



Second order Monte Carlo and Bayesian inference methods for assessing microbiological risks and nutritional benefits in transformation of vegetable

Clémence Rigaux, Clémence Rigaux

► To cite this version:

Clémence Rigaux, Clémence Rigaux. Second order Monte Carlo and Bayesian inference methods for assessing microbiological risks and nutritional benefits in transformation of vegetable. Life Sciences. AgroParisTech, 2013. French. <NNT : 2013AGPT0015>. <pastel-00967496>

HAL Id: pastel-00967496

<https://pastel.archives-ouvertes.fr/pastel-00967496>

Submitted on 28 Mar 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Doctorat ParisTech

THÈSE

pour obtenir le grade de docteur délivré par

**L'Institut des Sciences et Industries
du Vivant et de l'Environnement**

(AgroParisTech)

Spécialité : Biostatistique

présentée et soutenue publiquement par

Clémence RIGAUX

le 14 février 2013

**Méthodes de Monte Carlo du second ordre et d'inférence bayésienne
pour l'évaluation des risques microbiologiques et des bénéfices
nutritionnels dans la transformation des légumes**

Directeur de thèse : **Frédéric CARLIN**
Co-directeur de thèse : **Isabelle ALBERT**

Jury

Mme Marie-Laure DELIGNETTE-MULLER, Professeur, VetAgroSup, Université Lyon 1
M. Marcel ZWIETERING, Professeur, Laboratory of Food Microbiology, Wageningen University
M. Eric PARENT, ICPEF HDR, MIA 518, ENGREF-AgroParisTech
Mme Nathalie PERROT, Directeur de recherche, INRA, Unité MALICES
M. Frédéric CARLIN, Directeur de Recherche, INRA UMR 408, Université d'Avignon
Mme Isabelle ALBERT, Chargée de Recherche, INRA, Unité Met@risk

Rapporteur
Rapporteur
Examineur
Examineur
Directeur de thèse
Codirecteur de thèse



Doctorat ParisTech

THÈSE

pour obtenir le grade de docteur délivré par

**L'Institut des Sciences et Industries
du Vivant et de l'Environnement**

(AgroParisTech)

Spécialité : Biostatistique

présentée et soutenue publiquement par

Clémence RIGAUX

le 14 février 2013

Méthodes de Monte Carlo du second ordre et d'inférence bayésienne pour l'évaluation des risques microbiologiques et des bénéfiques nutritionnels dans la transformation des légumes

Directeur de thèse : **Frédéric CARLIN**
Co-directeur de thèse : **Isabelle ALBERT**

Jury

Mme Marie-Laure DELIGNETTE-MULLER, Professeur, VetAgroSup, Université Lyon 1
M. Marcel ZWIETERING, Professeur, Laboratory of Food Microbiology, Wageningen University
M. Eric PARENT, ICPEF HDR, MIA 518, ENGREF-AgroParisTech
Mme Nathalie PERROT, Directeur de recherche, INRA, Unité MALICES
M. Frédéric CARLIN, Directeur de Recherche, INRA UMR 408, Université d'Avignon
Mme Isabelle ALBERT, Chargée de Recherche, INRA, Unité Met@risk

Rapporteur
Rapporteur
Examinateur
Examinateur
Directeur de thèse
Codirecteur de thèse

Remerciements

Je tiens à remercier toutes les personnes qui ont contribué à cette thèse, qui est une grande aventure scientifique mais aussi humaine. Tout d'abord, un grand merci à mes deux directeurs de thèse, Frédéric Carlin et Isabelle Albert, qui m'ont accordé leur confiance et m'ont accompagnée dans mes travaux tout au long de cette thèse. Merci Isabelle pour le sérieux de ton suivi. Frédéric, tes encouragements ont été précieux et ça a été un véritable plaisir de découvrir le monde de la microbiologie et du risque alimentaire à tes côtés. Mes brefs séjours à l'INRA d'Avignon m'ont aussi permis de découvrir l'univers de la paillasse, les boîtes de pétri, le festival d'Avignon et les beaux remparts de cette ville ensoleillée !

Un grand merci également à mes collaborateurs statisticiens : Sophie Ancelet, qui a partagé mes soucis et mes espoirs sur la purée de courgette, ainsi que son expérience de jeune chercheuse, et Jean-Baptiste Denis, pour ses encouragements, ses discussions scientifiques et son aide sur les haricots verts... Merci également aux microbiologistes et aux chimistes qui ont travaillé à nos côtés et dont les connaissances sont indispensables pour la modélisation, l'interprétation des données et des résultats : Christophe Nguyen-thé, Stéphane André, Stéphane Georgé, Catherine Renard, ainsi que les autres membres du projet Ribenut. J'en profite pour remercier l'Agence Nationale de la Recherche qui a financé ma thèse.

Merci à Marie-Laure Delignette-Muller et à Marcel Zwietering d'avoir accepté d'être rapporteurs de ma thèse. Merci également à Eric Parent d'avoir accepté de présider mon jury et d'avoir suivi mon travail au cours des différents comités de thèse, et à Nathalie Perrot d'avoir accepté d'être examinatrice de ma thèse.

Je tiens également à exprimer ma gratitude aux autres membres de mon comité de thèse, pour avoir régulièrement suivi mon travail et pour leurs conseils avisés : David Makowski, Jeanne-Marie Membré, Sophie Donnet et Francis Courtois.

Un grand merci à tous les membres de l'unité Met@risk et aux stagiaires de passage pour le partage de moments conviviaux comme les pauses-café ou les repas à notre excellente cantine. En particulier, un grand merci à Laura pour son efficacité et son amitié, à Jessica pour ses conseils et sa bonne humeur et à Eve dont l'évolution de la thèse a coïncidé avec la mienne. Que ce soit en sociologie ou en statistique, le quotidien des thésards se ressemble et ce fut un plaisir de partager nos expériences, nos peines et nos joies !

Merci également aux autres doctorants et (jeunes) chercheurs rencontrés au cours de cette thèse pour nos discussions scientifiques, le partage de nos expériences et à l'occasion de bons repas

(merci Sophie !). En particulier merci à Natalie Commeau et Aurore Lavigne, 2 étages ça fait une différence mais les problématiques sont parfois similaires...

Merci à mes amis pour tous les bons moments partagés, les soirées, les bons repas, les rencontres sportives etc.

Un grand merci à mes parents et à toute ma famille, qui m'ont soutenue et encouragée tout au long de mes études et de cette thèse. Le goût des sciences, pour nous c'est de famille ! Merci en particulier à Isabelle et Gilles qui m'ont accueillie chaleureusement à chacun de mes passages à Avignon.

Enfin, merci Rodolphe pour tout ton amour et ton soutien inconditionnel, puissions-nous toujours naviguer ensemble, vers de nombreuses et belles aventures !

Résumé

L'objectif du travail est de mettre en place des méthodes d'évaluation des risques microbiologiques et des bénéfices nutritionnels lors de la transformation des légumes, en vue d'une analyse risque-bénéfice. Le risque (industriel) considéré est l'altération des conserves de haricots verts due à la bactérie thermophile *Geobacillus stearothermophilus*, et le bénéfice nutritionnel est la teneur en vitamine C des haricots verts appertisés. Des paramètres de référence ont d'abord été acquis, par une méta-analyse par approche bayésienne pour la partie risque. Les paramètres de résistance thermique D à 121.1°C et pH 7, z_T et z_{pH} de *G. stearothermophilus* ont ainsi été estimés respectivement à 3.3 min, 9.1°C et 4.3 unités pH en moyenne en milieu aqueux. Les modèles risque et bénéfice ont ensuite été analysés par la méthode de simulation de Monte Carlo à deux dimensions, permettant une propagation séparée de la variabilité et de l'incertitude. Les pertes en vitamine C prédites par le modèle entre la matière première et les légumes appertisés sont de 86% en moyenne, et le taux de non-stabilité à 55°C prédit est de 0.5% en moyenne, globalement en accord avec la réalité. Une analyse risque-bénéfice a ensuite visé à optimiser le bénéfice tout en gardant le risque à un niveau acceptable, par exploration de scénarios d'interventions possibles basés sur les résultats d'analyses de sensibilité. Enfin, un modèle d'évaluation des risques dus à la bactérie pathogène *Bacillus cereus* dans une purée de courgettes a été confronté à des données de contamination dans des produits incubés, par le biais d'une inférence bayésienne.

Mots clés : Statistique bayésienne, simulation de Monte-Carlo du second ordre, analyse risque-bénéfice, évaluation des risques alimentaires, évaluation du bénéfice nutritionnel, *Geobacillus stearothermophilus*, *Bacillus cereus*, Vitamine C, légumes, conserves.

Abstract

The aim of this work is to set up microbiological risk and nutritional benefit assessment methods in the transformation of vegetables, in view of a risk-benefit analysis. The considered (industrial) risk is the alteration of green bean cans due to thermophilic bacteria *Geobacillus stearothermophilus*, and the nutritional benefit is the vitamin C content in appertized green beans. Reference parameters have first been acquired, by a meta-analysis using Bayesian inference for the risk part. Thermal resistance parameters D at 121.1°C and pH 7, z_T and z_{pH} of *G. stearothermophilus* have been respectively estimated at 3.3 min, 9.1°C and 4.3 pH units on average in aqueous media. The risk and benefit models have then been analyzed by a two-dimensional Monte Carlo simulation method, allowing a separated propagation of uncertainty and variability. The vitamin C losses between fresh and appertized green beans predicted by the model are of 86% on average, and the predicted non-stability at 55°C rate is of 0.5% on average, in good accordance with reality. A risk-benefit analysis has then been carried out to optimize benefit while keeping risk at an acceptable level, by exploring possible intervention scenarios based on some sensibility analysis results. Finally, a risk analysis model involving pathogenic bacteria *Bacillus cereus* in a courgette puree has been confronted to incubated product contamination data, by means of a Bayesian inference.

Keywords: Bayesian statistics, second order Monte Carlo simulation, risk-benefit analysis, food risk assessment, nutritional benefit assessment, *Geobacillus stearothermophilus*, *Bacillus cereus*, Vitamin C, vegetables, canned food.

Title: Second order Monte Carlo and Bayesian inference methods for assessing microbiological risks and nutritional benefits in transformation of vegetable.

Laboratoire d'accueil

Institut national de recherche agronomique (INRA)

UR1204 Unité de Méthodologies d'Analyses du Risque Alimentaire (Met@risk)

16 rue Claude BERNARD

75 231 Paris cedex 05

Publications scientifiques issues de ce travail de thèse

Articles publiés

Rigaux C., Ancelet S., Carlin F., Nguyen-Thé C., Albert I., 2012. *Inferring an augmented Bayesian network to confront a complex quantitative microbial risk assessment model with durability studies: application to Bacillus cereus on a courgette purée production chain*. *Risk Analysis*, doi: 10.1111/j.1539-6924.2012.01888.x.

Rigaux C., Denis J-B., Albert I., Carlin F., 2013. *A meta-analysis accounting for sources of variability to estimate heat resistance reference parameters of bacteria using hierarchical Bayesian modeling: Estimation of D at 121.1 °C and pH 7, z_T and z_{pH} of Geobacillus stearothermophilus*. *International Journal of Food Microbiology*, 161, p 112-120.

Article soumis

Rigaux C., André S., Albert I., Carlin F. *Quantitative assessment of the risk of microbial spoilage in foods. Prediction of non-stability at 55°C caused by Geobacillus stearothermophilus in canned green beans*. Soumis à *International Journal of Food Microbiology*.

Table des matières

1	Introduction.....	19
1.1	Risque et bénéfice dans les aliments	19
1.1.1	L'analyse du risque alimentaire.....	20
1.1.2	L'analyse du bénéfice nutritionnel	22
1.1.3	L'analyse risque-bénéfice	23
1.2	Méthodes statistiques utilisées.....	29
1.2.1	Modélisation, modèles probabilistes, incertitude et variabilité	29
1.2.2	La simulation de Monte Carlo (1D et 2D)	30
1.2.3	L'approche bayésienne.....	34
1.2.4	Analyse de sensibilité	35
1.3	Cas d'études de la thèse.....	36
1.3.1	Risque microbiologique et bénéfice nutritionnel dans les haricots verts appertisés ...	36
1.3.1.a	La chaîne de production « haricots verts appertisés ».....	36
1.3.1.b	Le risque d'altération dans la conserve de haricots verts	37
1.3.1.c	Le bénéfice nutritionnel dans la conserve de haricots verts.....	39
1.3.2	Risque microbiologique dans la purée de courgettes.....	40
1.4	Organisation du mémoire	41
2	Une méta-analyse bayésienne de paramètres de résistance thermique d'une bactérie de référence en microbiologie des aliments : <i>Geobacillus stearothermophilus</i>.....	43
2.1	Estimation de paramètres microbiologiques par méta-analyse	43
2.2	Article : A meta-analysis accounting for sources of variability to estimate heat resistance reference parameters of bacteria using hierarchical Bayesian modeling: estimation of D at 121.1°C and pH 7, z_T and z_{pH} of <i>Geobacillus stearothermophilus</i>	45
2.2.1	Introduction.....	46

2.2.2	Material and Methods.....	47
2.2.2.a	Inactivation models.....	47
2.2.2.b	Data.....	47
2.2.2.c	Three competitive models.....	49
2.2.2.d	Statistical analysis.....	51
2.2.3	Results.....	54
2.2.3.a	Convergence of the MCMC chains toward posterior distributions.....	54
2.2.3.b	Determination of D_{ref} , z_T and z_{pH} using the basic model.....	54
2.2.3.c	Determination of D_{ref} , z_T and z_{pH} using the intermediate model.....	55
2.2.3.d	Determination of D_{ref} , z_T and z_{pH} using the complete model.....	58
2.2.4	Discussion.....	60
2.3	Compléments.....	66
2.3.1	Variations des estimations entre les études individuelles.....	66
2.3.2	Différents types d'ajustements ou de prévisions.....	68
2.3.3	Comparaison des estimations obtenues en milieux de laboratoire et en milieux alimentaires.....	72
3	Modèle d'évaluation du risque microbiologique de non-stabilité à 55°C due à <i>Geobacillus stearothermophilus</i> dans les haricots verts appertisés.....	75
3.1	Construction d'un modèle d'évaluation des risques.....	75
3.2	Article: Quantitative assessment of the risk of microbial spoilage in foods. Prediction of non-stability at 55°C caused by <i>Geobacillus stearothermophilus</i> in canned green beans.....	78
3.2.1	Introduction.....	79
3.2.2	Material and Methods.....	79
3.2.2.a	Overview of the model.....	79
3.2.2.b	Determination of <i>G. stearothermophilus</i> concentrations at different processing steps.....	80
3.2.2.c	Estimation of the green bean pH at different processing steps.....	81
3.2.2.d	Two-dimensional Monte Carlo simulation.....	81
3.2.2.e	Model and parameters.....	81
3.2.2.f	Sensitivity analysis.....	89
3.2.2.g	What-if scenarios.....	90
3.2.3	Results.....	91

3.2.3.a	Fitted distributions of <i>G. stearotherophilus</i> from microbiological analyses at several processing steps.....	91
3.2.3.b	Predicted changes in concentrations during processing and non-stability prevalence	91
3.2.3.c	Validation of the model with independent data.....	94
3.2.3.d	Determination of the most influential factors by sensitivity analysis.....	95
3.2.3.e	What-if scenarios to test the influence of microbiological phenomena and options of spoilage risk management	96
3.2.4	Discussion	97
3.3	Compléments à l'article.....	99
3.3.1	Des « what-if-scenarios » complémentaires	99
3.3.2	Contamination croisée due aux débris.....	100
4	Modélisation de l'évolution de la concentration en vitamine C au cours de la fabrication et de l'entreposage de conserves de haricots verts	103
4.1	Introduction.....	103
4.1.1	Contexte	103
4.1.2	La vitamine C	104
4.2	Modélisation de l'évolution de la vitamine C au cours du procédé de fabrication des conserves de haricots verts.....	106
4.2.1	Processus modélisés.....	106
4.2.2	Simulation, incertitude et variabilité des facteurs, modèle global	109
4.2.3	Modélisation des durées et températures du traitement dans le procédé de transformation et de la conservation.....	112
4.2.4	Modélisation de la concentration initiale en vitamine C	113
4.2.5	Modélisation de l'évolution physiologique	115
4.2.5.a	Équations.....	115
4.2.5.b	Détermination de la vitesse de réaction k_{ev} :	116
4.2.6	Modélisation de la réaction chimique de dégradation de la vitamine C	118
4.2.6.a	Équations.....	118
4.2.6.b	Modélisation de la dégradation de la vitamine C pendant le blanchiment et l'étape de transfert - concentration d'O ₂ supposée constante.....	120

4.2.6.c	Modélisation de la dégradation de la vitamine C lors de la stérilisation et lors de l'entreposage – concentration en O ₂ variable.....	121
4.2.6.d	Les paramètres chimiques de la réaction.....	124
4.2.6.e	Détermination du temps équivalent pour la dégradation de DHA lors de la stérilisation	127
4.2.7	Modélisation de la diffusion des vitamines dans l'eau de blanchiment et dans le jus de couverture	130
4.2.7.a	Le phénomène de diffusion.....	130
4.2.7.b	Modélisation de la diffusion au blanchiment par la loi de Fick.....	131
4.2.7.c	Modélisation de la diffusion au cours de l'entreposage des boîtes de haricots verts	133
4.3	Résultats du modèle.....	133
4.3.1	Évolution des concentrations de vitamine C prédites par le modèle	133
4.3.2	Confrontation avec des données indépendantes pour étudier la validité du modèle	136
4.3.3	Analyse de sensibilité	139
4.3.3.a	Impact de la variabilité des facteurs	139
4.3.3.b	Impact de toutes les sources de variation des facteurs	141
4.3.4	Tests de scénarios.....	143
4.4	Conclusion	144
5	Analyse risque-bénéfice dans la chaîne de fabrication des haricots verts appertisés	147
5.1	Introduction.....	147
5.2	Variables communes aux modèles risque et bénéfice.....	148
5.2.1	Lien entre le temps standard de dégradation à 85°C de DHA F_1 et la valeur stérilisatrice F_0	148
5.2.2	Graphe du modèle global risque-bénéfice	150
5.3	Résultats simultanés des modèles risque et bénéfice	152
5.4	Optimisation du compromis risque-bénéfice.....	154
5.4.1	Choix des facteurs modifiables.....	154
5.4.2	Tests de scénarios et dépassements de seuils	155
5.5	Conclusion	157

6	Confrontation d'un modèle d'analyse du risque microbiologique dû à la bactérie <i>Bacillus cereus</i> dans une purée de courgettes à des données de contamination dans des produits finis....	159
6.1	Considérer un modèle d'évaluation du risque comme un réseau bayésien.....	159
6.2	Article: Inferring an augmented Bayesian network to confront a complex quantitative microbial risk assessment model with durability studies: Application to <i>Bacillus cereus</i> on a courgette purée production chain	162
6.2.1	Introduction.....	163
6.2.2	Materials and methods	165
6.2.2.a	The QMRA model	165
6.2.2.b	Data from durability studies.....	171
6.2.2.c	The likelihood model.....	172
6.2.2.d	Bayesian inference of the augmented BN.....	173
6.2.2.e	Prior and posterior predictive concentrations in the QMRA model	174
6.2.3	Results	175
6.2.3.a	Update of prior beliefs	175
6.2.3.b	Comparison of the prior and posterior predicted <i>B. cereus</i> concentrations through the "farm-to-fork" food chain	177
6.2.4	Discussion	179
7	Conclusion et perspectives	183
7.1	Bilan.....	183
7.1.1	Bilan des travaux	183
7.1.2	Pluridisciplinarité et collaboration industrielle	185
7.2	Limites des travaux.....	186
7.2.1	Limites sur les données	186
7.2.2	Limites des méthodes.....	187
7.3	Perspectives.....	189

Liste des acronymes et abréviations

AA : Acide Ascorbique

ABC : *Approximate bayesian computation*

AFSSA : Agence française de sécurité sanitaire des aliments (devenue ANSES)

AJR : Apports journaliers recommandés

ANC : Apports nutritionnels conseillés

ANSES : Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail

AQR : Appréciation quantitative des risques

ATS : *Aerobic thermophilic spores*

CI : Intervalle de crédibilité

CTCPA : Centre technique de la conservation des produits agricoles

DAG : *Directed acyclic graph*

DALY : *Disability-adjusted life year*

DHA : Acide déshydroascorbique

EFSA : Autorité européenne de sécurité des aliments

FAO : *Food agricultural organisation*

HACCP : *Hazard analysis critical control point*

HV : Haricot vert

INRA : Institut national de recherche agronomique

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*

LOD : *Limit of detection*

MCMC : Méthodes de Monte-Carlo par Chaîne de Markov

UFC : Unité formant colonies

OGM : Organismes génétiquement modifiés

OMC : Organisation mondiale du commerce

OMS : Organisation mondiale de la santé

PCB : Polychlorobiphényles

PHIA : *Probabilistic health impact assessment*

PNNS : Programme national nutrition-santé

QALY : *Quality-adjusted life year*

TIAC : Toxi-infection alimentaire collective

VTR : Valeur toxicologique de référence

1D : Une dimension

2D : Deux dimensions

Chapitre 1

Introduction

1.1 Risque et bénéfice dans les aliments

L'alimentation correspond à un besoin fondamental chez l'homme, puisqu'elle apporte à son organisme les nutriments essentiels dont il a besoin pour se construire et l'énergie qui lui permet de fonctionner. En plus de son caractère indispensable, l'alimentation est porteuse de nombreux enjeux, relatifs à la santé publique et la sécurité sanitaire, mais aussi à la sociologie et l'économie. Dans le domaine de la sécurité sanitaire, l'alimentation est notamment un vecteur de risques chimiques et microbiologiques liés à la présence éventuelle de contaminants ou d'agents biologiques pathogènes dans les aliments. Dans le domaine de la santé publique, la nutrition est un déterminant majeur de maladies chroniques comme les maladies cardiovasculaires, le diabète, l'obésité ou le cancer. Les bienfaits d'une alimentation saine sont relayés par les politiques nutritionnelles mises en place au niveau gouvernemental ou mondial, comme le Programme National Nutrition Santé (PNNS) lancé en 2001 en France, avec des slogans du type « Manger cinq fruits ou légumes par jour » ou « Limiter les produits gras, salés, sucrés ». D'autres témoignages de la montée des préoccupations alimentaires sont le succès grandissant des produits « Bio », dont le marché a quadruplé en dix ans, ou l'apparition des « alicaments », ces nouveaux produits au croisement entre nutrition et santé, sensés posséder une action bénéfique supérieure à celle des produits classiques. Les débats sur les aliments contenant des Organismes Génétiquement Modifiés (OGM) et les diverses crises sanitaires récentes (scandale du lait de vache frelaté en Chine en 2008 ; épidémie à *Escherichia coli* en Allemagne en mai 2011) soulignent également l'importance des enjeux sanitaires, économiques, sociologiques liés à l'alimentation.

La consommation d'un aliment est source de divers types de risques et bénéfiques, qui peuvent être quantifiés et comparés. Dans ce contexte, cette thèse porte sur l'analyse statistique du risque microbiologique et du bénéfice nutritionnel dans des chaînes de transformations alimentaires. Des modèles d'évaluation quantitative du risque et du bénéfice sont construits, décrivant l'évolution des concentrations en bactéries et en vitamines au cours de la fabrication de légumes transformés. Ces modèles décrivent le comportement des bactéries et des vitamines dans la chaîne de production,

CHAPITRE 1. Introduction

dans le but d'aider les industries alimentaires à concilier deux objectifs : préserver les bénéfices nutritionnels et assurer un niveau élevé de sécurité microbiologique. La construction de ces modèles a notamment requis une étape d'acquisition de paramètres de référence, effectuée par méta-analyse dans le cas des paramètres microbiologiques. Un outil complémentaire à la construction de modèles a aussi été mis en place, portant sur la confrontation d'un modèle d'évaluation du risque avec des données de contamination sur chaîne.

1.1.1 L'analyse du risque alimentaire

Un *risque sanitaire* lié à l'alimentaire se définit comme la probabilité de survenue d'un effet néfaste sur la santé humaine en réaction à la consommation d'un aliment. Les risques sanitaires associés à la consommation d'un aliment sont principalement des risques biologiques (liés à la présence de bactéries pathogènes, virus, moisissures, parasites, toxines, d'allergènes), des risques chimiques/toxicologiques (liés à la présence de pesticides, d'additifs, de résidus, de dioxines), ou encore des risques physiques dus à la présence de corps étrangers type morceau de verre ou de bois dans le produit. Les risques physiques présentent des dangers immédiats. Les risques chimiques présentent en général un danger sur le long terme, type cancer, lié à l'accumulation dans le corps de faibles doses de contaminants. Les risques biologiques sont généralement associés à un danger à court terme, type toxi-infection alimentaire. Une toxi-infection alimentaire est une maladie contractée à la suite de l'ingestion de nourriture ou de boisson contaminées par des agents pathogènes (bactéries, virus, parasites ou prions). La période d'incubation pour une toxi-infection alimentaire varie entre quelques heures et quelques jours. En France entre 2006 et 2008, 3127 foyers de toxi-infections alimentaires collectives (TIAC) ont été recensés, impliquant 33 404 malades, 2 302 hospitalisations et 15 décès (ANSES, 2011). Par ailleurs, la prolifération de microorganismes non pathogènes peut mener à une altération de la qualité marchande d'un aliment, lorsqu'elle conduit à une modification de son aspect ou de ses propriétés organoleptiques (goût, odeur, saveur, texture, couleur). Il s'agit alors d'un *risque industriel*, plutôt d'ordre économique, car ces altérations peuvent nuire à la vente des aliments. Dans cette thèse, les risques microbiologiques étudiés concernent la qualité sanitaire des produits étudiés, aussi bien leur altération (risque industriel) que leurs effets potentiels sur la santé des consommateurs.

Différents organismes sont chargés de l'établissement de normes alimentaires et de l'analyse des risques alimentaires. En France, il s'agit de l'ANSES (Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail, anciennement AFSSA), dont la mission principale est de contribuer à la sécurité sanitaire humaine dans les domaines de l'environnement, du travail et de l'alimentation. Au niveau européen, l'EFSA (*European Food Safety Authority*) est chargée de l'évaluation des risques relatifs à la sécurité des aliments destinés à l'alimentation humaine et animale. Au niveau international, la Commission du *Codex alimentarius* créée en 1963 par l'Organisation des nations unies pour l'alimentation et l'agriculture (FAO) et par l'Organisation Mondiale de la Santé (OMS), est chargée de mettre au point des normes alimentaires, des lignes directrices et des codes d'usages internationaux et harmonisés visant à protéger la santé des consommateurs et à assurer des pratiques loyales dans le commerce des aliments. Différents comités d'experts sont interrogés par le *Codex Alimentarius*, tels que le JECFA (*Joint FAO/WHO Expert Committee on Food Additives and contaminants*) chargé des risques liés aux additifs et contaminants

CHAPITRE 1. Introduction

chimiques, le JMPR (*Joint FAO/WHO Meetings on Pesticide Residues*) chargé des risques liés aux résidus de pesticides et le JEMRA (*Joint FAO/WHO Meetings on Microbiological Risk Assessment*) chargé des risques microbiologiques. Les experts basent leur avis sur des analyses de risque qui comportent trois composantes, telles que définies par la commission du *Codex alimentarius* en 2003 : l'évaluation (ou appréciation) des risques, la gestion des risques, et la communication à propos des risques.

L'*évaluation scientifique du risque* est une discipline consistant à définir les conséquences sanitaires sur la santé publique résultant de l'exposition à un danger. Le risque peut se définir par la probabilité de survenue de l'effet indésirable sur la santé (maladie particulière, mort, hospitalisation, etc.) après exposition à l'aliment contaminé. L'évaluation du risque repose sur quatre grandes étapes : l'identification du danger (description de l'agent pathogène, la manière dont l'homme et les denrées peuvent être contaminés), la caractérisation du danger (qui consiste à évaluer les effets indésirables sur la santé), l'appréciation de l'exposition (qui dépend d'une part, de la fréquence, du niveau et du type de contamination des denrées alimentaires et d'autre part, de la quantité consommée par la population) et la caractérisation du risque est l'étape finale qui intègre l'information collectée aux étapes précédentes. L'appréciation du risque peut être réalisée soit par une approche qualitative, utilisant la perception des experts dans une approche graduée du danger, soit par une approche quantitative, tentant de chiffrer le nombre de malades ou la probabilité de maladie résultant de l'exposition au danger. Cette dernière approche nécessite de disposer d'informations nombreuses (données de contamination, relations sur les effets de la température sur l'inactivation des bactéries, relation entre la dose ingérée et la maladie, etc.).

La *gestion des risques* vise à identifier des mesures de réduction du risque et à les mettre en balance pour choisir les mesures de prévention des risques appropriées. Le *Codex Alimentarius* a ainsi défini la méthode HACCP (*Hazard Analysis Critical Control Point*), qui est un système qui identifie, évalue et maîtrise les dangers biologiques, chimiques ou physiques significatifs au regard de la sécurité des aliments. L'HACCP est apparu aux Etats-Unis dans les années 1960 et a été développé ensuite par la FDA (*Food and Drug Administration*), l'OMS et la commission du *Codex Alimentarius*. L'HACCP est aujourd'hui une méthode de référence pour les litiges internationaux au sein de l'Organisation Mondiale du Commerce (OMC). L'HACCP est basé sur 7 principes : identifier des dangers à tous les stades de la chaîne alimentaire, déterminer les points critiques pour la maîtrise de ces dangers, fixer les limites critiques, établir des actions de surveillance, établir des actions correctives, vérifier l'efficacité du système et établir un système documentaire.

La *communication sur les risques* consiste en un échange interactif d'informations et d'opinions entre les responsables de l'évaluation et de la gestion du risque, les milieux professionnels, les consommateurs et les autres parties intéressées.

L'analyse des risques sanitaires peut avoir des conséquences économiques importantes. En effet, les mesures sanitaires nécessaires visant à protéger la santé et la vie des personnes et animaux peuvent entraîner des restrictions au commerce. Cependant ces mesures ne doivent pas être détournées en instrument protectionniste. La réglementation internationale de l'OMC impose en effet que la libre circulation des marchandises ne peut être entravée par l'application de mesures sanitaires que si ces mesures sont fondées sur des principes scientifiques suffisants (accords de Marrakech, 1994). Par exemple, l'interdiction de l'importation de viande traitée avec des activateurs

CHAPITRE 1. Introduction

de croissance dans l'Union Européenne en 1989 a été contestée par les Etats-Unis, et l'Union Européenne a finalement été condamnée par l'OMC pour pratique protectionniste en 1997, en absence de preuve suffisante de l'existence de risques.

1.1.2 L'analyse du bénéfice nutritionnel

Un *bénéfice sanitaire* se définit comme une augmentation de l'amplitude d'un effet positif sur la santé humaine et/ou comme une diminution d'un effet délétère pour la santé humaine. Les *bénéfices nutritionnels* sont associés aux bénéfices provenant de la consommation d'un aliment contenant des nutriments bénéfiques pour la santé. Un exemple de bénéfice nutritionnel est la réduction du taux d'infarctus du myocarde associée à la consommation d'acides gras omega-3, présents notamment dans les poissons gras (Marckmann and Grønbaek, 1999). Un autre exemple est la protection contre les risques de maladies cardio-vasculaires et de cancer attribuée aux composés antioxydants type vitamine C, caroténoïdes ou polyphénols, présents dans les fruits et légumes (Dragsted et al., 1993; Wang et al., 1996). De manière générale, un aliment est considéré comme source de bénéfice quand il permet de couvrir une partie de nos besoins nutritionnels. Les besoins nutritionnels correspondent à la quantité de nutriments qu'il faut absorber pour couvrir les besoins nets d'une personne en bonne santé : entretien, fonctionnement métabolique et physiologique. En particulier, les nutriments *essentiels* sont des composants élémentaires nécessaires pour couvrir les besoins physiologiques de l'organisme. Les nutriments *indispensables* sont des nutriments essentiels qui ne peuvent pas être synthétisés par l'organisme en quantité suffisante pour assurer ses besoins, et qui doivent donc être apportés par l'alimentation. Parmi les nutriments essentiels se trouvent certains acides gras indispensables (type oméga-3 ou 6), certains acides aminés indispensables, des vitamines (comme la vitamine C, indispensable) ou encore des minéraux alimentaires (comme le magnésium ou le fer). Les besoins nutritionnels varient avec l'âge, le sexe, la corpulence, l'activité etc. Les apports nutritionnels conseillés (ANC) sont élaborés par les autorités nationales comme l'ANSES, et représentent la quantité suffisante de différents paramètres nutritionnels nécessaires pour assurer la couverture des besoins nutritionnels de la quasi-totalité de la population (97.5%). Un ANC est établi sur une base de sécurité de 130% des besoins nutritionnels moyens et constitue une valeur conseillée, supérieure aux besoins nutritionnels minimaux. L'ANC se différencie des AJR (Apports Journaliers Recommandés) qui sont des valeurs-repères réglementaires européennes de références pour l'étiquetage, correspondant à la contribution de l'aliment à la couverture des besoins *moyens* de la population en vitamines et minéraux.

Les étiquetages et les allégations nutritionnelles permettent aux industriels de mettre en avant les qualités de leurs produits et d'informer le consommateur. Ils sont réglementés par des décrets et des règlements européens (ANIA, 2009; Verhagen et al., 2010). Les *allégations nutritionnelles* correspondent à toute allégation qui affirme, suggère ou implique qu'une denrée alimentaire possède des propriétés nutritionnelles bénéfiques particulières (énergie fournie, nutriments contenus en proportion moindre ou plus élevée). Des exemples d'allégations nutritionnelles sont « allégé en sucres », « enrichi en fibres », « sans cholestérol », « source de vitamine C ». Les *allégations de santé* correspondent à toute allégation qui affirme, suggère ou implique l'existence d'une relation entre, d'une part, une catégorie de denrées alimentaires, une denrée alimentaire ou l'un de ses composants et, d'autre part, la santé (ANIA, 2009). Des exemples

CHAPITRE 1. Introduction

d'allégation santé sont « contribue à faire baisse le cholestérol » ou « le calcium est bon pour la croissance des enfants ». De manière générale l'étiquetage nutritionnel est facultatif ; il devient obligatoire si une allégation nutritionnelle ou une allégation de santé figure sur le produit, ou si des vitamines, minéraux ou autres substances ont été ajoutés au produit. Pour pouvoir mentionner sur l'étiquette la présence de certaines vitamines ou minéraux, leur teneur doit dépasser un seuil significatif, d'au moins 15% des AJR pour 100 g ou 100 ml de produit ou par unité d'emballage s'il ne contient qu'une portion. À titre d'exemple, le seuil d'allégation nutritionnelle pour la vitamine C est de 12 mg/100g et pour la vitamine B9 il est de 30 µg/100g (Commission des Communautés Européennes, 2008).

Contrairement à l'analyse du risque, l'analyse du bénéfice nutritionnel est un domaine assez récent. Les organismes chargés des normes alimentaires et en particulier de l'analyse des risques, l'ANSES, l'EFSA et la Commission du *Codex alimentarius*, sont également chargés de l'analyse des bénéfices nutritionnels. Bien que les effets bénéfiques des aliments aient été étudiés depuis longtemps, il existe peu de recommandations sur la façon de procéder à une évaluation du bénéfice dans des aliments. L'EFSA a proposé en 2010 un guide sur l'évaluation risque-bénéfice dans les aliments sur la santé humaine. Il y est proposé d'adopter une démarche pour l'évaluation du bénéfice similaire à celle adoptée pour l'évaluation du risque : identification du bénéfice, caractérisation de l'effet positif avec évaluation de la dose-réponse, évaluation de l'exposition, et caractérisation finale du bénéfice. Un bénéfice est généralement mesuré directement comme une réduction d'un risque de maladie, ou indirectement comme un changement sur des marqueurs de maladie ou des marqueurs de bonne santé. L'évaluation du bénéfice dans les aliments est notamment utilisée pour les demandes d'allégations nutritionnelles sur des produits existants, lors de la reformulation de produits par l'industrie de manière à augmenter un bénéfice (réduction des teneurs en sel, en sucre ou en mauvais acides gras) (Van Raaij et al., 2009; Verhagen et al., 2010), ou pour l'étude de certains régimes alimentaires ou de certains types d'aliments (régimes méditerranéen, pain complet).

Parmi les conseils nutritionnels généraux issus des objectifs du Programme national nutrition-santé (PNNS) figure l'augmentation de la consommation de fruits et légumes. Le PNNS recommande en particulier de consommer au moins « Cinq fruits et légumes par jour », sous toutes leurs formes : surgelés, frais, en conserve, cuits, crus, etc. Les fruits et légumes frais ayant une faible durée de vie et étant soumis à la saisonnalité, la transformation des fruits et légumes permet de s'affranchir de ces contraintes et facilite le suivi des conseils nutritionnels du PNNS. Différents modes de transformation de fruits et légumes existent. Les produits 1^{ère} gamme sont les produits frais, les produits 2^{ème} gamme sont les produits stérilisés (conserve), les produits 3^{ème} gamme sont les produits surgelés, les produits 4^{ème} gamme sont les produits frais prédécoupés (type salade en sachet) et les produits 5^{ème} gamme sont les produits cuits sous vide (pasteurisés puis conservés au froid). Dans cette thèse, un produit 2^{ème} gamme et un produit 5^{ème} gamme seront étudiés.

1.1.3 L'analyse risque-bénéfice

L'évaluation *risque-bénéfice* peut être définie comme une approche qui pondère la probabilité d'un danger (incidence et sévérité) provenant d'une exposition, d'un produit, d'une

CHAPITRE 1. Introduction

décision etc., avec la probabilité et l'ampleur d'un bénéfice (EFSA, 2010). Les analyses risques-bénéfices sont pratiquées dans de nombreuses disciplines : la médecine, l'ingénierie, l'agronomie, la nutrition, etc. La notion de risque-bénéfice est par exemple au cœur du développement de nouveaux médicaments, où les effets bénéfiques d'un traitement sont très souvent accompagnés d'effets indésirables. En ingénierie, agroalimentaire ou géo-ingénierie se pose par exemple la question du risque-bénéfice de l'énergie nucléaire, des OGM, ou encore des aliments issus de l'agriculture biologique (AFSSA, 2003). Les analyses risques-bénéfices en alimentation se développent de manière croissante car elles permettent de proposer des recommandations de consommation plus complètes, plus cohérentes et plus pertinentes que les analyses du risque seul ou du bénéfice seul. En effet, prendre en compte seulement le risque peut conduire à des recommandations entraînant une baisse de bénéfice, et prendre en compte seulement le bénéfice peut conduire à des recommandations entraînant un risque accru. Idéalement, pour être complètement pertinente, une analyse risque-bénéfice devrait prendre en compte l'ensemble des risques et l'ensemble des bénéfices associés à une exposition.

Les recommandations pour effectuer une évaluation risque-bénéfice sont encore en cours d'établissement. Suite au colloque organisé par l'EFSA sur les méthodes d'analyses risques-bénéfices dans les aliments en juillet 2006, la démarche recommandée pour les analyses risque-bénéfice est similaire à celle adoptée pour l'analyse du risque (EFSA, 2006; EFSA, 2010). Ainsi, trois étapes doivent être prises en compte: évaluation du risque-bénéfice, gestion du risque-bénéfice et communication sur le risque-bénéfice. L'évaluation du risque-bénéfice est elle-même composée de 3 étapes : évaluation du risque, évaluation du bénéfice et comparaison risque-bénéfice. La Figure 1.1 présente la procédure recommandée par l'EFSA pour une évaluation risque-bénéfice. Il est par ailleurs recommandé de procéder à une première évaluation visant à savoir si le risque dépasse largement le bénéfice ou vice-versa, puis à procéder si besoin à une évaluation plus raffinée du risque/bénéfice en fournissant des estimations semi-quantitatives ou quantitatives du risque et du bénéfice à des niveaux d'exposition pertinents. Il est recommandé d'utiliser si possible une métrique commune (ou composite) qui permet de décrire les effets bénéfiques et délétères sur la santé dans la même unité de mesure (type morbidité, mortalité ou charge de la maladie). La sortie d'une évaluation risque-bénéfice est alors exprimée à l'aide d'une valeur unique d'impact sur la santé.

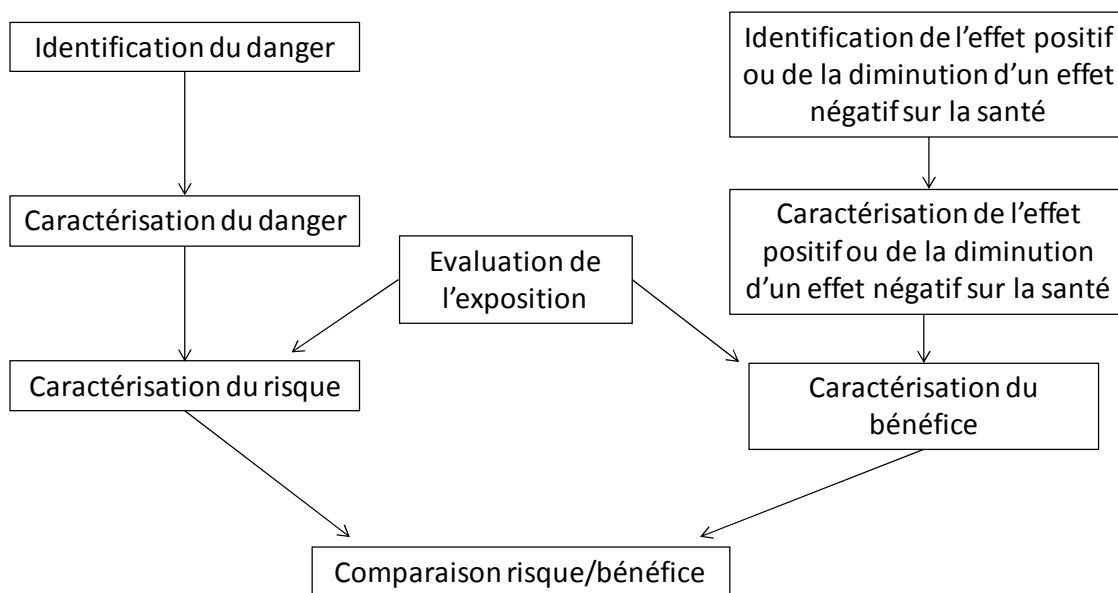


Figure 1.1. Procédure pour l'évaluation risque-bénéfice dans les aliments recommandée par le Comité Scientifique de l'EFSA (EFSA, 2010).

Une grande partie des analyses risque-bénéfice se base sur la comparaison des expositions ou des apports en nutriment à des seuils toxicologiques ou nutritionnels de référence, déjà établis. Ces seuils sont par exemple les valeurs toxicologiques de référence (VTR), les apports nutritionnels conseillés (ANC), les seuils de dangerosité des concentrations microbiennes. La VTR est un seuil toxicologique couramment utilisé pour les risques chimiques ; cet indice caractérise le lien entre l'exposition de l'homme à une substance toxique et l'occurrence ou la sévérité d'un effet nocif observé. Certaines analyses risque-bénéfice comparent simultanément les concentrations en nutriments et en contaminants aux seuils nutritionnels et toxicologiques de référence. Une approche consiste à calculer la consommation « sans risque » permettant d'être sous le seuil de dangerosité et de comparer l'apport nutritionnel correspondant aux références nutritionnelles. L'approche inverse est de calculer la consommation permettant d'atteindre les recommandations nutritionnelles, et de comparer l'exposition correspondante aux VTR. Des indices mesurant le ratio bénéfice/risque sont parfois calculés. Des exemples seront donnés plus loin.

D'autres approches se basent sur les relations dose-réponse établissant des relations entre les expositions en contaminants ou les apports en nutriments à des niveaux de maladies. Des ratios risque-bénéfice peuvent alors être calculés. Les indicateurs types QALY (*Quality-adjusted life year*) ou DALY (*Disability-adjusted life year*) utilisent de telles relations dose-réponse. Le QALY est un indicateur santé combinant la durée et la qualité de la vie, et le DALY est un indicateur mesurant les inconvénients d'une maladie (Murray, 1994; Gold et al., 2002). Les DALY ou QALY sont plutôt applicables pour classer les effets négatifs sur la santé pouvant être traduits en tant qu'années de vie perdues ou années de vie de qualité altérée. Certains dangers sont cependant difficilement traductibles en termes de DALY ou QALY, comme certains risques toxicologiques. La construction de

CHAPITRE 1. Introduction

DALY ou QALY repose en effet sur des informations parfois difficilement disponibles, comme les relations dose-réponse, l'exposition de la population à un contaminant ou la prévalence des maladies, l'estimation de l'altération de la qualité de la vie due à une maladie. Les DALY et QALY sont couramment utilisés à l'échelle internationale pour évaluer les interventions de santé et les traitements des maladies. Ils sont alors souvent associés à des études du coût des interventions et traitements, permettant de calculer des rapports coût/efficacité. Du côté des bénéfices, il n'existe pas vraiment à ce jour de métrique convenue relative aux effets positifs sur la santé et le bien-être (EFSA, 2010; Magnússon et al., 2012). Bos et al. (2009) ont récemment introduit un critère d'impact sur la santé (HIC), qui combiné à un modèle d'évaluation des risques, donne un modèle probabiliste d'évaluation de l'impact sur la santé (PHIA) pouvant être utilisé pour comparer certains risques (Bos et al., 2009).

