaux est de construire un modèle d'évaluation des risques de listériose par consommation de ce produit. Nous opérons en deux temps : estimation des paramètres d'entrée du modèle par inférence bayésienne puis, à partir des distributions obtenues, simulation de l'exposition et du risque par méthodes de Monte-Carlo de second ordre. Ces techniques permettent de propager séparément la variabilité et l'incertitude le long de la chaîne alimentaire. En particulier, un modèle d'estimation de la distribution de la contamination microbiologique d'un aliment, tenant compte des faibles concentrations est développé. L'effet sur le risque de différents scénarios comme le plafonnement de la croissance de *L. monocytogenes* ou l'élimination du chlore du procédé industriel est évalué.

Mots-clés : Statistique bayésienne, simulation de Monte-Carlo de second ordre, *Listeria monocytogenes*, évaluation des risques alimentaires, modèle non linéaire de croissance microbiologique et données incomplètes.

Abstract

The bacterial foodborne pathogen *Listeria monocytogenes* is a concern in minimally processed fresh vegetables for its ability to grow on fresh vegetables, even at low temperatures. The aim of this work is to build a model for the assessment of microbial risk on minimally processed vegetables. Firstly, the input parameters of the model are estimated through a bayesian inference. Secondly, second order Monte Carlo simulations are performed to estimate the exposure and the risk of listeriosis of the consumers. These techniques allows to propagate separately the variability and the uncertainty along the food chain. In particular a model estimating the distribution of microbial contaminations specifically accounting for low concentrations has been developed. The effect of several scenarios (maximal *L. monocytogenes* growth depending on initial concentration, consequence of the elimination of chlorine from the industrial process) on the risk of listeriosis following consumption of minimally processed fresh salads has been evaluated and discussed.

Keywords: Bayesian statistic, second order Monte-Carlo simulations, *Listeria monocytogenes*, food risk assessment, microbiological non linear growth model and incomplete data.