La grande majorité des analyses risques-bénéfices en alimentation porte sur les bénéfices *nutritionnels* et les risques *chimiques* d'un aliment (Tijhuis et al., 2012). Ces approches risques-bénéfices permettent de proposer des recommandations de consommation cohérentes, tenant compte à la fois des risques liés aux contaminants, et des bénéfices nutritionnels du même aliment (Renwick et al., 2004). Une analyse risque-bénéfice peut être nécessaire quand un même aliment contient différents composants porteurs chacun d'effets positifs ou négatifs (par exemple, les végétaux fournissent des micronutriments prévenant certains types de cancer mais peuvent contenir des substances potentiellement dangereuses type pesticides ou nitrates), ou bien quand un seul composant alimentaire peut avoir des effets à la fois positifs et potentiellement négatifs selon la dose consommée (acide folique, zinc, vitamine A, fer) (EFSA, 2010). Les changements de mode de consommation (remplacement du sucre par des édulcorants), des interventions (fortification en acide folique) ou des stratégies alternatives de transformation des aliments (stratégies pour limiter les teneurs en acrylamide) sont également des causes possibles d'analyses risques chimiques-bénéfices nutritionnels. L'aliment ayant donné lieu au plus grand nombre d'analyses risque chimique-bénéfice nutritionnel est le poisson (Rissanen et al., 2000; Cohen et al., 2005; Foran et al., 2005; Antonijevic et al., 2007; Domingo et al., 2007; Sioen et al., 2007; Van der Voet et al., 2007; Guevel et al., 2008; Verger et al., 2008; AFSSA, 2010; EFSA, 2010; Sirot, 2010). Les poissons sont en effet une source privilégiée de vitamines liposolubles et hydrosolubles (A, D, E, B6, B12) et de minéraux et oligo-éléments, et d'acides gras à longue chaîne qui contribuent à la réduction du risque cardiovasculaire et au bon fonctionnement cérébral. Cependant ils sont également des contributeurs majeurs de l'exposition alimentaire aux dioxines, aux PCB (Polychlorobiphényles) et au méthylmercure, dont l'action est toxique notamment sur le système nerveux central. Dans son rapport relatif aux risques/bénéfices (AFSSA, 2010), l'AFSSA donne un ensemble de recommandations visant généralement à limiter la consommation hebdomadaire de poissons à deux portions par semaine et à privilégier la consommation de poissons gras, afin de couvrir les besoins en nutriments bénéfiques comme les acides gras tout en limitant le risque de surexposition aux contaminants chimiques. Un exemple d'analyse risque-bénéfice utilisant des probabilités de dépassement de seuils est l'étude de Van der Voet et al. (2007), portant sur la consommation de poissons, où le risque est représenté par les dioxines et le bénéfice par l'apport d'acides gras (Figure 1.2). Un exemple d'analyse risque-bénéfice sur le poisson faisant appel aux QALY est donné dans Guevel et al. (2008). Dans cette étude, le risque considéré était l'impact du méthylmercure sur le développement cognitif, et le bénéfice considéré était le rôle protecteur d'acides gras polyinsaturés sur les maladies cardiovasculaires et sur le développement cognitif. La conclusion de cette étude

CHAPITRE 1. Introduction

était l'impact plutôt positif (mais très incertain) sur la santé d'une augmentation de la consommation de poisson.

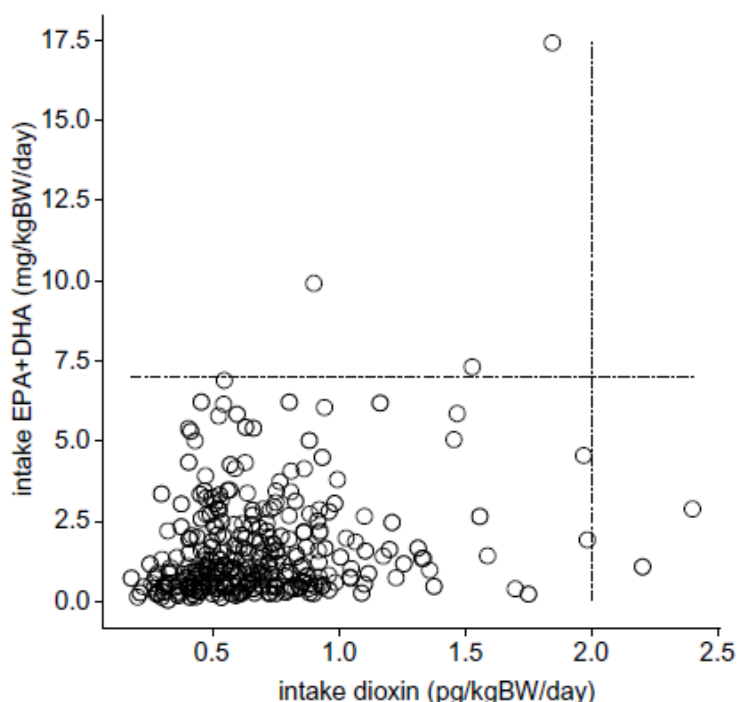


Figure 1.2. Exposition simultanée à des composants avec des effets néfastes et avec des effets bénéfiques pour la santé pour des niveaux typiques de consommation et de concentrations aux Pays-Bas. Echantillon aléatoire de 500 simulations sur un total de 10 000. Source : Van der Voet et al. (2007)

Un autre grand type d'analyses risques-bénéfices dans l'alimentation est rattaché au *risque microbiologique* (Magnússon et al., 2012). La sécurité microbiologique des aliments ne requiert en général pas d'évaluation du bénéfique, car les contaminants microbiens (en dehors de ceux spécifiquement utilisés pour la préparation d'un aliment, tels les ferments) ont généralement seulement des effets négatifs. Cependant, les interventions entraînant des réductions des risques microbiologiques peuvent être considérées comme étant source d'effets bénéfiques sur la santé (Magnússon et al., 2012). Ces méthodes pour contrôler le risque microbiologique incluent l'utilisation de conservateurs, de désinfectants, de pesticides, de fongicides, et d'antibiotiques pour les animaux d'élevage. Le bénéfique lié à la diminution des risques microbiologiques peut alors être contrebalancé par la création d'un risque chimique. La balance entre un risque microbiologique et un risque chimique donne lieu à des analyses « risques-bénéfice » ou « risques-risques ». L'utilisation de pesticides ou de désinfectants a ainsi donné lieu à de nombreuses études. Des exemples d'analyses risque-bénéfice portent sur la désinfection de l'eau potable par l'ozone (Havelaar et al., 2000), l'utilisation de fongicides contre les mycotoxines (Muri et al., 2009), ou l'utilisation d'un antibiotique contre les maladies bactériennes dans un élevage de volailles (Cox Jr. and Popken, 2004; Cox Jr., 2005). Ces analyses risques-bénéfices sont basées sur des analyses quantitatives du risque et les comparaisons ont été exprimées en DALY, QALY ou par un modèle PHIA. Dans ces trois cas, le

CHAPITRE 1. Introduction

bénéfice lié à l'intervention pour réduire le risque microbiologique a été jugé supérieur au risque chimique associé, comme illustré dans le cas de la désinfection de l'eau potable par l'ozone sur la Figure 1.3. Un autre type d'analyse risque-bénéfice possible en microbiologie concerne certaines bactéries minoritaires associées à la fois à un risque et à un bénéfice, comme les probiotiques (Magnússon et al., 2012). Les probiotiques sont des microorganismes vivants pouvant conférer un effet bénéfique à leur hôte quand ils sont administrés en quantité adéquate. Chez une certaine catégorie de population fragilisée, les probiotiques pourraient néanmoins être associés à un risque.

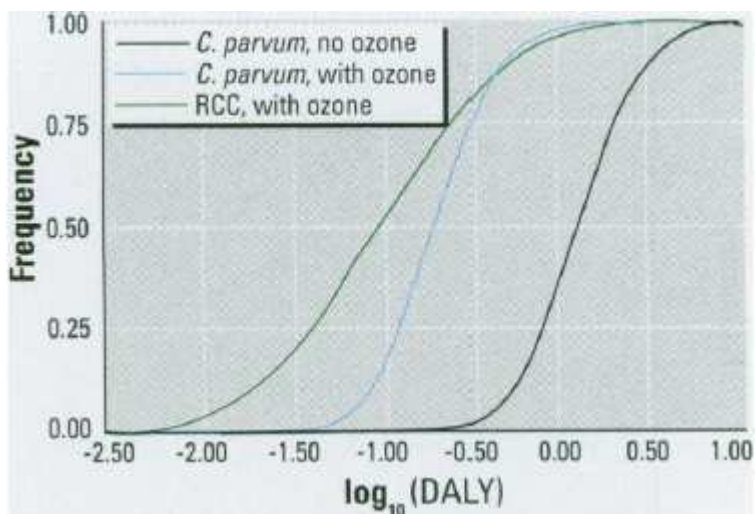


Figure 1.3. Effets néfastes sur la santé exprimés en termes de DALY sur une population d'un million d'habitants dus à la bactérie pathogène *Cryptosporidium parvum* et au bromate (sous-produit cancérigène de la désinfection par l'ozone) dans l'eau potable, avec ou sans désinfection par l'ozone. RCC : cancer cellulaire rénal. Source : Havelaar et al. (2000)

Dans une optique légèrement différente, certaines analyses *coûts-bénéfices* sur des mesures de réduction de risques microbiologiques posent des problématiques similaires aux analyses risques-bénéfices présentées ci-dessus. Ainsi, l'efficacité de différentes mesures de maîtrise du risque microbiologique associé à la présence de *Campylobacter* sur une chaîne de fabrication et de consommation de poulets a été étudiée par Havelaar et al. (2007). L'effet de différentes mesures telles qu'une réduction des disséminations fécales, de la contamination des poulets ou des carcasses, ou les effets de campagnes d'information pour prévenir la contamination croisée chez le consommateur, a été estimé à partir d'un modèle quantitatif d'analyse des risques depuis la ferme jusqu'au risque de maladie. Une analyse coût-bénéfice a de plus été réalisée, afin de comparer le coût des interventions et l'utilité des interventions (mesurée à partir des réductions des maladies exprimées en DALY et des réductions de coûts de prise en charge des maladies).

De manière générale, une différence notable entre le bénéfice nutritionnel et le risque microbiologique est que le risque microbiologique correspond à une situation extrême (dépassement d'un seuil élevé de bactéries, ce qui arrive rarement) tandis que le bénéfice nutritionnel correspond généralement à une situation moyenne (apport de nutriments suffisants dans la vie quotidienne). Le risque chimique lui est généralement étudié comme la probabilité de dépasser un seuil (événement

CHAPITRE 1. Introduction

extrême), suite à l'accumulation de faibles doses sur une longue période. Par ailleurs, contrôler le risque revient à limiter la dose de contaminant (chimique ou biologique) alors qu'optimiser le bénéfice revient soit à augmenter la dose de nutriments, soit à atteindre une dose optimale (certains nutriments sont nécessaires à faible dose mais peuvent être toxiques à haute dose, comme le zinc ou la vitamine C) (EFSA, 2010; Tjihuis et al., 2012).

Une grande étude menée par l'institut national de la santé publique aux Pays-Bas a comparé les bénéfices apportés par une alimentation saine avec les risques sanitaires liés aux aliments. Il en ressort que l'amélioration des régimes alimentaires apporterait un bénéfice en général 100 fois supérieur à l'amélioration de la sécurité alimentaire (Van Kreijl et al., 2006; Tjihuis et al., 2012). Cette étude montrerait que les impacts positifs d'une alimentation saine ne sont pas assez exploités.

L'analyse risque-bénéfice réalisée dans cette thèse porte sur l'optimisation des procédés de transformation des aliments dans le but d'augmenter un bénéfice nutritionnel, tout en contrôlant un risque industriel de non-stabilité des boîtes de conserves.

1.2 Méthodes statistiques utilisées

1.2.1 Modélisation, modèles probabilistes, incertitude et variabilité

De très nombreux domaines, comme la météorologie, l'hydrologie, la dynamique des populations, la pharmacocinétique, l'économie, la sociologie etc., font appel à des modèles. Un *modèle* peut être défini comme une « représentation théorique simplifiée d'un système réel, comportant une structure, la formulation de lois (relations déterministes ou probabilistes) reliant les impulsions ou entrées du système à ses réponses ou sorties, et la définition de paramètres (constantes ou variables pouvant prendre des valeurs numériques) entrant dans ces relations » (définition extraite du *Dictionnaire français d'Hydrogéologie*). Un modèle est supposé reproduire la réalité, mais parce qu'il est approché, *un modèle est toujours faux...* L'intérêt d'un modèle va résider dans sa capacité à apporter une réponse satisfaisante aux questions qui se posent sur le processus modélisé. Différents modèles (statistiques, mécanistes, conceptuels, etc.) peuvent être utiles pour approcher une même réalité, mais selon l'objectif recherché (prédire, expliquer, explorer, communiquer, etc.), ils auront plus ou moins d'utilité. Dans cette thèse, les modèles utilisés sont des modèles quantitatifs stochastiques traduisant des mécanismes biologiques impliqués dans la transformation d'un aliment. Le choix des mécanismes biologiques à modéliser s'est fait en collaboration des experts en microbiologie et en chimie des aliments.

En analyse quantitative des risques, la prise en compte de l'incertitude et de la variabilité est une question essentielle (Delignette-Muller and Rosso, 2000; Nauta, 2000; Vose, 2000). L'un des principes de travail recommandé par la Commission du *Codex alimentarius* est d'ailleurs que « l'expression de l'incertitude ou de la variabilité dans le résultat de l'estimation des risques peut

CHAPITRE 1. Introduction

être qualitative ou quantitative mais doit être quantifiée dans la mesure où cela est scientifiquement réalisable » (Codex Alimentarius Commission, 1999).

Un modèle d'analyse des risques fait intervenir de nombreux paramètres (ou variables), tels que la contamination bactérienne du produit ou la température d'un traitement thermique. Ces grandeurs ne possèdent généralement pas une valeur fixe ; afin de tenir compte de la variation des valeurs possibles qu'elles peuvent prendre, elles sont considérées comme des variables aléatoires, auxquelles sont affectées des distributions de probabilité. Des modèles probabilistes sont donc utilisés en analyse du risque. La prise en compte de manière probabiliste du risque microbiologique est d'autant plus importante qu'il correspond généralement à un risque extrême (de faible probabilité), qui n'apparaîtrait donc généralement pas via un modèle déterministe correspondant à une situation usuelle (moyenne).

La *variabilité* correspond à l'hétérogénéité naturelle dans une population qui est irréductible par des mesures additionnelles (Anderson and Hattis, 1999; Vose, 2000). Cette hétérogénéité est par exemple liée aux diversités entre micro-organismes, hôtes, produits, conditions de conservation ou de préparation. La variabilité a aussi été décrite comme l'« incertitude aléatoire » ou la « variabilité stochastique » ou la « variabilité interindividuelle ». L'*incertitude*, elle, représente un manque de connaissance lié à des erreurs de mesures, des erreurs d'échantillonnage, à un manque de données etc. L'incertitude est réductible par des mesures additionnelles. L'incertitude est parfois appelée « incertitude fondamentale » ou « incertitude épistémique » ou « degré de connaissance ». La combinaison de l'incertitude et de la variabilité est parfois appelée « incertitude totale ».

1.2.2 La simulation de Monte Carlo (1D et 2D)

La méthode de simulation de Monte Carlo est très utilisée en analyse des risques, ainsi que dans de nombreux domaines scientifiques (finance, ingénierie, etc.). Cette méthode probabiliste simple permet d'obtenir un échantillon de la distribution de la sortie du modèle à partir d'échantillons des variables d'entrée du modèle. Un grand nombre de tirages aléatoires est effectué dans les distributions de probabilité des variables d'entrée, et les sorties correspondantes du modèle sont calculées (de manière déterministe ou probabiliste selon les modèles). Chaque simulation est un calcul numérique sensé correspondre à une situation possible, plus ou moins probable, d'un système réel. Des statistiques sont ensuite calculées sur les résultats des simulations. La méthode donne des résultats approchés mais il est possible de contrôler la précision des calculs en fonction du nombre d'itérations effectuées (Tuffin, 2010).

Dans une simulation de Monte Carlo, les distributions de probabilité sur les variables d'entrée du modèle représentent de l'incertitude, de la variabilité, ou un mélange des deux. Dans la simulation de Monte Carlo « d'ordre 1 » ou à « 1 dimension » (1D), la nature des distributions de probabilité n'est pas distinguée, et les deux types de distributions sont mises sur un même plan. Ces types de simulation sont assez répandus, car leur mise en place est simple et ne nécessite pas de dissocier la part de variabilité et la part d'incertitude quand les deux types d'aléa affectent une variable. Cependant un inconvénient est la difficile interprétation des résultats de la simulation, car la distribution sur la sortie du modèle sera un mélange d'incertitude et de variabilité. Il ne sera par

CHAPITRE 1. Introduction

exemple pas possible de savoir si la variation de la sortie est plus due à des sources d'incertitudes importantes dans le modèle, ou bien à une véritable hétérogénéité (Vose, 2000). L'utilisation des résultats de la simulation de Monte Carlo, pour la gestion du risque par exemple, s'en trouve contrariée, car les mesures de gestion du risque à prendre ne seront pas les mêmes dans les deux cas. Dans le 1^{er} cas il s'agira de chercher à renforcer les connaissances pour diminuer l'incertitude, et dans le 2^{ème} cas il s'agira par exemple de mettre en place des mesures de maîtrise du risque afin de réduire les possibilités de forte contamination. De plus, le traitement de la variabilité sur le même plan que l'incertitude peut induire des biais dans les résultats de l'analyse de risque (Nauta, 2000; Vose, 2000). Les organisations internationales recommandent maintenant plutôt l'utilisation de la simulation de Monte Carlo « d'ordre deux », ou à « deux dimensions » (2D), qui permet d'évaluer séparément la variabilité et l'incertitude dans les modèles d'analyses des risques (Codex Alimentarius Commission, 1999; FAO/WHO, 2006; European Commission, 2003). Pour les variables présentant de la variabilité et dont la distribution de variabilité est connue à incertitude près, une modélisation hiérarchique de l'incertitude et de la variabilité est souvent utilisée (Pouillot et al., 2003), comme illustré dans la Figure 1.4. Différentes méthodes permettent d'ajuster de telles distributions, comme les statistiques inférentielles classiques, l'inférence bayésienne, le ré-échantillonnage par *bootstrap*, ou l'élicitation de données d'experts (Vose, 2000; Pouillot and Delignette-Muller, 2010). Les méthodes de simulations de Monte Carlo 2D sont de plus en plus utilisées dans l'analyse des risques alimentaires (Frey, 1992; Cullen and Frey, 1999; Mokhtari and Frey, 2005; Delignette-Muller et al., 2006; Pouillot et al., 2007; Membré et al., 2008; Mataragas et al., 2010; Pouillot et al., 2010; Rimbaud et al., 2010). Les modèles conçus pour la simulation de Monte Carlo 1D sont parfois appelés modèles d'ordre 1 (ou à 1 dimension), et ceux conçus pour la simulation de Monte Carlo 2D sont parfois appelés modèles d'ordre 2 (ou à 2 dimensions) (Vose, 2000).

Dans cette thèse, les modèles d'évaluation du risque et du bénéfice ont été étudiés par simulation de Monte Carlo 2D. Les termes de « variable », « paramètre » et parfois « facteur » d'un modèle ont généralement été utilisés de manière assez interchangeable, comme souvent dans l'évaluation des risques microbiologiques. Une exception apparaît dans l'article présenté au Chapitre 6, où le vocabulaire de la statistique bayésienne a été privilégié : le terme « paramètre » y est utilisé pour désigner les variables dont la distribution de probabilité représente seulement de l'incertitude. Par ailleurs, le terme d'« hyperparamètre » a parfois été utilisé pour désigner les variables dont la distribution de probabilité représente de l'incertitude seule sur un paramètre définissant une distribution de variabilité (voir Figure 1.4 et Pouillot et al. (2003)).

La Figure 5.1 illustre la méthode de simulation de Monte Carlo à 2 dimensions (Frey, 1992; Cullen et al., 1999; Pouillot et al., 2010). L'incertitude et la variabilité sont propagées dans deux dimensions différentes. Cette méthode revient à effectuer une simulation de Monte Carlo à une dimension avec les facteurs variables du modèle (conditionnellement à des valeurs d'incertitude tirées), puis à répéter cette analyse de risque pour différentes valeurs d'incertitude. Si on note N_u le nombre d'itérations dans la dimension incertitude, et N_v le nombre d'itérations dans la dimension variabilité, alors les étapes de la simulation de Monte Carlo 2D sont :

1. Échantillonner les paramètres incertains à partir de leurs distributions respectives (N_u itérations).

CHAPITRE 1. Introduction

2. Échantillonner les facteurs variables à partir de leurs distributions respectives (N_v itérations). Dans le cas des facteurs dont la distribution de variabilité est connue avec incertitude, cet échantillonnage se fait conditionnellement aux paramètres incertains précédemment échantillonnés ($N_u \times N_v$ itérations).

3. Calculer le risque correspondant aux $N_u \times N_v$ simulations.

4. Étudier ce risque en tenant compte des deux dimensions. Pour chaque valeur d'incertitude, des statistiques sont calculées dans la dimension variabilité (exemple : moyenne ou quantile). Puis l'incertitude sur ces statistiques est estimée dans la dimension incertitude, en prenant par exemple la médiane de ces statistiques comme estimation, et les quantiles 0.025 et 0.975 de ces statistiques comme bornes de l'intervalle de crédibilité à 95%.

Un exemple typique de résultat de simulation de Monte Carlo à 2 dimensions est un graphe comme illustré en Figure 1.6 (Delignette-Muller et al., 2006). Le cas d'étude porte sur la croissance de *Listeria monocytogenes* dans du saumon fumé.

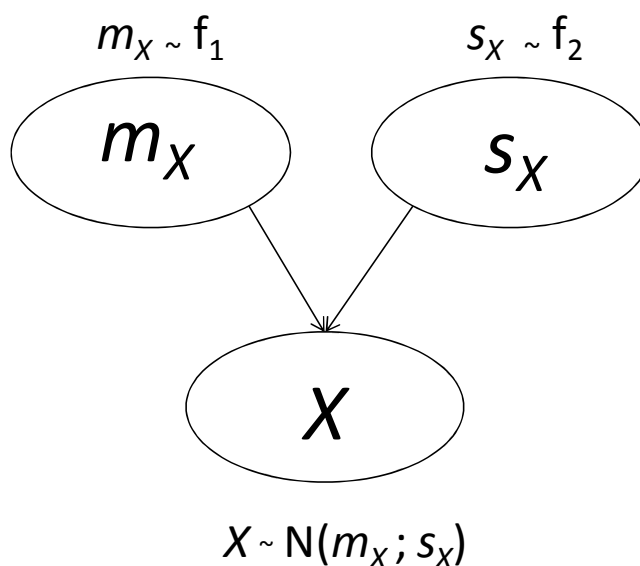


Figure 1.4. Graphe représentant un exemple de modélisation hiérarchique de l'incertitude et de la variabilité. La variable X suit une loi normale de moyenne m_X et d'écart-type s_X . m_X et s_X (parfois appelés paramètres, ou hyperparamètres) suivent chacun une distribution de probabilité (f_1 ou f_2) représentant de l'incertitude. La loi normale modélise la variabilité. La distribution marginale sur X est un mélange d'incertitude et de variabilité.

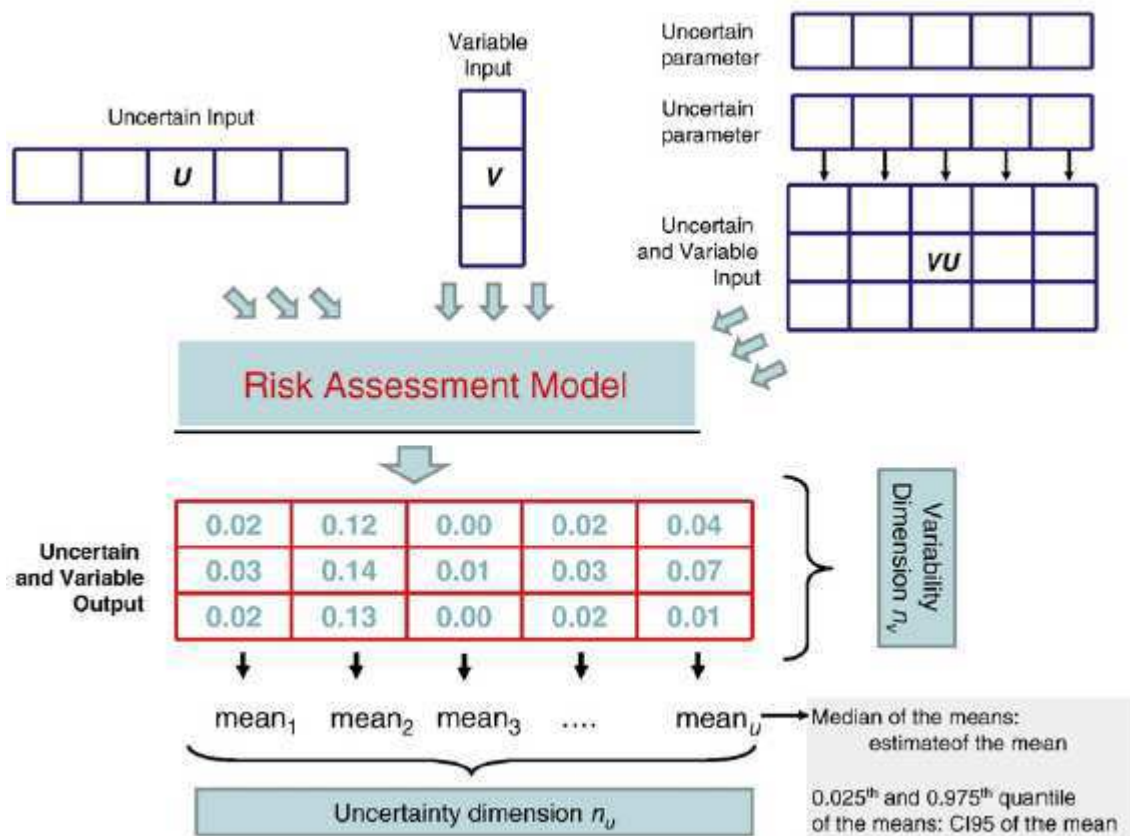


Figure 1.5. Illustration de la simulation de Monte Carlo à 2 dimensions. Source : Pouillot et Delignette-Muller, 2010.

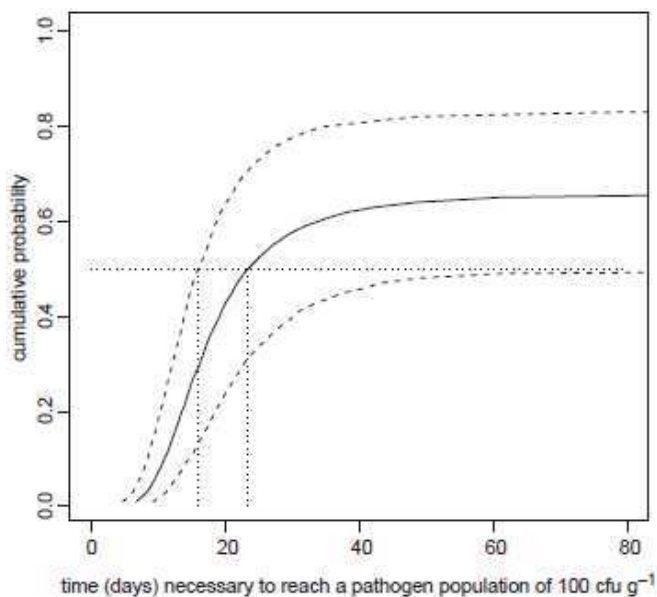


Figure 1.6. Résultat de simulation d'ordre 2 de la durée (en jours) nécessaire pour que la population de *Listeria monocytogenes* atteigne un niveau de 100 UFC/g dans du saumon fumé à 4°C. Pour

chaque valeur de la fonction de répartition (ordonnée), il est possible de lire sur la ligne continue l'estimation du percentile correspondant dans la variabilité (abscisse), et les deux lignes en trait pointillé délimitent l'intervalle de crédibilité à 95% reflétant l'incertitude sur le percentile estimé. Source : Delignette-Muller et al. (2006)

1.2.3 L'approche bayésienne

Comme de nombreux modèles, les modèles d'évaluation des risques (ou des bénéfices) peuvent se représenter graphiquement, par exemple sous la forme de réseaux bayésiens (Jensen, 2001; Parent and Bernier, 2007). Un réseau bayésien est un graphe encodant des indépendances conditionnelles d'une loi de probabilité jointe sur des variables aléatoires. Il peut être défini par la donnée d'un graphe acyclique orienté (DAG) et de lois conditionnelles par rapport aux parents de chaque nœud du graphe. Les nœuds représentent des variables et sont reliés entre eux par des arcs orientés, indiquant les dépendances entre les variables. Ces dépendances peuvent être déterministes ou aléatoires. Les DAG permettent de représenter facilement la structure hiérarchique d'un modèle, c'est-à-dire différents niveaux de regroupements possibles (par exemple regrouper les informations relatives à un même groupe génétique, ou à un même lot de légumes).

Les modèles d'évaluation du risque pouvant être considérés comme des réseaux bayésiens, ils peuvent être étudiés par inférence bayésienne. L'approche bayésienne est de plus en plus utilisée en analyse de risques, notamment pour estimer l'incertitude et la variabilité des paramètres des modèles (Pouillot et al., 2003; Delignette-Muller et al., 2006; Albert et al., 2008; Crépet et al., 2009; Spor et al., 2010; Jaloustre et al., 2011). Elle est basée sur le théorème de Bayes et permet d'estimer des paramètres, ou d'améliorer la connaissance de paramètres, à partir de données (Gelman et al., 2003; Robert, 2007).

Le théorème de Bayes s'écrit :

$$f(\theta|X) = \frac{\pi(\theta)l(X|\theta)}{\int \pi(\theta)l(X|\theta)d\theta}$$

θ est le paramètre à estimer, X les données, $\pi(\theta)$ la *loi a priori* de θ qui représente la connaissance sur θ avant l'observation des données X , $l(X|\theta)$ est la *vraisemblance* des données, c'est-à-dire la probabilité d'observer les données sachant θ , et $f(\theta|X)$ est la *loi a posteriori* de θ , c'est-à-dire la distribution du nouvel état de connaissance sur le paramètre θ , issu de la combinaison de la connaissance a priori et de l'information apportée par les données. La loi a priori décrit l'état de connaissance préalable sur θ et est plus ou moins informative : elle est dite « vague » ou encore « peu informative » si la connaissance sur le paramètre est très faible (par exemple modélisée par une loi avec une grande variance), ou bien elle est plus informative si la connaissance sur le paramètre est assez grande (par exemple modélisées par une loi normale de faible écart-type). Un intérêt de l'approche bayésienne est justement d'apporter de l'information via la définition de la loi a priori contenant la connaissance avant inférence, pour une meilleure estimation des paramètres quand les données sont rares. Le calcul de la loi a posteriori est en général difficile analytiquement, et il est fréquent de recourir à des méthodes de simulation basées sur des chaînes de Markov (MCMC), qui vont permettre de simuler des échantillons dans les lois a posteriori, en utilisant des

CHAPITRE 1. Introduction

algorithmes de Gibbs et de Metropolis-Hastings. Il convient néanmoins de s'assurer que les chaînes de Markov ont atteint leur distribution stationnaire (la loi a posteriori) pour être sûr que les valeurs simulées sont issues de cette loi. Des logiciels performants ont été développés afin de faciliter l'usage des méthodes MCMC, comme Winbugs, OpenBugs (Lunn et al., 2000; Lunn et al., 2009) ou Jags (Plummer, 2007). L'approche bayésienne a été utilisée dans cette thèse pour l'estimation de paramètres de référence par méta-analyse (Chapitre 2) et pour la confrontation d'un modèle d'évaluation du risque avec des données de contamination (Chapitre 6).

1.2.4 Analyse de sensibilité

L'analyse de sensibilité globale consiste à analyser un modèle mathématique en étudiant l'impact de la variation des variables d'entrée du modèle sur la variation de la sortie du modèle. L'analyse de sensibilité est utilisée dans de nombreux domaines d'application faisant appel à la modélisation, tels que l'ingénierie des systèmes complexes, les sciences sociales, économiques, physiques (Frey and Patil, 2002). Elle est de plus en plus utilisée en analyse des risques alimentaires (Zwietering and Van Gerwen, 2000; Frey et al., 2003; Mokhtari et al., 2005; Pouillot et al., 2007; Membré et al., 2008; Ellouze et al., 2010). C'est un outil précieux pour identifier les variables d'entrée les plus influentes et les moins influentes sur la sortie du modèle, pour identifier les interactions entre variables, ou encore pour évaluer la qualité d'un modèle. Il peut en découler des possibilités de simplification du modèle, de priorisation de la recherche de données additionnelles pour diminuer les sources d'incertitude majeures du modèle, d'orientation vers les mesures de maîtrise de risques les plus pertinentes.

Frey et Patil (2002) proposent une revue des méthodes d'analyses de sensibilité utilisées en analyse des risques alimentaires, telles que les méthodes graphiques, les coefficients de corrélation, l'analyse par régression, les méthodes par surface de réponses, et les méthodes basées sur l'analyse de la variance (ANOVA, indices FAST, indices de Sobol). Les méthodes d'analyse de sensibilité les plus pertinentes en analyse des risques alimentaires sont celles qui prennent en compte la non-monotonie, la non-linéarité des modèles, les seuils et les interactions, très fréquents dans les modèles QMRA. Par exemple les coefficients de corrélation de Spearman prennent en compte la non-monotonie et la non-linéarité des modèles, mais ne prennent pas en compte les interactions. La méthode ANOVA permet de prendre en compte les différentes contraintes mentionnées ci-dessus, mais nécessite une discrétisation des variables continues, ce qui entraîne une perte d'information. Une des méthodes les plus intéressantes est le recours aux indices de sensibilité de Sobol (Saltelli, 2002), qui ne nécessitent pas d'hypothèse sur le modèle ni de discrétisation des variables, et prennent en compte les interactions.

Par ailleurs, l'analyse de sensibilité globale peut être complétée par des analyses de scénarios, qui permettent de tester de nouvelles situations dans le modèle. Cette approche peut être intéressante pour étudier l'impact d'éventuelles mesures de maîtrise du risque, ou l'impact de l'incertitude de certains paramètres ou de certaines hypothèses.

1.3 Cas d'études de la thèse

1.3.1 Risque microbiologique et bénéfice nutritionnel dans les haricots verts appertisés

Cette thèse entre dans le cadre du projet ANR Ribenut, qui a pour objet la mise en place de nouvelles approches pour une évaluation du compromis risque microbiologique – bénéfice nutritionnel pour les légumes traités thermiquement. Les partenaires du projet sont l'Unité INRA Met@risk, l'UMR 408 Sécurité et Qualité d'Origine Végétale de l'INRA d'Avignon, l'UMR GENIAL (Génie des Procédés Alimentaires) constituée entre AgroParisTech, l'INRA et le Cnam, le LUBEM de l'Université de Bretagne, deux partenaires industriels et des centres techniques : CTCPA (Avignon), Aérial (Strasbourg) et Adria (Quimper).

Le projet ANR Ribenut comporte plusieurs groupes de travail, dont celui intitulé « Développement d'un modèle probabiliste évaluant le compromis risque/bénéfice », qui est le sujet principal de ma thèse. Dans ce cadre, un modèle risque-bénéfice a été construit autour de la chaîne de fabrication de conserves de haricots verts.

1.3.1.a La chaîne de production « haricots verts appertisés »

La Figure 1.7 présente le parcours suivi par un haricot vert lors du procédé d'appertisation. Les étapes principales sont la récolte des légumes, puis des étapes de lavage-éboutage-calibrage, puis le blanchiment, l'emboîtement doublé d'un ajout de jus (constitué d'eau et de sel) et la stérilisation.

Deux étapes particulièrement importantes dans le procédé industriel sont le blanchiment et la stérilisation. Le blanchiment est un traitement thermique de quelques minutes entre 70°C et 100°C. Son rôle est principalement d'inactiver les enzymes afin d'assurer une meilleure préservation des légumes (pour éviter des modifications de couleurs et/ou de saveurs), et de réduire la charge microbienne. Il permet également d'attendrir le légume de manière à faciliter sa mise en boîte. La stérilisation est un traitement thermique entre 110°C et 130°C visant à inactiver les enzymes et la majeure partie des microorganismes, notamment des bactéries naturellement présentes dans les légumes, pour stabiliser les produits. La stérilisation des légumes dans une boîte hermétique se nomme appertisation et permet de conserver les légumes plusieurs années. Blanchiment et stérilisation assurent également la cuisson du légume.

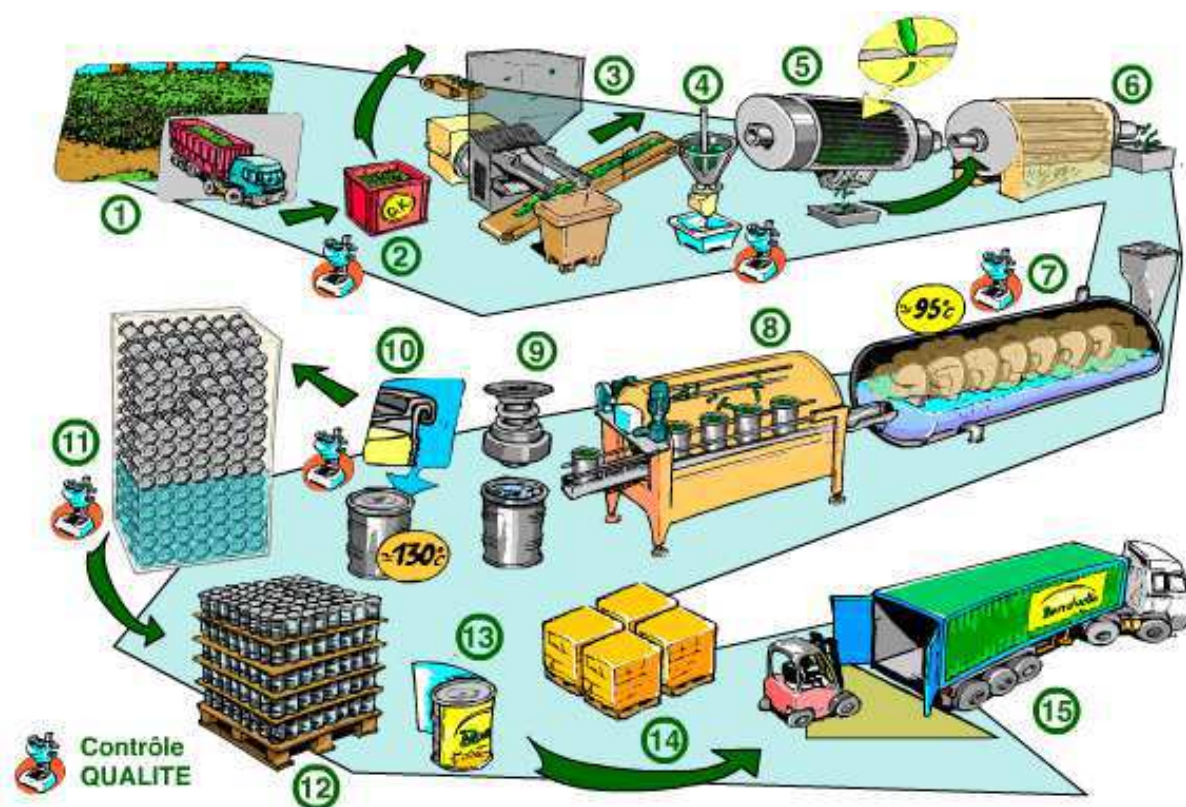


Figure 1.7. Schéma de la chaîne de production d'une conserve de haricots verts. Etapes : 1. Parcelles de production - 2. Réception - 3. Tri par ventilation - 4. Épierrage/lavage - 5. Éboutage - 6. Calibrage - 7. Blanchiment - 8. Emboitage - 9. Jutage : remplissage des boîtes avec un jus de couverture - 10. Sertissage - 11. Stérilisation - 12. Stockage des boîtes blanches - 13. Étiquetage - 14. Conditionnement et stockage - 15. Expédition. Source : www.bonduelle.com.

1.3.1.b Le risque d'altération dans la conserve de haricots verts

Les conserves présentent généralement une très bonne sécurité sanitaire, grâce au procédé de stérilisation qui permet d'inactiver la grande majorité des bactéries pathogènes. Le risque principal en cas de défaillance lors de la stérilisation est la présence de *Clostridium botulinum*, une bactérie pouvant former des toxines dont l'ingestion peut provoquer une maladie grave, le botulisme. Mal diagnostiquée ou mal traitée, cette maladie peut entraîner la mort en quelques jours. Des mesures de sécurité sont donc mises en place afin de contrôler les produits appertisés. Les industriels effectuent des contrôles réguliers afin de respecter les différentes normes sanitaires. Les cas de stérilisation défectueuse des conserves apparaissent cependant majoritairement dans les conserves familiales, moins contrôlées que dans l'industrie agro-alimentaire.

Un test de stabilité, réalisé suivant la norme AFNOR V08-408 (AFNOR, 1997), permet de s'assurer de la *stabilité biologique* d'un lot de produits stérilisés, c'est-à-dire l'absence de

CHAPITRE 1. Introduction

microorganismes susceptibles de se développer dans l'aliment dans les conditions non réfrigérées normalement prévues pour la fabrication, la distribution et l'entreposage. Ce test consiste à placer pendant 7 jours à 37°C et à 55°C des boîtes prélevés à l'issue du traitement thermique. À la fin de ce test, le produit est considéré stable si aucun bombement ou fuite n'est constaté ; si la variation de pH entre les unités étuvées et les unités non étuvées témoins ne dépasse pas 0,5 unités de pH. L'aspect et l'odeur du produit sont également vérifiés, et des examens microbiologiques peuvent compléter ces tests. À l'issue de ces tests, un défaut de stabilité biologique à 37°C doit entraîner la consignation du lot correspondant et faire l'objet d'un contrôle renforcé, car le produit peut alors présenter un risque sanitaire. Par contre, un défaut de stabilité biologique à 55°C n'entraîne pas de consignation du lot, le produit ne présentant pas de risque sanitaire pour le consommateur si le test à 37°C était négatif. Il doit cependant conduire le professionnel à prendre des mesures correctives nécessaires pour améliorer l'hygiène des fabrications. Le test d'incubation à 55°C doit en effet être considéré comme un indicateur de la qualité hygiénique du produit, et une augmentation anormale du taux de non-stabilité à 55°C pourrait être suivie d'un défaut de stabilité à 37°C. Il est habituel chez les industriels d'observer un taux non nul de non-stabilité à 55°C, de l'ordre de quelques pourcents. La non-stabilité à 55°C peut être due à différentes espèces de bactéries sporulées thermophiles hautement résistantes. Les trois espèces les plus fréquemment retrouvées dans les cas de non-stabilité sont *Moorella thermoacetica/thermoautotrophica*, *Geobacillus stearothermophilus* et *Thermoanaerobacterium* spp. (André et al., 2012b).

Dans la chaîne haricots verts appertisés, le risque modélisé dans cette thèse est le risque de non-stabilité à 55°C due à la bactérie *G. stearothermophilus* (Figure 1.8), un agent majeur des cas d'altération de conserves (en particulier des conserves de légumes), et par conséquent un problème persistant en industrie agroalimentaire (Logan and De Vos, 2009; Burgess et al., 2010; Prevost et al., 2010). Cette bactérie est aérobie (se multiplie en présence d'oxygène), thermophile (capable de vivre à une température élevée : 50-70°C), sporulé et mobile, de température optimale de croissance entre 39°C et 65°C (CTCPA, 2010). Cette bactérie non pathogène peut être responsable de l'altération « flat sour » des conserves, caractérisée par un surissement (les produits prennent une saveur aigre) sans bombage par production d'acides et sans gaz. Les aliments appertisés concernés sont principalement les produits laitiers, les légumes et les plats cuisinés.

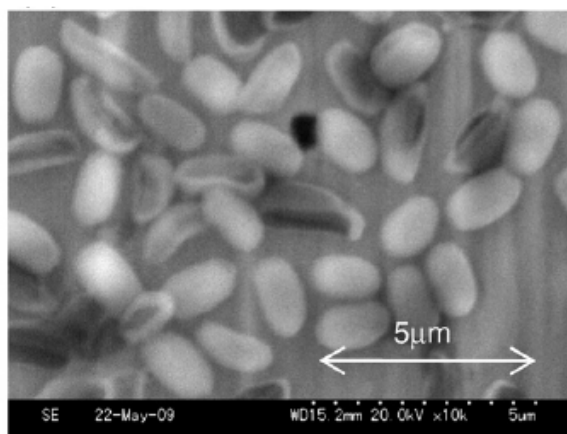


Figure 1.8. Image au microscope électronique à balayage de spores non traitées de *Geobacillus stearothermophilus*. Source : Zhao et al. (2010)

1.3.1.c Le bénéfice nutritionnel dans la conserve de haricots verts

Les haricots verts sont sources de nombreux bénéfices nutritionnels : vitamines (C, A, E, B dont la vitamine B9 regroupant l'acide folique et ses dérivés), minéraux (potassium, calcium, magnésium, fer, oligo-éléments), fibres, etc. De plus, c'est un légume peu calorique.

Le bénéfice nutritionnel modélisé dans cette thèse est la vitamine C (Figure 1.9), une vitamine qui a été très étudiée jusqu'à présent et dont le rôle sur la santé est primordial. La vitamine C est une vitamine hydro-soluble qui participe à de nombreux processus dans l'organisme. Elle possède en particulier des pouvoirs anti-oxydants (capture des radicaux libres), elle contribue au maintien des fonctions immunitaires, à la cicatrisation, à la fabrication du collagène, la protection de la paroi des vaisseaux sanguins, l'assimilation du fer et la détoxification de substances cancérigènes. Une pathologie associée à une forte carence en vitamine C sur plusieurs mois est le scorbut, maladie grave pouvant entraîner des œdèmes, des hémorragies puis la mort. Cette pathologie était fréquente chez les marins du XV^{ème} au XVIII^{ème} siècle. Des situations de carence modérée en vitamine C peuvent être responsables de perte d'appétit, d'amaigrissement et de fatigue. Chez l'Homme, la vitamine C est apportée uniquement par l'alimentation (il s'agit d'un nutriment indispensable) et les principales sources de vitamine C sont les fruits (agrumes et fruits rouges) et les légumes. Les apports nutritionnels conseillés (ANC) journaliers varient entre 110 mg et 140 mg chez l'adulte selon les catégories de population (classes d'âge, femmes enceintes). Par ailleurs, une surconsommation de vitamine C peut amener des effets indésirables, tels que des troubles intestinaux et des diarrhées. L'apport maximal tolérable est fixé à 2 g/jour chez les adultes selon les autorités canadiennes et américaines (Food and Nutrition Board, 2000). Cette vitamine étant très fragile, car sensible à l'eau, la chaleur, l'air et la lumière, il est important de favoriser au maximum sa conservation dans les aliments au cours de leur fabrication et de leur conservation.

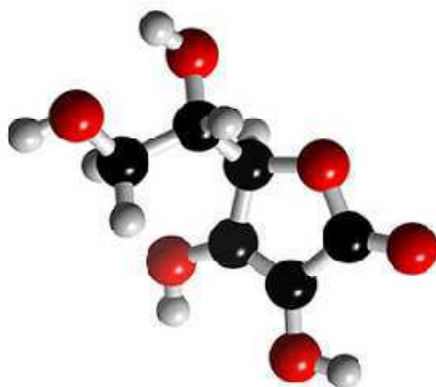


Figure 1.9. Représentation 3D de la molécule de vitamine C (acide ascorbique C₆H₈O₆).
Source : <http://2011-acideascorbique.blogspot.fr/p/i-quest-ce-que-la-vitamine-c.html>

1.3.2 Risque microbiologique dans la purée de courgettes

Complémentairement à l'analyse risque-bénéfice, de nouveaux outils ont été développés dans cette thèse dans le but d'améliorer un modèle à partir de données d'observation. Un modèle d'analyse quantitative du risque microbiologique déjà existant (Afchain et al., 2008) a été utilisé. Ce modèle décrit l'évolution de la concentration en *Bacillus cereus* le long d'une chaîne de production de purée de courgettes, puis jusque dans le réfrigérateur du consommateur. Un tel modèle est dit «de la fourche à la fourchette». Il prend en compte la diversité génétique de *B. cereus*, en modélisant quasiment distinctement l'évolution des 6 groupes génétiques de *B. cereus*.

Bacillus cereus est une bactérie pathogène sporulante aérobie-anaérobie à coloration Gram positive (Figure 1.10). L'ingestion de cellules ou de spores de *B. cereus*, ou d'une toxine produite dans l'aliment lors de sa croissance, sont responsables de toxi-infections caractérisées par des symptômes diarrhéiques ou d'intoxinations caractérisées par des symptômes émétiques (vomissements). *B. cereus* a été entre 1996 et 2005 la 4^{ème} cause de TIAC en France, avec 1 766 cas répartis sur 94 foyers confirmés (ANSES, 2011). Cette bactérie se retrouve naturellement sous forme de spores dans le sol, et peut contaminer une très grande variété d'aliments, en particulier les aliments d'origine végétale. La population de *B. cereus* présente une grande diversité génétique qui se traduit par une grande variété de comportements (Guinebretiere et al., 2008). Parmi les 6 groupes génétiques déjà associés à des toxi-infections alimentaires, les groupes III et IV sont mésophiles (se développent entre 10°C et 50°C), le groupe VII est thermotolérant (se développe entre 15°C et 50°C) et les groupes II et VI sont psychrotrophes (se développent entre 5°C et 40°C). Le seuil de dangerosité est d'environ 5 log₁₀UFC/g dans les aliments consommés. Les aliments à risque sont les plats cuisinés, les produits épicés ou aromatisés, les aliments déshydratés ou cuits à l'eau, surtout s'ils sont consommés de manière différée (ANSES, 2011). Les spores thermorésistantes rendent en effet *B. cereus* particulièrement adapté aux aliments traités thermiquement et consommés après un délai ayant permis la multiplication de la bactérie, comme les produits 5^{ème} gamme type purée de courgettes.

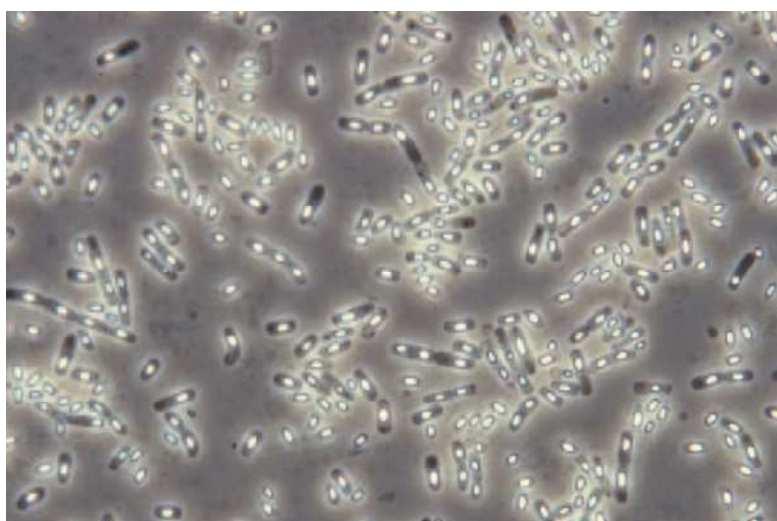


Figure 1.10. Photographie de cellules végétatives de *Bacillus cereus*. INRA

1.4 Organisation du mémoire

La majeure partie des travaux présentés dans cette thèse a été effectuée dans le cadre de l'évaluation du compromis risque microbiologique – bénéfique nutritionnel sur la chaîne de transformation des haricots verts appertisés.

Un premier travail a été l'acquisition de paramètres de référence (Chapitre 2). Les paramètres microbiologiques de résistance thermique de *G. stearothermophilus* en milieu aqueux ont été estimés par le biais d'une méta-analyse à partir d'un ensemble de 430 données issues principalement de la littérature. L'approche bayésienne mise en place a permis l'estimation de l'incertitude et de la variabilité des paramètres microbiologiques. Des sources possibles de variabilité sur ces paramètres sont par exemple les souches ou les milieux de traitement. De plus, plusieurs modèles ont été étudiés et comparés, afin de tester l'influence de diverses hypothèses de variabilité sur les paramètres de résistance thermique.

Les paramètres de résistance thermique de *G. stearothermophilus* ont ensuite été utilisés lors de la construction d'un modèle quantitatif d'évaluation des risques (Chapitre 3). Ce modèle décrit l'évolution de la concentration en *G. stearothermophilus* le long de la chaîne de fabrication de la conserve de haricots verts et modélise le risque industriel de non-stabilité à 55°C dans ces conserves. Les processus modélisés sont la contamination initiale de la matière première, une contamination croisée (par le biais de l'eau de blanchiment, le jus de couverture, et éventuellement les débris de haricots verts), l'inactivation des spores due aux traitements thermiques, ainsi que la germination et la croissance de *G. stearothermophilus* lors de l'étape d'incubation. La méthode de simulation de Monte Carlo 2D a été utilisée pour analyser les résultats du modèle. Une analyse de sensibilité a ensuite été mise en place afin d'identifier les paramètres-clés du modèle (les variables contribuant le plus à la variabilité de la réponse du modèle). Enfin, des scénarios ont été explorés, afin de tester l'influence sur les résultats du modèle de certaines hypothèses entachées d'une certaine incertitude, ou bien afin d'évaluer l'effet de certaines mesures possibles de maîtrise du risque.

Une autre partie de cette thèse a été consacrée à la construction d'un modèle quantitatif d'évaluation de la teneur en vitamine C dans les haricots verts au cours de la fabrication des boîtes de conserve de haricots verts, depuis la récolte des légumes jusqu'à la conservation des boîtes (Chapitre 4). Les processus modélisés sont la concentration initiale en vitamine C, les pertes par réactions chimiques et les pertes par transfert de matière (diffusion dans l'eau de blanchiment et dans le jus de couverture). La modélisation des pertes en vitamine C a été distinguée selon qu'il s'agissait de la forme réduite (acide ascorbique) ou de la forme oxydée (acide déshydroascorbique) de la vitamine C, car selon la forme de la vitamine les pertes ne sont pas toujours régies par les mêmes phénomènes. Les méthodes utilisées pour la construction de ce modèle d'évaluation du bénéfique sont similaires à celles utilisées pour la construction du modèle d'évaluation du risque, et la chaîne de production modélisée est identique dans les deux études. La méthode de simulation de Monte Carlo 2D a de nouveau été utilisée, ainsi que des tests de scénarios et une analyse de sensibilité pour rechercher les facteurs-clés du modèle.

CHAPITRE 1. Introduction

Une analyse risque-bénéfice a ensuite été menée sur la chaîne haricots verts appertisés, en considérant le risque de non-stabilité à 55°C due à *G. stearothermophilus* et le bénéfice nutritionnel associé aux teneurs en vitamine C (Chapitre 5). L'analyse risque-bénéfice s'est basée sur les modèles d'évaluation du risque et du bénéfice développés dans les chapitres 3 et 4. L'évolution simultanée des concentrations en *G. stearothermophilus* et en vitamine C dans les boîtes de conserve a été étudiée, ainsi que leur dépendance. Des optimisations possibles du compromis risque/bénéfice ont été recherchées au moyen de tests de scénarios basés sur les résultats des analyses de sensibilité.

Enfin, un travail complémentaire à l'évaluation du compromis risque-bénéfice dans les haricots verts a été réalisé sur une autre chaîne de transformation d'un aliment, la purée de courgette (Chapitre 6). Ce travail consistait à développer de nouveaux outils par approche bayésienne, dans le but d'améliorer l'estimation des paramètres d'un modèle d'évaluation quantitative du risque microbiologique à partir d'informations issues d'observations réelles acquises sur une chaîne de transformation d'un aliment. Un modèle d'évaluation des risques déjà existant et établi a été utilisé ; il concerne les risques associés à la bactérie pathogène *B. cereus* dans une chaîne de production de purée de courgette (Afchain et al., 2008). Le modèle a d'abord été retravaillé afin de construire un modèle d'« ordre 2 », c'est-à-dire un modèle distinguant la modélisation de la variabilité et de l'incertitude. Puis ce modèle a été « augmenté » afin de relier la sortie du modèle aux données de vieillissement. Les distributions des paramètres du modèle ont ensuite été mises à jour à partir de l'information apportée par les données par le biais d'une inférence bayésienne appliquée au modèle global d'évaluation du risque (calcul des lois a posteriori). Une simulation de Monte Carlo 2D a enfin été réalisée sur le modèle, d'abord à partir des lois a priori puis ensuite à partir des lois a posteriori des paramètres, et les résultats a priori et a posteriori de l'évaluation des risques ont été comparés.

Les chapitres 2, 3 et 6 sont présentés sous la forme d'articles. L'article présenté au chapitre 2 a été accepté en décembre 2012 par la revue *International Journal of Food Microbiology*. L'article présenté au chapitre 3 a été soumis à la revue *International Journal of Food Microbiology*, et l'article présenté au chapitre 6 a été accepté en août 2012 dans la revue *Risk Analysis*. Dans cette thèse, des introductions et des compléments ont été rajoutés à ces articles.

Chapitre 2

Une méta-analyse bayésienne de paramètres de résistance thermique d'une bactérie de référence en microbiologie des aliments : *Geobacillus stearothermophilus*

Ce chapitre présente une méthode d'estimation de paramètres microbiologiques de résistance thermique de *G. stearothermophilus* en milieu aqueux, dont les estimations seront utilisées dans un modèle d'analyse des risques dans les haricots verts appertisés présenté au chapitre 3. Ce travail a fait l'objet d'un article accepté en décembre 2012 dans la revue *International Journal of Food Microbiology*, reproduit dans la partie 2.2 de ce chapitre.

2.1 Estimation de paramètres microbiologiques par méta-analyse

De manière générale, la précision et la fiabilité des estimations d'un paramètre augmentent avec la quantité et la qualité des données disponibles. Il est donc intéressant de baser ces estimations sur la plus grande quantité de données possibles. Ces données peuvent provenir de différentes études individuelles provenant de la littérature, de données individuelles ou de bases de données déjà existantes. L'enjeu va alors être d'obtenir une estimation globale en prenant en compte les sources d'hétérogénéité des données dans la méthode d'estimation. La *méta-analyse* répond à ces objectifs ; c'est une méthode d'analyse permettant d'obtenir une synthèse rigoureuse et quantifiée de résultats d'études individuelles réalisées sur un même sujet. La méta-analyse permet

CHAPITRE 2. Méta-analyse

aussi de mettre en évidence les sources de variabilité entre les études et d'augmenter la puissance statistique des tests sur les paramètres. L'inclusion ou l'exclusion dans la méta-analyse d'études portant sur le sujet d'intérêt doit être clairement justifiée.

Très utilisée en médecine depuis les années 1980, les méta-analyses sont de plus en plus utilisées en microbiologie des aliments, en particulier pour l'estimation des paramètres microbiologiques de résistance thermique ou de croissance des bactéries (Pouillot et al., 2003; Delignette-Muller et al., 2006; van Asselt and Zwietering, 2006; Crépet et al., 2009; Halder et al., 2010; Den Besten and Zwietering, 2012; Jaloustre et al., 2012). Les paramètres des modèles primaires et/ou secondaires d'inactivation ou de croissance des bactéries sont alors estimés à partir de différentes sources de données. L'utilisation de différentes sources de données peut cependant amener de l'hétérogénéité, puisque la résistance thermique ou la croissance d'une bactérie peut varier en fonction des procédés, des types de produit, des différentes souches de bactérie à l'intérieur d'une même espèce, etc. La méthode la plus simple de méta-analyse est une estimation globale à partir d'une simple régression linéaire sur un ensemble de données relatives au micro-organisme d'intérêt pour des produits plus ou moins variés. van Asselt et al. (2006) ont par exemple utilisé cette approche pour déterminer la résistance thermique de *Listeria monocytogenes* sur un ensemble très varié de produits (bœuf, poulet, œuf, lait, saumon, carottes, champignons, etc.). Halder et al. (2010) ont eux adopté une démarche un peu différente, puisqu'ils ont construit des groupes de procédés et de produits similaires (produits laitiers ; fruits et légumes ; herbes et épices ; etc.) et ont ensuite procédé à des estimations de paramètres sur chacun des groupes. Cette dernière démarche permet de tenir compte en partie des sources de variabilité entre les données. D'autres méthodes possibles pour prendre en compte les sources de variabilité dans un ensemble de données ont été proposées : la modélisation hiérarchique associée à une inférence bayésienne et l'utilisation de modèles mixtes en statistique fréquentiste. Pouillot et al. (2003), Delignette-Muller et al. (2006), Crépet et al. (2009) ou encore Jaloustre et al. (2011) ont ainsi utilisé une approche bayésienne prenant en compte des sources de variabilité sur un même produit ou sur un groupe de produits proches (respectivement dans du lait, du saumon fumé, des salades et du boeuf). Jaloustre et al. (2012) ont utilisé les modèles mixtes ; ils montrent que la non prise en compte des sources de variabilité peut mener à des estimations biaisées des paramètres.

Les études individuelles incluses dans la méta-analyse présentée dans ce chapitre portent sur différents types de milieux aqueux (eau, tampon phosphate, haricot vert, etc.) et différentes souches de *G. stearothermophilus*. La résistance thermique de *G. stearothermophilus* en milieu aqueux a été étudiée en prenant en compte les différentes sources de variabilité entre ces études (type de milieu de traitement, souche, milieu de sporulation, milieu de recouvrement, expérimentateur). Une modélisation hiérarchique bayésienne a été utilisée. De plus, différentes hypothèses de variabilité sur les paramètres de résistance thermique ont été étudiées, ainsi que leur impact sur les estimations de ces paramètres.

2.2 Article : A meta-analysis accounting for sources of variability to estimate heat resistance reference parameters of bacteria using hierarchical Bayesian modeling: estimation of D at 121.1°C and pH 7, z_T and z_{pH} of *Geobacillus stearothermophilus*

Cette partie est la reproduction d'un article écrit en collaboration entre Clémence Rigaux^a, Jean-Baptiste Denis^b, Isabelle Albert^a et Frédéric Carlin^{c,d}. Il a été accepté en décembre 2012 dans la revue International Journal of Food Microbiology.

^aINRA, UR 1204, Met@risk, Food Risk Analysis Methodologies, F-75005 Paris, France

^bINRA, Unité MIA, Domaine de Vilvert, 78352 Jouy-en-Josas Cedex, France

^cINRA, UMR 408, Sécurité et Qualité des Produits d'Origine Végétale, F-84000 Avignon, France

^dUniversité d'Avignon et des Pays de Vaucluse, UMR 408, Sécurité et Qualité des Produits d'Origine Végétale, F-84000 Avignon, France

Abstract

Predicting microbial survival requires reference parameters for each micro-organism of concern. When data are abundant and publicly available, a meta-analysis is a useful approach for assessment of these parameters, which can be performed with hierarchical Bayesian modeling. *Geobacillus stearothermophilus* is a major agent of microbial spoilage of canned foods and is therefore a persistent problem in the food industry. The thermal inactivation parameters of *Geobacillus stearothermophilus* (D_{ref} , i.e. the decimal reduction time D at the reference temperature 121.1°C and pH 7.0, z_T and z_{pH}) were estimated from a large set of 430 D values mainly collected from scientific literature. Between-study variability hypotheses on the inactivation parameters D_{ref} , z_T and z_{pH} were explored, using three different hierarchical Bayesian models. Parameter estimations were made using Bayesian inference and the models were compared with a graphical and a Bayesian criterion. Results show the necessity to account for random effects associated with between-study variability. Assuming variability on D_{ref} , z_T and z_{pH} , the resulting distributions for D_{ref} , z_T and z_{pH} led to a mean of 3.3min for D_{ref} (95% Credible Interval CI=[0.8; 9.6]), to a mean of 9.1°C for z_T (CI=[5.4; 13.1]) and to a mean of 4.3 pH units for z_{pH} (CI=[2.9; 6.3]), in the range pH 3 to pH 7.5. Results are also given separating variability and uncertainty in these distributions, as well as adjusted parametric distributions to facilitate further use of these results in aqueous canned foods such as canned vegetables.

Key words

Meta-analysis, Hierarchical Bayesian models, Variability, *Geobacillus stearothermophilus*, Thermal inactivation parameters, Canning

2.2.1 Introduction

Survival prediction of micro-organisms in food uses parameters describing (i) the intrinsic resistance of micro-organisms and (ii) the effect of changes in the physical and chemical characteristics of the environment. The most widely used parameters in thermal processing are the decimal reduction time (D) and the z parameters. z_T and z_{pH} express the increase in temperature or the decrease in pH resulting in a tenfold reduction in D . A reference D at a reference temperature and pH can be used for convenience and is denominated as D_{ref} . Many D and z values are available in the literature for most pathogenic and many food spoilage bacteria. A simple examination of available D and z values for a given micro-organism reveals a high variance. This variance, or diversity, can be attributed either to uncertainty (lack of perfect knowledge of the parameter value, due for instance to a lack of experimental data), or to uncertainty together with variability (intrinsic heterogeneity of a population (Vose, 2000)) due for instance to differences in sensitivity of tested strains. A meta-analysis of this abundant literature is then a relevant approach to determine reference thermal inactivation parameters integrating different sources of variability and which can be used for further risk assessments (Gonzales Barron et al., 2009; Den Besten et al., 2012). In food microbiology, meta-analyses are increasingly used to evaluate growth or resistance parameters of micro-organisms of concern in food (van Asselt et al., 2006; Halder et al., 2010; Santillana Farakos and Zwietering, 2011) and also increasingly integrate between-study variability (Pouillot et al., 2003; Crépet et al., 2009; Jaloustre et al., 2012; Van Lieverloo et al., 2012).

In this work, a meta-analysis is used to determine the thermal inactivation parameters D_{ref} , z_T and z_{pH} of *Geobacillus stearothermophilus* in aqueous media, and the effect of different variability hypotheses on the determination of D_{ref} , z_T and z_{pH} is explored. *G. stearothermophilus* is a non pathogenic aerobic bacteria that forms highly heat-resistant endospores. It is a major agent of microbial spoilage of canned foods and in particular of canned vegetables and is therefore a persistent problem in the food industry (Logan et al., 2009; Burgess et al., 2010; Prevost et al., 2010). Numerous papers have been dedicated to the fate of *G. stearothermophilus* during heat-treatment in relation to vegetable canning in particular and the effect of different factors on its survival, such as heating temperature, nature or pH of the heating medium, or composition of the recovery medium. In this paper, three models were considered to study the impact of considering (or not) a between-study variability. In these different models, D_{ref} , z_T and z_{pH} were considered as only uncertain, or were considered as both uncertain and variable. Such variability hypotheses were based on the diversity of reported D_{ref} values and z_T values (for instance z_T estimated at 6.5°C in Periago et al. (1998c) and at 10.5°C in Tejedor et al. (2001)). A Bayesian approach (Robert, 2007) was used to evaluate separately microbial variability and uncertainty (Pouillot et al., 2003), as recommended by international agencies in microbiological risk assessment (Codex alimentarius Commission, 1999; European Commission, 2003).

2.2.2 Material and Methods

2.2.2.a Inactivation models

Spore inactivation was assimilated to first order kinetics and survival curves were classically considered as log-linear. The primary inactivation model is therefore described by:

$$\log_{10}(N_t) = \log_{10}(N_o) - \frac{t}{D(T,pH)} \quad (2-1)$$

where N_o (CFU/g) is the initial population size at time 0 (min), t (min) is the processing time, N_t (CFU/g) is the population size at time t , and $D(T, pH)$ (min) is the decimal reduction time at temperature T (°C) and pH of treatment.

The secondary model is an extension of the classical Bigelow model (Couvert et al., 2005). In addition to the effect of a change in temperature on D , this model also describes the effect on D of a change in pH. It is expressed as

$$D(T, pH) = D_{ref} \cdot 10^{(T_{ref}-T)/z_T} \cdot 10^{(pH-pH_{ref})/z_{pH}} \quad (2-2)$$

which can be reformulated into:

$$\log_{10}D(T, pH) = \left(\log_{10} D_{ref} + \frac{T_{ref}}{z_T} - \frac{pH_{ref}}{z_{pH}} \right) + \left(-\frac{1}{z_T} \right) \cdot T + \left(\frac{1}{z_{pH}} \right) \cdot pH \quad (2-3)$$

where T (°C) is the temperature and pH is the pH of the product during processing, T_{ref} (°C) is the reference temperature, here fixed at the reference sterilization temperature of 121.1°C, pH_{ref} (pH unit) is the pH of reference, fixed at 7.0, z_T (°C) is the increase in temperature resulting in a ten-fold reduction in D , z_{pH} (pH unit) is the decrease in pH resulting in a ten-fold reduction in D , and $D_{ref} = D_{121.1^\circ\text{C}, pH7}$ (min) is the decimal reduction time at 121.1°C and pH 7.

2.2.2.b Data

Data consist in 430 D -values at temperatures ranging between 104°C and 140°C and pH ranging between 3 and 7.5, obtained from 18 different *G. stearothermophilus* (formerly *Bacillus stearothermophilus*) strains and from 20 different heating media. An overview of the data is shown on Figure 2.1, where $\log_{10}D$ values are plotted against temperature, and different symbols are used for different ranges of pH. The full set of data is also available as an Excel spreadsheet (Supplementary material, Table 2.7).

Data were collected from 30 sources: 27 peer-reviewed scientific publications, 1 communication as poster, and 2 personal communications of unpublished data. All D -values were the ones communicated and calculated by the authors on the hypothesis of a log-linear inactivation model (Eq. (1)). These D -values have often been determined by the fitting of the rapid inactivation phase of the survival curve. The collected data were selected from various criteria. Only those

CHAPITRE 2. Méta-analyse

compatible with an application to canned vegetables were retained, and so only experiments performed with spores heat-treated in aqueous media have been considered. These heating media were water, McIlvaine buffer, Butterfield phosphate buffer, phosphate buffer, Sorensen buffer, acetate buffer, Tris buffer, some vegetable media (mushroom purée, green beans, mushrooms, asparagus), and some other food substrates. Data obtained at $\text{pH} > 7.5$ have been excluded, because of scarcity of data at $\text{pH} > 7.5$ and a poor relevance with food pH (and so the model is only valid for $\text{pH} \leq 7.5$). To account for the sources of diversity on the whole data-set, a notion of “*study*” has been defined as a collection of data obtained from relatively close experimental conditions detailed hereafter. Data were attributed to different studies when:

- they were extracted from different papers, or
- they were obtained with different strains, or
- different heating media were used, or
- spores were obtained in highly different media, or
- surviving spores were enumerated in highly different media.

Within a study, temperature and pH can vary: their changes will be taken into account in the inactivation secondary model (Eq. (3)). At least one source of variability exists between two studies. The whole dataset was then divided into 179 studies. Sometimes the distinction between studies was not obvious. For instance the heating media acidified with two different products in Fernandez et al. (1994) have been considered as two different media. The composition of every study can be found in the supplementary material, Table 2.7.

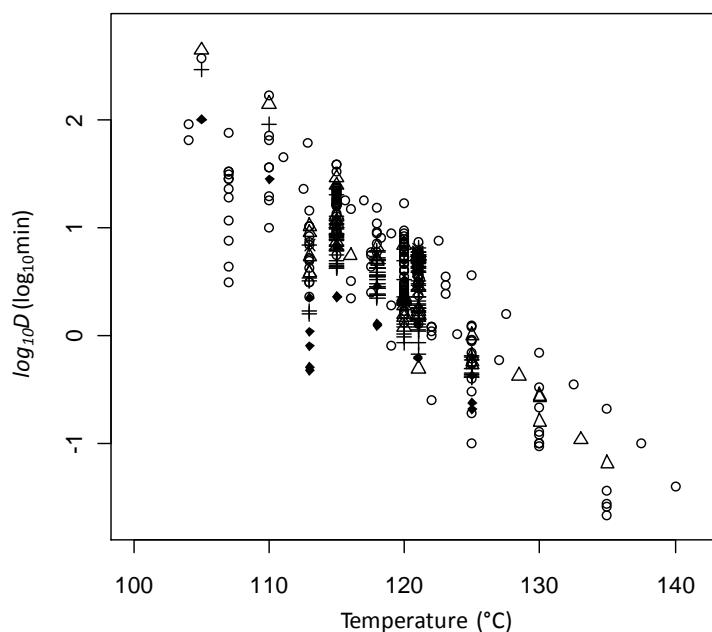


Figure 2.1. *G. stearothermophilus* $\log_{10} D$ values extracted from the scientific literature plotted against temperature; symbols denote five classes of pH: (\blacklozenge) $3 \leq \text{pH} < 5$, ($+$) $5 \leq \text{pH} < 6$, (\triangle) $6 \leq \text{pH} < 7$, (\circ) $\text{pH} = 7$, ($*$) $7 < \text{pH} \leq 7.5$.

In this work, uncontrollable (source, strain) and controllable (heating, sporulation and enumeration media) variables were used as covariates together. As not enough data were available to estimate the effects of all the covariates, a hierarchical approach was used, accounting for random effects for all the covariates, and allowing the estimation of uncertainty and variability of the heat resistance parameters D_{ref} , z_T and z_{pH} in a general aqueous medium.

2.2.2.c Three competitive models

Determination of D_{ref} , z_T and z_{pH} was done using three models of increasing complexity developed under different assumptions. The three models may be considered as Bayesian networks (Jensen, 2001), and are represented in Figure 2.2 by means of directed acyclic graphs (DAG). Basically a DAG represents quantities involved in a model, such as variables and data, called nodes, and the links between these nodes. Terminal nodes will be called “parameters” in this paper (i.e., here, all the nodes without index i or j) and the other nodes, except data, will be called “intermediate variables”. The links between the nodes of the three DAGs are given in Table 2.1.

- In the first model (Figure 2.2 a), called the “*basic model*”, D_{ref} , z_T and z_{pH} are estimated based on a simple linear regression, assuming no between-study variability. Under this assumption, D_{ref} , z_T and z_{pH} take one value each able to represent the thermal inactivation for all studies.

- The second model (Figure 2.2 b), called the “*intermediate model*”, assumes that D_{ref} is impacted by between-study variability, but not z_T and z_{pH} . For instance different strains tested in the same conditions (same pH, temperature, etc.) may exhibit different D_{ref} values, but keep identical z_T and z_{pH} values. To account for the variability of D_{ref} in the DAG (Figure 2.2 b), a hierarchical structure is introduced: a loop on the studies indexed by j is added, as well as nodes μ_{LDref} and σ_{LDref} which represent the mean and the standard deviation of $\log_{10}D_{ref}$, and inter-study variability of $\log_{10}D_{ref}$ is modeled by a normal law (for each study j , $\log_{10}D_{ref}^j \sim N(\mu_{LDref}, \sigma_{LDref})$).

- The third model (Figure 2.2 c), called the “*complete model*”, assumes that D_{ref} , z_T and z_{pH} are all impacted by between-study variability. A hierarchical structure on the three variables is introduced to account for these sources of variability in the DAG.

Both z_T and z_{pH} have a similar significance in the model: they describe the effect of an environmental factor (heating temperature and heating medium pH, respectively) on the heat-resistance parameter. Consequently and for consistency, in each of the three models z_T and z_{pH} were assigned the same assumptions on their between-study variability, and thus were either both variable, or both non variable.

In the statistical Bayesian approach (Robert, 2007), all the nodes of a Bayesian network, except the data and the covariates, are considered as random variables. The distributions on these variables represent some uncertainty as to the true value of the variables. The Bayesian inference allows estimation of this uncertainty. In addition, in the intermediate and complete models, variability and uncertainty are separately modeled by means of the hierarchical structure.

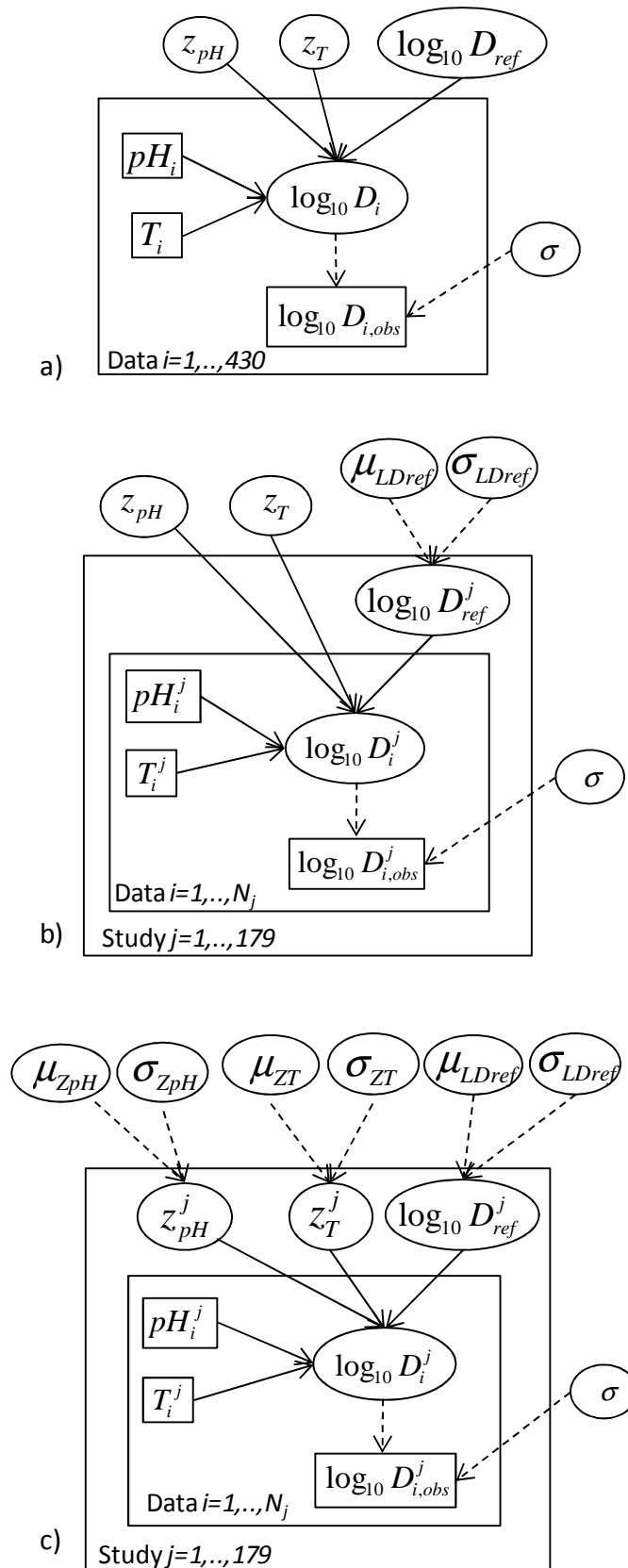


Figure 2.2. Directed acyclic graphs of the basic model (a), of the intermediate model, assuming D_{ref} variable (b), and of the complete model assuming D_{ref} , z_T and z_{pH} variable (c). Model quantities are represented by nodes. Ellipses indicate random variables and rectangles indicate data or covariates

CHAPITRE 2. Méta-analyse

(T, pH) . Solid edges indicate deterministic links and dashed edges indicate stochastic links between nodes. These links are described in Table 2.1. $\log_{10} D_i$ and $\log_{10} D_{i,obs}$ (resp. $\log_{10} D_i^j$ and $\log_{10} D_{i,obs}^j$) are the decimal logarithms of the modeled and of the observed decimal reduction times at a constant pH pH_i (resp. pH_i^j) and a constant temperature T_i (resp. T_i^j), in data i (resp. coming from study j). μ_{zT} (resp. $\mu_{z_{pH}}, \mu_{LDref}$) is the mean of z_T (resp. $z_{pH}, \log_{10} D_{ref}$). σ_{zT} (resp. $\sigma_{z_{pH}}, \sigma_{LDref}$) is the standard deviation of z_T (resp. $z_{pH}, \log_{10} D_{ref}$). σ is the residual model error. N_j is the quantity of data included in study j .

Table 2.1. Description of the links between variables in the different models (see also Figure 2.2).

Model	Node	Type	Definition ^a
Basic model	$\log_{10} D_i$	deterministic	Eq. (3): secondary model
	$\log_{10} D_{i,obs}$	stochastic	$N(\log_{10} D_i, \sigma)$
Intermediate and complete model	$\log_{10} D_i^j$	deterministic	Eq. (3): secondary model
	$\log_{10} D_{i,obs}^j$	stochastic	$N(\log_{10} D_i^j, \sigma)$
	$\log_{10} D_{ref}^j$	stochastic	$N(\mu_{LDref}, \sigma_{LDref})$
Complete model	z_T^j	stochastic	$N(\mu_{zT}, \sigma_{zT})$
	z_{pH}^j	stochastic	$N(\mu_{z_{pH}}, \sigma_{z_{pH}})$

^a $N(a,b)$ denotes the normal distribution of mean a and standard deviation b .

2.2.2.d Statistical analysis

Computing posterior distributions

Estimations are made using Bayesian inference. Prior distributions on the parameters, reflecting the *a priori* knowledge, are combined with observed data with the help of Bayes' theorem to compute the posterior distribution on the same parameters, providing the new state of knowledge.

In this paper, rather poorly informative prior distributions were used, with supports based on microbiological considerations. Prior distributions are described in Table 2.2. Attention was paid to choosing fairly "compatible" priors in order to obtain sound model comparisons (Dawid and Lauritzen, 2000). Indeed, some authors argue that priors of the complete model and of the nested models must be compatible, in order not to favor arbitrarily one model over another. We took this difficulty into account in a simple way, checking that with the chosen priors the distance between the

CHAPITRE 2. Méta-analyse

(marginal) prior distributions of $\log_{10}D_{ref}$, z_T and z_{pH} in the three models is small (Figure 2.3). The induced marginal prior distributions on the 430 data $\log_{10}D_{i,obs}$ (or $\log_{10}D_{i,obs}^j$) were also compared and were very similar in the three models.

Table 2.2. Prior distributions of the heat resistance parameters of *G. stearothermophilus* used for the three different models. See Section 2.2.2.a and Figure 2.2 for the definition of variables. $U(a,b)$ denotes the uniform distribution on $[a,b]$.

Variables	Unit	Basic model	Intermediate model	Complete model
$\mu_{z_{pH}}$	pH unit	-	-	$U(0, 20)$
$\sigma_{z_{pH}}$	pH unit	-	-	$U(0, 5)$
μ_{z_T}	°C	-	-	$U(0, 20)$
σ_{z_T}	°C	-	-	$U(0, 5)$
μ_{LDref}	\log_{10} min	-	$U(-2, 2)$	$U(-2, 2)$
σ_{LDref}	\log_{10} min	-	$U(0, 1)$	$U(0, 1)$
z_{pH}	pH unit	$U(0, 20)$	$U(0, 20)$	-
z_T	°C	$U(0, 20)$	$U(0, 20)$	-
$\log_{10}D_{ref}$	\log_{10} min	$U(-2, 2)$	-	-
σ	\log_{10} min	$U(0, 5)$	$U(0, 5)$	$U(0, 5)$

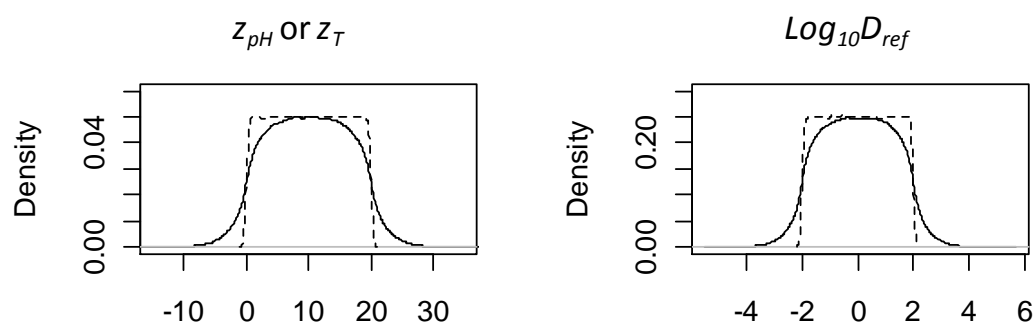


Figure 2.3. Comparison of *a priori* distributions on parameters $\log_{10}D_{ref}$, z_T and z_{pH} in the basic model (dashed lines) and marginal *a priori* distributions on intermediate variables $\log_{10}D_{ref}$, z_T and z_{pH} in the complete model (solid lines).

CHAPITRE 2. Méta-analyse

The posterior distributions were computed using Markov Chain Monte Carlo Simulations (MCMC) (Robert and Casella, 2004) using the software Jags 3.2.0 (Plummer, 2007). After an adaptation phase of 20 000 iterations with the basic and intermediate models and 2 million iterations for the complete model, the convergence was checked by visually analyzing two independent chains with different initial values, and by calculating the Gelman-Rubin convergence statistics (Brooks and Gelman, 1998). Samples of the posterior distributions were then obtained from 20 000 iterations with a space interval of 10 for the basic and intermediate models, and from 1 million iterations with a space interval of 500 for the complete model, leading in both cases to posterior samples of size 4000.

Statistics of estimated parameters and adjustment to parametric distributions

Several descriptive statistics were calculated for the estimated parameters. The distributions of the intermediate variable D_{ref} in the intermediate model, and of D_{ref} , z_T and z_{pH} in the complete model, were also considered by sampling in the *a posteriori* joint distributions of μ_{zT} , $\mu_{z_{pH}}$, μ_{LDref} , σ_{zT} , $\sigma_{z_{pH}}$, σ_{LDref} and using the relationships described in Table 2.1.

Some standard parametric distributions were fitted to the simulated posterior distributions of the parameters and of the intermediate variables D_{ref} , z_T and z_{pH} of the complete model to facilitate further use of these results. They were fitted by the maximum-likelihood method, using the R package fitdistrplus (Pouillot et al., 2010). The distributions that minimized the Akaike Information Criterion (AIC) were chosen.

Quality of fit

The Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) was calculated to compare the three models. The DIC measures the relative goodness of fit of a statistical model including a penalization by an estimation of the effective number of parameters. It is well adapted to Bayesian hierarchical model selection and is easily calculated from the MCMC samples. The model with the smallest DIC is generally supposed to be the better fitting model. As this criterion suffers from bias (using twice the data, for estimation and predictions) and may under-penalize complex models (Ando, 2007; Plummer, 2008), predictions obtained from cross-validation were also performed.

The different models were compared for their ability to predict observations in the framework of a 10-fold cross-validation process. It consisted in (i) randomly partitioning the data set into 10 equally-sized-groups of data, (ii) for each group of data G_i , performing a Bayesian inference from the training sample made up of the 9 other groups of data and calculating predictions for the data of group G_i (the validation sample). The results of each Bayesian inference are based on posterior samples of size 4000. For each $\log_{10}D_{i,obs}$, the prediction $\log_{10}D_{i,pred}$ of $\log_{10}D_{i,obs}$ was then defined by the median of the 4000 predicted values obtained, and the 95% credible interval (95%CI) was calculated. Graphical comparisons between predicted and observed values were then performed. As indicators of the relative goodness of fit, the square of the Pearson correlation coefficient between $\log_{10}D_{i,pred}$ and $\log_{10}D_{i,obs}$ (R^2) was calculated, as well as a more complete

criterion, the continuous ranked probability score (CRPS) (Gschlößl and Czado, 2005; Gneiting and Raftery, 2007), defined as:

$$CRPS(F_i, \log_{10}D_{i,obs}) = - \int_{-\infty}^{+\infty} (F_i(y) - 1[y \geq \log_{10}D_{i,obs}])^2 dy$$

$$CRPS = \frac{1}{N} \sum_{i=1}^N CRPS(F_i, \log_{10}D_{i,obs})$$

where F_i is the cumulative distribution function of the predicted values of datum $\log_{10}D_{i,obs}$, N is the number of data, and $1[\cdot]$ is the indicator function. The empirical version of F_i was used to estimate the CRPS. The CRPS is a numerical score assessing the quality of forecast, based on the distance between the predictions and the data, and also on the sharpness of the predictive distributions. A smaller CRPS in absolute value means a better fit.”

2.2.3 Results

2.2.3.a Convergence of the MCMC chains toward posterior distributions

A good convergence is reached with the three models (overlapping of the two independent chains and Gelman-Rubin statistic ≈ 1.0). However, reaching convergence required more iterations and longer calculation time for the complete model than for the other two models, especially for estimating $\mu_{z_{pH}}$ and $\sigma_{z_{pH}}$, which were quite autocorrelated. This can be explained by the more complex structure of the complete model, and by the rather low quantity of data available for estimating $\mu_{z_{pH}}$ or $\sigma_{z_{pH}}$ (compared to temperature, pH does not vary much in the whole set of data and in studies). Furthermore, it was observed in the three models that $\mu_{z_{pH}}$, $\sigma_{z_{pH}}$ or z_{pH} were quite sensitive to the choice of priors, and so their estimation is less reliable than the estimation of other variables.

2.2.3.b Determination of D_{ref} , z_T and z_{pH} using the basic model.

Descriptive statistics of empirical posterior distributions for the basic model are presented in Table 2.3. D_{ref} is estimated at 2.7min, z_T at 10.0°C and z_{pH} at 5.5 pH units. The residual error σ is rather high, with a mean of 0.32 \log_{10} min, compared to the range of $\log_{10}D$ values in the whole data set ([-1.7, 2.6] \log_{10} min). The credible intervals on D_{ref} , z_T and z_{pH} are narrow. They do not seem representative of the diversity of possible values that can be found in the different individual studies. Indeed, they do not contain certain values of D_{ref} , z_T and z_{pH} estimated (independently from the meta-analysis) in some studies j , such as $\widehat{z}_{T,j=32} = 6.5^\circ\text{C}$ (95%CI=[5.7°C, 7.4°C]), or $\widehat{z}_{pH,j=14} = 2.9$ (95%CI=[2.4, 3.7]), or $\widehat{D}_{ref,j=33} = 5.7$ min (95%CI=[4.2 min, 7.8min]).

CHAPITRE 2. Méta-analyse

Overall, the cross-validation predictive D -values moderately fit the observed D values, with high 95%CI for all data, as shown in Figure 2.4 a and Table 2.4 ($R^2=0.74$ and CRPS=-0.18).

In addition, as fairly non informative priors and a rather large number of data were used, the Bayesian estimations of this model are very close to estimations obtained with the classical ordinary least square method.

Table 2.3. Descriptive statistics of empirical posterior distributions of the heat resistance parameters of *G. stearothermophilus* estimated with the basic model. See Material and Methods for details. ^a Q_x stands for the X^{th} quantile.

Parameter	Unit	Mean	SD	$Q_{0.025}^a$	$Q_{0.975}^a$
D_{ref}	min	2.7	0.1	2.5	2.9
z_T	°C	10.0	0.3	9.5	10.6
z_{pH}	pH unit	5.5	0.6	4.5	6.8
σ	\log_{10} min	0.32	0.01	0.30	0.34

Table 2.4. DIC, R^2 , and CRPS obtained with the three models (R^2 and CRPS were obtained from the cross-validation predictions). The lowest DIC values and the highest R^2 and CRPS values correspond to the best fit of the model to data.

Model	DIC	R^2	CRPS
Basic	247	0.74	-0.18
Intermediate	-621	0.88	-0.10
Complete	-751	0.89	-0.09

2.2.3.c Determination of D_{ref} , z_T and z_{pH} using the intermediate model

Descriptive statistics of empirical posterior distributions for the intermediate model are presented in Table 2.5. The residual error σ is much smaller than in the basic model, and its mean is

CHAPITRE 2. Méta-analyse

estimated at $0.10 \log_{10} \text{min}$, showing a better fit with the data. The estimations of D_{ref} , z_T and z_{pH} are different from their estimations using the basic model: D_{ref} is estimated at 3.4min, z_T at 8.2°C and z_{pH} at 4.0 pH unit. Smaller estimations of z_T and z_{pH} are obtained, whereas a higher estimation is obtained for D_{ref} , with a very much higher standard deviation (3.3 min vs 0.1 min). Such increase in standard deviation is logical since the intermediate model accounts for between-study variability on D_{ref} . The credible intervals of z_T and z_{pH} are still narrow, and as in the basic model, do not contain certain values of z_T and z_{pH} individually estimated from particular studies. The DIC is much better than in the basic model (see Table 2.4).

Overall, the cross-validation predictive D -values fit the observed D values much better than in the basic model, as illustrated in Figure 2.4 b and Table 2.4, with a higher R^2 value (0.88) and a higher CRPS (-0.10), and with globally lower 95%CI on the predicted values. Due to the presence of a correlation structure between observations in the intermediate model (as well as in the complete model), it performed better taking better advantage of the part of the data used to infer the model. A few data are not as well predicted as the others. The reason could be the very low number of data in the related studies, making difficult or impossible the estimation of the random effect in each of these studies when using 10-fold cross-validation.

Table 2.5. Descriptive statistics of empirical posterior distributions of the heat resistance parameters of *G. stearothermophilus* estimated with the intermediate model. See Material and Methods for details.

Type of variables	Node	Unit	Mean	SD	$Q_{0.025}^a$	$Q_{0.975}^a$
Parameters	μ_{LDref}	$\log_{10} \text{min}$	0.39	0.03	0.34	0.45
	σ_{LDref}	$\log_{10} \text{min}$	0.35	0.02	0.31	0.39
	z_T	°C	8.2	0.1	8.1	8.4
	z_{pH}	pH unit	4.0	0.2	3.7	4.4
	σ	$\log_{10} \text{min}$	0.10	0.00	0.09	0.11
Intermediate variable	D_{ref}	min	3.4	3.3	0.5	12.0

^a Q_x stands for the X^{th} quantile.

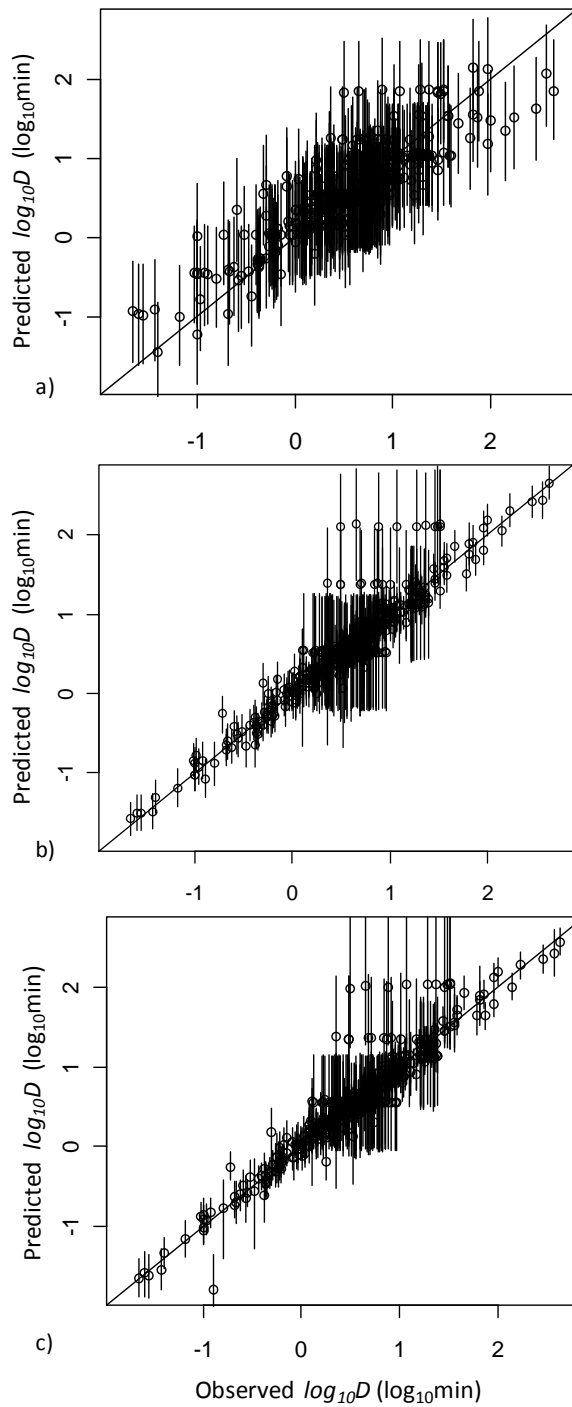


Figure 2.4. Comparison of observed $\text{Log}_{10}D$ values (x-axis) and $\text{Log}_{10}D$ values predicted by cross-validation knowing their study j , using the basic linear model (a), the intermediate model assuming D_{ref} variable (b), and the complete random model assuming D_{ref} , z_T and z_{pH} variable (c). Points represent the median predicted values $\log_{10}D_{i,pred}$, and segments represent the 95% credible intervals of the predicted values. Perfect fitting is obtained when points are on the drawn line $Y=X$.

2.2.3.d Determination of D_{ref} , z_T and z_{pH} using the complete model

Table 2.6 gives some descriptive statistics of empirical posterior distributions for the complete model and adjusted distributions. The residual error σ , estimated at $0.08 \log_{10} \text{min}$, is smaller than in the other models, showing a better fit. The global estimations of D_{ref} , z_T and z_{pH} (means equal to 3.3 min, 9.1°C and 4.3 pH units, respectively) are different from the estimations given by the basic and the intermediate models, with notably a much larger variation range of z_T and z_{pH} (respectively standard deviation $\text{SD} = 1.9^\circ\text{C}$ vs $\text{SD}=0.3^\circ\text{C}$ or 0.1°C , and $\text{SD}=0.8$ vs $\text{SD}=0.6$ or 0.2). The uncertainty attached to the estimation of $\mu_{z_{pH}}$ and $\sigma_{z_{pH}}$ ($\mu_{z_{pH}} = 4.3 \pm 0.4$ and $\sigma_{z_{pH}} = 0.6 \pm 0.4$) is relatively higher than the uncertainty attached to the estimation of the other parameters (e.g. $\mu_{LDref} = 0.43 \pm 0.02$ and $\sigma_{LDref} = 0.28 \pm 0.02 \log_{10} \text{min}$), and $\sigma_{z_{pH}}$ is not significantly different from zero (Table 2.6). The densities of prior, empirical posterior and adjusted posterior distributions of the inactivation parameters show a large reduction of uncertainty for each parameter: posterior distributions are much more peaked than the fairly uninformative prior distributions (Figure 2.5). A significant correlation between some posterior parameters distributions was detected, such as the correlation between $\mu_{z_{pH}}$ and $\sigma_{z_{pH}}$ or between μ_{z_T} and σ_{z_T} . The corresponding Spearman correlation matrix is given in Table 2.7. However, no significant correlation between the intermediate variables D_{ref} , z_T and z_{pH} has been detected. These factors can therefore be considered as independent. The DIC is better than in the basic and intermediate models (see Table 2.4).

The cross-validation predictions of the data are really satisfactory (except for a few data as with the intermediate model) (Figure 2.4 c) and are better, although marginally, than with the intermediate model ($R^2 = 0.89$ and $\text{CRPS} = -0.09$, see Table 2.4).

Table 2.6. Descriptive statistics of empirical posterior distributions of the heat resistance parameters of *G. stearothermophilus* estimated with the complete model, and corresponding adjusted distributions. See Material and Methods section for details and corresponding adjusted distributions.

Type of variables	Node	Unit	Mean	SD	$Q_{0.025}^a$	$Q_{0.975}^a$	Adjusted distribution ^b or estimated value
Parameters	μ_{LDref}	$\log_{10} \text{min}$	0.43	0.02	0.38	0.48	$N(0.432, 0.024)$
	σ_{LDref}	$\log_{10} \text{min}$	0.28	0.02	0.24	0.32	$\text{Ln}N(-1.277, 0.073)$
	μ_{z_T}	$^\circ\text{C}$	9.1	0.3	8.5	9.9	$N(9.13, 0.34)$
	σ_{z_T}	$^\circ\text{C}$	1.9	0.4	1.3	2.6	$\text{Ln}N(0.60, 0.19)$
	$\mu_{z_{pH}}$	pH unit	4.3	0.4	3.8	5.3	$\text{Ln}N(1.45, 0.09)$
	$\sigma_{z_{pH}}$	pH unit	0.6	0.4	0.0	1.5	$G(1.99, 3.29)$
	σ	$\log_{10} \text{min}$	0.08	0.00	0.07	0.09	0.08
Intermediate variables	D_{ref}	min	3.3	2.4	0.8	9.6	$\text{Ln}N(1.00, 0.65)$
	z_T	$^\circ\text{C}$	9.1	1.9	5.4	13.1	$N(9.1, 1.9)$
	z_{pH}	pH unit	4.3	0.8	2.9	6.3	$N(4.3, 0.8)$

^a Q_x stands for the X^{th} quantile.

CHAPITRE 2. Méta-analyse

${}^bN(a,b)$ (resp. $\text{Ln}N(a,b)$) stands for the normal distribution (resp. the lognormal distribution based on a normal law) of mean a and standard deviation b . $G(a,b)$ stands for the gamma distribution with shape parameter a and rate parameter b .

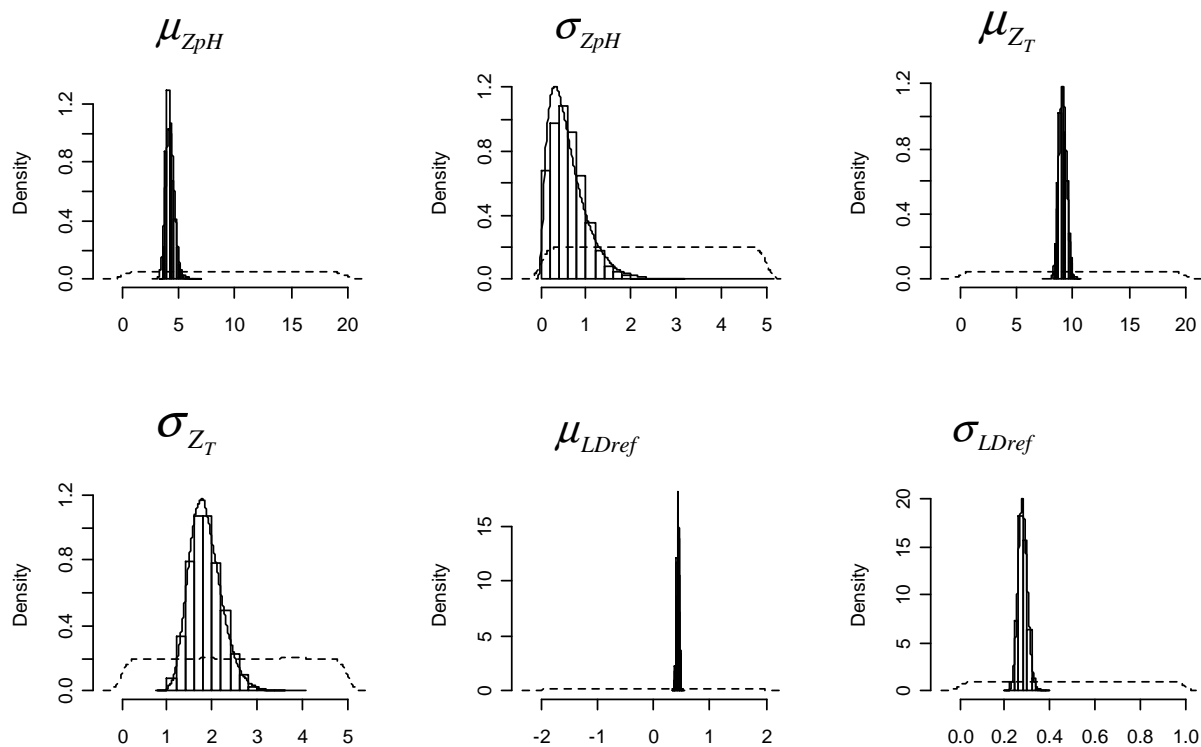


Figure 2.5. Prior distributions (dashed lines), empirical posterior distributions (histograms) and adjusted posterior distributions (solid lines) of μ_{ZpH} , σ_{ZpH} , μ_{ZT} , σ_{ZT} , μ_{LDref} and σ_{LDref} , obtained with the complete model.

Table 2.7. Spearman rank correlation between the parameters of the complete model for their posterior distributions.

Correlations	μ_{ZpH}	σ_{ZpH}	μ_{ZT}	σ_{ZT}	μ_{LDref}	σ_{LDref}
μ_{ZpH}	1	0.70	0.01	0.01	-0.11	-0.06
σ_{ZpH}	0.70	1	0.00	0.01	-0.07	-0.03
μ_{ZT}	0.01	0.00	1	0.62	0.36	-0.39
σ_{ZT}	0.01	0.01	0.62	1	0.33	-0.49
μ_{LDref}	-0.11	-0.07	0.36	0.33	1	-0.21
σ_{LDref}	-0.06	-0.03	-0.39	-0.49	-0.21	1

2.2.4 Discussion

The basic model (based on a simple linear regression) does not seem satisfactory because of a high residual error ($\sigma = 0.32 \log_{10} \text{min}$), an overestimation of z_T and credible intervals not representative of the diversity of estimations found in literature. The intermediate and complete models, which respectively assume D_{ref} variable, and D_{ref} , z_T and z_{pH} variable between studies, show much better adjustments to data with lower residual errors (respectively $\sigma = 0.10 \log_{10} \text{min}$ and $\sigma = 0.08 \log_{10} \text{min}$), lower DIC, and better cross-validation predictions (graphical comparison, higher R^2 and CRPS). The credible intervals of z_T and z_{pH} in the complete model are more representative of the possible values found in literature than the ones in the basic and intermediate models. For these reasons, the complete model seems slightly better for the determination of D_{ref} , z_T and z_{pH} than the intermediate model, and the basic model seems not acceptable. There is a strong necessity to account for random effects corresponding to between-study variability, in particular for D_{ref} , and in a reasonable way for z_T and z_{pH} with a more subtle influence on the final result. This means that the heat resistance or the influence of heating temperature or medium pH on the heat resistance cannot be described by a unique parameter, and may be strain-dependent for instance. However the inter-study variability of z_{pH} cannot be demonstrated from current data because $\sigma_{z_{pH}}$ is not significantly different from zero (Table 2.6). A model assuming the inter-variability on D_{ref} and z_T alone was also explored; it did not give significantly different estimations of D_{ref} and z_T but gave a slight difference in z_{pH} (4.0 pH units (SD=0.1) versus 4.3 pH units (SD=0.8)), with similar cross-validation predictions ($R^2=0.89$ and CRPS value=-0.09) (results not shown).

In the complete model, D_{ref} , z_T and z_{pH} were estimated at 3.3min (SD = 2.4 min), 9.1°C (SD = 1.9°C) and 4.3 pH unit (SD = 0.8) (Table 2.6). The high value of the upper bound of the 95% CI of z_T (13 min) is likely due to high z_T values of some studies. The data set was large, but rather unbalanced. Some test inactivation temperatures or pHs were overrepresented (126 D values out of 430 measured at 120°C-122°C and pH 6.7-7.3 for instance). The range of pH is only partially explored in many studies, and consequently the estimation of z_{pH} is uncertain and quite sensitive to the choice of priors. The use of D values in the present paper implicitly suggests the linearity of survival curves, although shoulders and tails were sometimes observed on the few available survival curves. However, it is by far the most common assumption in the food industry. Finally, for simplification and owing to the unavailability of experimental details on the whole data set, the uncertainty on each observed D -value has not been taken into account. If available, the best solution to account for this uncertainty is to use as data the inactivation curves instead of the D -values, and then use the primary and secondary models in the global model, as has been done for growth curves in Pouillot et al. 2003. Another solution can be to perform a weighted linear regression with weights inversely proportional to the precision of the data (Gonzales Barron et al., 2009). Standard deviations can be used as a measure of the precision but were not available on the whole data set in this work. Despite these inevitable limits, the interest of this work is to propose a robust estimation of reference inactivation parameters of *G. stearothersophilus* from an extensive set of data and sources.

These estimations may be used to predict the whole range of possible D -values (distinguishing the part of variation coming from uncertainty and from variability) in an aqueous medium, for a range of temperature between 100°C and 140°C, and pH between 3 and 7.5. To facilitate further use, parametric posterior distributions were adjusted on both parameters and

CHAPITRE 2. Méta-analyse

intermediate variables of the complete model (Table 2.6). Proposed distributions on D_{ref} , z_T and z_{pH} may be used directly in a one dimension Monte Carlo simulation for instance. If separation of uncertainty and variability is necessary, the proposed distributions on parameters μ_{zT} , $\mu_{z_{pH}}$, μ_{LDref} , σ_{zT} , $\sigma_{z_{pH}}$ and σ_{LDref} may be used in a second order Monte Carlo simulation for instance (Vose, 2000; Pouillot et al., 2007). Some correlations between the posterior uncertainty distributions of the parameters μ_{zT} , $\mu_{z_{pH}}$, μ_{LDref} , σ_{zT} , $\sigma_{z_{pH}}$ and σ_{LDref} were detected. If needed, correlated samples can be generated from the adjusted distributions and from the Spearman correlation matrix using the method of Iman and Conover (Iman and Conover, 1982), which is implemented in the R package mc2d (Pouillot et al., 2010). After generating a sample of possible values for D_{ref} , z_T and z_{pH} , D -values will be obtained using the secondary model (Eq. 3), leading to uncertainty and/or variability distributions on D (and to prediction distributions if adding an error $\epsilon \sim N(0, \sigma)$ in Eq.3).

Meta-analyses have been increasingly performed to estimate inactivation or growth parameters of micro-organisms of concern for food safety and quality. Den Besten and Zwietering (2012) discussed the relevance of meta-analyses to generate sufficiently generic parameters -with their variability- that can be used in further quantitative risk assessments. Van Asselt and Zwietering (2006) and Farakos and Zwietering (2011) determined inactivation parameters for a range of bacteria and influencing factors such as temperature and pressure using a global approach from a large range of products without integrating between-study variability. This global approach has the advantage to provide a first global parameter estimation, but Jaloustre et al. (2012) showed that not taking into account uncontrolled between-study variability in *Clostridium perfringens* heat resistance parameters may lead to high underestimation of temperature effect over D_{ref} , as also shown in the present work. A possible way to account for sources of variability is to define food groups including like products or like processes and to determine group-specific parameters, as proposed by Halder et al. (2010) in an attempt to balance the large variability coming from a global approach. With their method, variability between the food groups is taken into account, and sources of variability between the data of a food group are reduced, leading to more specific parameters estimations. Another way to account for sources of variability is to use multiple regression models, mixed linear models or hierarchical Bayesian models in order to model the sources of variability within a data set (which can be more or less specific to a combination of food, pathogen and extrinsic environments). Pouillot et al. (2003), Delignette-Muller et al. (2006) and later Crépet et al. (2009) thereby proposed to estimate uncertainty and variability in bacterial growth of *Listeria monocytogenes* respectively in milk, cold-smoked salmon and in fresh leafy salads using hierarchical Bayesian inference. Van Lieverloo et al. (2012) used a multiple regression model to assess the thermal inactivation of *Listeria monocytogenes* in various liquid food products. Jaloustre et al. (2012) used mixed effect models to estimate inactivation parameters of *C. perfringens* in beef, and described uncontrolled between-study variability on D_{ref} (but not on z_T). In the present work, we also worked with a specific data set (aqueous media), and used hierarchical Bayesian modeling to account for between-study variability on the three parameters D_{ref} , z_T and z_{pH} . The effects of these assumptions and the separate estimation of uncertainty and variability were relatively easily explored within the hierarchical Bayesian framework.

In this work, D -values were obtained from different types of heating media which can be classified into laboratory media or food-based substrates. An interesting question is the comparison of the D -values obtained from laboratory media and from food-based substrates. This question was explored with an additional model and inference, concluding that there was an absence of significant

CHAPITRE 2. Méta-analyse

difference in the heat resistance parameters of *G. stearothermophilus* in laboratory media and in food-based substrates (results not shown).

The definition of safe heat-processing relies on the prediction of inactivation of a target micro-organism. Assuming that (i) natural microbial contaminants behave in the same way as in laboratory environments and (ii) in the absence of current feasible alternatives (*in situ* estimation of the resistance of contaminating cells for instance), taking decisions based on scientific and grey literature is therefore the soundest option. This literature can be relatively extensive for some micro-organisms. Such abundance can be seen as a complication (what selection of data should be the most appropriate for this application?). In our opinion it is an opportunity: the diversity in the published data reflects uncertainty and variability in the behavior of natural contaminants. The aim is to find adequate methods accounting for different sources of uncertainty and variability to take advantage of this quantity of data. Meta-analyses also highlight the necessity to obtain new data, in particular at a distance from reference conditions, pH 7.0 and 121.1°C. The heat-resistance parameters of *G. stearothermophilus* determined in this work could find a large application in the canning industry (in particular for canned vegetables). The generic character of the method offers possibilities of extension to other micro-organisms of concern for the food industry.

Acknowledgments

This work is a partial fulfillment of author C. Rigaux' PhD Thesis and has been supported by *Agence National de la Recherche* Paris, France under contract ANR-09-ALIA-014 (Ribenu project). Thanks are due to Dr Olivier Couvert (LUBEM, Quimper, France) and to Stéphane André (CTCPA, Avignon, France) for kind communication of data and helpful scientific discussions.

CHAPITRE 2. Méta-analyse

Supplementary material

Table 2.7. Decimal reduction *D* times of *G. stearothermophilus* used in the present work. Sources, strains and heating media are given here. Other features defining a study are given in the original papers.

Reference	Study	Strain	Heating medium	pH	T (°C)	D (min)	Reference	Study	Strain	Heating medium	pH	T (°C)	D (min)	
Fernandez et al., 1995	1	ATCC12980	Water	7	115	22.4		14	ATCC12980	Natural mushroom	4.7	115	2.3	
	1	ATCC12980	Water	7	118	8.9		14	ATCC12980	Natural mushroom	4.7	118	1.3	
	1	ATCC12980	Water	7	121	3.1		14	ATCC12980	Natural mushroom	4.7	121	0.6	
	1	ATCC12980	Water	7	125	0.9		14	ATCC12980	Natural mushroom	4.7	125	0.2	
Lopez et al., 1996a	2	ATCC7953	Mcllvaine Buffer	7	115	10.2		15	ATCC12980	Natural mushroom	6.2	115	9.0	
	2	ATCC7953	Mcllvaine Buffer	7	120	2.1		15	ATCC12980	Natural mushroom	6.2	118	5.0	
	2	ATCC7953	Mcllvaine Buffer	7	125	0.6		15	ATCC12980	Natural mushroom	6.2	121	1.5	
	2	ATCC7953	Mcllvaine Buffer	7	130	0.1		15	ATCC12980	Natural mushroom	6.2	125	0.7	
	2	ATCC7953	Mcllvaine Buffer	7	135	0.0		15	ATCC12980	Natural mushroom	5.3	115	4.2	
	3	ATCC12980	Mcllvaine Buffer	7	115	11.4		15	ATCC12980	Natural mushroom	5.3	118	2.3	
	3	ATCC12980	Mcllvaine Buffer	7	120	2.3		15	ATCC12980	Natural mushroom	5.3	121	0.9	
	3	ATCC12980	Mcllvaine Buffer	7	125	0.6		15	ATCC12980	Natural mushroom	5.3	125	0.5	
	3	ATCC12980	Mcllvaine Buffer	7	130	0.1		15	ATCC12980	Natural mushroom	4.7	115	2.3	
	3	ATCC12980	Mcllvaine Buffer	7	135	0.0		15	ATCC12980	Natural mushroom	4.7	118	1.3	
	4	ATCC15951	Mcllvaine Buffer	7	115	8.9		15	ATCC12980	Natural mushroom	4.7	121	0.6	
	4	ATCC15951	Mcllvaine Buffer	7	120	2.5		15	ATCC12980	Natural mushroom	4.7	125	0.2	
	4	ATCC15951	Mcllvaine Buffer	7	125	0.6		André et al., 2009	16	CTCPA2804138	Green bean	5.5	115	7.5
	4	ATCC15951	Mcllvaine Buffer	7	130	0.1			16	CTCPA2804138	Green bean	5.5	118	3.0
	4	ATCC15951	Mcllvaine Buffer	7	135	0.0			16	CTCPA2804138	Green bean	5.5	121	1.4
	5	ATCC15952	Mcllvaine Buffer	7	115	9.1			17	CTCPA2804138	Green bean	5.5	115	10.0
5	ATCC15952	Mcllvaine Buffer	7	120	1.6	17	CTCPA2804138		Green bean	5.5	118	5.2		
5	ATCC15952	Mcllvaine Buffer	7	125	0.4	17	CTCPA2804138		Green bean	5.5	121	1.7		
5	ATCC15952	Mcllvaine Buffer	7	130	0.1	18	CTCPA2804138		Green bean	5.5	115	7.9		
5	ATCC15952	Mcllvaine Buffer	7	135	0.0	18	CTCPA2804138		Green bean	5.5	118	2.8		
Le Jean et al., 1994	6	ATCC7553	Water	7	111	45.2	18		CTCPA2804138	Green bean	5.5	121	1.8	
	6	ATCC7553	Water	7	115	18.7	19		CTCPA2804138	Green bean	5.5	115	7.0	
	6	ATCC7553	Water	7	117	18.0	19		CTCPA2804138	Green bean	5.5	118	3.7	
	6	ATCC7553	Water	7	119	9.0	19		CTCPA2804138	Green bean	5.5	121	1.4	
	6	ATCC7553	Water	7	121	5.7	20		CTCPA2804138	Green bean	5.5	115	10.9	
	6	ATCC7553	Water	7	123	2.9	20		CTCPA2804138	Green bean	5.5	118	3.6	
	6	ATCC7553	Water	7	125	1.2	20		CTCPA2804138	Green bean	5.5	121	1.4	
	6	ATCC7553	Water	7	127	0.6	21		CTCPA2804138	Green bean	5.5	115	6.1	
	Haas et al., 1996	7	Unspecified	Mcllvaine Buffer	5	115	4.4	21	CTCPA2804138	Green bean	5.5	118	3.3	
		7	Unspecified	Mcllvaine Buffer	5	121	0.7	21	CTCPA2804138	Green bean	5.5	121	1.3	
7		Unspecified	Mcllvaine Buffer	5	125	0.4	22	CTCPA2804138	Green bean	5.5	115	6.3		
8		Unspecified	Mcllvaine Buffer	7	121	1.8	22	CTCPA2804138	Green bean	5.5	118	2.8		
8		Unspecified	Mcllvaine Buffer	7	125	0.2	22	CTCPA2804138	Green bean	5.5	121	1.7		
8		Unspecified	Mcllvaine Buffer	7	130	0.1	23	CTCPA2804138	Green bean	5.5	115	8.5		
Rodrigo et al., 1999		9	ATCC12980	Food	5.3	115	8.7	23	CTCPA2804138	Green bean	5.5	118	4.9	
		9	ATCC12980	Food	5.3	118	3.5	23	CTCPA2804138	Green bean	5.5	121	2.6	
	9	ATCC12980	Food	5.3	121	1.4	24	CTCPA2804138	Green bean	5.5	115	8.8		
	9	ATCC12980	Food	5.3	125	0.4	24	CTCPA2804138	Green bean	5.5	118	4.3		
	9	ATCC12980	Food + vinegar	4.8	115	6.5	24	CTCPA2804138	Green bean	5.5	121	3.0		
	9	ATCC12980	Food + vinegar	4.8	118	2.9	25	CTCPA2804138	Green bean	5.5	115	7.8		
	9	ATCC12980	Food + vinegar	4.8	121	1.3	25	CTCPA2804138	Green bean	5.5	118	3.5		
	9	ATCC12980	Food + vinegar	4.8	125	0.4	25	CTCPA2804138	Green bean	5.5	121	2.3		
	Ocio et al., 1996	10	ATCC12980	Water	7	115	23.3	26	CTCPA2804138	Green bean	5.5	115	8.1	
10		ATCC12980	Water	7	118	6.0	26	CTCPA2804138	Green bean	5.5	118	5.9		
10		ATCC12980	Water	7	121	3.1	26	CTCPA2804138	Green bean	5.5	121	2.9		
10		ATCC12980	Water	7	125	0.9	27	CTCPA2804138	Green bean	5.5	115	11.7		
10		ATCC12980	Water	7	130	0.2	27	CTCPA2804138	Green bean	5.5	118	5.1		
11		ATCC12980	Mushroom purée	6.5	115	24.8	27	CTCPA2804138	Green bean	5.5	121	1.8		
11		ATCC12980	Mushroom purée	6.5	118	6.6	28	CTCPA2804138	Green bean	5.5	115	12.8		
11		ATCC12980	Mushroom purée	6.5	121	2.8	28	CTCPA2804138	Green bean	5.5	118	4.2		
11		ATCC12980	Mushroom purée	6.5	125	1.0	28	CTCPA2804138	Green bean	5.5	121	2.0		
11		ATCC12980	Mushroom purée	6.5	130	0.3	29	CTCPA2804138	Green bean	5.5	115	10.1		
Fernandez et al., 1994		12	ATCC12980	Water	7	115	23.4	29	CTCPA2804138	Green bean	5.5	118	4.7	
	12	ATCC12980	Water	7	118	9.0	29	CTCPA2804138	Green bean	5.5	121	2.1		
	12	ATCC12980	Water	7	121	3.1	30	CTCPA2804138	Green bean	5.5	115	11.7		
	12	ATCC12980	Water	7	125	0.9	30	CTCPA2804138	Green bean	5.5	118	6.1		
	13	ATCC12980	Natural mushroom	6.7	115	13.1	30	CTCPA2804138	Green bean	5.5	121	2.3		
	13	ATCC12980	Natural mushroom	6.7	118	5.2	Wescott et al., 1995	31	ATCC12980	Phosphate buffer	6.8	128.5	0.4	
	13	ATCC12980	Natural mushroom	6.7	121	1.8		31	ATCC12980	Phosphate buffer	6.8	130	0.3	
	14	ATCC12980	Natural mushroom	6.2	115	11.5		31	ATCC12980	Phosphate buffer	6.8	133	0.1	
	14	ATCC12980	Natural mushroom	6.2	118	5.1		31	ATCC12980	Phosphate buffer	6.8	135	0.1	
	Feeherry et al., 1987	14	ATCC12980	Natural mushroom	6.2	121	1.7		32	ATCC12980	Water	7	112.8	62.0
		14	ATCC12980	Natural mushroom	6.2	125	0.6		32	ATCC12980	Water	7	115.6	18.0
		14	ATCC12980	Natural mushroom	6.2	125	0.6		32	ATCC12980	Water	7	118.3	8.0
		14	ATCC12980	Natural mushroom	5.3	115	4.6		32	ATCC12980	Water	7	121.1	3.3
		14	ATCC12980	Natural mushroom	5.3	118	2.4		32	ATCC12980	Water	7	123.9	1.1
14		ATCC12980	Natural mushroom	5.3	121	1.1	Periago et al., 1998a		33	ATCC12980	Water	7	115	38.6
14		ATCC12980	Natural mushroom	5.3	125	0.4			33	ATCC12980	Water	7	118	15.3

CHAPITRE 2. Méta-analyse

Reference	Study	Strain	Heating medium	pH	T (°C)	D (min)	Reference	Study	Strain	Heating medium	pH	T (°C)	D (min)
	33	ATCC12980	Water	7	121	7.4		65	ATCC12980	Mcllvaine Buffer	7	120	3.0
	33	ATCC12980	Water	7	123	3.5		66	ATCC12980	Mcllvaine Buffer	7	120	4.7
	33	ATCC12980	Water	7	125	1.3		67	ATCC12980	Water	7	120	5.3
	34	ATCC12980	Water	7	115	33.3		68	ATCC15952	Mcllvaine Buffer	7	120	2.3
	34	ATCC12980	Water	7	118	11.0		69	ATCC15952	Mcllvaine Buffer	7	120	4.0
	34	ATCC12980	Water	7	121	4.2		70	ATCC15952	Mcllvaine Buffer	7	120	4.6
	34	ATCC12980	Water	7	123	2.5		71	ATCC15952	Mcllvaine Buffer	7	120	4.1
	34	ATCC12980	Water	7	125	0.7		72	ATCC15952	Water	7	120	9.4
Lopez et al., 1998	35	ATCC12980	Mcllvaine Buffer	6.5	120	2.3		73	ATCC7953	Water	7	120	8.3
	35	ATCC12980	Mcllvaine Buffer	6	120	2.1		74	ATCC15951	Water	7	120	8.9
	35	ATCC12980	Mcllvaine Buffer	5	120	1.0		75	ATCC12980	Water + NaCl	7	120	4.7
	36	ATCC7953	Mcllvaine Buffer	6.5	120	2.1		76	ATCC12980	Water + NaCl	7	120	3.0
	36	ATCC7953	Mcllvaine Buffer	6	120	1.9		77	ATCC15951	Water + NaCl	7	120	5.3
	36	ATCC7953	Mcllvaine Buffer	5	120	1.2		78	ATCC15951	Water + NaCl	7	120	3.6
	37	ATCC15951	Mcllvaine Buffer	6.5	120	2.2	Davies et al., 1977	79	NCDO1096	Water	7	120	16.7
	37	ATCC15951	Mcllvaine Buffer	6	120	2.2		79	NCDO1096	Water	7	122.5	7.6
	37	ATCC15951	Mcllvaine Buffer	5	120	1.0		79	NCDO1096	Water	7	125	3.6
	38	ATCC15952	Mcllvaine Buffer	6.5	120	1.6		79	NCDO1096	Water	7	127.5	1.6
	38	ATCC15952	Mcllvaine Buffer	6	120	1.5		79	NCDO1096	Water	7	130	0.7
	38	ATCC15952	Mcllvaine Buffer	5	120	0.9		79	NCDO1096	Water	7	132.5	0.4
Periago et al., 1998c	39	ATCC12980	Water	7	115	21.4		79	NCDO1096	Water	7	135	0.2
	39	ATCC12980	Water	7	118	7.1		79	NCDO1096	Water	7	137.5	0.1
	39	ATCC12980	Water	7	121	2.4		79	NCDO1096	Water	7	140	0.0
	39	ATCC12980	Water	7	125	0.6	Gauthier et al., 1978	80	ATCC7953	Water	7	121	3.1
Tejedor et al., 2001	40	ATCC12980	Food substrate	5	115	4.9		81	ATCC7953	Phosphate buffer	7.2	121	5.7
	40	ATCC12980	Food substrate	5	118	2.4		81	ATCC7953	Phosphate buffer	7.2	121	5.1
	40	ATCC12980	Food substrate	5	121	1.2		82	ATCC7953	Phosphate buffer	7.2	121	4.3
	40	ATCC12980	Food substrate	5	125	0.4		82	ATCC7953	Phosphate buffer	7.2	121	4.4
	40	ATCC12980	Food substrate	5.2	115	4.5		83	ATCC7953	Phosphate buffer	7.2	121	5.7
	40	ATCC12980	Food substrate	5.2	118	2.3		83	ATCC7953	Phosphate buffer	7.2	121	5.1
	40	ATCC12980	Food substrate	5.2	121	1.2		84	ATCC7953	Phosphate buffer	7.2	121	5.2
	40	ATCC12980	Food substrate	5.2	125	0.5		84	ATCC7953	Phosphate buffer	7.2	121	5.0
	40	ATCC12980	Food substrate	5.4	115	5.6		85	ATCC7953	Phosphate buffer	7.2	121	4.9
	40	ATCC12980	Food substrate	5.4	118	2.9		85	ATCC7953	Phosphate buffer	7.2	121	5.1
	40	ATCC12980	Food substrate	5.4	121	1.5		86	ATCC7953	Phosphate buffer	7.2	121	5.5
	40	ATCC12980	Food substrate	5.4	125	0.6		86	ATCC7953	Phosphate buffer	7.2	121	5.0
	40	ATCC12980	Food substrate	5.6	115	8.6		87	ATCC7953	Phosphate buffer	7.2	121	4.9
	40	ATCC12980	Food substrate	5.6	118	3.9		87	ATCC7953	Phosphate buffer	7.2	121	5.3
	40	ATCC12980	Food substrate	5.6	121	1.8		88	ATCC7953	Phosphate buffer	7.2	121	5.5
	40	ATCC12980	Food substrate	5.6	125	0.6		88	ATCC7953	Phosphate buffer	7.2	121	5.4
	40	ATCC12980	Food substrate	5.8	115	9.7		89	ATCC7953	Phosphate buffer	7.2	121	5.2
	40	ATCC12980	Food substrate	5.8	118	4.3		89	ATCC7953	Phosphate buffer	7.2	121	4.7
	40	ATCC12980	Food substrate	5.8	121	1.9		90	ATCC7953	Water	7	121	5.7
	40	ATCC12980	Food substrate	5.8	125	0.7		90	ATCC7953	Water	7	121	5.2
Rodrigo et al., 1997	41	ATCC12980	Water	7	121	2.8		91	ATCC7953	Ringer solution	7.2	121	6.0
	41	ATCC12980	Water	7	130	0.3		91	ATCC7953	Ringer solution	7.2	121	5.4
	42	ATCC12980	Natural mushroom	6.7	121	1.6		92	ATCC7953	Ringer solution	7.2	121	4.2
	42	ATCC12980	Natural mushroom	6.7	130	0.2		92	ATCC7953	Ringer solution	7.2	121	3.9
Lopez et al., 1997	43	ATCC12980	Mcllvaine Buffer	7	120	2.4		93	ATCC7953	Ringer solution	7.2	121	3.2
	44	ATCC12980	Mcllvaine Buffer	7	120	2.1		93	ATCC7953	Ringer solution	7.2	121	2.7
	45	ATCC12980	Mcllvaine Buffer	7	120	2.4		94	ATCC7953	Ringer solution	7.2	121	3.9
	46	ATCC12980	Mcllvaine Buffer	7	120	1.9		94	ATCC7953	Ringer solution	7.2	121	4.1
	47	ATCC7953	Mcllvaine Buffer	7	120	2.5		95	ATCC7953	Ringer solution	7.2	121	5.0
	48	ATCC7953	Mcllvaine Buffer	7	120	2.2		95	ATCC7953	Ringer solution	7.2	121	4.9
	49	ATCC7953	Mcllvaine Buffer	7	120	2.8		96	ATCC7953	Ringer solution	7.2	121	5.5
	50	ATCC7953	Mcllvaine Buffer	7	120	2.7		96	ATCC7953	Ringer solution	7.2	121	5.1
	51	ATCC15951	Mcllvaine Buffer	7	120	2.2		97	ATCC7953	Ringer solution	7.2	121	5.4
	52	ATCC15951	Mcllvaine Buffer	7	120	2.2		97	ATCC7953	Ringer solution	7.2	121	4.7
	53	ATCC15951	Mcllvaine Buffer	7	120	2.5		98	ATCC7953	Ringer solution	7.2	121	5.4
	54	ATCC15951	Mcllvaine Buffer	7	120	2.4		98	ATCC7953	Ringer solution	7.2	121	4.9
	55	ATCC15952	Mcllvaine Buffer	7	120	1.6		99	ATCC7953	Ringer solution	7.2	121	5.6
	56	ATCC15952	Mcllvaine Buffer	7	120	1.3		99	ATCC7953	Ringer solution	7.2	121	4.4
	57	ATCC15952	Mcllvaine Buffer	7	120	1.7	Etoa et al., 1988	100	CNCH5781	Water	7	121	3.5
	58	ATCC15952	Mcllvaine Buffer	7	120	1.7		101	CNCH5781	Water	7	121	3.5
Periago et al., 1998b	59	ATCC12980	Asparagus	6.1	115	6.6		102	CNCH5781	Water	7	121	4.0
	59	ATCC12980	Asparagus	6.1	120	1.2		103	CNCH5781	Water	7	121	3.6
	60	ATCC12980	Bean	6	116	5.5		104	CNCH5781	Water	7	121	3.6
	60	ATCC12980	Bean	6	121	0.5		105	CNCH5781	Nutrient broth	7.5	121	4.3
Lopez et al., 1996b	61	ATCC12980	Sorensen buffer	7	120	2.2		106	CNCH5781	Nutrient broth	7.5	121	6.0
	62	ATCC15952	Sorensen buffer	7	120	2.4		107	CNCH5781	Nutrient broth	7.5	121	6.5
	63	ATCC12980	Mcllvaine Buffer	7	120	2.2		108	CNCH5781	Nutrient broth	7.5	121	5.0
	64	ATCC12980	Mcllvaine Buffer	7	120	2.0		109	CNCH5781	Nutrient broth	7.5	121	5.9

CHAPITRE 2. Méta-analyse

Reference	Study	Strain	Heating medium	pH	T (°C)	D (min)	Reference	Study	Strain	Heating medium	pH	T (°C)	D (min)			
Cook and Gilbert, 1968a	110	NCIB8919	Water	7	115	24.4	150	A36	Phosphate buffer	7.0	122	0.3				
	111	NCIB8919	Water	7	115	20.5				7.0	125	0.1				
	112	NCIB8919	Water	7	115	22.3				3.0	113	0.5				
	113	NCIB8919	Water	7	115	8.5				3.5	113	0.5				
	114	NCIB8919	Water	7	115	22.9				4.0	113	1.1				
	115	NCIB8919	Water	7	115	18.9				4.5	113	0.8				
	116	NCIB8919	Water	7	115	21.2				5.0	113	1.6				
	116	NCIB8919	Water	7	115	20.4				5.5	113	1.7				
	116	NCIB8919	Water	7	115	20.5				6.0	113	3.8				
	116	NCIB8919	Water	7	115	16.4				6.5	113	3.8				
	116	NCIB8919	Water	7	115	18.7				7.0	113	4.3				
	Cook and Gilbert, 1968b	117	NCIB8919	Water	7	115				16.6	150	A36	Phosphate buffer	7.5	113	7.0
		118	NCIB8919	Water	7	115				16.7	151	A36	Phosphate buffer	7	107	11.6
		119	NCIB8919	Water	7	115				16.6	152	A36	Phosphate buffer	7	107	3.2
		120	NCIB8919	Water	7	115				17.4	153	A36	Phosphate buffer	7	107	23.2
		121	NCIB8919	Water	7	115				17.9	154	A36	Phosphate buffer	7	107	32.8
122		NCIB8919	Water	7	115	16.6	155	A36	Phosphate buffer	7	107	19.1				
123		NCIB8919	Water	7	115	17.4	156	A36	Phosphate buffer	7	107	28.7				
124		NCIB8919	Water	7	115	12.7	157	A36	Phosphate buffer	7	107	32.8				
125		NCIB8919	Water	7	115	16.5	158	A36	Phosphate buffer	7	107	7.7				
126		NCIB8919	Mcllvaine buffer	5	115	4.2	159	A36	Phosphate buffer	7	107	4.4				
Wallace et al., 1978	126	NCIB8919	Mcllvaine buffer	6	115	7.4	160	A36	Phosphate buffer	7	107	31.1				
	126	NCIB8919	Mcllvaine buffer	7	115	9.2	161	ATCC12980	Phosphate buffer	7	104	90.7				
	127	NCIB8919	Sorensen buffer	6	115	10.8	161	ATCC12980	Phosphate buffer	7	107	75.5				
	127	NCIB8919	Sorensen buffer	7	115	11.3	161	ATCC12980	Phosphate buffer	7	110	19.6				
	127	NCIB8919	Sorensen buffer	7	115	11.2	161	ATCC12980	Phosphate buffer	7	113	10.3				
	128	1518	Sorensen buffer	7	110	64.9	161	ATCC12980	Phosphate buffer	7	116	3.2				
	128	1518	Sorensen buffer	7	116	14.7	161	ATCC12980	Phosphate buffer	7	119	1.9				
	128	1518	Sorensen buffer	7	121	1.3	161	ATCC12980	Phosphate buffer	7	122	1.0				
	Anderson et al., 1974	129	NCTC10003	Acetate buffer	4	105	100.0	161	ATCC12980	Phosphate buffer	7	125	0.3			
		129	NCTC10003	Acetate buffer	4	110	28.0	161	ATCC12980	Phosphate buffer	6.7	113	9.0			
		129	NCTC10003	Acetate buffer	4	115	6.8	161	ATCC12980	Phosphate buffer	5.7	113	3.5			
		129	NCTC10003	Acetate buffer	4	120	3.3	161	ATCC12980	Phosphate buffer	4.5	113	2.3			
		129	NCTC10003	Acetate buffer	5	105	290.0	161	ATCC12980	Phosphate buffer	5.0	113	3.2			
		129	NCTC10003	Acetate buffer	5	110	91.0	161	ATCC12980	Phosphate buffer	6.1	113	5.4			
		129	NCTC10003	Acetate buffer	5	115	20.0	161	ATCC12980	Phosphate buffer	6.9	113	10.3			
		129	NCTC10003	Acetate buffer	5	120	4.9	162	ATCC12980	Phosphate buffer	7	113	2.3			
129		NCTC10003	Acetate buffer	6	105	440.0	163	ATCC12980	Phosphate buffer	7	113	3.1				
129		NCTC10003	Acetate buffer	6	110	140.0	164	ATCC12980	Phosphate buffer	7	113	3.1				
129		NCTC10003	Acetate buffer	6	115	29.0	165	ATCC12980	Phosphate buffer	7	113	5.1				
129		NCTC10003	Acetate buffer	6	120	7.0	166	ATCC12980	Phosphate buffer	7	113	4.9				
129		NCTC10003	Acetate buffer	7	110	170.0	167	ATCC12980	Phosphate buffer	7	113	6.8				
129		NCTC10003	Acetate buffer	7	115	38.0	168	ATCC12980	Phosphate buffer	7	113	7.4				
130		NCTC10003	Tris buffer	7	105	370.0	169	ATCC12980	Phosphate buffer	7	113	8.3				
Mallidis et Scholefield, 1986		130	NCTC10003	Tris buffer	7	110	72.0	170	ATCC12980	Phosphate buffer	7	113	10.3			
	130	NCTC10003	Tris buffer	7	115	16.0	171	ATCC12980	Phosphate buffer	7	113	14.6				
	130	NCTC10003	Tris buffer	7	120	5.4	172	2804138	Phosphate buffer	7	110	36.0				
	131	NCIB8923	Unspecified	7	120	4.1	172	2804138	Phosphate buffer	7	115	9.7				
	132	NCIB8923	Unspecified	7	120	5.4	172	2804138	Phosphate buffer	7	117.5	5.6				
	133	NCIB8923	Unspecified	7	120	6.4	172	2804138	Phosphate buffer	7	120	3.1				
	134	NCIB8923	Unspecified	7	120	5.3	172	2804138	Phosphate buffer	7	122	1.2				
	135	NCIB8923	Unspecified	7	120	6.3	173	2804 173	Green bean	5.2	120	1.3				
	136	NCIB8923	Unspecified	7	120	7.2	174	2804 173	Phosphate buffer	7	110	18.2				
	137	NCIB8923	Unspecified	7	120	5.4	174	2804 173	Phosphate buffer	7	115	5.6				
	138	NCIB8923	Unspecified	7	120	6.8	174	2804 173	Phosphate buffer	7	117.5	2.5				
	139	NCIB8923	Unspecified	7	120	5.3	174	2804 173	Phosphate buffer	7	120	1.7				
	140	NCIB8923	Unspecified	7	120	6.1	174	2804 173	Phosphate buffer	7	122	1.1				
	141	NCIB8923	Unspecified	7	120	6.0	175	2804 138	Green bean	5.2	120	2.3				
	142	NCIB8923	Unspecified	7	120	7.3	176	2804 138	Phosphate buffer	7	110	36.0				
	143	NCIB8923	Unspecified	7	120	6.7	176	2804 138	Phosphate buffer	7	115	9.7				
144	NCIB8923	Unspecified	7	120	7.2	176	2804 138	Phosphate buffer	7	117.5	4.4					
145	NCIB8923	Unspecified	7	120	6.2	176	2804 138	Phosphate buffer	7	120	2.8					
146	NCIB8923	Unspecified	7	120	5.4	176	2804 138	Phosphate buffer	7	122	1.2					
147	NCIB8923	Unspecified	7	120	5.1	177	2 804 168	Green bean	5.2	120	3.3					
148	NCIB8923	Unspecified	7	120	7.3	178	2 804 168	Phosphate buffer	7	112.5	22.8					
149	NCIB8923	Unspecified	7	120	7.8	178	2 804 168	Phosphate buffer	7	115	13.8					
Couvert, 2002	150	A36	Phosphate buffer	7	104	64.9	178	2 804 168	Phosphate buffer	7	117.5	5.9				
	150	A36	Phosphate buffer	7	107	28.7	178	2 804 168	Phosphate buffer	7	120	2.5				
	150	A36	Phosphate buffer	7	110	10.1	178	2 804 168	Phosphate buffer	7	122	1.2				
	150	A36	Phosphate buffer	7	113	5.2	Patazca et al., 2006	ATCC10149	Deminerlized water	7.5	121	5.5				
	150	A36	Phosphate buffer	7	116	2.2										
	150	A36	Phosphate buffer	7	119	0.8										

2.3 Compléments

2.3.1 Variations des estimations entre les études individuelles

Préalablement à la mise en place de la méta-analyse, les paramètres de résistance thermique de *G. stearothermophilus* avaient été estimés sur des études prises en compte individuellement et comportant un nombre de points expérimentaux suffisamment élevé pour permettre une estimation « assez précise » des paramètres de chacune des études. La comparaison des résultats obtenus entre ces diverses études permettait une première estimation de la variabilité possible de ces paramètres, à prendre éventuellement en compte dans la méta-analyse.

De manière classique en microbiologie, les estimations individuelles de ces paramètres sont obtenues à partir de régressions linéaires basées sur l'extension du modèle secondaire de Bigelow (partie 2.2.2.a, équation 2.3). Les intervalles de confiance à 95% sur z_T et z_{pH} sont obtenus par transformation des intervalles de confiance fréquentistes usuels obtenus sur les coefficients de la régression $-1/z_T$ et $1/z_{pH}$. Les intervalles de confiance à 95% sur D_{ref} ont été obtenus par transformation de l'intervalle de confiance des prévisions de $\log_{10}D$ pour $T=121.1^\circ\text{C}$ et $\text{pH}=7$. Les estimations des paramètres ont été réalisées sur les études avec un minimum de 3 ou 4 températures différentes, ou 3 ou 4 pH différents. Comme dans de nombreuses études la gamme des pH testés est nulle (un seul pH testé) ou étroite (inférieure à 0.5 unité pH), le nombre d'estimations individuelles de z_{pH} est faible. Les estimations dont les intervalles de confiances étaient démesurément larges (par exemple du type $] -\infty; +\infty[$) n'ont pas été représentées.

Les diverses estimations présentées en Figure 2.6 montrent la grande variabilité des estimations de D_{ref} sur les diverses études (estimations entre 0.4 et 11.3 min) et la variabilité des estimations de z_T (estimations entre 6,3 et 9,5 °C), visibles en particulier par le non-chevauchement de certains intervalles de confiance à 95%. Les différences entre les estimations de z_{pH} ne sont pas significatives d'après le chevauchement des intervalles de confiance, mais seulement 6 estimations assez précises ont été possibles. Ces résultats mettent en évidence les variations entre les estimations individuelles des paramètres entre les études (au moins pour D_{ref} et z_T), ce qui justifie la prise en compte dans la méta-analyse des sources de variabilité possibles entre les études.

Les estimations individuelles des paramètres pour chaque étude peuvent par ailleurs être retrouvés dans le cadre de la méta-analyse à partir du modèle complet ($\log_{10}D_{ref}^j$, z_T^j et z_{pH}^j estimés pour l'étude j , Figure 2.2). La précision de ces estimations est améliorée grâce à la plus grande quantité d'information disponible.

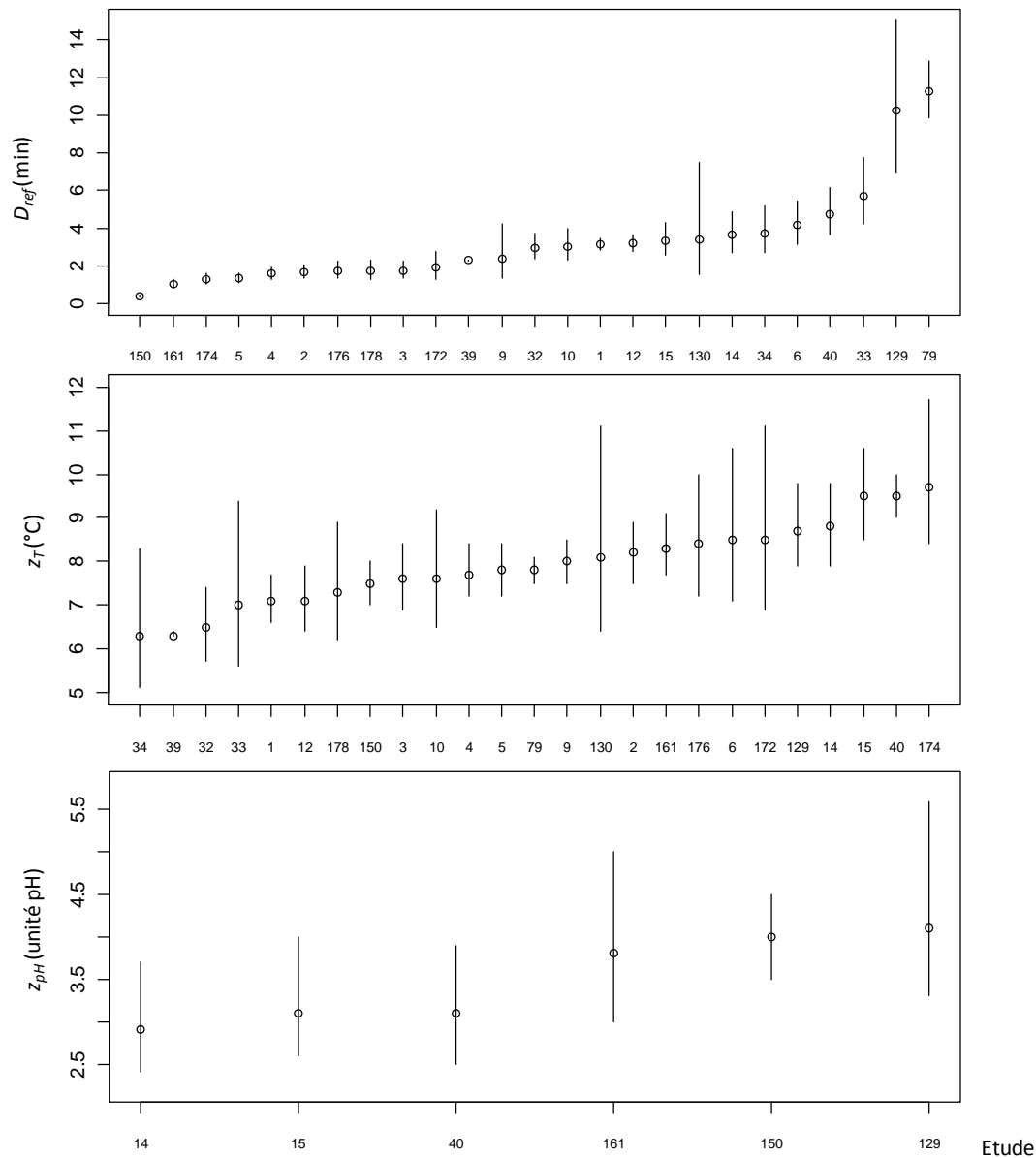


Figure 2.6. Estimations individuelles des paramètres de résistance thermique de *G. stearothermophilus* sur les études de taille suffisamment grandes. Les points représentent les estimations et les barres les intervalles de confiance à 95%.

2.3.2 Différents types d'ajustements ou de prévisions

Cette partie présente des travaux entrepris suite à des questions de *reviewers* lors de la soumission de l'article à la revue IJFM, et concerne des aspects techniques sur les différents types d'ajustements ou de prévisions à utiliser lors de la comparaison de modèles.

Ajustements

Il peut être intéressant de comparer la capacité des différents modèles à décrire les données observées, en comparant les graphes de qualité d'ajustements des modèles (Jaloustre et al., 2011). Chaque ajustement a été représenté par la médiane et par les intervalles de crédibilité à 95% (issus des quantiles d'ordre 0.025 et 0.975) des distributions empiriques a posteriori des variables $\log_{10}D_i$ ou $\log_{10}D_i^j$ représentées dans les DAG des modèles (Figure 2.2). La comparaison des ajustements et des données observées pour chaque modèle est présentée en Figure 2.7, ainsi que le critère CRPS en Tableau 2.8. La comparaison des graphes et du CRPS montre que les ajustements sont nettement meilleurs pour les modèles intermédiaire (CRPS=-0.04) et complet (CRPS=-0.03) que pour le modèle basique (CRPS=-0.23), et légèrement meilleurs pour le modèle complet que pour le modèle intermédiaire. Les données observées sont en effet rarement contenues dans les intervalles de crédibilité à 95% des ajustements pour le modèle basique, alors qu'elles le sont beaucoup plus pour les modèles intermédiaires et complets. Il est normal que la qualité de l'ajustement augmente avec la complexité du modèle et le nombre de paramètres, cependant ces comparaisons montrent que la structure des données est beaucoup mieux décrite en supposant D_{ref} variable entre les études, alors que le gain de qualité d'ajustement est moindre quand on suppose z_T et z_{pH} variables entre les études.

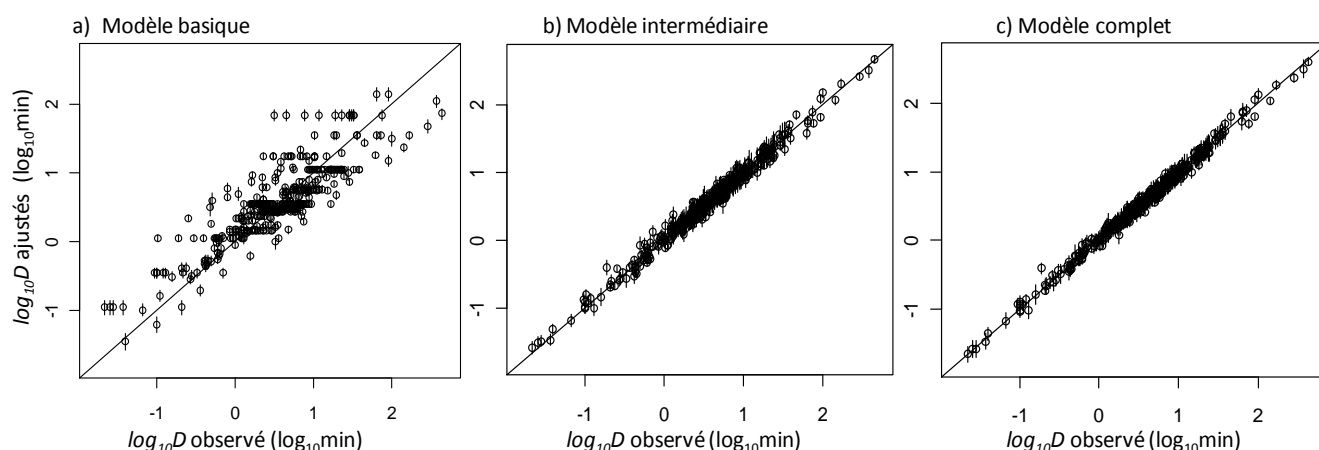


Figure 2.7. Graphes de comparaison entre les valeurs observées de $\log_{10}D$ et leurs ajustements (médiane et intervalles de crédibilité à 95%), pour les modèles basique (a), intermédiaire (b) et complet (c).

Prévisions de données d'étude connue

Dans l'article, les prévisions présentées en Figure 2.4 ont été obtenues à partir des distributions a posteriori des variables $\log_{10}D_{i,pred}^j \sim N(\log_{10}D_i^j, \sigma)$, par validation croisée *10-fold* (c'est-à-dire que la prévision de chaque donnée a été faite à partir de lois a posteriori sur les paramètres des modèles estimées à partir de 90% des autres données – voir partie 2.2.2.d). Les études des données à prédire étaient supposées connues, c'est-à-dire que les valeurs de la variable $\log_{10}D_{ref}^j$ (respectivement des variables $\log_{10}D_{ref}^j$, z_T^j et z_{pH}^j) estimée(s) sur chaque étude j dans le modèle intermédiaire (respectivement dans le modèle complet) ont été utilisées pour les prévisions. La validation croisée a été utilisée pour éviter le sur-ajustement. Comme l'atteste la comparaison graphique et la comparaison des critères CRPS (Figure 2.4 et Tableau 2.4), la qualité des prévisions est globalement bien meilleure dans le cas des modèles intermédiaires et complets (CRPS=-0.10 et -0.09) que dans le cas du modèle basique (CRPS=-0.18), et légèrement meilleure pour le modèle complet que pour le modèle intermédiaire.

Prévisions de données d'étude inconnue (« prédictives »)

La prévision d'une nouvelle valeur de D , à une température et un pH donnés mais pour une souche inconnue, un milieu de traitement aqueux inconnu, etc., est aussi intéressante. Ce type de prévision est parfois appelé « prédictive » ; ce sont les valeurs potentielles pour un nouvel échantillon Y' compte tenu du même modèle et de données anciennes (x,y) (x : variables connues explicatives, y : données correspondantes). Ici, ces prévisions sont obtenues à partir des distributions a posteriori des variables $\log_{10}D_{i,pred}^j \sim N(\log_{10}D_i^j, \sigma)$, où de nouvelles valeurs des variables $\log_{10}D_{ref}^j$, z_T^j et z_{pH}^j ont été simulées pour chaque étude à partir des lois a posteriori des hyperparamètres dans le cas des modèles intermédiaires et complets (par exemple $\log_{10}D_{ref}^j \sim N(\mu_{LDref}, \sigma_{LDref})$). Dans le cas du modèle basique, ce sont les mêmes prévisions que ci-dessus si la validation croisée est utilisée. Ici, les prévisions ont été réalisées par validation croisée et sans validation croisée, et les résultats étant quasiment identiques, seuls ceux obtenus sans validation croisée (prédictives) sont présentés ci-dessous. La comparaison graphique des prévisions (Figure 2.8) et du CRPS (Tableau 2.8) montre assez peu de différences entre les trois types de modèles (CRPS = -0.18, -0.19 ou -0.17). 94.4% des données sont comprises dans les intervalles de crédibilité à 95% dans le cas du modèle basique, 95.8% dans le cas du modèle intermédiaire et 95.4% dans le cas du modèle complet, ce qui montre que les prévisions obtenues avec les modèles intermédiaires ou complets, plus larges, sont légèrement meilleures que celles du modèle basique. Un autre critère intéressant possible est le pourcentage de données « extrêmes » par rapport à la distribution prédictive. Le caractère extrême peut être étudié à partir des « p-valeurs prédictives », c'est-à-dire le pourcentage de valeurs de la distribution prédictive supérieures à la donnée observée (Marshall and Spiegelhalter, 2003; Green et al., 2009). Le pourcentage de données de p-valeurs prédictives > 0.975 ou < 0.025 (c'est-à-dire de valeurs supérieures au quantile 0.975 ou inférieures au quantile 0.025 de la distribution prédictive) est de 10.0% pour le modèle basique, 9.3% pour le modèle intermédiaire, et 7.9% pour le modèle complet, indiquant qu'il y a moins de données extrêmes par rapport aux prévisions dans le cas du modèle complet. Finalement, les prédictives

CHAPITRE 2. Méta-analyse

obtenues avec les trois modèles sont donc assez semblables mais avec un petit avantage dans les extrêmes pour les modèles intermédiaire et complet.

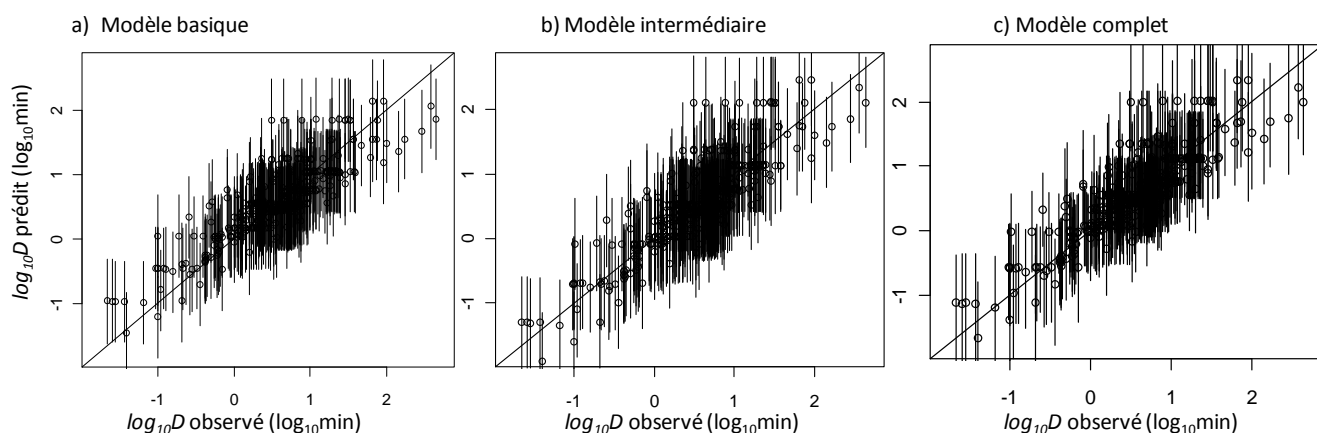


Figure 2.8. Graphes de comparaison entre les valeurs observées de $\log_{10}D$ et les prédictives (médiane et intervalles de crédibilité à 95%), pour les modèles basique (a), intermédiaire (b) et complet (c).

Décrire D d'étude inconnue dans un modèle AQR (« prévisions brutes »)

Dans les modèles d'évaluation quantitative du risque, il est fréquent d'utiliser les modèles primaires ou secondaires d'inactivation (ou de croissance) de manière déterministe, sans utiliser l'erreur σ du modèle (Delignette-Muller et al., 2000; Nauta, 2000; Pouillot et al., 2007; Afchain et al., 2008). Il s'agit toujours de prédire une nouvelle valeur de D , à une température et un pH donnés mais pour une souche inconnue, un milieu de traitement aqueux inconnu, etc. Dans ce cas, les distributions obtenues sur les valeurs de D dépendent seulement des distributions d'incertitude et/ou de variabilité sur D_{ref} , z_T et z_{pH} . Ici, cela correspond aux distributions sur $\log_{10}D_i^j$, où de nouvelles valeurs des variables $\log_{10}D_{ref}^j$, z_T^j et z_{pH}^j ont été simulées pour chaque étude à partir des lois a posteriori des hyperparamètres dans le cas des modèles intermédiaires et complets (par exemple $\log_{10}D_{ref}^j \sim N(\mu_{LDref}, \sigma_{LDref})$) (comme pour les prédictives). Dans le cas du modèle basique, ces valeurs correspondent aux ajustements. La comparaison des critères CRPS donne un avantage aux modèles intermédiaires et complets (Tableau 2.8). La comparaison graphique (Figure 2.9) des valeurs de D obtenues pour les trois modèles montre des intervalles de crédibilité à 95% très petits dans le cas du modèle basique, qui ne contiennent les données observées que dans 12.6% des cas. Dans le cas des modèles intermédiaires et complets les intervalles de crédibilité à 95% sont plus larges et englobent environ 95% des données, ce qui s'explique par le fait que ces modèles expliquent une part de la variabilité des données en supposant certains paramètres microbiologiques variables entre les études, alors que dans le cas du modèle basique toute la variabilité non expliquée est contenue dans une large erreur résiduelle σ , non prise en compte ici. Les valeurs de D et leurs intervalles de crédibilité obtenus avec le modèle basique ne sont pas satisfaisants, ce qui montre que si les sources de variabilité ne sont pas prises en compte dans les modèles, il faudrait au minimum les

CHAPITRE 2. Méta-analyse

prendre en compte par le biais de l'erreur résiduelle σ . Ces comparaisons soulignent de nouveau l'intérêt des modèles complets et intermédiaires, où la prise en compte des sources de variabilité dans le modèle permet d'obtenir des intervalles de crédibilité nettement plus corrects, et d'expliquer la provenance des sources de variabilité.

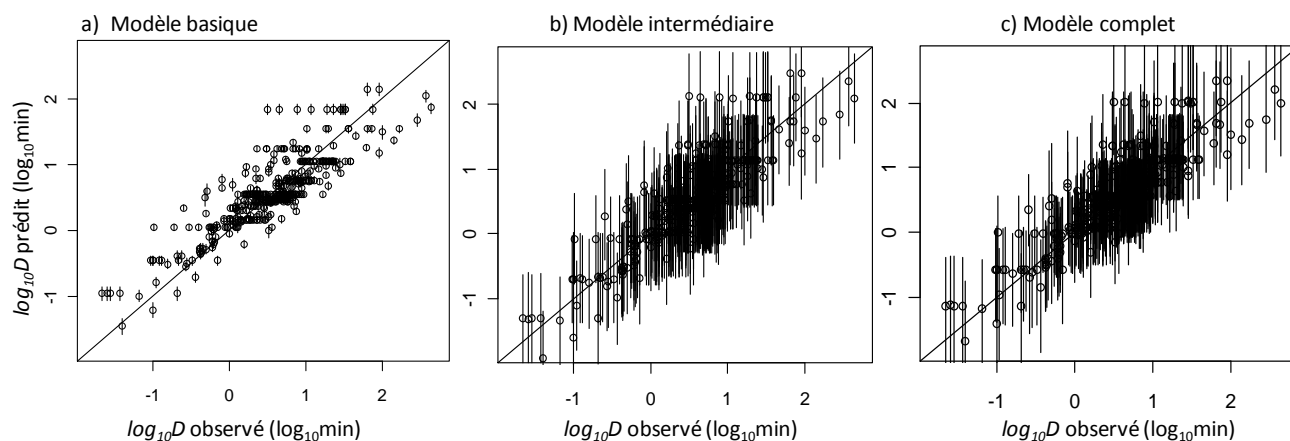


Figure 2.9. Graphes de comparaison entre les valeurs observées de $\text{Log}_{10}D$ et les « prévisions brutes » de $\text{Log}_{10}D$ (médiane et intervalles de crédibilité à 95%), pour les modèles basique (a), intermédiaire (b) et complet (c).

Tableau 2.8. CRPS obtenus avec les trois modèles pour les ajustements, les prédictives, ou les « prévisions brutes ». Les CRPS les plus proches de 0 correspondent aux meilleurs ajustements/prédictives/prévisions brutes.

CRPS	Modèle basique	Modèle intermédiaire	Modèle complet
Ajustements	-0.23	-0.04	-0.03
Prédictives	-0.18	-0.19	-0.17
"Prévisions brutes"	-0.23	-0.19	-0.18

En conclusion, l'intérêt de tenir compte de la variabilité inter-étude des paramètres de résistance thermique apparaît nettement, tant au niveau des ajustements qu'au niveau des différents types de prévisions.

2.3.3 Comparaison des estimations obtenues en milieu de laboratoire et en milieux alimentaires

Dans la méta-analyse, les milieux de traitement comprennent des milieux de laboratoire (type tampon phosphate) et des milieux alimentaires (type haricot vert). Une question intéressante est de rechercher si la résistance thermique de *G. stearothermophilus* est semblable dans les deux types de milieux. Il a en effet déjà été montré pour d'autres espèces que des spores produites dans un milieu alimentaire réel pouvaient être plus résistantes que des spores obtenues en milieu de laboratoire (Huemer et al., 1998).

Une analyse a été faite afin de comparer les valeurs de D obtenues dans les milieux de traitement de laboratoire (type de milieu $tm=1$) ou à partir de substrats alimentaires ($tm=2$). Il a été supposé que D_{ref} pouvait varier entre ces deux types de milieux (par simplification, les autres paramètres z_T et z_{pH} n'ont pas été supposés varier entre les deux types de milieux). Environ 70% des données ont été obtenues à partir d'un milieu de laboratoire, et 30% à partir de milieux alimentaires. Un DAG représentant le modèle utilisé pour la comparaison est présenté en Figure 2.10 (les relations entre les variables et les lois a priori sont similaires à ceux présentés dans l'article pour le modèle intermédiaire).

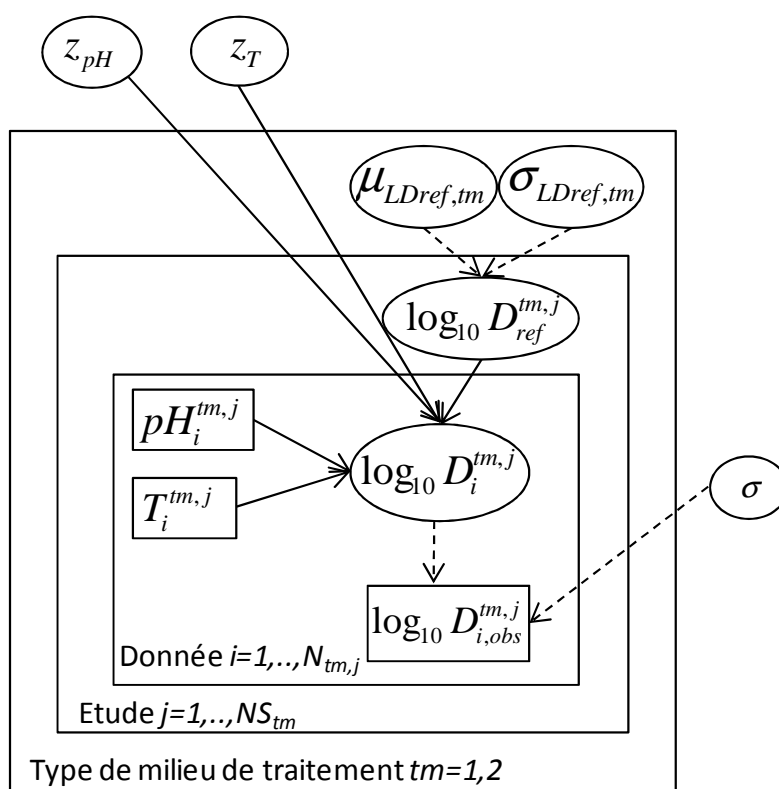


Figure 2.10. DAG du modèle de comparaison de la résistance thermique entre milieux de laboratoire et milieux alimentaires, en supposant D_{ref} variable entre les deux types de milieux.

CHAPITRE 2. Méta-analyse

Tableau 2.9. Statistiques descriptives des distributions a posteriori empiriques des paramètres de résistance thermique de *G. stearothermophilus* estimées avec le modèle de comparaison milieu de laboratoire ($t_m=1$) / milieu alimentaire ($t_m=2$).

Type de variables	Nœud	Unité	Moyenne	SD	$Q_{0.025}^a$	$Q_{0.975}^a$
Paramètres	$\mu_{LDref,1}$	\log_{10} min	0.35	0.03	0.29	0.42
	$\sigma_{LDref,1}$	\log_{10} min	0.37	0.02	0.33	0.42
	$\mu_{LDref,2}$	\log_{10} min	0.55	0.03	0.49	0.62
	$\sigma_{LDref,2}$	\log_{10} min	0.14	0.02	0.10	0.19
	z_T	°C	8.2	0.1	8.1	8.4
	z_{pH}	pH unit	4.0	0.2	3.7	4.3
	σ	\log_{10} min	0.10	0.00	0.09	0.10
Variables intermédiaires	$D_{ref,1}$	min	3.3	3.5	0.4	12.4
	$D_{ref,2}$	min	3.8	1.3	1.8	7.0

L'inférence bayésienne a été menée avec le logiciel Jags et la convergence du modèle a été vérifiée par analyse visuelle et à partir de la statistique de Gelman-Rubin. Les résultats sont présentés dans le Tableau 2.9. A partir des milieux de laboratoire D_{ref} est estimé en moyenne à 3.3 min, avec un intervalle de crédibilité à 95% de [0.4 ; 12.4] min, alors qu'à partir des milieux alimentaires D_{ref} est estimé en moyenne à 3.8 min, avec un intervalle de crédibilité à 95% de [1.8 ; 7.0] min. Ces résultats ne permettent pas de montrer de différence significative entre $D_{ref,1}$ et $D_{ref,2}$, et donc entre la résistance thermique de *G. stearothermophilus* dans les deux types de milieu. La variabilité sur D_{ref} est par ailleurs plus grande dans les milieux de laboratoires que dans les milieux d'origine alimentaires (estimations $\sigma_{Dref,1} = 0.37$ versus $\sigma_{Dref,2} = 0.14$ en moyenne, avec des intervalles de crédibilité à 95% distincts).

Chapitre 3

Modèle d'évaluation du risque microbiologique de non-stabilité à 55°C due à *Geobacillus stearothermophilus* dans les haricots verts appertisés

Ce chapitre présente la construction et l'évaluation d'un modèle d'évaluation des risques de non-stabilité à 55°C due à *G. stearothermophilus* dans les haricots verts appertisés. Ce travail a fait l'objet d'un article reproduit dans la partie 3.2, soumis à la revue *International Journal of Food Microbiology*.

3.1 Construction d'un modèle d'évaluation des risques

Ce modèle d'évaluation des risques a été construit en collaboration avec des experts en microbiologie des aliments. Sa construction a demandé un important travail de recueil de données, qui étaient plus ou moins disponibles, représentatives et précises. Ainsi, des données ont été recueillies à partir :

- de mesures de concentrations bactériennes effectuées sur les chaînes de transformation d'un aliment en usine au cours de plusieurs campagnes de production
- d'analyses expérimentales visant à préciser un phénomène mal connu, comme les valeurs du pH de l'aliment au cours de la stérilisation, ou les conditions d'apparition de la germination des spores de *G. stearothermophilus*

CHAPITRE 3. Modèle d'évaluation du risque

- de la littérature scientifique (voire de la littérature « grise »), comme les valeurs de résistance thermique de *G. stearothermophilus*
- d'informations fournis par les industriels sur les procédés impliqués dans l'appertisation des haricots verts.

Ces données sont entachées d'incertitudes, dues à la précision des mesures, à la présence de censures (concentrations bactériennes sous le seuil de détection), à des biais quand les données ont été obtenues dans des conditions légèrement différentes de celles du modèle, etc. Une fois le modèle construit, si l'analyse des résultats du modèle montre que certaines sources d'incertitudes ont un fort impact sur les résultats du modèle, il est envisageable de rassembler de nouvelles informations et données afin d'augmenter la précision du modèle.

L'information apportée par les données a besoin d'être traduite sous forme intégrable dans le modèle, par exemple, dans une approche probabiliste, sous forme de distributions de probabilité de variables du modèle. L'ajustement de distributions aux données passe par une étape de choix de distribution, d'estimation de paramètres, puis d'évaluation de la qualité de l'ajustement. Le choix de la distribution se base par exemple sur la connaissance du phénomène ou de l'objet modélisé (exemple : processus de Bernoulli) ou sur l'analyse des propriétés des données (discrète ou continue, plage de variation, paramètres de formes type coefficients d'aplatissement ou de dissymétrie) (Vose, 2000; Pouillot et al., 2010). Différentes méthodes d'estimation des paramètres sont possibles, comme les techniques statistiques inférentielles classiques (estimation par maximum de vraisemblance, par la méthode des moments), l'inférence bayésienne, ou l'optimisation d'un critère de qualité d'ajustement (minimisation de la somme des carrés des erreurs par un algorithme itératif). Enfin la qualité de l'ajustement peut être évaluée à l'aide de graphes ou de critères de qualité d'ajustement (statistiques du Chi², de Kolmogorov-Smirnoff, d'Anderson-Darling). Lors de la construction d'un modèle, ces différentes méthodes peuvent être appliquées pour quantifier l'incertitude sur les paramètres d'un modèle ou pour ajuster des distributions de variabilité à partir de données. L'incertitude sur les paramètres d'un modèle peut également être quantifiée par la méthode du *bootstrap*. L'ajustement à partir de données d'une distribution paramétrique représentant de la variabilité et de distributions représentant l'incertitude sur les paramètres de la distribution de variabilité (Figure 1.4) peut se faire avec les techniques statistiques classiques, l'inférence bayésienne ou le *bootstrap*. Ainsi dans ce modèle d'analyse des risques, des distributions de variabilité de concentrations bactériennes dans les haricots verts à différentes étapes de la chaîne de production ont été ajustées à partir de la méthode du maximum de vraisemblance et les incertitudes sur les paramètres des distributions ont été estimées par *bootstrap* non paramétrique. Le package R *fitdistrplus* (Pouillot et al., 2010) a été utilisé ; l'un de ses avantages est de pouvoir prendre en compte les données censurées. Des méthodes d'inférence bayésienne ont également été utilisées pour construire le modèle, lors de l'estimation des paramètres de résistance thermique de *G. stearothermophilus* par méta-analyse (voir Chapitre 2) ou pour l'estimation des proportions de *G. stearothermophilus* parmi les bactéries aérobies thermophiles.

Certains paramètres du modèle ne peuvent pas être informés par des données, ces données n'étant pas disponibles : trop difficiles ou chères à acquérir, ou bien jamais prises en considération. Il est alors possible de recourir à des opinions d'« experts » pour préciser leurs valeurs et les distributions possibles sur ces paramètres. Basées sur des jugements subjectifs, « l'élicitation de dires d'experts » doit être entreprise avec précaution pour éviter différentes sources de biais et

CHAPITRE 3. Modèle d'évaluation du risque

d'erreurs (difficultés pour un expert de quantifier un savoir qualitatif). Des distributions souvent utilisées pour modéliser les jugements d'experts sont les distributions triangulaires et BetaPert, paramétrées par leurs valeurs minimum, les plus vraisemblables et maximum, dont l'interprétation est assez aisée pour l'expert. Dans ce modèle, des avis d'experts ont par exemple été utilisés pour définir la concentration maximale en *G. stearothermophilus* atteignable dans le produit.

Une fois construit, ce modèle d'analyse des risques a été analysé par la méthode de simulation de Monte Carlo à deux dimensions (cf partie 1.2.2). Des distributions de variabilité du risque ont été obtenues, ainsi que des intervalles d'incertitudes sur ces estimations. Il est alors possible d'évaluer la part de la variabilité et de l'incertitude dans les résultats du modèle. La construction du modèle d'analyse des risques a été suivie d'une confrontation des résultats avec des données indépendantes, afin d'apporter des éléments de validation du modèle. Puis une analyse de sensibilité basée sur les indices de Sobol a été faite sur les facteurs variables du modèle, afin de déterminer les facteurs variables dont les variations influençaient le plus la variabilité de la sortie du modèle. Les indices de sensibilité des facteurs variables ont ainsi été calculés conditionnellement à des valeurs échantillonnées dans la dimension incertitude pour les paramètres incertains. La répétition des analyses de sensibilité pour différentes valeurs d'incertitude a alors permis d'estimer l'incertitude sur ces indices. Par cette analyse de sensibilité, les facteurs clés du modèle ont été déterminés, fournissant des indications sur les mesures de maîtrise du risque les plus performantes envisageables. Différents scénarios ont ensuite été testés afin d'explorer des mesures de maîtrise du risque possible ou d'analyser l'impact de certaines hypothèses assez incertaines du modèle.

3.2 Article: Quantitative assessment of the risk of microbial spoilage in foods. Prediction of non-stability at 55°C caused by *Geobacillus stearothermophilus* in canned green beans

Cette partie est la reproduction d'un article écrit en collaboration entre Clémence Rigaux^a, Stéphane André^b, Isabelle Albert^a et Frédéric Carlin^{c,d}. Il a été soumis dans la revue International Journal of Food Microbiology.

^aINRA, UR 1204, Met@risk, Food Risk Analysis Methodologies, F-75005 Paris, France

^bCTCPA, Site Agroparc, ZA de l'aéroport, F-84911 Avignon, France

^cINRA, UMR 408, Sécurité et Qualité des Produits d'Origine Végétale, F-84000 Avignon, France

^dUniversité d'Avignon et des Pays de Vaucluse, UMR 408, Sécurité et Qualité des Produits d'Origine Végétale, F-84000 Avignon, France

Abstract

Microbial spoilage of canned foods by thermophilic and highly heat-resistant spore-forming bacteria, such as *Geobacillus stearothermophilus*, is a persistent problem in the food industry. An incubation test at 55°C for 7 days, then validation of biological stability, is used as an indicator of compliance with good manufacturing practices. We propose a microbial risk assessment model predicting the percentage of non-stability due to *G. stearothermophilus* in canned green beans manufactured by a French company. The model accounts for initial microbial contaminations of fresh unprocessed green beans with *G. stearothermophilus*, cross-contaminations in the processing chain, inactivation processes and probability of survival and growth. The sterilization process is modeled by an equivalent heating time depending on sterilization value F_0 and on *G. stearothermophilus* resistance parameter z_T . Following the recommendations of international agencies, second order Monte-Carlo simulations are used, separately propagating uncertainty and variability on parameters.

As a result of the model, the mean predicted non-stability rate is of 0.5%, with a 95% uncertainty interval of [0.1%; 1.2%]. This is in great accordance with data reported in the French industry. A sensitivity analysis based on Sobol indices and some scenario tests underline the importance of cross-contamination at the blanching step, in addition to inactivation due to the sterilization process.

Key-words

Microbiological risk assessment; *Geobacillus stearothermophilus*; Food spoilage; Canned green beans; Second order Monte Carlo simulations

3.2.1 Introduction

Canned foods have an excellent safety record due to the inactivation of pathogenic bacteria. The highest risk, due to *Clostridium botulinum*, is controlled in low acid foods (pH > 4.5) by the application of the 'botulinum cook', which is defined as equivalent to 3 min heating at 121°C. However there is still a risk of spoilage of canned foods, because of the persistence of non-pathogenic thermophilic and highly heat-resistant spore-forming bacteria (Logan et al., 2009; Burgess et al., 2010; Prevost et al., 2010). The stability after incubation at 55°C for 7 days (French standard NF V08-408 (AFNOR, 1997)) of canned foods is used as a hygiene indicator and as a norm for exportation to hot-climate regions as required by some national regulations, such as that of France (NOR: AGRG9700991A, 1997). Among thermophilic bacteria, *Geobacillus stearothermophilus* is recognized as a major source of spoilage in canned foods and is frequently detected in cans presenting defects after a 7-day incubation at 55°C (André et al., 2012b).

The canning process combines several operations allowing contamination, inactivation and/or growth of micro-organisms. The final stability is the result of the individual contributions of the processing operations. In this work, the fate of *G. stearothermophilus* is modeled along a canned green bean processing chain to predict the rate of non-stability due to *G. stearothermophilus* after the standard incubation test at 55°C. A quantitative microbial risk assessment was used, with a probabilistic approach accounting for sources of variability and uncertainty. Variability represents the natural heterogeneity of a factor, coming for instance from heterogeneity between bacterial strains, and is irreducible by nature. Uncertainty comes from a reducible lack of knowledge on the true value of a parameter and for instance from a lack of information, sampling or measurement errors. Risk variability distributions and evaluation of uncertainty associated with this spoilage risk were obtained using a two-dimensional Monte Carlo simulation separately propagating uncertainty and variability through the model (Vose, 2000; Mokhtari et al., 2005; Pouillot et al., 2007; Pouillot et al., 2010). This distinction is useful for risk managers and has been recommended by international agencies for a few years (Codex Alimentarius Commission, 1999; European Commission, 2003).

The aim of this study was to build a model giving a reliable prediction of spoilage in canned foods. The model was further used to simulate the impact of some risk management options or the potential consequences of the uncertainty of some assumptions on the final predictions, using what-if scenarios. A sensitivity analysis using Sobol indices allowed the detection of the most influential risk factors.

3.2.2 Material and Methods

3.2.2.a Overview of the model

A modular (Nauta, 2001) process risk model made up of several steps following the contamination of *G. stearothermophilus* from fresh unprocessed green beans to the end of the industrial canning process has been developed. *G. stearothermophilus* growth leading to the spoilage of canned green beans is considered during a 7-day incubation test at 55°C. The different steps of the food chain and the basic microbial process impacting the fate of *G. stearothermophilus* at each step

CHAPITRE 3. Modèle d'évaluation du risque

are described in Table 3.1. Basically, the risk model accounts for (i) the inactivation of *G. stearothermophilus* during heating processes (blanching and sterilization), depending on processing time and temperature and on pH of canned food, (ii) cross-contamination at blanching and brining and (iii) the probability of survival and growth to cause spoilage after incubation. Expert opinion, literature data, information on the food process and specifically collected data, such as prevalence of *G. stearothermophilus* in green bean samples, have been used to build the model.

The modeled unit is one can containing 445g of green beans filled with 405g of covering brine, leading to a final weight of 850g. The outputs of the model are (i) the concentrations of *G. stearothermophilus* after sterilization and after incubation tests and (ii) the non-stability rate, i.e. the percentage of green bean cans presenting a defect at the end of the incubation test.

Table 3.1. Description of the steps of the food pathway (including the incubation test) and link with the modeled basic microbial processes impacting the fate of *G. stearothermophilus*.

Step X of the food pathway and of incubation test (abbreviation)	Microbiological process affecting <i>G.stearothermophilus</i>
Green bean harvesting (INIT)	Initial contamination
Blanching (BLAN)	Inactivation of spores, and cross contamination by blanching water
Canning and brining (BRIN)	Cross-contamination by recovery brine
Sterilization (STERI)	Inactivation of spores
Cooling (COOL)	none
Incubation test (INCUB)	Germination then growth

3.2.2.b Determination of *G. stearothermophilus* concentrations at different processing steps

Concentrations of Aerobic Thermophilic Spores (ATS) were determined in samples of green beans collected on the chains of two processing plants during the 2-month production periods of years 2011 and 2012. Samples (of 10 g or 100 ml) were collected from raw green beans before blanching (number of samples n =95), from blanched green beans immediately before can filling (n=93), from blanching water (n=45) and from covering brine (n =99). Samples were stored frozen at -18°C for 1 to 4 weeks, thawed at room temperature, homogenised, then treated at 100°C for 10 min to eliminate vegetative cells (French Standard NFV08-602 (AFNOR, 2011)). Serial decimal dilutions were made in tryptone salt buffer and 1 ml was mixed with 25 ml melted BCP agar and incubated at 55°C for 2 days. When no colonies were detected, ATS counts were considered as left-censored data (i.e., under the threshold of detection: 10 CFU/g or 1 CFU/ml). When colonies were detected, up to three colonies were randomly picked for identification of *G. stearothermophilus*

CHAPITRE 3. Modèle d'évaluation du risque

using the molecular tool SporesTraQ™ (Prevost et al., 2010), in order to estimate the proportion (or ratio) of *G. stearothermophilus* among ATS.

3.2.2.c Estimation of the green bean pH at different processing steps

The green bean pH at the blanching step was based on specific measures on the process chain before sterilization. The pH of the sterilized product, in which growth will occur, was assumed identical to the pH of the covering brine after sterilization, which was measured by the canning processor according to the French standard NFV08-408 (AFNOR, 1997).

To estimate the pH during sterilization, cans (1/4 US size) were filled with 83 g of frozen blanched green beans and with 110 ml of covering brine made up of tap water and sodium chloride at 1% (wt/vol). Cans were placed in an autoclave and heated at 120 °C. Cans were sampled at regular time intervals between 0 min and 15 min after the cut. The pH of the brine and of blended green beans was measured using a pHmeter calibrated at pH 4.0 and 7.0 buffer solutions (Sartorius, Aubagne, France). This procedure was applied on 5 samples of green beans for each time of treatment. A rapid equilibrium was shown between brine and green beans, no difference was observed immediately after the cut. Consequently the pH at equilibrium was considered as the pH to which *G. stearothermophilus* is subjected during sterilization.

3.2.2.d Two-dimensional Monte Carlo simulation

Within the two-dimensional Monte Carlo simulation framework, uncertainty and variability were separately propagated through the model. According to expert opinions and data and modeling choices, parameters of the model were classified into four categories: fixed, variable, uncertain and both variable and uncertain, as previously detailed (Pouillot et al., 2010). The simulations were made by first sampling in the uncertainty dimension (of sample size N_u), then by sampling in the variability dimension (of sample size N_v) conditionally to the sampled uncertain parameters (Vose, 2000; Pouillot et al., 2007; Pouillot et al., 2010). This procedure leads to a global sample of size $N_u \times N_v$. The simulations were performed with $N_u = 1500$ and $N_v = 3000$, using the R software (R Development Core Team, 2010).

3.2.2.e Model and parameters

The global model is represented by a directed acyclic graph (DAG) (Figure 3.1), which gives the conditional dependencies between the model parameters.

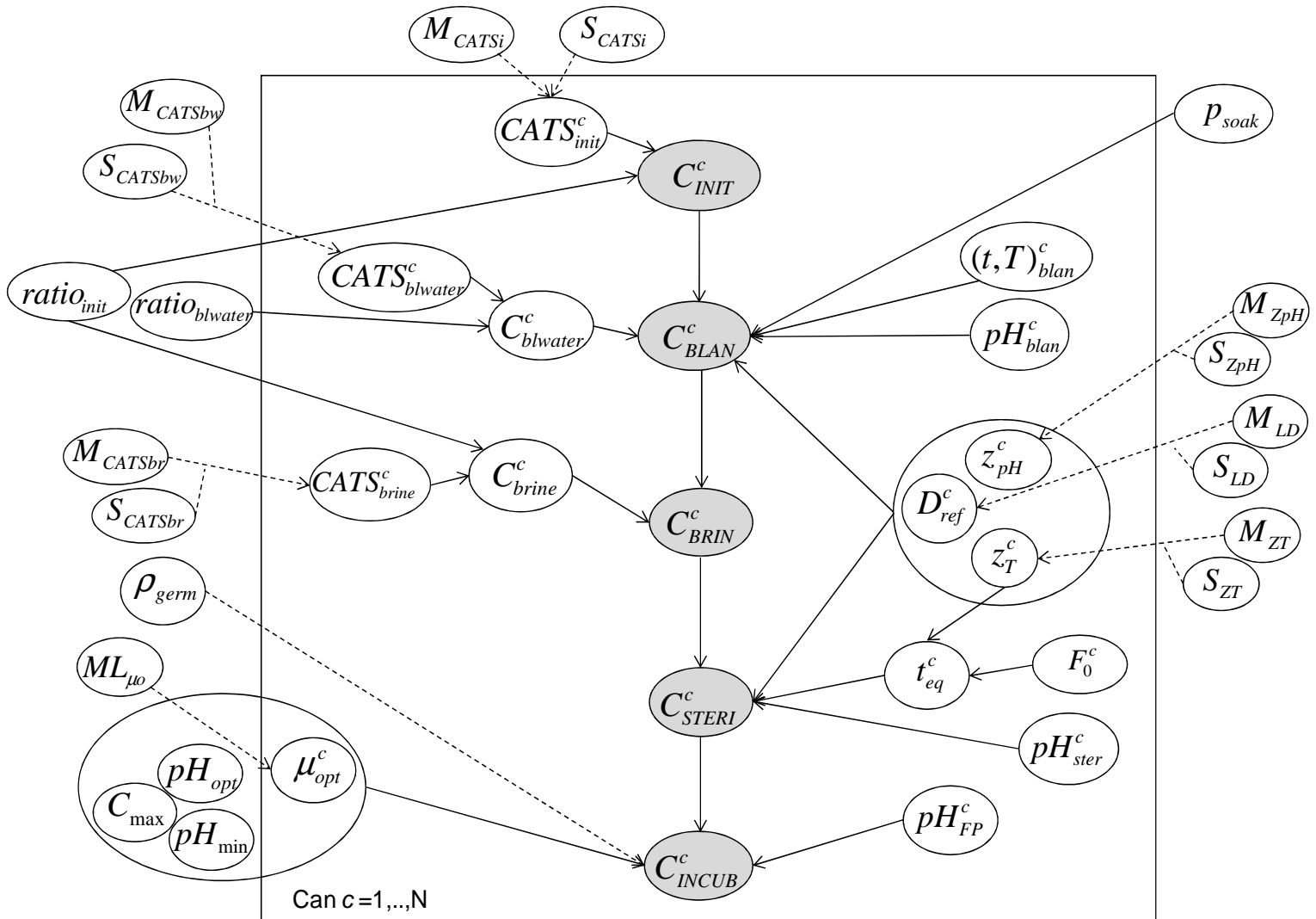


Figure 3.1. The global risk model represented by a directed acyclic graph (DAG). Model parameters are represented by nodes. The nodes are described in Table 3.2 and/or in Section 3.2.2.e. The principal nodes C_X , are colored in grey; they denote the *G. stearothermophilus* concentrations at steps X of the food chain (see Table 3.1). Solid edges indicate deterministic links and dashed edges indicate stochastic links between nodes. Index c denotes the can numbers and specifies the can-dependent-variables.

CHAPITRE 3. Modèle d'évaluation du risque

Table 3.2. Model parameters: description, variability and uncertainty distribution, mean and 95% CI of the global marginal distribution. $N(a; b)$ (resp. $N(a; b)T(c, d)$, $LnN(a; b)$) stands for the normal distribution (resp. the normal distribution truncated on $[c; d]$, the lognormal distribution based on a normal law) of mean a and standard deviation b . $\Gamma(a; b)$ stands for the gamma distribution with shape parameter a and rate parameter b . $BP(a; b; c)$ stands for a BetaPert distribution of minimum value a , most likely value b and maximum value c . $B(a; b)$ stands for a Beta distribution of parameters a and b .

Parameter	Description (unit)	Variability distribution	Uncertainty distribution	Mean and 95% probability range ^a	Source ^c	
Contamination param.	$CATS_{init}$	ATS concentrations in fresh green beans ($\log_{10}CFU/g$)	$N(M_{CATSi}; S_{CATSi})$	$M_{CATSi} \sim N(0.44; 0.24)$ $S_{CATSi} \sim LnN(-0.11; 0.39)$	0.4 [-1.7; 2.6]	Spec. data
	$CATS_{blwater}$	ATS concentrations in blanching water ($\log_{10}CFU/g$)	$N(M_{CATSbw}; S_{CATSbw})$	$M_{CATSbw} \sim N(2.26; 0.18)$ $S_{CATSbw} \sim LnN(0.17; 0.10)$	2.3 [-0.1; 4.6]	Spec. data
	$CATS_{brine}$	ATS concentrations in brine ($\log_{10}CFU/g$)	$N(M_{CATSbr}; S_{CATSbr})$	$M_{CATSbr} \sim N(0.37; 0.14)$ $S_{CATSbr} \sim LnN(0.13; 0.07)$	0.4 [-1.9; 2.6]	Spec. data
	$ratio_{init}$	Proportion of Gbs among ATS in fresh green beans or in brine	-	$B(78; 167)$	0.32 [0.26; 0.38]	Spec. data
	$ratio_{blwater}$	Proportion of Gbs among ATS in blanching water	-	$B(72; 39)$	0.65 [0.56; 0.73]	Spec. data
	p_{soak}	Water soak rate at blanching	-	$BP(1\%, 2.7\%, 6\%)$	0.03 [0.01; 0.05]	Spec. data
Microbial param. of <i>G. steatothermophilus</i>	D_{ref}	Time (min) to the first decimal reduction at $T=121,1^{\circ}C$ and $pH=7$	$10^{N(M_{LD}, S_{LD})}T(0.1; 30)$	$M_{LD} \sim N(0.432; 0.024)^b$ $S_{LD} \sim LnN(-1.277; 0.073)^b$	3.3 [0.8; 9.6]	Rigaux et al. 2013
	Z_T	Increase in temperature resulting in a ten-fold reduction in D_{ref} ($^{\circ}C$)	$N(M_{ZT}, S_{ZT})T(3; 15)$	$M_{ZT} \sim N(9.13; 0.34)^b$ $S_{ZT} \sim LnN(0.60; 0.19)^b$	9.1 [5.5; 12.9]	Rigaux et al. 2013
	Z_{pH}	Decrease in pH resulting in a ten-fold reduction in D_{ref} (pH unit)	$N(M_{ZpH}, S_{ZpH})T(1; 13)$	$M_{ZpH} \sim LnN(1.45, 0.09)^b$ $S_{ZpH} \sim \Gamma(1.99, 3.29)^b$	4.3 [2.9; 6.2]	Rigaux et al. 2013
	ρ_{germ}	Spore germination probability	-	$B(4.6; 1.4)$	0.8 [0.4; 1.0]	Expert op.
	C_{max}	Maximum bacterial concentration in the product ($\log_{10}CFU/g$)	-	$N(9.00; 0.25)$	9.0 [8.5; 9.5]	Expert op.
	μ_{opt}	Optimal growth rate (h^{-1})	$BP(2.01; ML_{\mu}; 2.72)$	$ML_{\mu} \sim BP(2.20; 2.31; 2.43)$	2.3 [2.1; 2.6]	Based on Llaudes et al. 2001
	pH_{min}	Minimal growth cardinal pH	-	$N(5; 0.1)$	5.0 [4.8; 5.2]	Expert op.
	pH_{opt}	Optimal growth cardinal pH	-	$N(7; 0.1)$	7.0 [6.8; 7.2]	Expert op.
Environmental param.	pH_{blan}	Green bean pH during blanching	$N(6.4; 0.2)T(4.8; 7)$	-	6.4 [6.0; 6.8]	Spec. data
	pH_{ster}	Green bean pH during sterilisation	$N(5.8; 0.2)T(4.8; 7)$	-	5.8 [5.4; 6.2]	Spec. data
	pH_{FP}	Finished product pH	$N(5.55; 0.1)T(4.8; 7)$	-	5.5 [5.3; 5.7]	Spec. data
	t_{blan}	Blanching duration (min)	$BP(3.5; 6; 10)$	-	6.3 [4.2; 8.7]	Company
	T_{blan}	Blanching temperature ($^{\circ}C$)	$BP(85; 91; 98)$	-	91.1 [86.7; 95.8]	Company
	F_0	Value of sterilisation (min)	$LnN(3.40, 0.11)$	-	30.2 [24.3; 37.2]	Company + model

^aValues are based on the confounded analysis of variability and uncertainty and were determined by Monte Carlo simulation (see Section 2.2). ^bSome correlations between the hyperparameters (mean

CHAPITRE 3. Modèle d'évaluation du risque

and SD) of D_{ref} , z_T and z_{pH} were observed in the work of Rigaux et al. 2013, and are defined in Table 3.3. They may be taken into account using the Iman & Conover method (Iman et al., 1982) which is implemented in the R package mc2d (Pouillot et al., 2010). Nevertheless no significant correlation is observed between the parameters D_{ref} , z_T and z_{pH} . ^cAs sources, "Spec. data" means data specifically collected for this work, and "Expert op." means expert opinion.

Initial contamination and cross-contamination sources

Log-normal variability distributions were adjusted on the measured ATS concentrations based upon maximum likelihood estimation, and parameter uncertainty was evaluated by non-parametric bootstrap using the R package fitdistrplus (Pouillot et al., 2010). *G. stearothermophilus* ratio at blanching was significantly different from the ratio in blanching water and in covering brine, and this was considered in the model parameters. The corresponding adjusted distributions are presented in Table 3.2. The concentrations in green beans before filling were used to validate the model. As no significant correlation between ATS concentrations and *G. stearothermophilus* proportions was detected, the *G. stearothermophilus* concentration is given by:

$$C_x = CATS_x \cdot ratio_x \quad (1)$$

where C_x (resp. $CATS_x$) (CFU/g) is the *G. stearothermophilus* (resp. ATS) concentration and $ratio_x$ is the proportion of *G. stearothermophilus* among ATS in the contamination source x . x represents either the fresh green beans ($x=init$) or the blanching water ($x=blwater$) or brine ($x=brine$).

Inactivation models

Spore inactivation was assimilated to first order kinetics and survival curves were classically considered as log-linear. The primary inactivation model is therefore described by:

$$\log_{10}(C(t)) = \log_{10}(C(0)) - \frac{t}{D(T,pH)} \quad (2)$$

where $C(0)$ (CFU/g) is the initial population at time 0 (min), t (min) is the processing time, $C(t)$ (CFU/g) is the population at time t , and $D(T, pH)$ (min) is the decimal reduction time at temperature T (°C) and pH at time of treatment.

The secondary model is an extension of the Bigelow model (Couvert et al., 2005). In addition to the effect of a temperature change on D , this model also describes the effect on D of a pH change. It is expressed as

$$D(T, pH) = D_{ref} \cdot 10^{(T_{ref}-T)/z_T} \cdot 10^{(pH-pH_{ref})/z_{pH}} \quad (3)$$

where T (°C) is the temperature and pH is the pH of the product during processing, T_{ref} (°C) is the reference temperature, here fixed at the reference sterilization temperature of 121.1°C, pH_{ref} (pH unit) is the pH of reference, fixed at 7.0, z_T (°C) is the increase in temperature resulting in a ten-fold

CHAPITRE 3. Modèle d'évaluation du risque

reduction in D , z_{pH} (pH unit) is the decrease in pH resulting in a ten-fold reduction in D , and $D_{ref} = D_{121.1^{\circ}\text{C}, \text{pH}7}$ (min) is the decimal reduction time at 121.1°C and pH 7.

The heat resistance parameters D_{ref} , z_T and z_{pH} were estimated from a meta-analysis performed on 430 D values mainly coming from literature, using a hierarchical Bayesian model (Rigaux et al., 2013). D_{ref} , z_T and z_{pH} were considered as both uncertain and variable parameters (Table 3.2). Correlations between the hyperparameters (mean and standard deviation (SD) of D_{ref} , z_T and z_{pH}) (Table 3.3) were taken into account in the simulation by the Iman & Conover method (Iman et al., 1982).

Blanching duration and time-temperature profiles during the sterilization of canned green beans were communicated by the company and considered as variable parameters as done by Pouillot et al. (2007) (Table 3.2). The temperature was considered as homogeneous in the canned green beans under the assumption of convective heat diffusion.

The product pH during blanching and sterilization was considered variable. The pH distributions were based on specific measurements (Section 2.3) and assumed to be normal with mean 6.4 and standard deviation (SD) 0.2 (respectively mean 5.8 and SD 0.2), truncated on [4.8; 7] (see Table 3.2).

Table 3.3. Spearman rank correlation between the hyperparameters of the microbial heat resistance parameters D_{ref} , z_T and z_{pH} of *G. stearothermophilus*, defined in the work of Rigaux et al., 2013. See Table 3.2 for the definitions of the hyperparameters.

Correlations	$M_{z_{pH}}$	$S_{z_{pH}}$	M_{z_T}	S_{z_T}	M_{LD}	S_{LD}
$M_{z_{pH}}$	1	0.70	0.01	0.01	-0.11	-0.06
$S_{z_{pH}}$	0.70	1	0.00	0.01	-0.07	-0.03
M_{z_T}	0.01	0.00	1	0.62	0.36	-0.39
S_{z_T}	0.01	0.01	0.62	1	0.33	-0.49
M_{LD}	-0.11	-0.07	0.36	0.33	1	-0.21
S_{LD}	-0.06	-0.03	-0.39	-0.49	-0.21	1

Calculation of a sterilization equivalent time

A sterilization equivalent time as a function of the sterilization value F_0 (the reference descriptor of the applied heat treatment determined with $z_T = 10^{\circ}\text{C}$) and of the resistance parameter z_T of the specific target bacterium (here *G. stearothermophilus*) was built. Six sterilization time-temperature profiles communicated by the company were first converted into equivalent heat treatments of duration $t=t_{eq}(z_T)$ at the target temperature $T_{ref}=121.1^{\circ}\text{C}$ using Eq. (4) (Fernandez and Peck, 1997) as follows:

$$t_{eq}(z_T) = \sum_{i=1}^n (t_i - t_{i-1}) 10^{\frac{1}{z_T} \left(\frac{T_i + T_{i-1}}{2} - T_{ref} \right)} \quad (4)$$

where t_i (min) and T_i (°C) are the observed time and temperature at the i^{th} measurement point of a time-temperature profile of n points $(t_i, T_i)_{i=1, \dots, n}$. The resulting equivalent time $t_{eq}(z_T)$ depends on $G. stearrowthermophilus$ z_T , which varies at each simulation according to its probability distributions (Figure 3.2a, Table 3.2). A conversion model accounting for the dependence between z_T and t_{eq} and separating the effect of the bacteria resistance and of the heat treatment was built as follows. An almost linear dependence (Figure 3.2.b, $R^2=0.88$) was observed between $\ln(t_{eq})$ and $1/z_T$ (which can be explained by the insignificance of the terms obtained in Eq. (4) at temperatures $< 100^\circ\text{C}$). Then, following the envelope method (Vose, 2000), a least square regression of $\ln(t_{eq})$ on $1/z_T$ was performed and modeled by:

$$\ln(t_{eq}) = a + b \left(\frac{1}{z_T} \right) + \epsilon \quad (5)$$

$\epsilon \sim N(0, c)$ stands for a normal distribution of mean 0 and SD c and is the model error, which here mainly represents the variation in the applied time-temperature profiles. Homogeneity of the error variance was assumed for simplification and to separate the effect of the intensity of the heat treatment (described by F_0) and that of z_T (describing microbial resistance). Eq. (5) was then reformulated into Eq. (6) (Figure 3.2.b):

$$t_{eq} = \exp \left(b \left(\frac{1}{z_T} - \frac{1}{z_{TREF}} \right) \right) \cdot \exp(X) \quad (6)$$

$$\text{where } X = a + \frac{b}{z_{TREF}} + \epsilon \sim N \left(a + \frac{b}{z_{TREF}}, c \right).$$

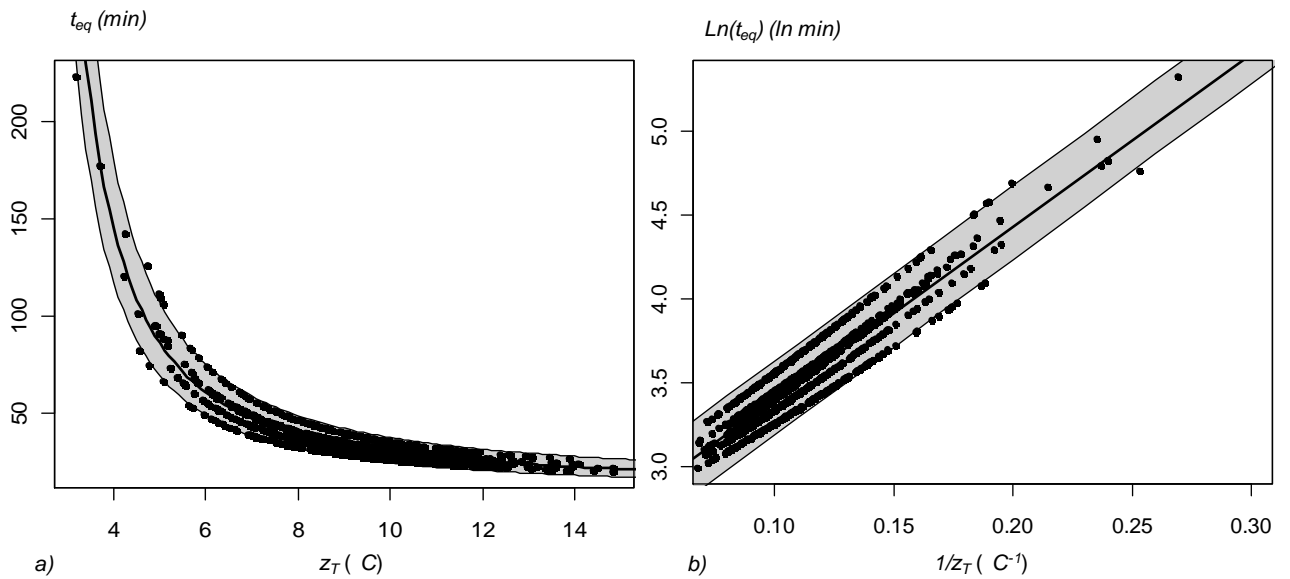


Figure 3.2. Sterilization equivalent time at 121.1°C t_{eq} as a function of the heat resistance parameter z_T (a), $\text{Ln}(t_{eq})$ as a function of $1/z_T$ (b), and envelopes of the adjusted models. Points correspond to a

CHAPITRE 3. Modèle d'évaluation du risque

sample of 1000 z_T (or $1/z_T$) values and their corresponding equivalent times t_{eq} (or $\ln(t_{eq})$) determined from Eq.(4). The bold middle lines represent the adjusted models (Eq. 7 and 5), and the bounding lines represent the 95% CI of the envelope (colored in grey). t_{eq} (or $\ln(t_{eq})$) values corresponding to points in the envelope with a fixed z_T (or $1/z_T$) value represent the 95% variability interval of the heat treatment.

By definition, F_0 corresponds to the sterilization equivalent time at 121.1°C determined with reference value $z_T = z_{T,REF} = 10^\circ\text{C}$. Thus, assuming that $F_0 = \exp(X)$, Eq. (6) becomes Eq.(7):

$$t_{eq}(z_T, F_0) = \exp\left(b\left(\frac{1}{z_T} - \frac{1}{z_{T,REF}}\right)\right) \cdot F_0 \quad (7)$$

The resulting F_0 variability distribution is a lognormal distribution of mean 3.40 and standard deviation 0.11 (corresponding to estimations $\hat{a}=2.36 \ln(\text{min})$, $\hat{b}=10.41^\circ\text{C}$ and $\hat{c}=0.11 \ln(\text{min})$), leading to F_0 values of mean 30.2min and 95% variability interval [24.1, 37.2]min, which is fully consistent with the F_0 determined by the company from the time-temperature profiles.

Cross-contamination models

Cross-contamination was simply assumed to bring a quantity of new spores in the can, following Eq. (8):

$$N_X = N_{X-1} + N_{Xcc} \quad (8)$$

where N_X (resp. N_{X-1}) (CFU) is the quantity of *G. stearothermophilus* at the food chain step X (resp. at the step before step X) and N_{Xcc} is the quantity of *G. stearothermophilus* brought by cross-contamination at step X .

The first source of cross-contamination is the blanching water. The quantity of spore contamination from the blanching water was considered as proportional to the amount of water impregnating green beans during this process step. The impregnating rate ρ_{soak} was estimated at around 2.7% using data obtained from an experimental measure in laboratory consisting in weighing some green bean samples before and after a cooking process (data not shown)(see Table 3.2 for the adjusted distribution). The quantity of spores added (CFU) by cross-contamination at blanching is then defined by

$$N_{BLANcc} = w_{GB} \cdot \rho_{soak} \cdot C_{blwater} \quad (9)$$

where $w_{GB} = 445\text{g}$ is the green bean weight at blanching, and $C_{blwater}$ (CFU/g) is the *G. stearothermophilus* concentration in blanching water.

The second source of contamination comes from the recovery liquid at filling. The quantity of spores added (CFU) by cross-contamination from the recovery liquid is simply defined by

$$N_{JUICC} = w_{juice} \cdot C_{juice} \quad (10)$$

CHAPITRE 3. Modèle d'évaluation du risque

where $w_{juice} = 405$ g is the juice weight in a can and C_{juice} (CFU/g) is the *G. stearothermophilus* concentration in juice defined in Section 2.3.1.

Incubation model

The incubation module is made of 3 steps: germination of spores, growth of vegetative cells and non-stability. Germination of spores may occur after sterilization. Each spore was assumed to have a probability p_{germ} of germinating. The model is defined by Eq. (11):

$$N_{Germination} = Bin(N_{STERI}, p_{germ}) \quad (11)$$

where $Bin(a, b)$ stands for a binomial distribution of parameters a and b , and N_x (CFU) is the (rounded down) number of spores at step X .

The primary growth model is the logistic model with delay (Rosso et al., 1995), defined by:

$$C(t) = \begin{cases} C(0) & \text{if } t < \lambda \\ C_{max} / [1 + \left(\frac{C_{max}}{C_0} - 1\right) \exp(-\mu_{max}(t - \lambda))] & \text{if } t \geq \lambda \end{cases} \quad (12)$$

where t (h) is the duration of the growth step, $C(0)$ and $C(t)$ (CFU/g) are respectively the modeled concentrations at time $t=0$ and at time t , C_{max} (CFU/g) is the maximum bacterial concentration in the product, μ_{max} (h^{-1}) the maximum growth rate and λ (h) the lag time. $\lambda=0$ was assumed in this work as growth occurs in rather optimal conditions and as a fail-safe assumption.

The secondary growth model is based on the "Gamma" model developed by Zwietering et al., 1992, and describes μ_{max} as a function of environmental conditions such as temperature, pH and water activity:

$$\mu_{max} = \mu_{opt} \cdot \gamma(T) \cdot \gamma(pH) \cdot \gamma(aw) \quad (13)$$

where μ_{opt} (h^{-1}) is the optimal growth rate in the product, $\gamma(T)$, $\gamma(pH)$ and $\gamma(aw)$ describe the effect of the temperature T , the pH pH and the water activity aw on the growth rate. As water activity in green beans and the incubation temperature ($T=55^\circ\text{C}$) are almost optimal for the growth of *G. stearothermophilus*, it was assumed $\gamma(T) = 1$ and $\gamma(aw) = 1$. The pH effect is described by:

$$\gamma(pH) = \frac{(pH - pH_{min})(pH + pH_{min} - 2 pH_{opt})}{(pH - pH_{min})(pH + pH_{min} - 2 pH_{opt}) - (pH - pH_{opt})^2} \quad (14)$$

where pH_{min} and pH_{opt} (pH unit) are the minimal and optimal cardinal pH of *G. stearothermophilus*.

Changes in pH were detected in food or medium containing populations higher than 10^7 *G. stearothermophilus* ml^{-1} (Llaudes et al., 2001; Yoo et al., 2006), and within 24h incubation at 55°C (Llaudes et al., 2001). Non-stability was therefore assumed to occur if *G. stearothermophilus*

CHAPITRE 3. Modèle d'évaluation du risque

concentration exceeds $7 \log_{10}\text{CFU/g}$ before 72 h incubation at 55°C . This assumption is rather conservative considering the reported observations.

The probability distributions on all the parameters used in the germination and growth models are described in Table 3.2. Specific laboratory experiments to determine the growth rate of spores of *G. stearothermophilus* strains (CTCPA 2804 173, 2804 138, 2804 168) were produced according to the French standard method NF T 72-231 (AFNOR, 1988) and separately spiked at 10 spores/g and 10^2 spores/g in 50g of green beans in 50ml of brine. Despite suboptimal pH, μ_{max} was of the same order of magnitude as that observed in other media and foods (Thompson and Thames, 1967; Ng and Schaffner, 1997; Ng et al., 2002; Yoo et al., 2006; Heinrich et al., 2008). We therefore assumed that canned green beans were a highly suitable medium for *G. stearothermophilus* growth and set the most likely value of μ_{opt} at 2.3h^{-1} , based on the cell doubling time T_d reported by Llaudes et al. (2001) in tryptic soy broth ($\mu_{opt} = \ln(2)/T_d$). The probability of germination p_{germ} was estimated at around 90% from expert opinion. The pH of the sterilized product in which growth will occur was estimated at 5.5 pH unit on average, based on specific measurements (Section 2.3). *G. stearothermophilus* C_{max} values are between 10^7 and 10^{11} (Thompson et al., 1967; Ng et al., 1997; Yoo et al., 2006; Heinrich et al., 2008), André unpublished data 2011). Mean C_{max} was taken at 10^9 , which is rather conservative considering that alteration may occur from $C_{max} = 10^7$. pH_{opt} was taken as 7.0, which agrees with the optimal growth of *G. stearothermophilus* at pH 6.2 – 7.5 (Logan et al., 2009). pH_{min} was set at 5.0 from André (Personal communication) showing no growth at pH 5.0 in green beans and possible growth at pH 5.2 in diverse instances.

3.2.2.f Sensitivity analysis

Sobol sensitivity indices were determined by the Saltelli variance-based method (Saltelli, 2002) and computed with the R package sensitivity. This method has been selected for its model independence, its capacity of integrating non-linearity, thresholds and interactions, and of treating untransformed variables (no need of discretization) (Frey et al., 2003; Ellouze et al., 2010).

The sensitivity analysis was performed on the variability parameters from samples of size $N_v=10000$ (drawn from their variability distributions conditionally to a set of uncertain values), and was independently repeated for $N_v=1500$ realizations of uncertainty, leading to an uncertainty distribution on the sensitivity indices. This approach, initially developed by Mokhtari and Frey (2005) and further applied by Pouillot et al. (2007) and Membré et al. (2008), was used with a computation of Sobol indices instead of ANOVA. The model response was *G. stearothermophilus* concentration after sterilization, C_{STERI} (in $\log_{10}\text{CFU/g}$). To focus on the most relevant factors and for computational ease, the *G. stearothermophilus* concentrations C_X were used instead of factors $CATS_x$ and $ratio_x$ (see Eq. (1) in 2.5.1). The validity of the Monte-Carlo estimation of the sensitivity indices was guaranteed by the independence of all factors. Sobol indices are comprised between 0 and 1. The highest Sobol indices indicate factors having the highest influence on the variability of the model response.

3.2.2.g What-if scenarios

What-if scenarios were explored to test the effect of some decisions on the management of the risk of spoilage (scenarios 1 to 12) or of model assumptions (scenarios 13 to 16) on the estimated % of non-stability (Table 3.4). In scenarios 1 to 3, different F_0 values were explored in the range [10; 25] min to test the effect of a decrease in the heat treatment at sterilization (mean F_0 in the reference model = 30.3min). In scenarios 4 and 5, some better hygiene measures were tested, assuming no cross-contamination at the blanching or brining step, in combination with different F_0 values (scenarios 6 and 7). Poor hygiene scenarios corresponding to higher microbial loads at the blanching and/or brining steps were explored (scenarios 8 to 10). The effect of a pH drop, obtained for instance by organic acid addition at brining as applied on some canned vegetables, was tested (scenarios 11 and 12). Scenarios 13 to 15 tested the impact of different germination rates. Finally, the effect of assuming independence between the hyperparameters of D_{ref} , z_T and z_{pH} (see Section 3.2.2.e and Tables 3.2 and 3.3) was tested in scenario 16.

Table 3.4. What-if-scenarios and corresponding non-stability rates (estimation [95% CI]). Parameters are described in Table 3.2 or in Section 3.2.2.e. The non-stability rates were evaluated on the variability dimension. The point estimate and the 95%CI were respectively provided through the 50th, 2.5th and 97.5th percentiles evaluated in the uncertainty dimension.

	What-if-scenario	Non-stability rate (%)
0	Reference model	0.5 [0.1; 1.2]
1	$F_0 = 25$ min	0.9 [0.3; 2.1]
2	$F_0 = 20$ min	2.0 [0.8; 4.1]
3	$F_0 = 10$ min	13.9 [8.4; 20.6]
4	$N_{BLANcc} = 0$ CFU	0.2 [0.0; 0.7]
5	$N_{JUICEcc} = 0$ CFU	0.4 [0.1; 1.1]
6	$F_0 = 20$ min and $N_{BLANcc} = 0$ CFU	1.1 [0.4; 2.5]
7	$F_0 = 25$ min and $N_{BLANcc} = 0$ CFU	0.5 [0.1; 1.2]
8	$CATS_{blwater} \sim N(4.3; 1.2) \log_{10}CFU/g$	1.6 [0.6; 3.1]
9	$CATS_{brine} \sim N(2.4; 1.1) \log_{10}CFU/g$	1.1 [0.4; 2.2]
10	$CATS_{blwater} \sim N(4.3; 1.2) \log_{10}CFU/g$ and $CATS_{brine} \sim N(2.4; 1.1) \log_{10}CFU/g$	1.9 [0.9; 3.5]
11	$pH_{steri} = 5.55$ and $pH_{FP} = 5.25$	0.2 [0.0; 0.7]
12	$pH_{steri} = 5.45$ and $pH_{FP} = 5.15$	0.1 [0.0; 0.6]
13	$p_{germ} = 0.90$	0.5 [0.1; 1.3]
14	$p_{germ} = 0.10$	0.2 [0.0; 0.7]
15	$p_{germ} = 0.01$	0.1 [0.0; 0.3]
16	Independence between the hyperparameters of D_{ref} , z_T and z_{pH} (see Tables 2 and 3)	0.5 [0.1; 1.4]

3.2.3 Results

3.2.3.a Fitted distributions of *G. stearothermophilus* from microbiological analyses at several processing steps

Specific microbiological analyses performed at several processing steps (Section 3.2.2.b) consistently showed the presence of significant populations of spores of aerobic thermophilic bacteria (see adjusted distributions in Table 3.2). Among those, a high percentage of *G. stearothermophilus* (often $\geq 30\%$ of the ATS) was detected in the surveyed samples. For instance, the concentration of *G. stearothermophilus* spores in raw unprocessed green beans was estimated at a mean of $-0.1 \log_{10}\text{CFU/g}$ with a 95 % variation range $[Q_{0.025}, Q_{0.975}] = [-1.9, 1.7] \log_{10}\text{CFU/g}$ (Table 3.5).

3.2.3.b Predicted changes in concentrations during processing and non-stability prevalence

Two processing steps have a major impact on changes in the concentrations in *G. stearothermophilus* (Figure 3.3). These are the blanching step, which results in a global increase in the bacteria concentration due to cross contamination (the inactivation during this process operation is negligible: -1% on average), and the sterilization, which results in a high microbial inactivation. Clear-cut situations were observed during the 7-day incubation at 55°C : either growth leading to spoilage or no growth at all.

Table 3.5. Statistics on the *G. stearothermophilus* evolution along the food pathway, distinguishing uncertainty and variability, in all cans and in positive (i.e. containing at least one CFU) cans.

	Initial	Blanching	Brining	Sterilisation		Incubation	
Prevalence (%)	99.8 [91.4; 100.0]	100.0 [99.9; 100.0]	100.0 [100.0; 100.0]	0.5 [0.1; 1.3]		0.5 [0.1; 1.3]	
	all cans	all cans	all cans	all cans	positive cans	all cans	positive cans
$Q_{0.025}$ concentration ($\log_{10}\text{CFU/g}$)	-1.9 [-3.0; -0.8]	-0.7 [-1.2; -0.2]	-0.5 [-0.9; -0.2]	-3.2 [-3.2; -3.2]	-2.9 [-2.9; -2.6]	-3.2 [-3.2; -3.2]	2.8 [-2.9; 9.4]
Mean concentration ($\log_{10}\text{CFU/g}$)	-0.1 [-0.5; 0.4]	1.0 [0.6; 1.4]	0.9 [0.6; 1.3]	-3.2 [-3.2; -3.2]	-2.1 [-2.5; -1.6]	-3.2 [-3.2; -3.1]	8.4 [4.8; 9.4]
$Q_{0.975}$ concentration ($\log_{10}\text{CFU/g}$)	1.7 [0.6; 3.7]	3.0 [2.3; 3.8]	2.8 [2.2; 3.6]	-3.2 [-3.2; -3.2]	-0.8 [-2.0; 0.6]	-3.2 [-3.2; -3.2]	9.0 [8.5; 9.5]

Notes: Estimation [95% CI] of the Prevalence, the Mean concentration, the Quantiles 0.025 and 0.975 of the concentration. The 4 statistics (prevalence, mean, $Q_{0.025}$ and $Q_{0.975}$ of the concentration) were evaluated on the variability dimension. The point estimate and the 95%CI were respectively provided through the 50th, 2.5th and 97.5th percentiles evaluated in the uncertainty dimension. The

CHAPITRE 3. Modèle d'évaluation du risque

concentration in $\log_{10}\text{CFU/g}$ in a can containing less than 1 spore (or vegetative cell) was assumed to be the concentration corresponding to 0.5 spore (or vegetative cell), to avoid undetermined log concentrations, and assumed to correspond to $-2.9 \log_{10}\text{CFU/g}$ at the initial and blanching steps, and to $-3.2 \log_{10}\text{CFU/g}$ at the brining, sterilisation or incubation steps.

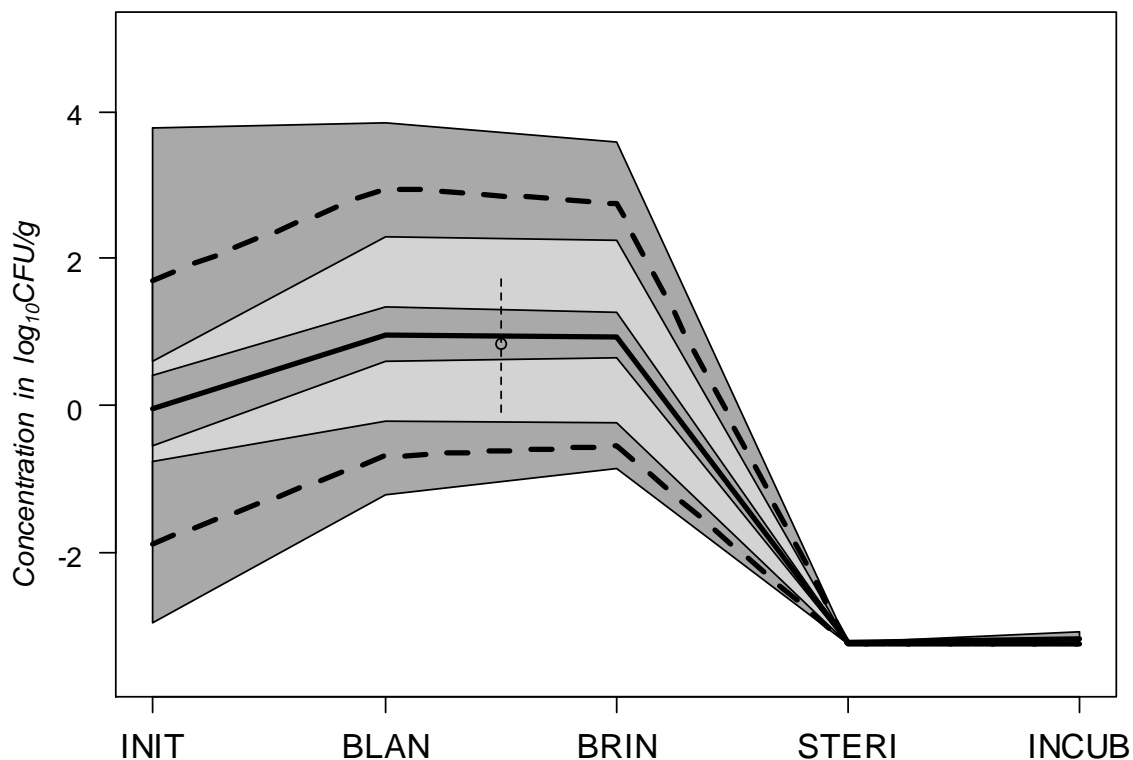


Figure 3.3. Concentration in *G. stearothermophilus* modeled at the different steps *X* of the food chain (distinguishing uncertainty and variability). The solid bold line represents the mean concentration, the dashed bold lines represent the limits of the 95% variability interval, and the dark grey zones represent the 95% uncertainty intervals on these statistics. The small circle represents the observed concentration measured in green beans just before filling (see Section 2.3.1), and the vertical dashed segment represents the corresponding 95%CI. The concentration in $\log_{10}\text{CFU/g}$ in a can containing less than 1 spore (or vegetative cell) was assumed to be the concentration corresponding to 0.5 spore (or vegetative cell), to avoid undetermined log concentrations, and assumed to correspond to $-2.9 \log_{10}\text{CFU/g}$ at the initial and blanching steps, and to $-3.2 \log_{10}\text{CFU/g}$ at the brining, sterilisation or incubation steps.

CHAPITRE 3. Modèle d'évaluation du risque

Table 3.5 gives statistics on the evolution of the *G. stearothermophilus* concentration, distinguishing uncertainty and variability in the results. Prevalence, which measures the proportion of cans containing at least one CFU is equal to 99.8% in unprocessed green beans, increases to 100.0% after blanching and decreases to 0.5% after sterilization. The variability of the concentrations in 95% of the cans is high before sterilization (e.g., 95% variation range= $[-0.7; 3.0]$ \log_{10} CFU/g at blanching]). After sterilization, the expected concentration in 95% of the cans is $-3.2 \log_{10}$ CFU/g, corresponding to less than one spore per can. In positive cans (i.e. cans containing at least one spore), the *G. stearothermophilus* concentration is quite variable (95% variation range= $[-2.9; -0.8]$ \log_{10} CFU/g, which corresponds to 1 to 135 CFU per can) (Table 3.5 and Figure 3.4). The uncertainty associated with all these different estimations is relatively low, except for the concentrations in the positive cans after incubation (95% credible interval on the mean concentration of $[4.8, 9.4]$ \log_{10} CFU/g). The weight of variability is higher than the weight of uncertainty in the dispersion of concentrations after sterilization (Figure 3.4), justifying the interest of performing a sensitivity analysis with regards to variability sources.

The model predicts that about 92% of the cans containing at least one spore after sterilization will develop spoilage. This is favored by the high growth ability of germinated spores, since 99.9% of the cans containing at least one vegetative cell after spore germination will develop spoilage. Before 72h and mostly around 24h, *G. stearothermophilus* concentration in positive cans exceeds $7 \log_{10}$ CFU/g, which corresponds to the minimal concentration required for spoilage (Figure 3.5). Finally, the predicted non-stability rate is of 0.5% with a 95% uncertainty interval of $[0.1\%; 1.2\%]$. This rate is very close to the prevalence rate after sterilization (0.5%, 95%CI= $[0.1\%; 1.3\%]$).

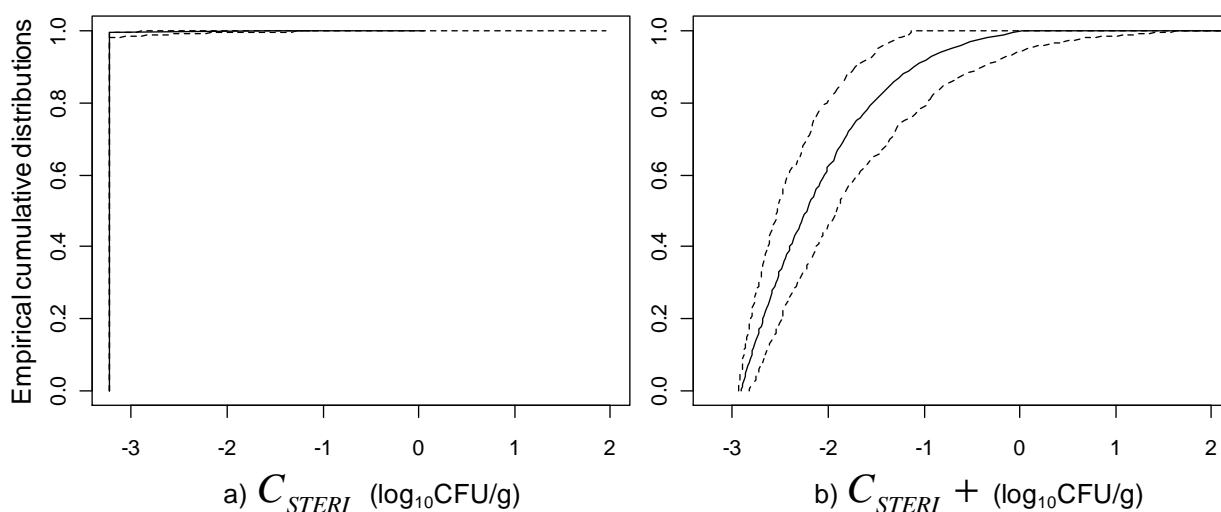


Figure 3.4. Variability Empirical Cumulative Distributions (ECD) of the *G. stearothermophilus* concentrations just after sterilization (solid line) and 95% uncertainty interval (range between dashed lines), in all cans (a) and in positive cans (b). Quantiles were evaluated on the variability dimension; the x-range between dashed lines corresponds to the 95% uncertainty range on each quantile of variability. The point estimate and the 95%CI were respectively provided through the 50th, 2.5th and

CHAPITRE 3. Modèle d'évaluation du risque

97.5th percentiles evaluated in the uncertainty dimension. The plots were obtained with $N_u=500$ and $N_v=10\,000$. To avoid undetermined log concentrations, the concentration in $\log_{10}\text{CFU/g}$ in a can containing less than 1 spore was assumed to be the concentration corresponding to 0.5 spore, which corresponds to $-3.2 \log_{10}\text{CFU/g}$ at the sterilisation step.

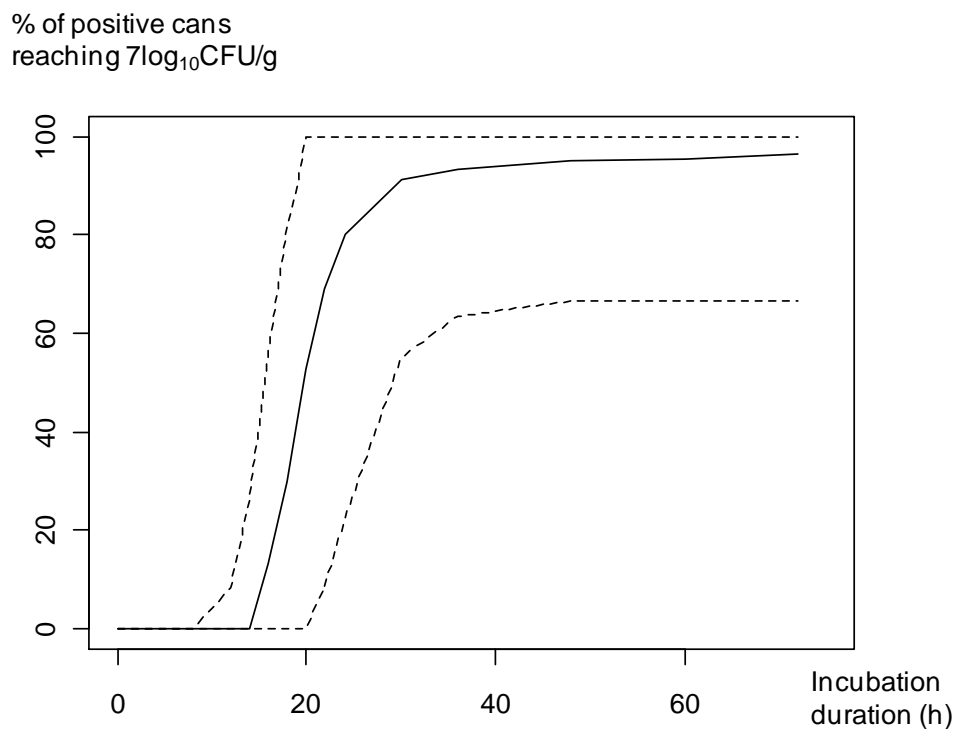


Figure 3.5. Percentage of positive cans reaching $7 \log_{10}\text{CFU/g}$ (corresponding to spoilage) as a function of the incubation duration. Solid lines represent the percentage estimation and dashed lines indicate the 95%CI, representing uncertainty around the estimation. The percentages were evaluated on the variability dimension. The point estimate and the 95%CI were respectively provided through the 50th, 2.5th and 97.5th percentiles evaluated in the uncertainty dimension.

3.2.3.c Validation of the model with independent data

These results can be compared with two types of observed data, independently collected, as elements of model validation. The first data is the *G. stearothermophilus* concentration in green beans immediately before can filling, defined as the product of the measured ATS concentration and the measured *G. stearothermophilus* ratio (see Section 2.3.1) (Figure 3.3). This concentration ($0.9 \log_{10}\text{CFU/g}$, $95\%CI=[-0.1; 1.8]$) is slightly under the modeled mean concentration before filling ($1.0 \log_{10}\text{CFU/g}$) and is totally included in the predicted 95% variation range ($[-0.7; 3.0] \log_{10}\text{CFU/g}$), which is quite satisfying.

CHAPITRE 3. Modèle d'évaluation du risque

The other type of data compared with the predictions is the observed percentage of non-stability at 55°C due to *G. stearothermophilus* reported by green bean processors. The general percentage of non-stability (AFNOR, 1997) recorded on a ten-year survey by 13 processors was equal to 1.5% with a 95% Clopper-Pearson proportion confidence interval (CPpCI) of [1.3%; 1.6%] for a total of approximately 63,000 tested green bean cans (André, CTCPA, unpublished data). The observed percentage of non-stability specifically due to *G. stearothermophilus* was obtained by multiplying the global percentage of non-stability by the percentage of non-stability specifically due to *G. stearothermophilus* in canned green beans. This latter percentage is approximately equal to 58% (with a 95% CPpCI of [33%; 80%]) (André et al., 2012b). Consequently, the general percentage of non-stability due to *G. stearothermophilus* was estimated at 0.9%, with a 95%CI of [0.5%; 1.2%]. The specific observed percentage of non-stability at 55°C of the 643 tested green bean cans at the period and in the plants where the microbiological analyses were performed was slightly less (not mentioned for confidentiality reasons) than the general percentage of non-stability. Consequently the specific observed percentage of non-stability specifically due to *G. stearothermophilus* was slightly less than 0.9%. This latter observed rate has to be compared to the modelled non-stability rate which was 0.5% (95%CI = [0.1%; 1.2%]). The two rates are very close and the credible intervals largely overlap.

3.2.3.d Determination of the most influential factors by sensitivity analysis

The sensitivity analysis based on Sobol index calculation (Table 3.6) allows among variable factors the determination of those having a major influence on the variability of the *G. stearothermophilus* concentration in green bean cans after sterilization. The variability range of each factor is the one defined in the model. The *G. stearothermophilus* heat resistance parameter D_{ref} is the most influential factor. By itself it explains 70% of the response variability ($S_i = 0.70$ with 95%CI=[0.51; 0.87]) and also has a high influence when interacting with other factors as shown by a high St_i at 0.90. The second most influential parameter is the *G. stearothermophilus* heat resistance parameter z_T . Its influence is linked to interactions with other factors as St_i (which accounts for interactions between factors) is equal to 0.18 and S_i is equal to only 0.05. The computation of second-order indices (data not shown) logically showed that the major interaction was between D_{ref} and z_T ($S_{ij}^2 \approx 0.10$). Four factors also have a significant (but much less pronounced) influence on the variability of the model response: the value of sterilization F_0 , the product pH during sterilization pH_{steri} , the heat resistance parameter z_{pH} and the *G. stearothermophilus* contamination in blanching water $C_{blwater}$ (respectively $St_i = 0.04$, $St_i = 0.04$, $St_i = 0.02$ and $St_i = 0.02$). The sources of uncertainty of the model parameters do not question the major influence of D_{ref} and z_T because the 95%CI on their St_i do not (or only marginally) overlap with other factors. However a noticeable exception is z_{pH} , with a large 95%CI on St_i suggesting a rather uncertain influence (from null to moderate) on the variability of the model.

CHAPITRE 3. Modèle d'évaluation du risque

Table 3.6. Sensitivity analysis: estimation [95% CI] of the total effect indices St_i , the first-order indices S_i . The indices were evaluated on the variability dimension. The point estimate and the 95%CI of the indices were respectively provided through the 50th, 2.5th and 97.5th percentiles evaluated in the uncertainty dimension. Values higher than 1 or negative are a consequence of calculation approximations.

Factors	St_i	S_i
D_{ref}	0.90 [0.69; 1.14]	0.70 [0.51; 0.87]
z_T	0.18 [0.07; 0.32]	0.05 [-0.05; 0.15]
F_0	0.04 [0.00; 0.10]	0.02 [-0.04; 0.08]
pH_{steri}	0.04 [0.00; 0.10]	0.02 [-0.03; 0.08]
z_{pH}	0.02 [-0.02; 0.15]	0.01 [-0.03; 0.08]
$C_{blwater}$	0.02 [-0.01; 0.05]	0.01 [-0.02; 0.05]
C_{brine}	0.00 [-0.02; 0.02]	0.01 [-0.01; 0.03]
C_{INIT}	0.00 [-0.01; 0.03]	0.01 [-0.01; 0.03]
t_{blan}	-0.01 [-0.01; 0.00]	0.01 [0.00; 0.01]
T_{blan}	-0.01 [-0.01; 0.00]	0.01 [0.00; 0.01]
pH_{blan}	-0.01 [-0.01; 0.00]	0.01 [0.00; 0.01]

3.2.3.e What-if scenarios to test the influence of microbiological phenomena and options of spoilage risk management

Model predictions satisfactorily fitted with some observations (percentage of non-stability, concentrations at one processing step) (Section 3.2). Consequently the model can be used to test the consequences of some new scenarios not initially described by the model (Table 3.4).

The non-stability rate markedly decreases when the F_0 increases, with for instance a non-stability rate of 13.9% for a F_0 of 10min and of 0.9% for a F_0 of 25min (Table 3.4, scenarios 1 to 3). Scenarios 4 and 5 show that assuming no cross-contamination during filling results in very little change in the response, whereas the impact of no cross-contamination at blanching results in a decrease in non-stability by a half (medians of 0.2% versus 0.5%). Scenarios 6 to 7 show that improvement of hygiene (i.e. absence of cross-contamination at blanching) could allow a reduction of the sterilization value F_0 with no or only a slight increase in the non-stability rate. Conversely, higher microbial loads at the blanching and/or the brining steps (hypotheses of a 2-log increase in ATS contaminations) caused up to four-times higher rates of non-stability (scenarios 8 to 10). A slight decrease in pH by 0.25 or 0.35 on average resulted in a marked decrease in the non-stability rates (scenarios 11 and 12). Scenarios 13 to 15 show that only a high decrease in the germination rate causes a significant decrease in the non-stability rate. The impact of assuming independence between the hyperparameters of the heat resistant parameters D_{ref} , z_T and z_{pH} is also very small, with only a slight increase in the top limit of the 95%CI representing uncertainty.

3.2.4 Discussion

The framework usually developed for the assessment of the microbial risk of foodborne poisoning was applied to the assessment of the risk of spoilage of canned food. The model estimated the changes in *G. stearothermophilus* concentration along a canned green-bean processing chain and gave a satisfactory prediction of the risk of non-stability reported by green bean canners. The predicted non-stability rate of 0.5%, with 95%CI = [0.1%; 1.2%]) was close to the estimation of the industrial non-stability risk due to *G. stearothermophilus*, which was slightly less than 0.9%.

Consequently the model was considered suitable for testing the consequences of some risk management options or of some biological assumptions. Increasing the hygiene during processing (corresponding to an absence of *G. stearothermophilus* contamination at the blanching step) reduced the % of non-stability by a half or kept it at the same level with a simultaneous reduction in the F_0 value. Approximately the same reduction was obtained by an increase in F_0 of 5 min in the range $F_0 = 10$ to $F_0 = 25$ min. A pH decrease could also interestingly reduce the percentage of non-stability. On the contrary, a poor respect of hygiene allowing high microbial loads in blanching water or in covering brine may lead to significant increase in non-stability rates. Moreover, the model hypothesized a high germination rate, meaning that most spores of *G. stearothermophilus* contaminating a can of green beans after sterilization would be able to germinate and grow. Germination and outgrowth of spores of *Bacillus* species and of *G. stearothermophilus* is generally impaired by heat stress, as shown by difficult recovery in suboptimal growth conditions (Feeherry et al., 1987; Leguerinel et al., 2005). That is why the consequences of different germination assumptions were tested. As a result, only scientific evidence of a very low germination rate of surviving spores after heat treatment could lead to an integration of a more sophisticated model considering a correlation between heat shock intensity and probability of germination. The influence of the growth parameters could also have been tested, as only a few data are currently available. However these few data consistently show a rapid growth of *G. stearothermophilus* in foods and media at a range of pH similar to the one considered in this work, followed by a rapid spoilage (pH decrease).

Additionally, the model could be improved by new data and/or scientific knowledge. A possible additional cross-contamination source is the release of green bean debris during the transfer of green beans along the food production chain. This debris has a much higher ATS concentration (mean at $1.7 \log_{10}$ CFU/g, with a 95% variation range = [-2.0, 5.4] according to experimental measures) than do fresh unprocessed green beans (Table 3.2). It sporadically cross-contaminates cans, but the rate of cross-contamination (% of contaminated cans and % of debris in those cans) is completely unknown. Consequently this was not integrated into the model. The primary inactivation model was assumed to be log-linear, although shoulders or tails are sometimes observed on survival curves; this is the most common assumption in the food industry (Rigaux et al. 2012). The absence of a selective medium for *G. stearothermophilus* led to dealing with ATS concentrations and confirmation tests.

The results from the concentration evolution, the sensitivity analyses and the what-if-scenarios are consistent and identify important steps and risk factors. The sterilization process is crucial (Figure 3.3, Table 3.4 and 3.6), resulting in a very high spore inactivation. Consequently, it is not surprising that the factors implicated in the sterilization model are the most influential factors

CHAPITRE 3. Modèle d'évaluation du risque

according to the sensitivity analysis (D_{ref} , z_T , pH_{steri} , F_0 , z_{pH}) (Table 3.6). The *G. stearothermophilus* decimal reduction time D_{ref} is actually so influential that alone or in interaction with the other factors, it explains 90% of the variability of the *G. stearothermophilus* concentration after sterilization. The second and the fifth most influential factors are also heat resistance parameters of *G. stearothermophilus* (z_T and z_{pH}). These parameters cannot be modified, which consequently restricts possible management actions to limit spoilage risk. The fourth most influential factor, pH_{steri} , would also be difficult to modify. Nevertheless, the F_0 , as well as the cross-contamination from blanching water ($C_{blwater}$), appears to be quite influential and constitutes possible points of management intervention (as illustrated in what-if-scenarios 1 to 7).

Sensitivity analysis is increasingly used in risk assessment, in one-dimensional models (Zwietering et al., 2000; Frey et al., 2003; Mokhtari et al., 2005; Ellouze et al., 2010) or in two-dimensional models (Pouillot et al., 2007; Membré et al., 2008; Mataragas et al., 2010; Busschaert et al., 2011). Sensitivity analysis or what-if-scenarios may be used to identify the highest sources of variability and of uncertainty in the final model. They can also be used right from the start of the construction of the model for prioritizing additional research, improving or simplifying the model. For instance in this work the sensitivity analysis showed the importance of the product pH during sterilization (pH_{steri}), and led to additional pH measures to improve the knowledge about this environmental parameter. What-if-scenarios 13 to 16 explored the effect of some modeling assumptions and in particular showed that a complex germination model was certainly unnecessary. Based on the results of the sensitivity analysis (Table 3.6), it would be possible to fix some parameters which have little influence, such as the blanching time and temperature.

To conclude, we underline the interest of constructing a model to learn a lot about a risk in a food chain. Indeed, model construction encourages questioning about the microbial processes involved in the food chain. It also encourages additional research to improve the model and reduce the sources of uncertainty, particularly when some parameters are shown (via a sensitivity analysis) to have a strong influence in the model. Once constructed, this model can be used to study the risk factors and to help risk managers to control the microbial risk of foodborne poisoning or spoilage.

Acknowledgments

This work is a partial fulfillment of author C. Rigaux's PhD Thesis and has been supported by Agence National de la Recherche Paris, France under contract ANR-09-ALIA-014 (Ribenuit project). Thanks are due to Dr Olivier Couvert (LUBEM, Quimper, France) and to Jean-Baptiste Denis (INRA, Jouy-en-Josas, France) for kind communication of data and helpful scientific discussions. Thanks are also due to Rachel Galland (Bonduelle, France) and to other industrial partners for kindly providing technical information about process.

3.3 Compléments à l'article

3.3.1 Des « what-if-scenarios » complémentaires

De nouveaux scénarios, non inclus dans l'article, ont été explorés dans cette partie. Le premier scénario testé permet de tester une hypothèse du modèle entachée d'une certaine incertitude : la valeur du taux d'imprégnation ou de « mouillage » des haricots verts dans l'eau de blanchiment p_{soak} (scénario 1). Les scénarios 2 à 8 testent l'effet de contaminations en ATS plus élevées que dans le modèle de référence dans les haricots frais ou pour les sources de contamination croisées. L'intérêt est de voir si de telles hausses de contaminations pourraient expliquer les taux de non-stabilité élevés observés sur certaines usines certaines années.

Les résultats de ces scénarios sont donnés dans le Tableau 3.7. L'effet d'une diminution d'un facteur 10 du taux d'absorption p_{soak} ne diminue que légèrement le taux final de non-stabilité (0.3% en moyenne au lieu de 0.5%), ce qui relativise l'importance de ce facteur (scénario 1). Les augmentations des diverses sources de contamination (initiale et croisées) mènent logiquement à des augmentations des taux de non-stabilité (scénarios 2 à 8). Des concentrations en ATS à 3 \log_{10} UFC/g dans les haricots verts frais ou dans le jus mènent notamment à des taux de non-stabilité de 1.2% en moyenne, proches du taux de non-stabilité général de 1.5% observé sur 13 usines sur une dizaine d'années. Une hygiène améliorée a une faible influence sur les taux de non-stabilité. Cependant de trop grandes négligences sur ce point augmentent le taux de non-stabilité de manière sensible.

Tableau 3.7. *What-if*-scénarios et taux de non-stabilité correspondants (estimation, [95% CI]). Les paramètres sont décrits dans le tableau 3.2 ou en section 2.3.3. Les taux de non-stabilité sont évalués dans la dimension variabilité. L'estimation est le 95%CI proviennent respectivement du 50th, 2.5th et 97.5th percentiles évalués dans la dimension incertitude.

	<i>What-if</i> -scénario	Taux de non-stabilité (%)
0	Modèle de référence	0.5 [0.1; 1.2]
1	$p_{soak} = 0.27\%$	0.3 [0.1 ; 0.9]
2	$CATS_{init} = 3 \log_{10}$ CFU/g	1.2 [0.4 ; 2.6]
3	$CATS_{init} = 5 \log_{10}$ CFU/g	3.8 [1.8 ; 6.8]
4	$CATS_{brine} = 3 \log_{10}$ CFU/g	1.2 [0.4 ; 2.6]
5	$CATS_{brine} = 5 \log_{10}$ CFU/g	3.6 [1.8 ; 6.2]
6	$CATS_{blwater} = 3 \log_{10}$ CFU/g	0.5 [0.1 ; 1.2]
7	$CATS_{blwater} = 5 \log_{10}$ CFU/g	2.0 [0.8 ; 4.0]
8	les 3 sources de contamination à $5 \log_{10}$ CFU/g	4.2 [2.1 ; 7.3]

3.3.2 Contamination croisée due aux débris

Une source de contamination croisée possible, non prise en compte dans le modèle présenté dans l'article, provient des débris de haricots verts. Ces débris sont des fragments de haricots verts qui peuvent stagner le long de la chaîne de production, puis être distribués ponctuellement dans le flot de haricots verts. Contrairement aux autres sources de contamination croisée modélisées (jus et eau de blanchiment), il s'agit d'une contamination croisée ponctuelle, qui va affecter seulement un nombre réduit de boîtes. Un sous-modèle a été proposé pour traduire ce phénomène, mais il n'a pas été pris en compte dans l'article car il reposait sur des hypothèses trop incertaines. Une partie des données nécessaires à la construction de ce sous-modèle étaient en effet manquantes ou rares, et difficiles à collecter. Des analyses microbiologiques sur des débris ont cependant montré des niveaux de contamination en *ATS* et en *G. stearothermophilus* élevés, plus élevés notamment que ceux du flot principal de haricots verts, ce qui justifie l'intérêt de s'intéresser à cette source de contamination croisée. La structure théorique imaginée pour ce sous-modèle à partir d'opinions d'experts est présentée ici, et pourrait être enrichie à l'avenir avec des données.

La présence de débris dans une boîte a été considérée comme aléatoire, ainsi que la quantité w_{frag} (g) de débris dans une boîte. Le modèle représentant la distribution des débris a été conçu en se représentant d'abord la contamination de la chaîne de production tout entière. Ainsi, la proportion $p_{frag\ chain}$ de débris de haricots verts dans la masse entière de haricots verts présents sur la chaîne de production a été supposée varier entre 0.001% et 0.1%, avec une valeur la plus vraisemblable de 0.01% (proportion modélisée par une loi BetaPert). Ainsi sur 1000 kg de haricots verts, le plus probable sera de retrouver 100 g de haricots verts sous forme de débris. Un facteur de dilution f_{dil} modélise le degré de dispersion des débris dans les boîtes lors de l'étape d'emboitage. Une valeur du facteur f_{dil} de 1 correspond à une situation où tous les débris de la chaîne se retrouvent dans le plus petit nombre n_{min} de boîtes de conserves possible (par exemple 3 boîtes de contenance de 445g chacune si il y a 1100g de débris à répartir), et une valeur de f_{dil} de 20 correspond à une répartition de débris dans $20 \times n_{min}$ boîtes de conserve. La valeur la plus vraisemblable pour le facteur de dilution f_{dil} a été supposée de 10. $1/f_{dil}$ correspond approximativement à la proportion moyenne de débris dans une boîte contenant des débris, le poids moyen de débris dans ces boîtes est alors d'environ $w_{frag} = w_{GB}/f_{dil}$ où $w_{GB} = 445g$ est le poids de haricots verts dans une boîte de conserve. La présence de débris dans une boîte prise en hasard est alors de probabilité $p_{frag\ chain} \times f_{dil}$. La présence de débris dans une boîte a donc été modélisée par une loi de Bernoulli de probabilité $p_{frag\ chain} \times f_{dil}$. La concentration de *G. stearothermophilus* dans les débris étant notée C_{frag} (UFC/g), la quantité $N_{TRANScc}$ de spores (UFC) apportées par les débris dans une boîte quelconque prise au hasard est finalement définie par :

$$N_{TRANScc} = Ber(p_{frag\ chain} \cdot f_{dil}) \cdot C_{frag} \cdot w_{frag}$$

où $Ber(p)$ est la loi de Bernoulli de probabilité p .

Par ailleurs, une boîte ayant reçu un poids w_{frag} de débris recevra une masse de $445 - w_{frag}$ g de haricots verts normaux, afin que le poids total de haricots verts dans cette boîte soit égal au poids des haricots verts dans toutes les boîtes (445g). Enfin, une troncature a été ajoutée afin de s'assurer que la concentration en *G. stearothermophilus* des débris était supérieure ou égale à la concentration en *G. stearothermophilus* des haricots verts.

CHAPITRE 3. Modèle d'évaluation du risque

La concentration C_{frag} en *G. stearothermophilus* dans les débris est définie par le produit de la concentration en ATS dans les débris ($CATS_{frag}$) et de la proportion de *G. stearothermophilus* parmi les ATS ($ratio_{frag}$). Comme pour les autres sources de contamination, $CATS_{frag}$ et $ratio_{frag}$ ont été estimés à partir d'analyses microbiologiques effectuées sur les chaînes de productions des haricots verts (partie 3.2.2.b) (n=41 données). Le Tableau 3.8 présente les distributions attribuées aux paramètres impliqués dans ce sous-modèle de contamination croisée.

Tableau 3.8. Paramètres du sous-modèle de contamination croisée par les débris : description, distributions de variabilité et d'incertitude, moyenne et 95% CI de la distribution marginale globale (voir Tableau 3.2).

Paramètre	Description (unité)	Distribution de variabilité	Distribution d'incertitude	Moyenne et plage de variation à 95%
$CATS_{frag}$	Concentrations en ATS dans les débris de haricots (\log_{10} UFC/g)	$N(M_{CATSf}; S_{CATSf})$	$M_{CATSf} \sim N(1.69; 0.29)$ $S_{CATSf} \sim LnN(0.59; 0.17)$	1.7 [-2.0; 5.4]
$ratio_{frag}$	Proportion de <i>Gbs</i> parmi les ATS dans les débris de haricots verts	-	$B(24; 29)$	0.45 [0.32; 0.59]
$p_{frag\ chain}$	Proportion de débris dans les haricots de la chaîne de production	$BP(0.001\%; ML_{pf}; 0.1\%)$	$ML_{pf} \sim$ $BP(0.005\%; 0.01\%; 0.05\%)$	0.03% [0.00%; 0.07%]
f_{dil}	Facteur de dilution des débris	-	$BP(5; 10; 20)$	10.7 [6.2; 16.2]

Quand ce sous-modèle est intégré au modèle global juste avant l'étape de jutage, les résultats de non-stabilité changent très peu (taux de non-stabilité toujours de 0.5% en moyenne avec un intervalle d'incertitude à 95% de [0.1%; 1.2%]). Ce faible impact des débris semble dû à la très faible proportion de boîtes recevant des débris dans le modèle. Si la proportion $p_{frag\ chain}$ est augmentée à 0.1%, le taux de non-stabilité reste inchangé et si $p_{frag\ chain}$ est augmentée à 1%, le taux de non-stabilité augmente seulement très légèrement (0.5%, 95%CI=[0.1%; 1.3%]). L'impact des débris semble donc faible mais la proportion réelle de débris et le facteur de dilution sont cependant très difficiles à évaluer sans données et il semble donc difficile de tirer des conclusions de ce sous-modèle avant d'avoir pu l'enrichir par de nouvelles informations plus fiables. Il constitue néanmoins une base pour de futurs travaux sur la prise en compte de la contamination croisée.

Chapitre 4

Modélisation de l'évolution de la concentration en vitamine C au cours de la fabrication et de l'entreposage de conserves de haricots verts

Ce chapitre traite de la modélisation du bénéfice nutritionnel associé à la consommation de haricots verts appertisés. Le produit et la chaîne de production sont les mêmes que ceux présentés dans le modèle d'analyse de risque microbiologique au chapitre 3.

4.1 Introduction

4.1.1 Contexte

L'analyse quantitative du bénéfice nutritionnel le long de chaînes de production alimentaires est nettement moins développée que l'analyse quantitative du risque microbiologique. Néanmoins, il existe de nombreuses études sur les processus de dégradation des vitamines ou sur l'impact de certains procédés de fabrication comme le blanchiment ou la surgélation (Selman, 1994; Dewettinck et al., 2006; Sablani et al., 2006; Tosun and Yucecan, 2007), ainsi que des travaux de modélisation des pertes associées à certains traitements thermiques ou à de la diffusion (Arroqui et al., 2002; Corradini and Peleg, 2006; Leskova et al., 2006; Van Boekel, 2008; Pénicaud, 2009). Un objectif crucial de ces études est la recherche de procédés assurant la meilleure préservation possible des nutriments du produit. Les bénéfices nutritionnels d'un produit ont des conséquences importantes en santé publique et dans l'industrie agro-alimentaire, comme en témoignent les campagnes de santé publique type PNNS et la multiplication des allégations nutritionnelles (Verhagen et al., 2010).

Le modèle construit dans ce travail modélise l'évolution de la teneur en vitamine C au cours du procédé de fabrication des conserves de haricots verts, depuis la récolte des légumes jusqu'à

CHAPITRE 4. Modèle d'évaluation du bénéfice

l'entreposage des conserves. Des méthodes de modélisation similaires à celles utilisées pour le modèle d'analyse quantitative du risque microbiologique ont été utilisées, comme la simulation de Monte Carlo à 2 dimensions et l'analyse de sensibilité. Les sources d'incertitudes dans ce modèle étaient plus grandes que dans le modèle d'analyse du risque microbiologique, du fait d'un plus grand manque de données et de la nouveauté de la démarche dans ce domaine. Une fois le modèle construit, ses résultats ont été comparés avec des données indépendantes afin d'évaluer sa validité. Deux types d'analyses de sensibilité ont alors permis d'identifier les facteurs-clés du modèle, et des scénarios ont été explorés afin d'évaluer l'effet de certaines hypothèses ou d'identifier des mesures possibles pour augmenter les teneurs en vitamine C.

4.1.2 La vitamine C

La vitamine C se présente sous deux formes bioactives, toutes deux sources de bénéfices nutritionnels :

- L'acide ascorbique (AA), la forme réduite de la vitamine C, sa forme principale.
- L'acide déshydroascorbique (DHA), la forme oxydée de la vitamine C.

Les pertes en vitamine C proviennent d'une part de *transferts de matière*, provenant de la diffusion des vitamines dans les eaux de cuisson et dans le jus de couverture, et d'autre part de *réactions chimiques de dégradation*.

Les mécanismes de dégradation chimiques de la vitamine C sont très complexes (Lin and Agalloco, 1979; Manso et al., 2001). De nombreux facteurs influencent la dégradation de la vitamine C, tels la température, le pH, la concentration en oxygène, l'intensité de la lumière, l'humidité relative, la présence d'ions métalliques dans le produit et la composition propre de la matrice alimentaire (Lin et al., 1979). De manière générale, la vitesse de la réaction chimique d'oxydation augmente avec la concentration en oxygène dissous, la température et le pH des aliments. La réaction n'est cependant pas simple en raison des possibles interactions entre la température et l'oxygène : plus la température est élevée, plus la concentration en oxygène dissous est faible. Il existe donc une température à laquelle la vitesse apparente de la réaction de dégradation de la vitamine C est maximale. L'équilibre de la concentration en oxygène à une certaine température peut cependant mettre du temps à être atteint. Lors de la dégradation chimique de la vitamine C, un enchaînement de plusieurs réactions parallèles est possible comme décrit dans la Figure 4.1 (Yuan and Chen, 1998). Cet enchaînement de réactions chimiques est présenté en Figure 4.2 sous une forme simplifiée qui sera utilisée dans la modélisation. Ainsi :

- en présence d'oxygène, AA se transforme en DHA par une réaction d'oxydation (vitesse k_1)
- DHA se transforme de manière irréversible en un produit noté P_1 par voie anaérobie, à la vitesse k_2
- DHA peut éventuellement se recycler en AA par réduction (vitesse k_3)

CHAPITRE 4. Modèle d'évaluation du bénéfice

- AA peut éventuellement se dégrader de manière irréversible par voie anaérobie en un produit noté P₂ (vitesse k₄). Toute activité vitaminique est irrémédiablement perdue dans les produits P₁ et P₂.

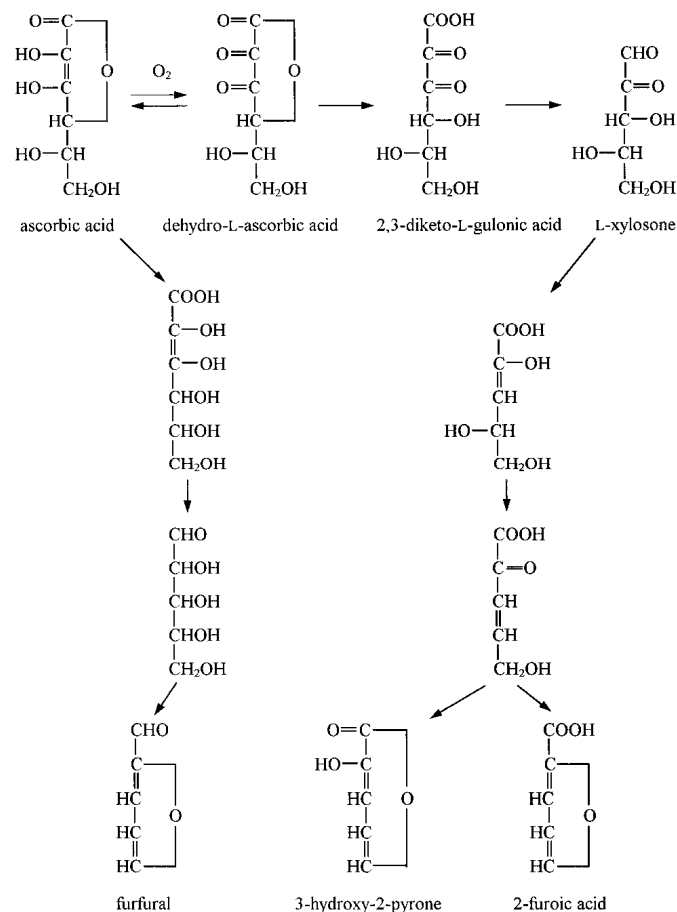


Figure 4.1. Voies de réactions possibles pour la dégradation de l'acide ascorbique. Source : Yuan and Chen (1998).

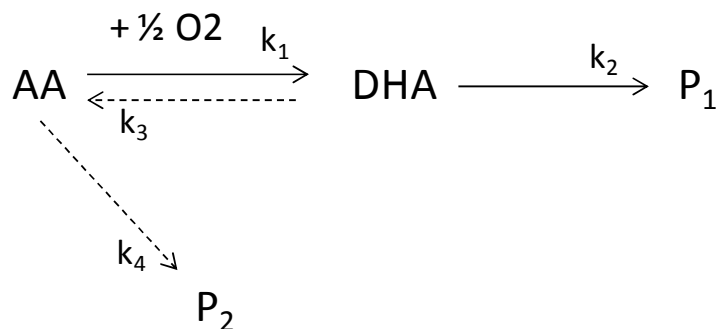


Figure 4.2. Schéma simplifié des réactions possibles de dégradation de l'acide ascorbique prises en compte dans le modèle.

Des données de concentrations en vitamine C ont été utilisées pour construire le modèle. En particulier lors du projet ANR Ribenut, de nombreux échantillons de haricots verts ont été analysés par différents laboratoires en utilisant des méthodes différentes. Des incertitudes, souvent non négligeables, étaient attachées aux mesures des concentrations en vitamine C réalisées et aux proportions de AA et de DHA. En effet, ces données proviennent du dosage de la vitamine C par chromatographie en phase liquide à haute performance (CLHP). Cette méthode permet de séparer les constituants d'un mélange par fractionnement à travers une colonne. Des réactions chimiques sont préalablement provoquées pour faciliter les mesures, par exemple pour transformer tout le AA en DHA ou bien tout le DHA en AA. Ces réactions chimiques ne sont pas facilement maîtrisables et sont une source possible de biais dans les mesures. De plus selon le procédé expérimental utilisé, tel le mode de stabilisation des échantillons, de conservation, de broyage du produit ou la méthode de dosage, une possible oxydation du AA et/ou une dégradation de DHA peuvent modifier les concentrations en vitamine C ou le ratio AA/DHA, entraînant des variations importantes des mesures. Les incertitudes attachées à ces mesures ont été prises en compte dans la mesure du possible dans le modèle (distributions d'incertitudes, confrontation avec des données de diverses sources en partie 4.3.2).

4.2 Modélisation de l'évolution de la vitamine C au cours du procédé de fabrication des conserves de haricots verts

4.2.1 Processus modélisés

La dégradation de AA et de DHA est modélisée de manière séparée, car leur réactivité chimique n'est pas identique. Dans la modélisation, seules sont considérées l'oxydation de AA en DHA et la dégradation anaérobie de DHA (réactions n°1 et n°2 représentées par deux flèches pleines sur la Figure 4.2). Les deux autres réactions possibles n'ont pas été modélisées car elles semblent se produire plus rarement dans les conditions du produit considéré (pH peu acide, milieu non réducteur), et car il y a peu de modèles disponibles dans la littérature jusqu'à présent. Cependant, l'ensemble des différentes réactions possibles a été pris en compte dans une certaine mesure, car les paramètres chimiques des réactions ont été estimés à partir de mesures expérimentales de la vitesse apparente de disparition de AA ou DHA, résultat de l'ensemble des réactions de dégradation (partie 4.2.6.c).

Dans le modèle, la réaction chimique va être modélisée de deux façons selon l'état des haricots verts et les conditions auxquels ils sont exposés :

CHAPITRE 4. Modèle d'évaluation du bénéfice

- La première modélisation correspond à l'évolution de la vitamine C lors de l'évolution physiologique des haricots, c'est-à-dire quand les haricots sont à température ambiante et n'ont pas encore été soumis aux températures élevées qui modifieront leur structure et leur composition, notamment par une inactivation des enzymes. Les haricots sont encore « vivants ».
- La deuxième modélisation correspond à une réaction chimique après inactivation des enzymes, une fois que le haricot a été altéré car soumis à des températures élevées ($\geq 80^{\circ}\text{C}$), lors du blanchiment et de la stérilisation notamment. Par la suite on parlera simplement de « réaction chimique ».

Le Tableau 4.1 résume les processus modélisés, ainsi que les hypothèses principales de modélisation. L'unité de production modélisée est une boîte de conserve de 850 g contenant 445 g de haricots verts et 405 g de jus de couverture. Dans le modèle, la teneur en vitamine C diminue sous l'effet d'une évolution physiologique pendant les phases de préparation des haricots verts après la récolte, puis sous l'effet de réactions chimiques aux différents points chauds de la chaîne et enfin sous l'effet de phénomènes de diffusion dans l'eau de blanchiment et dans le jus de couverture. L'étape de préparation des haricots verts correspond au regroupement de différentes étapes du procédé : temps de séjour de la matière première non transformée, lavage, parage (éboutage, calibrage) et temps d'attente de la matière parée. Les facteurs environnementaux pris en compte dans le modèle sont la température, le temps et la concentration en oxygène. L'effet du pH a été négligé dans ce travail en raison des faibles variations de pH des haricots et à cause du manque de données sur l'effet de ces variations de pH sur l'évolution de la vitamine C.

Tableau 4.1. Procédés et processus chimiques modélisés. Synthèse des différentes hypothèses du modèle.

Etape du procédé (abréviation)	Processus chimiques impactant le taux de vitamine C	Détails et hypothèses
Récolte (INIT)	Concentration initiale	Quantité de vitamine C totale entre 3 et 33mg/100g; proportion de AA proche de 90%
Préparation (WAIT)	Evolution physiologique	<ul style="list-style-type: none"> • Dégradation de AA linéaire en fonction du temps et indépendante de la température ($k_{ev} \approx 0.004\text{mg}/100\text{g}/\text{min}$) • DHA stable (compensation entre pertes propres et production par oxydation de AA) (Howard et al., 1999, Toledo et al., 2003)
Blanchiment (BLAN)	Réaction chimique et diffusion	<ul style="list-style-type: none"> • Diffusion modélisée par la loi de Fick (Crank, 1975; Selman, 1994); comportement supposé identique pour AA et DHA ; vitesse de diffusion dépendant de la température via la loi d'Arrhénius • Réaction chimique : modèles de décroissance exponentielle de AA et DHA, vitesse des réactions dépendant de la température via la loi d'Arrhénius et prise en compte du taux d'O₂ lors de l'oxydation de AA (Pénicaud, 2009) ; taux d'O₂ supposé constant (déterminé par la température via la table de Winkler)
Transfert (TRANS)	Réaction chimique	Réaction chimique : modèles de décroissance exponentielle de AA et DHA, vitesse des réactions dépendant de la température via la loi d'Arrhénius et prise en compte du taux d'O ₂ lors de l'oxydation de AA (Pénicaud, 2009) ; taux d'O ₂ supposé constant (déterminé par la température via la table de Winkler)
Stérilisation (STERI)	Réaction chimique	<ul style="list-style-type: none"> • AA supposé stable en absence d'O₂ • DHA : modèles de décroissance exponentielle ; vitesse de réaction dépendant de la température via la loi d'Arrhénius - détermination d'un temps équivalent de chauffage à la température de référence 85°C
Entreposage (STOCK)	Réaction chimique et diffusion	<ul style="list-style-type: none"> • Réaction chimique : AA consommé par l'O₂ présent dans l'espace de tête ; DHA se dégrade par un modèle de décroissance exponentielle et une vitesse de réaction dépendant de la température via la loi d'Arrhénius • Diffusion homogène entre haricots et jus au terme de l'entreposage (perte des vitamines diffusées dans le jus)

4.2.2 Simulation, incertitude et variabilité des facteurs, modèle global

La méthode de simulation de Monte Carlo à deux dimensions a été utilisée, afin de séparer la propagation dans le modèle de l'incertitude et de la variabilité (voir partie 1.2.2). Les variables d'entrée du modèle sont décrites dans le Tableau 4.2, ainsi que leurs distributions de probabilité, en distinguant les distributions d'incertitude et de variabilité. La Figure 4.3 présente le graphe acyclique dirigé (DAG) du modèle vitamine, donnant un aperçu synthétique du modèle global.

La concentration initiale en vitamine C peut varier en fonction de la variété de haricot vert et des conditions pédoclimatiques (croisement de considérations sur les sols et sur le climat). Elle a donc été supposée variable. La distribution de variabilité a été supposée connue à une incertitude près liée aux incertitudes de mesures et au nombre relativement faible de données utilisées pour ajuster une distribution de probabilité sur cette concentration initiale. En absence d'information fiable laissant penser que la proportion de forme réduite (AA) dans les haricots verts frais est variable, elle a été supposée seulement incertaine.

Les paramètres chimiques comme les vitesses de réactions chimiques à une température de référence, les énergies d'activation, l'ordre de la réaction chimique et les coefficients de diffusivité ont été supposés fixes (non variables) et connus à incertitude près (liée au faible nombre de données). Cette hypothèse correspond à des avis d'experts. Seule la vitesse de la réaction de dégradation lors de l'évolution physiologique a été supposée à la fois variable et incertaine, car les conditions environnementales (température, soleil, humidité etc.) de l'étape de préparation des haricots verts sont assez variables et peuvent influencer notablement la vitesse de dégradation.

Les paramètres comme les durées et températures de process industriel, la quantité d'air emprisonnée dans l'espace de tête de la boîte de conserve, le rayon des haricots verts ont eux été supposés variables d'une boîte de conserve à une autre, ce qui correspond à une réalité industrielle. On a supposé la distribution de variabilité de ces derniers paramètres parfaitement connue (incertitude négligeable car bonne connaissance de l'industriel permettant par ailleurs de simplifier le modèle).

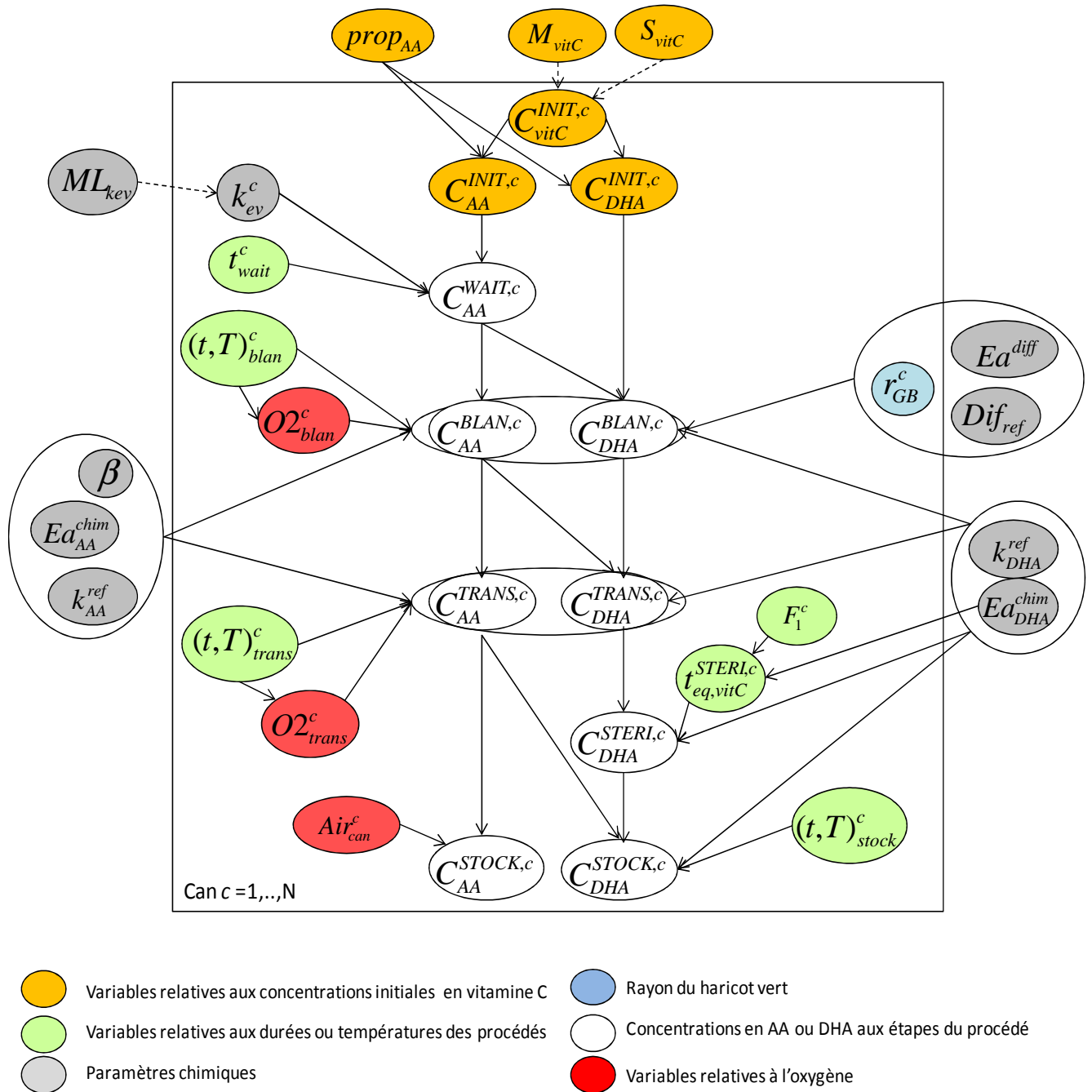


Figure 4.3. DAG du modèle de dégradation des vitamines. Les variables du modèle sont représentées par des ellipses et les dépendances entre les variables sont représentées par des flèches. Une flèche pleine indique un lien déterministe et une flèche pointillée un lien stochastique. Une boucle sur les boîtes (« can » en anglais) représente un niveau de hiérarchie dans le modèle. Les variables indicées par c prennent des valeurs différentes selon les boîtes (dimension variabilité). Les définitions des variables sont données en Tableau 4.2.

CHAPITRE 4. Modèle d'évaluation du bénéfice

Tableau 4.2. Description, et distributions d'incertitude et/ou de variabilité des paramètres d'entrée du modèle.

Paramètre	Description (unité)	Distribution de variabilité	Distribution d'incertitude	Moyenne et intervalle de variation à 95% ^a	
Paramètres chimiques	C_{vitC}^{INIT}	Concentration en vitamine C dans les haricots verts frais (mg/100g)	$N(M_{vitC}, S_{vitC})T(0;100)$	$M_{vitC} \sim N(16.8; 0.8)$ $S_{vitC} \sim LnN(1.93; 0.08)$	17.0 [4.0; 30.6]
	$prop_{AA}$	Proportion de AA sur la vitamine C totale dans les haricots verts frais	-	$BP(0.5; 0.9; 1.0)$	0.85 [0.66; 0.98]
	k_{ev}	Vitesse d'évolution physiologique (mg/100g/min)	$BP(0.001, ML_{kev}, 0.006)$	$MI_{kev} \sim$ $BP(0.002, 0.004, 0.005)$	0.004 [0.002; 0.006]
	β	Ordre de la réaction chimique par rapport à l'oxygène	-	$BP(0.6, 0.9, 1.2)$	0.90 [0.69; 1.11]
	Ea_{AA}^{chim}	Energie d'activation de AA pour la réaction chimique (kJ/mol)	-	$N(44; 4)T(0; 200)$	44.1 [36.1; 52.2]
	k_{AA}^{ref}	Vitesse de la réaction d'oxydation de AA à la température de référence de 85°C ($\text{mol}^{-\beta} \text{L}^{\beta} \text{min}^{-1}$)	-	$BP(0.07; 0.18; 0.43)$	0.20 [0.09; 0.34]
	Ea_{DHA}^{chim}	Energie d'activation de DHA pour la réaction chimique (kJ/mol)	-	$N(80; 14)T(0; 200)$	80.0 [53.1; 108.1]
	k_{DHA}^{ref}	Vitesse de la réaction chimique de DHA à la température de référence de 85°C (min^{-1})	-	$BP(0.05; 0.07; 0.09)$	0.07 [0.05; 0.09]
	Dif_{ref}	Coefficient de diffusivité de la vitamine C dans les haricots verts à la température de référence de 85°C ($10^{-9} \text{m}^2 \text{s}^{-1}$)	-	$BP(1.6; 1.8; 2.1)$	1.82 [1.65; 2.01]
Ea^{diff}	Energie d'activation de la vitamine C pour la diffusion (kJ/mol)	-	$N(20; 5)T(0; 200)$	20.0 [9.7; 29.7]	
Paramètres environnementaux	r_{GB}	Rayon d'un haricot vert fin (mm)	$BP(4; 4.25; 4; 5)$	-	4.3 [4.1; 4.4]
	air_{can}	Volume de l'espace de tête (mL)	$BP(3; 4.5; 6)$	-	4.5 [3.5; 5.6]
	t_{wait}	Durée d'attente (min)	$BP(140; 700; 1260)$	-	694 [295; 1096]
	t_{blan}	Durée de blanchiment (min)	$BP(3.5; 6; 10)$	-	6.3 [4.2; 8.7]
	T_{blan}	Température de blanchiment (°C)	$BP(85; 91; 98)$	-	91.2 [86.8; 96.0]
	t_{trans}	Durée de transfert (min)	$BP(0.05; 0.5; 1)$	-	0.5 [0.2; 0.9]
	T_{trans}	Température de transfert (°C)	$BP(70; 80; 85)$	-	79.1 [73.4; 83.9]
	F_1	Temps standard de dégradation à 85°C de DHA (min)	$LnN(5.30, 0.05)$	-	199.9 [181.1; 219.9]
	t_{stock}	Durée d'entreposage (jour)	$BP(10; 547.5; 1095)$	-	542 [160; 937]
T_{stock}	Température d'entreposage (°C)	$BP(14; 16; 30)$	-	18.0 [14.4; 24.3]	

Notes: $N(a,b)$ (respectivement $N(a,b)T(c,d)$) est la distribution normale de moyenne a et d'écart type

CHAPITRE 4. Modèle d'évaluation du bénéfice

b (resp. tronquée sur $[c,d]$), $LnN(a,b)$ est la distribution lognormale basée sur une loi normale de moyenne a et d'écart type b , et $BP(a,b,c)$ est la loi BetaPert de minimum a , de valeur la plus vraisemblable b et de valeur maximum c . ^aIntervalles déterminés par simulation de Monte Carlo à incertitude et variabilité confondus.

4.2.3 Modélisation des durées et températures du traitement dans le procédé de transformation et de la conservation

Étapes de préparation, blanchiment et transfert

À partir du diagramme de fabrication du produit fourni par l'industriel, des lois Beta-Pert ont été ajustées sur les durées et températures de ces étapes. Elles représentent la variabilité des conditions de fabrication possibles. Ces lois sont définies à partir d'une valeur minimum, d'une valeur la plus vraisemblable et d'une valeur maximum.

Pour l'étape de « préparation » des haricots verts, les temps d'attente de la matière première, du lavage, du parage (éboutage, calibrage) et les temps d'attente des matières parées ont été sommés, en supposant que les haricots ne subissent qu'une évolution physiologique indépendante de la température (cf partie 4.2.5). Le temps d'attente de la matière première varie de 2 h à 16 h ; le lavage prend moins d'une minute ; le temps de parage varie entre 20 min et 1 h ; et le temps d'attente de la matière première parée varie de 0 h à 4 h. La durée totale de l'étape de préparation est ainsi comprise entre 2h30 et 21h.

L'étape de blanchiment dure entre 3,5 min et 10 min, à une température comprise entre 85°C et 98°C. L'étape de transfert dure entre quelques secondes et 1 min, à une température comprise entre 70°C et 85°C.

Étape de stérilisation

Un modèle complexe a été utilisé pour caractériser l'étape cruciale de la stérilisation (en particulier pour la modélisation du risque microbiologique, cf Chapitre 3). À partir de profils temps-températures fournis par l'industriel, un temps équivalent de stérilisation à la température de référence de 85°C a été calculé, puis exprimé en fonction du traitement thermique appliqué et d'un paramètre de résistance thermique de la vitamine C (partie 4.2.6.e).

Étapes de conservation

La durée de conservation des conserves peut varier de 10 jours à 3 ans (informations transmises par l'industriel) et a été modélisée par une loi BetaPert de minimum 10 jours, de valeur la

CHAPITRE 4. Modèle d'évaluation du bénéfice

plus vraisemblable 1,5 an et de valeur maximale 3 ans. En raison d'un manque de données, la température de conservation a simplement été supposée varier entre 14°C et 30°C, et les variations de températures au cours de la conservation n'ont pas été modélisées car leur effet dans le modèle sur la dégradation de AA et DHA a été supposé faible.

4.2.4 Modélisation de la concentration initiale en vitamine C

La concentration initiale en vitamine C est modélisée à partir de la concentration initiale en vitamine C totale dans les haricots verts frais, notée $C_{vit\ C}^{INIT}$, et de la proportion de forme réduite (AA) dans les haricots verts frais, notée $prop_{AA}$. Les quantités de AA et de DHA à l'étape initiale sont données par les formules :

$$C_{AA}^{INIT} = C_{vit\ C}^{INIT} \cdot prop_{AA}$$

$$C_{DHA}^{INIT} = C_{vit\ C}^{INIT} \cdot (1 - prop_{AA})$$

$C_{vit\ C}^{INIT}$ varie en fonction des variétés de haricots, des conditions pédoclimatiques lors de la culture et de la maturité des légumes. Des mesures de concentrations de vitamine C dans les haricots frais ont été faites dans le cadre du projet ANR Ribenut. D'autres mesures existent dans la littérature. Une synthèse des différentes mesures, faites sur une dizaine de variétés de haricots verts, plusieurs lieux et plusieurs années, est présentée dans le Tableau 4.3 (Tee et al., 1988; Selman, 1994; Favell, 1998; Oruña-Concha et al., 1998; Souci et al., 2000; Sanchez-Mata et al., 2003; Tosun et al., 2007) :

Tableau 4.3. Concentrations en vitamines C dans des haricots verts frais.

Référence	Information	Taux de vitamine C en mg/100g (poids frais)				Proportion de AA
		Moy	Sd	Min	Max	
Projet ANR Ribenut	France (Normandie), 2010, 2011, 2012, 10 variétés, 65 données	17.8	7.0	6.6	33.8	Selon méthodologie : Entre 82% et 100% Entre 30% et 63%
Souci et al. 2000 (Food composition and Nutrition Table)	-	19	-	9	27	-
D.J. Favell, 1998	Italie, 2 échantillons (début et fin de saison)	13.5	2.3	-	-	-
Tee et al., 1988	Malaisie, Marché local, Duplicat	16.7	0.8	-	-	61%
Selman et al., 1994	-	12	-	-	-	-
Tosun et Yucesan, 2007	Turquie, Marché local ouvert, 3 échantillons	12.7	0.8	-	-	-
Sanchez-Mata et al., 2003	Espagne, 1997, Triplicata	6.19	0.8	-	-	40%
Oruna-Concha et al., 1998	Espagne, 1994	7.5	-	-	-	-

CHAPITRE 4. Modèle d'évaluation du bénéfice

Estimation du paramètre C_{vitC}^{INIT} :

Une base de données constituée à partir de 77 données répertoriées dans le tableau ci-dessus a été constituée. Au moins 9 variétés de haricots sont représentées dans cette base (Samoa, Porto, Xanthos, Archipel, Maritime, Lugos, Fusion, AS1554, Proton), ainsi que plusieurs lieux et années de production. Quand des articles ne mentionnaient qu'une moyenne et un écart-type sur n répétitions, n données de moyenne et écart-type semblables ont été simulées à partir d'une loi normale et introduites dans la base (représentant 8 données sur 77). Une distribution de variabilité normale a été ajustée sur C_{vitC}^{INIT} à partir de cette base de données, et l'incertitude attachée aux paramètres de la distribution a été estimée par bootstrap non paramétrique, en utilisant le package `fitdistrplus` de R. Une troncature sur $[0, 100]$ mg/100g a été prise en compte dans la modélisation pour restreindre la distribution à des valeurs cohérentes :

$$C_{vitC}^{INIT} \sim N(M_{vitC}, S_{vitC})T(0; 100) \text{ mg/100g}$$

Une distribution normale représente l'incertitude sur la moyenne M_{vitC} et une distribution lognormale représente l'incertitude sur l'écart-type S_{vitC} (de valeur moyenne 6.9 mg/100g) :

$$M_{vitC} \sim N(16.8; 0.8) \text{ mg/100g}$$

$$S_{vitC} \sim LnN(1.93; 0.08) \text{ mg/100g}$$

Estimation du paramètre $prop_{AA}$:

AA est généralement présumé majoritaire dans les fruits et légumes. Cette proportion semble varier d'un fruit et d'un légume à l'autre. À titre d'exemple cette proportion a été estimée à environ 87% dans les épinards frais (Toledo et al., 2003), 51% dans les pommes de terre, 20% dans la laitue (Tee et al., 1988), et environ 92% dans le kiwi, la fraise ou la mangue (Gil et al., 2006).

Dans les haricots verts, cette proportion a pu être estimée à environ 61%, 40%, entre 82% et 100%, et entre 30% et 63%, selon les sources et méthodes (Tableau 4.3). Selon les opinions d'experts, les différences observées entre les différentes proportions mesurées dans les différents travaux seraient plutôt plus dues à des incertitudes liées à différentes méthodologies d'analyses ou de stabilisation qu'à des variations réelles de proportions de AA dans les haricots frais. Les analyses de la proportion de AA dans les haricots verts frais qui ont été obtenues lors du projet Ribenut semblent relativement fiables. En effet, ces échantillons ont été prélevés en usine et stabilisés immédiatement par mise en contact rapide avec de l'acide métaphosphorique suivie d'une surgélation immédiate. D'après ces analyses, la proportion de AA dans les haricots verts frais serait proche de 90%. Dans l'étude de Sanchez-Mata et al. (2003), une faible proportion (40%) a été estimée mais une oxydation a pu être possible lors des broyages et de la lyophilisation des haricots verts. En absence d'information fiable laissant penser que cette proportion peut varier dans les haricots verts (HV) frais, cette proportion a été supposé fixe mais connue à incertitude près. L'incertitude sur $prop_{AA}$ est finalement modélisée par une distribution BetaPert de valeur la plus vraisemblable 90% et de bornes minimum et maximum 50% et 100%.

$$prop_{AA} \sim BP(min = 0.50; ml = 0.90; max = 1.00)$$

4.2.5 Modélisation de l'évolution physiologique

4.2.5.a Équations

L'évolution physiologique a lieu pendant l'étape de préparation des légumes. On observe en effet habituellement des pertes en vitamine C dans les fruits ou légumes pendant une étape de conservation après récolte (Favell, 1998; Howard et al., 1999; Sanchez-Mata et al., 2003; Toledo et al., 2003; Sablani et al., 2006). Cette étape n'est pas simple à modéliser car la dégradation de la vitamine C peut varier en fonction de la température, la luminosité, l'altération ou de la découpe éventuelle du produit (Lin et al., 1979). Elle varie aussi selon le type de légume ou de fruit.

Des hypothèses de modélisation assez simples ont été retenues : une hypothèse de dégradation linéaire de la vitamine C au cours du temps, d'ordre 0 (les pertes ne dépendent pas de la concentration initiale), et qui ne dépend ni de la concentration en oxygène, ni de la température. La dégradation dépendra uniquement de la durée de l'étape. La vitesse de dégradation n'a pas été supposée dépendre de la température car l'effet de la température est mal connu et la gamme de température à cette étape est relativement étroite. L'hypothèse d'une dégradation linéaire de AA sur une période de 16 jours dans des haricots verts a été émise par Howard et al., 1999. Une dégradation de AA de type linéaire a été observée dans les épinards pendant un conservation sur plusieurs jours à 8°C, ainsi qu'une concentration de DHA assez stable durant cette période (Toledo et al., 2003).

La concentration en mg/100g du composé X (X=AA, DHA ou oxygène O₂) dans le légume à l'instant t (min) est notée $C_{X,t}$. Les hypothèses suivantes sont émises :

- AA se transforme en DHA avec une vitesse $k_1 = -k_{ev}$ (avec $k \geq 0$, en mg.100g.⁻¹min⁻¹ i.e. en mg par 100g de haricots et par min)
- DHA se dégrade (à la vitesse k_2) mais cette perte est compensée par la production de DHA provenant de la dégradation de AA (à la vitesse k_1). Il y a une régulation de l'homéostasie du végétal (capacité du système à conserver son équilibre de fonctionnement). Dans l'évolution physiologique, ces deux mécanismes sont supposés équilibrés, c'est-à-dire que la concentration de DHA reste constante.

Ces hypothèses s'écrivent :

$$\frac{dC_{AA}}{dt} = k_1 = -k_{ev}$$

$$\frac{dC_{DHA}}{dt} = k_1 - k_2 \approx 0$$

Ces hypothèses sont équivalentes, en supposant des concentrations toujours positives, à :

$$C_{AA}^t = \begin{cases} C_{AA}^{t=0} - k_{ev} \cdot t & \text{si } C_{AA}^{t=0} \geq k_{ev} \cdot t \\ 0 & \text{si } C_{AA}^{t=0} < k_{ev} \cdot t \end{cases}$$

$$C_{DHA}^t = C_{DHA}^{t=0}$$

4.2.5.b Détermination de la vitesse de réaction k_{ev} :

L'étape de préparation des haricots verts correspondant à une conservation à température ambiante sur une durée variant entre 2h30 à 21h. La vitesse de réaction k_{ev} a donc été estimée à partir de données correspondant à ces conditions (températures ambiante, première heures de conservations).

Estimation à partir d'analyses expérimentales effectuées dans le cadre du projet ANR :

Ces analyses ont été effectuées par l'INRA d'Avignon, le CTCPA et Aérial, et consistaient à mesurer les concentrations de vitamines C dans des haricots verts prélevés à l'issue de la récolte et conservés à température ambiante à l'usine pendant 24h (prélèvements d'août 2010). 21 mesures ont été faites, à $t=0h$, $t=1h$, $t=2h$, $t=3.5h$, $t=5h$, $t=6.5h$ et $t=24h$. Des pertes globales de 30% de la vitamine C lors du stockage pendant 24h de haricots verts à température ambiante ont été observées. La Figure 4.4 représente les mesures obtenues au cours du suivi pendant 24h. Par régression linéaire, k_{ev} est estimé à partir des données à $0.002 \text{ mg}\cdot 100\text{g}\cdot\text{min}^{-1}$ (intervalle de confiance à 95% : [0.001 ; 0.003]).

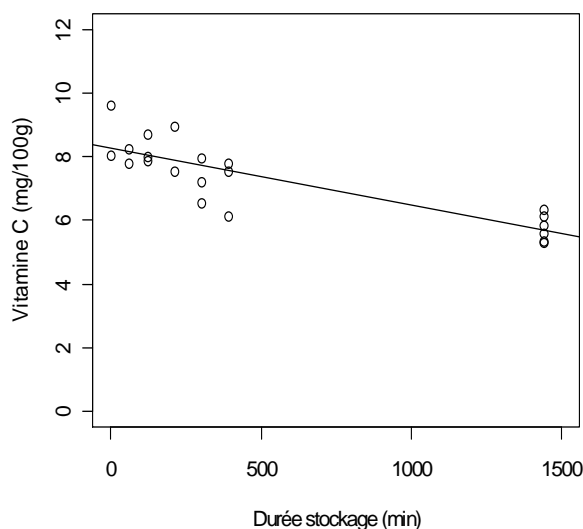


Figure 4.4. Concentrations en vitamine C (AA+DHA) mesurées au cours du suivi pendant 24h à température ambiante en usine (points), et droite de régression linéaire ajustée ($R^2=0.72$)

Confrontation avec les données bibliographiques :

Le Tableau 4.4 présente les pertes en vitamine C estimées durant la conservation de légumes frais. Selon les légumes et les conditions expérimentales (température, humidité,...), les pertes de vitamine C sont assez variables. Une étude a montré que 8°C était la température optimale de préservation de la vitamine C dans les haricots verts (Torija et al. 1997). Les pertes observées dans la

CHAPITRE 4. Modèle d'évaluation du bénéfice

littérature sont donc plutôt en accord avec les pertes observées lors du projet ANR Ribenut, surtout celles observées par Favell, 1998. Son étude montre qu'une perte en vitamine C de 30% en 24h a été observée pendant une conservation à 20°C de haricots verts. La vitesse de dégradation diminue ensuite : les pertes après 14 jours sont d'environ 60%. Des pertes plus faibles ont été observées par Pfeiffer et al. 1994 et Unilever Nutrition Center, 1995. Sanchez-Mata et al. (2003) reportent une concentration de AA plutôt stable et une concentration de DHA décroissante, ce qui ne correspond pas à ce qui a été observé dans Howard et al. (1999), et Toledo et al. (2003). Des pertes plus faibles ont pu être observées dans des tomates, d'environ 4% en 24h, que ce soit à la température de 25°C ou de 4°C (Sablani et al., 2006), et de 10,7% en 4 jours à 25°C.

Tableau 4.4. Pertes de vitamines C durant le stockage de légumes frais.

Référence	Légume	Température	Pertes estimées en vitamine C
D.J. Favell, 1998	Haricots verts	20°C	30% en 24h 60% en 14 jours
Sanchez-Mata et al. 2003	Haricots verts	8°C	42% en 4 jours AA plutôt stable, DHA décroissant
Howard et al. 1999	Haricots verts	4°C	20% de pertes en AA en 4 jours 66% de pertes de AA en 8 jours
Pfeitter et al. 1994*	Haricots verts	Ambiante	50% en 7 jours
Unilever Nutrition Center, 1995*	Haricots verts	Ambiante	40% de pertes après 14 jours
Toledo et al. 2003	Epinards (feuilles mures)	8°C	4% en AA et 0% en DHA en 4 jours
Sablani et al., 2006	Tomates	4°C ou 25°C	4% en 24h

*articles non accessibles mais principaux résultats donnés dans l'étude bibliographique de Dewettinck et al. 2006 (Dewettinck et al., 2006)

L'hypothèse globale retenue est celle de pertes globales de 30% de la vitamine C lors du stockage pendant 24h de haricots verts à température ambiante, observée par Favell (1998) et correspondant aux pertes observées dans nos mesures. Le calcul de k_{ev} fait intervenir les concentrations initiales en vitamine C et le pourcentage de pertes par unité de temps. Pour voir l'influence de ces facteurs sur l'estimation de k_{ev} , la valeur de k_{ev} a été calculée pour différentes quantités initiales et pour différentes pertes. La valeur de k_{ev} correspondant à X% de pertes au cours d'une étape d'évolution physiologique de durée t est donnée par :

$$\frac{C_{AA}^{t=0} - C_{AA}^t}{C_{AA}^{t=0}} = X \leftrightarrow \frac{k_{ev} \cdot t}{C_{AA}^{t=0}} = X \leftrightarrow k_{ev} = \frac{C_{AA}^{t=0} \cdot X}{t} \text{ en } mg \cdot 100g^{-1} \cdot min^{-1}$$

CHAPITRE 4. Modèle d'évaluation du bénéfice

Le Tableau 4.5 rapporte des vitesses d'évolution physiologiques variant entre 0.0003 et 0.0056 mg.100g⁻¹min⁻¹. Dans la modélisation, l'hypothèse de 30% de pertes en 24h à température ambiante a été privilégiée, ainsi qu'une concentration initiale moyenne en vitamine C proche de 17.5 mg/100g en moyenne. Une loi BetaPert a été choisie pour modéliser la variabilité des vitesses d'évolution physiologiques k_{ev} possibles, et une incertitude sur la valeur la plus vraisemblable de k_{ev} a été prise en compte :

$$k_{ev} \sim BP(0.001; ML_{k_{ev}}; 0.006) \text{ mg.100g.}^{-1}\text{min}^{-1}$$

$$ML_{k_{ev}} \sim BP(0.002; 0.004; 0.005)$$

A titre d'exemple, avec $k=0.004 \text{ mg.100g}^{-1}\text{min}^{-1}$, la perte est de 0,6 mg de vitamine C par 100 g de haricot vert en 2h30 et de 5,0 mg de vitamine C par 100g de haricot vert en 21h (ce qui correspond aux durées min et max de l'évolution physiologique).

Tableau 4.5. Vitesse d'évolution physiologique k_{ev} calculée selon différentes hypothèses sur le pourcentage de pertes en AA en 24h et la concentration initiale en AA.

Pertes en 24h (%)	30%	30%	30%	10%	10%	20%	10%
$C_{AA,t=0}$ (mg/100g)	17.5	5	27	27	5	17.5	17.5
k_{ev} (mg/100g/min)	0.0036	0.0010	0.0056	0.0019	0.0003	0.0024	0.0012

4.2.6 Modélisation de la réaction chimique de dégradation de la vitamine C

4.2.6.a Équations

La prévision des pertes en vitamine au cours d'un traitement thermique a été étudiée dans plusieurs articles et de nombreux modèles existent (Lin et al., 1979; Eison-Perchonok and Downes, 1982; Manso et al., 2001; Karhan et al., 2004; Corradini et al., 2006; Leskova et al., 2006; Pénicaud, 2009).

Un modèle de décroissance exponentielle des vitamines est utilisé, dans lequel la vitesse de la réaction augmente en fonction de la concentration en vitamine, de la température et, pour la dégradation de AA, de la concentration en O₂ dont le rôle a été mis en évidence par plusieurs travaux (Eison-perchonok and Downes, 1982; Pénicaud, 2009). Le modèle primaire de la réaction s'écrit (Pénicaud, 2009) :

CHAPITRE 4. Modèle d'évaluation du bénéfice

$$\frac{dC_{AA}}{dt} = -k_1 \cdot C_{AA}^\alpha \cdot C_{O_2}^\beta \quad (4.1)$$

$$\frac{dC_{DHA}}{dt} = k_1 \cdot C_{AA}^\alpha \cdot C_{O_2}^\beta - k_2 \cdot C_{DHA} \quad (4.2)$$

où :

- C_{O_2} (mol/L) est la concentration en oxygène dissous dans le milieu

- C_{AA} et C_{DHA} (mol/L) sont les concentrations en AA et DHA

- α et β sont les ordres des réactions par rapport à AA et par rapport à O_2

- k_1 ($\text{mol}^{-\beta} \text{l}^\beta \text{min}^{-1}$) et k_2 (min^{-1}) sont des coefficients de vitesse de réaction. Ils dépendent de la température T via la loi d'Arrhénius (modèle secondaire) :

$$k_1 = k_{AA}^{ref} \cdot \exp\left(-\frac{Ea_{AA}^{chim}}{R} \left(\frac{1}{T} - \frac{1}{T_{ref}}\right)\right) \quad (4.3)$$

$$k_2 = k_{DHA}^{ref} \cdot \exp\left(-\frac{Ea_{DHA}^{chim}}{R} \left(\frac{1}{T} - \frac{1}{T_{ref}}\right)\right) \quad (4.4)$$

où

- R est la constante universelle de gaz parfaits ($R=8.314 \text{ J}/(\text{mol} \cdot \text{K})$)
- T est la température en kelvin ($\text{K} = ^\circ\text{C} + 273,15$)
- T_{ref} (K) est la température de référence du modèle vitamine, fixée ici à 358.15 K (soit 85°C).
- t (min) est la durée de la réaction.
- k_{AA}^{ref} ($\text{mol}^{-\beta} \text{l}^\beta \text{min}^{-1}$) (resp. k_{DHA}^{ref} (min^{-1})) est le coefficient de vitesse de la réaction de AA (resp. DHA) à la température de référence T_{ref} (K).
- Ea_{AA}^{chim} (J/mol) est l'énergie d'activation liée à l'oxydation de AA dans les haricots verts
- Ea_{DHA}^{chim} (J/mol) est l'énergie d'activation liée à la dégradation de DHA dans les haricots verts.

L'oxydation de l'acide ascorbique au cours d'un traitement thermique est souvent supposée suivre une réaction d'ordre 1 par rapport à l'acide ascorbique (Leskova et al., 2006; Pénicaud, 2009). On supposera donc que $\alpha = 1$.

La température de référence dans le modèle « vitamine » sera établie à $T_{ref} = 85^\circ\text{C}$, ce qui correspond à la température atteinte au blanchiment. Les données permettant d'estimer les vitesses de réactions autour de cette température de 85°C sont également disponibles. Notons que dans la modélisation du risque microbiologique, la température de référence est différente et égale à la température stérilisatrice de référence 121.1°C .

On note que k_1 et k_2 , ainsi que k_{AA}^{ref} et k_{DHA}^{ref} n'ont pas la même unité. La loi d'Arrhénius décrit l'évolution des constantes de réaction en fonction de la température, qu'il y ait de l'oxygène impliqué ou non. k_2 est ici quantifié par défaut comme la quantité de DHA manquante par rapport à la disparition du DHA. L'énergie d'activation E_o mesure l'effet de la température sur la vitesse des

CHAPITRE 4. Modèle d'évaluation du bénéfice

réactions : plus l'énergie d'activation est faible, plus les réactions accélèrent quand la température augmente.

La solution de ces équations différentielles est différente selon que la concentration d'O₂ est constante ou pas pendant l'étape concernée. Des hypothèses ont été émises : la concentration d'O₂ est supposée constante au blanchiment, au refroidissement et au transfert, mais pas à la stérilisation, au refroidissement et à la conservation où elle est supposée décroissante.

4.2.6.b Modélisation de la dégradation de la vitamine C pendant le blanchiment et l'étape de transfert - concentration d'O₂ supposée constante

Equations

La concentration en oxygène est supposée constante au blanchiment et au transfert. Les solutions des équations différentielles 4.1 et 4.2 s'écrivent alors :

$$C_{AA}^t = C_{AA}^{t=0} \cdot \exp(-k_1 \cdot C_{O_2}^\beta \cdot t) \quad (4.5)$$

$$C_{DHA}^t = C_{DHA}^{t=0} \cdot \exp(-k_2 \cdot t) + C_{AA}^{t=0} \cdot [\exp(-k_1 \cdot C_{O_2}^\beta \cdot t) - \exp(-k_2 \cdot t)] \cdot \left(\frac{k_1 C_{O_2}^\beta}{k_2 - k_1 C_{O_2}^\beta} \right) \quad (4.6)$$

Le premier terme correspond à la dégradation propre du DHA initialement présent au temps $t=0$, et le deuxième terme correspond à une production de DHA due à l'oxydation de AA, et à la dégradation de ce DHA produit. Dans ces équations les concentrations sont exprimées en mol/L. Si les concentrations sont exprimées en mg/100g, l'équation 4.5 reste inchangée mais l'équation 4.6 devient :

$$C_{DHA}^t = C_{DHA}^{t=0} \cdot \exp(-k_2 \cdot t) + C_{AA}^{t=0} \cdot \frac{Mm_{DHA}}{Mm_{AA}} \cdot [\exp(-k_1 \cdot C_{O_2}^\beta \cdot t) - \exp(-k_2 \cdot t)] \cdot \left(\frac{k_1 C_{O_2}^\beta}{k_2 - k_1 C_{O_2}^\beta} \right) \quad (4.7)$$

$Mm_{DHA} = 174\text{g/mol}$ est la masse molaire de DHA (C₆H₆O₆), et $Mm_{AA} = 176\text{g/mol}$ est la masse molaire de AA (C₆H₈O₆).

Détermination des concentrations en oxygène dans les haricots verts au cours du blanchiment et du transfert

Dans les tissus d'un haricot vert frais, il y a très peu d'O₂, de l'ordre de 1% à 2%. Mais quand les membranes biologiques sont déstructurées par la chaleur, un phénomène de diffusion de l'O₂ présent dans le milieu environnant le haricot a lieu. Le milieu environnant le haricot est de l'eau, au blanchiment tout comme à l'étape de transfert. De plus, le haricot vert absorbe les liquides avec

CHAPITRE 4. Modèle d'évaluation du bénéfice

lesquels il est en contact. La concentration d' O_2 dans les tissus du haricot vert a donc simplement été supposée égale à la concentration en O_2 dans son environnement (l'eau de blanchiment puis l'eau qui imbibe le haricot pendant les étapes de transfert). La concentration en O_2 dans de l'eau à une température donnée, correspondant à une pression partielle par rapport à l'air, est donnée par la table de Winkler (Winkler, 1888) (Figure 4.5). Plus la température augmente, plus la concentration en O_2 diminue. Par exemple, à 80°C et à pression atmosphérique, la concentration en O_2 dans de l'eau est de 0.09 mol/L , à 90°C elle est de 0.05 mol/L , et au-delà de 100°C elle est nulle.

Il n'existe pas de formule liant la concentration en oxygène à la température, mais nous avons ajusté une droite de régression linéaire sur la portion de températures qui nous intéresse, $40^\circ\text{C} \leq T \leq 100^\circ\text{C}$ (coefficient de détermination $R^2=0.99$) (Figure 4.5). Cette droite donne un résultat jugé suffisamment précis pour notre modélisation. La concentration en O_2 en mol/L dans les tissus du haricot vert a donc été supposée égale à la concentration en O_2 dans l'eau à une température supérieure à 40°C , approchée par la formule :

$$C_{O_2,HV}(T) = C_{O_2,eau}(T) = \begin{cases} 3.33 \cdot 10^{-1} - 3,15 \cdot 10^{-3} T & \text{si } T \leq 100^\circ\text{C} \\ 0 & \text{si } T > 100^\circ\text{C} \end{cases}$$

Ainsi, la concentration en O_2 dans l'eau de blanchiment sera en moyenne de 0.05 mol/L (température moyenne d'environ 90°C), et dans l'eau absorbée par le haricot vert au transfert elle sera en moyenne de 0.09 mol/L (température moyenne de 80°C).

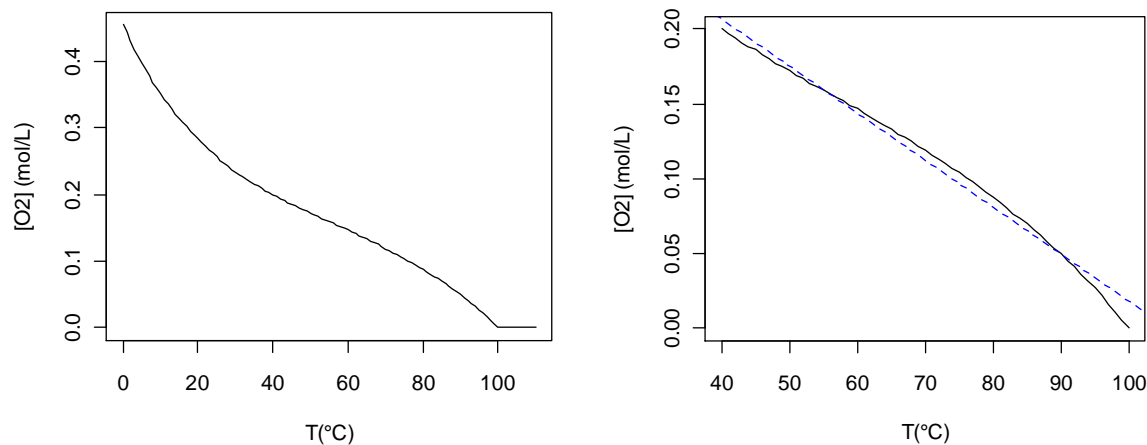


Figure 4.5. Table de Winkler : concentration en oxygène à pression atmosphérique et en présence d'air en fonction de la température de l'eau. A gauche : $0^\circ\text{C} \leq T \leq 110^\circ\text{C}$, à droite : $40^\circ\text{C} \leq T \leq 100^\circ\text{C}$, avec droite de régression ajustée.

4.2.6.c Modélisation de la dégradation de la vitamine C lors de la stérilisation et lors de l'entreposage – concentration en O_2 variable

CHAPITRE 4. Modèle d'évaluation du bénéfice

Après sertissage, la quantité d'O₂ dans la conserve est celle de la bulle d'air piégée entre le couvercle de la boîte et le jus de couverture. Comme la température du produit à l'emboitage est très élevée, la quantité d'oxygène apporté par le haricot vert et le jus de couverture est en effet supposée quasiment nulle. Après le sertissage, la faible quantité d'O₂ présente dans la conserve va se dissoudre dans le jus puis être consommée par l'oxydation de AA jusqu'à éventuelle disparition au bout d'un certain temps t . La concentration d'oxygène n'est donc pas supposée constante pendant ces étapes et les solutions exposées ci-dessus ne sont donc pas valides. Par ailleurs, les autres sources de consommation de l'O₂ (telles que l'oxydation d'autres constituants du haricot vert ou la consommation par la croissance éventuelle de bactéries) sont aussi négligées.

Stérilisation

Par hypothèse, l'O₂ présent dans la conserve ne diffuse pas dans le mélange haricot vert + jus à la stérilisation, par manque de temps et car la solubilité de l'O₂ est quasiment nulle dans l'eau à température supérieure à 100°C (même si la boîte est soumise à une forte pression à la stérilisation, ce qui augmente la solubilité, mais la concentration en O₂ reste quasiment nulle à ces températures très élevées). La concentration en O₂ dans les haricots verts est donc supposée très faible à la stérilisation. Donc l'hypothèse retenue est que AA ne s'oxyde pas à la stérilisation et reste en concentration constante (par manque d'O₂ disponible).

DHA se dégrade selon l'équation :

$$\frac{dC_{DHA}}{dt} = -k_2 \cdot C_{DHA} \Leftrightarrow C_{DHA}^t = C_{DHA}^{t=0} \cdot \exp(-k_2 \cdot t) \quad (4.8)$$

$$\text{avec } k_2 = k_{DHA}^{ref} \cdot \exp\left(-\frac{Ea_{DHA}^{chim}}{R} \left(\frac{1}{T} - \frac{1}{T_{ref}}\right)\right) \quad (4.9)$$

Entreposage

La quantité d'O₂ présente dans la conserve va se dissoudre dans le jus puis être consommée par l'oxydation de AA jusqu'à éventuelle disparition au bout d'un certain temps t . Si l'O₂ était en excès par rapport à AA au début de la conservation, alors tout AA va être oxydé lors de la conservation et il va rester de l'O₂. Si AA était en excès par rapport à l'O₂ au début de l'entreposage, alors tout l'O₂ va être consommé et il va rester de l'AA dans la boîte, qui sera stable au cours de l'entreposage. La concentration de DHA augmentera au cours de l'oxydation de AA, puis DHA se dégradera lentement (par voie anaérobie) au cours du temps.

Dégradation de AA durant l'entreposage :

L'hypothèse est que tout l'O₂ présent dans la boîte est consommé par la dégradation de AA, et cela dans un délai très court par rapport aux durées d'entreposage. On ne fait donc pas dépendre la dégradation de AA des durées d'entreposage. L'hypothèse d'une dissolution rapide de l'oxygène dans le jus est étayée par des mesures faites en usine lors du projet Ribenut. La concentration en

CHAPITRE 4. Modèle d'évaluation du bénéfice

oxygène a été mesurée dans l'air contenu dans l'espace de tête d'une des boîtes de conserve, prélevée quelques heures après la stérilisation. Cet air contenait 10.8% d'O₂ et 3.3% de CO₂, ce qui montre qu'une partie de l'oxygène a été réduit ou s'est dissous dans le jus peu après la stérilisation. De plus, la concentration en O₂ dans le jus d'une boîte 4/4 a été mesurée environ 15h après sa stérilisation, à 0.58 mg/L (soit 18.1 µmol/L), ce qui est faible et correspond à la solubilité de l'O₂ vers 97°C d'après la table de Winkler. La quantité d'O₂ dans le jus montre que l'O₂ a déjà été en partie dissous dans le jus après un laps de temps assez court.

La réaction chimique exacte de l'oxydation de AA est (Oey et al., 2006; Pénicaud, 2009):



Donc le taux de consommation r_{O_2} de l'O₂ devrait être deux fois moins élevé que le taux de consommation r_{AA} de AA. Cependant il est fréquemment observé de manière empirique un taux de consommation de l'O₂ un peu plus élevé, qui correspondrait à une autre réaction simultanée (voir le schéma des réactions possibles en 4.1.2). Ainsi, un coefficient de $\beta \approx 0.9$ a été estimé à la place de $\frac{1}{2}$ (Pénicaud et al., 2009). Cet ordre de valeur a été utilisé pour modéliser la réaction chimique ; par cohérence cette valeur sera également utilisée pour modéliser les pertes lors de la conservation. Donc on supposera que la consommation de 1 mole de AA correspond à la consommation d'environ 0.9 mole d'O₂ (et à la production d'1 mole de DHA):

$$r_{AA} = r_{O_2}/\beta \approx r_{O_2}/0.9 \approx 1.11 r_{O_2}.$$

Soit V_{air} le volume d'air en mL dans l'espace de tête. Soit V_m le volume molaire d'un gaz parfait en L/mol ; dans les conditions standards de température et de pression $V_m = 22,4$ L/mol. L'air contient 21% d'O₂. La masse molaire Mm_{AA} de AA (C₆H₈O₆) est de 176 g/mol et celle de DHA (C₆H₆O₆) est $Mm_{DHA} = 174$ g/mol. La masse molaire Mm_{O_2} de O₂ est de 32 g/mol.

La masse d'O₂ apportée par l'air dans la boîte est donnée par la formule :

$$m_{O_2} = \frac{Mm_{O_2} \cdot V_{air} \cdot 0,21}{V_m}$$

La masse de AA consommée est donnée par la formule :

$$m_{AA \text{ consommé}} = \frac{1}{\beta} \cdot \frac{Mm_{AA} \cdot V_{air} \cdot 0,21}{V_m}$$

La masse de DHA produit est donnée par la formule :

$$m_{DHA \text{ produit}} = \frac{1}{\beta} \cdot \frac{Mm_{DHA} \cdot V_{air} \cdot 0,21}{V_m}$$

Le volume V_{air} d'air dans l'espace de tête de boîtes de 800 g a été mesuré. Des boîtes de conserve prélevées après stérilisation ont été ouvertes dans de l'eau et le volume de la bulle d'air formée après ouverture a été mesuré. Deux valeurs de V_{air} ont pu être mesurées : 4 mL et 5 mL. Une distribution BetaPert a été utilisée dans le modèle pour représenter une variabilité des valeurs possibles de ce volume d'air :

$$air_{can} \sim BP(min = 3; ml = 4.5; max = 6) \text{ mL}$$

CHAPITRE 4. Modèle d'évaluation du bénéfice

A titre d'exemple, 4,5 mL d'air apporte $m_{O_2} = 32 \times 4,5 \times 0,21/22,4 = 1,4$ mg d' O_2 (équation 4-17). La quantité correspondante de AA pouvant être consommé sera $m_{AA\text{ consommé}} = 176 \times 4,5 \times 0,21/(22,4 \times 0,9) = 8,25$ mg (équation 4-18) et la quantité de DHA potentiellement produit sera $m_{DHA\text{ produit}} = 174 \times 4,5 \times 0,21/(22,4 \times 0,9) = 8,16$ mg. Ces quantités ne sont pas négligeables sachant que la quantité totale de vitamine C dans les 445 g de haricots frais contenus dans une boîte 4/4 est, avant toute perte, de 75 mg en moyenne.

Dégradation de DHA durant l'entreposage :

L'oxydation de AA dans la conserve produit du DHA. La vitesse de cette oxydation est supposée très rapide par rapport à la durée d'entreposage (entre 10 jours et 3 ans), et pour simplifier, cette durée d'oxydation est supposée négligeable. La quantité totale de DHA susceptible de se dégrader lors de l'entreposage est donc la somme du DHA présent après stérilisation et du DHA produit par l'oxydation de AA :

$$m_{DHA\text{ stock},t=0} = m_{DHA\text{ steri}} + m_{DHA\text{ produit}} \quad (\text{mg})$$

La dégradation de DHA est ensuite modélisée par la même équation qu'à la stérilisation.

4.2.6.d Les paramètres chimiques de la réaction

Les paramètres chimiques de la réaction sont l'ordre β de la réaction par rapport à l'oxygène, et les paramètres intervenant dans la loi d'Arrhénius : k_{AA}^{ref} , k_{DHA}^{ref} , Ea_{AA}^{chim} et Ea_{DHA}^{chim} . Le paramètre α a été supposé égal à 1 en 4.2.6.a.

La loi d'Arrhénius permet de calculer une vitesse de réaction en fonction de la température. Pour avoir une meilleure validité, il est préférable de l'utiliser sur une plage de température plutôt restreinte. En effet les valeurs des énergies d'activation de la réaction chimique (Ea_{AA}^{chim} et Ea_{DHA}^{chim}) dépendent elles-mêmes un peu de la température et sans doute un peu aussi de la concentration en oxygène. Or dans notre modèle, les plages de température utilisées sont assez larges (entre 70°C et 125°C pour les étapes de blanchiment, transfert, stérilisation, et vers 20°C lors de la conservation). Cependant, en absence de données à certaines températures, et comme l'incertitude sur l'énergie d'activation est quasiment plus élevée que sa possible variation en fonction de la température et de la concentration en oxygène, une valeur unique d'énergie d'activation pour Ea_{AA}^{chim} et une valeur unique pour Ea_{DHA}^{chim} serviront à décrire la réaction chimique tout au long du processus. Les valeurs des différents paramètres chimiques ont été estimées dans la mesure du possible à partir de données obtenues dans des plages de températures relativement proches des températures des procédés les plus impactant a priori (blanchiment, transfert, stérilisation), et donc à température assez élevée (vers 85°C).

Valeur de l'ordre β de la réaction par rapport à l'oxygène dans la dégradation de AA

L'ordre par rapport à l'oxygène de cette réaction a été estimé dans la thèse de C.Pénicaud (2009), dans une gamme de température entre 3°C et 33°C dans de l'eau pure. Cette gamme de température est faible par rapport aux températures qui nous intéressent, mais en absence d'autre information nous avons utilisé pour β les estimations issues de ces travaux. β a été estimé à 1.1 ± 0.9 dans la thèse de Pénicaud, 2009, mais de nouvelles valeurs obtenues par analyses multivariées (non présentées dans la thèse) nous ont été personnellement communiquées par C.Pénicaud. β a alors été estimé à 0.9, avec une plage d'incertitude entre 0.69 et 1.07. Dans ce modèle, une loi BetaPert a été utilisée pour représenter l'incertitude autour de la valeur de β :

$$\beta \sim BP(\min = 0.6; ml = 0.9; \max = 1.2)$$

Calcul de Ea_{DHA}^{chim} et k_{DHA}^{ref}

Dans l'article de Mertz et al. (2010) (Table 4), les pertes de DHA dans un nectar de Tamarillo ont été étudiées après une étape de chauffage de 10 min à 80°C, 90°C ou 95°C. Ces températures sont proches de la température de référence vitamine $T_{ref} = 85^\circ\text{C}$ utilisée dans notre modèle. Afin d'estimer les paramètres correspondant à la dégradation propre de DHA, les résultats obtenus en nectar dégazé, pour lesquels la concentration en O_2 est très faible et donc AA est assez stable, ont été utilisés. Les données sont constituées des pourcentages de DHA restant au temps $t=10\text{min}$, à 80°C, 90°C et 95°C.

En combinant le modèle primaire et la loi d'Arrhénius qui régissent les pertes en DHA (équations 4.8 et 4.9), on obtient un modèle global :

$$C_{DHA}^t = C_{DHA}^{t=0} \cdot \exp\left(-k_{DHA}^{ref} \cdot \exp\left(-\frac{Ea_{DHA}^{chim}}{R} \left(\frac{1}{T} - \frac{1}{T_{ref}}\right)\right) \cdot t\right) \quad (4.10)$$

Les paramètres Ea_{DHA}^{chim} et k_{DHA}^{ref} ont été estimés par régression non linéaire à partir de cette équation, ce qui permet d'obtenir une estimation plus juste et plus précise des paramètres qu'en procédant en deux étapes (Van Boekel, 1996; Van Boekel, 2008). L'estimation a été faite en utilisant le package nls2 de R (Grothendieck, 2012), par minimisation de la somme des carrés des erreurs par un algorithme itératif. Les données étant reportées sous la forme de moyenne et écarts-types et les données brutes n'étant plus disponibles, nous avons utilisé des données simulées à partir des données reportées pour calculer nos estimations (simulation de 3 répétitions comme dans l'article, de même moyenne et écart-type), afin d'améliorer la précision des estimations et de tenir compte de l'incertitude sur ces concentrations mesurées.

Les estimations obtenues sont :

$$Ea_{DHA}^{chim} \approx 80.2 \text{ kJ/mol et } 95\%CI = [55.9; 106.9]$$

$$k_{DHA}^{ref} \approx 0.069 \text{ min}^{-1} \text{ et } 95\%CI = [0.058; 0.082]$$

CHAPITRE 4. Modèle d'évaluation du bénéfice

Les distributions suivantes ont été choisies pour le modèle :

$$Ea_{DHA}^{chim} \sim N(80; 14)T(0; 200) \text{ kJ. mol}^{-1}$$

$$k_{DHA}^{ref} \sim BP(\min = 0.05; ml = 0.07; \max = 0.09) \text{ min}^{-1}$$

Calcul de Ea_{AA}^{chim} et k_{AA}^{ref} à partir de données obtenues dans le cadre du projet ANR Tempantiox (INRA Avignon, Unité SQPOV ; CTCPA)

Un des objectifs du projet ANR Tempantiox était de prédire l'évolution de molécules d'intérêt dans des milieux modèles et des produits végétaux soumis à des traitements technologiques. En particulier des analyses ont été effectuées sur l'évolution de AA en milieu modèle (tampon malate à pH 3.8) à températures comprises entre 50°C et 90°C, qui sont proches de la température de référence vitamine ($T_{ref} = 85^\circ\text{C}$). Les données utilisées sont constituées des valeurs brutes des concentrations en AA au cours du temps (de 0 à 180 min) pour des chauffages effectués à 3 températures différentes (50°C, 70°C et 90°C), avec 3 répétitions à chaque instant et température.

En combinant le modèle primaire et la loi d'Arrhénius qui régissent les pertes en AA (équations 4.3 et 4.5), on obtient un modèle global :

$$C_{AA}^t = C_{AA}^{t=0} \cdot \exp\left(-k_{AA}^{ref} \cdot \exp\left(-\frac{Ea_{AA}^{chim}}{R}\left(\frac{1}{T} - \frac{1}{T_{ref}}\right)\right) \cdot C_{O_2}^\beta \cdot t\right) \quad (4.11)$$

La concentration en oxygène C_{O_2} dépend de la température T et peut être estimée à partir de la table de Winkler (voir 4.5.2). Le coefficient β , ordre de réaction par rapport à l'oxygène, a été estimé dans la thèse de C.Pénicaud autour de 0.9 ; en absence de meilleure estimation on a retenu cette valeur, en prenant en compte l'incertitude sur cette valeur. Les paramètres Ea_{AA}^{chim} et k_{AA}^{ref} ont ensuite été estimés par régression non linéaire à partir de l'équation 4.11 en utilisant le package nls2 de R (Grothendieck, 2012). Selon que l'incertitude sur β est prise en compte ou pas (sous forme de la distribution d'incertitude donnée ci-dessus), les estimations suivantes sont obtenues :

❖ *Estimations 1 (avec $\beta = 0.9$) :*

$$Ea_{AA}^{chim} = 44.3 \text{ kJ/mol, IC95\%} = [38.3; 50.4]$$

$$k_{AA}^{ref} = 0.18 \text{ mol}^{-\beta} \text{ l}^\beta \text{ min}^{-1}, \text{ IC95\%} = [0.16; 0.21]$$

❖ *Estimations 2 (incertitude sur β prise en compte par simulation de Monte Carlo) :*

$$Ea_{AA}^{chim} = 44.1 \text{ kJ/mol, IC95\%} = [38.1; 51.0]$$

$$k_{AA}^{ref} = 0.18 \text{ mol}^{-\beta} \text{ l}^\beta \text{ min}^{-1}, \text{ IC95\%} = [0.10; 0.34]$$

CHAPITRE 4. Modèle d'évaluation du bénéfice

La prise en compte de l'incertitude sur β augmente donc un peu la largeur de l'intervalle de confiance sur k_{AA}^{ref} mais quasiment pas celui sur Ea_{AA}^{chim} .

A titre de comparaison, dans la thèse de Pénicaud (2009), Ea_{AA}^{chim} a été estimé entre 66 et 82 kJ/mol dans la gamme de température plus faible (3°C – 33°C). Dans l'étude de Eison-perchonok et al. (1982), les énergies d'activation ont été estimées dans la gamme de température 33°C – 55°C en fonction de la concentration d'oxygène, à 40 ± 1 , 57 ± 1 et 69 ± 1 kJ mol⁻¹ à 10%, 15% and 21% d'oxygène respectivement. Les valeurs obtenues dans le projet Tempantiox sont du même ordre de grandeur mais légèrement plus basses, ce qui pourrait s'expliquer par la gamme de température plus élevée (70°C – 90°C). Par ailleurs, d'autres valeurs d'énergie d'activation de AA existent dans la littérature mais ont été déterminées sans prendre en compte l'influence de l'oxygène, ce qui modifie les estimations des paramètres. Les paramètres estimés dans ces articles ne correspondent donc par directement avec eux estimés pour notre modèle. L'ordre de grandeur des énergies d'activation est cependant semblable (entre 31 et 105 kJ/mol selon la température, le milieu, la concentration en oxygène etc.) (Manso et al., 2001; Karhan et al., 2004; Heldman and Lund, 2007; Pénicaud, 2009).

Les distributions suivantes ont été choisies pour le modèle, représentant de l'incertitude autour des « vraies » valeurs des paramètres. Pour prendre en compte l'incertitude sur β , les intervalles de confiance les plus larges ont été privilégiés, mais la corrélation observée entre Ea_{AA}^{chim} et k_{AA}^{ref} (dans le cas n°2) n'a pas été prise en compte par simplification, étant donné la relativement faible différence entre les deux types d'intervalles de confiance, et en raison des diverses sources de biais (concentration en O₂ estimée par la table de Winkler ; analyses faites en milieu modèle différent du milieu haricot vert ; ajout de fer dans le milieu modèle pour maîtriser l'auto-oxydation de la vitamine C due aux traces métalliques entraînant des vitesses plus rapides).

$$Ea_{AA}^{chim} \sim N(44; 4)T(0; 200) \text{ kJ. mol}^{-1}$$

$$k_{AA}^{ref} \sim BP(\min = 0.07; ml = 0.18; \max = 0.43) \text{ mol}^{-\beta} l^{\beta} \text{ min}^{-1}$$

4.2.6.e Détermination du temps équivalent pour la dégradation de DHA lors de la stérilisation

Détermination d'un temps équivalent « vitamines »

Un temps équivalent de chauffage à une température de référence a été construit, permettant de décrire la dégradation de DHA. Ce temps équivalent « vitamines » est basé sur le même principe que le temps équivalent utilisé classiquement en microbiologie des aliments (Chapitre 3).

Seule la dégradation de DHA est prise en compte à la stérilisation (partie 4.2.5.3) ; elle est modélisée par les équations 4.8 et 4.9 rappelées ci-dessous :

CHAPITRE 4. Modèle d'évaluation du bénéfice

$$C_{DHA}^t = C_{DHA}^{t=0} \cdot \exp(-k_2 \cdot t)$$

$$k_2 = k_2(T) = k_{DHA}^{ref} \cdot \exp\left(-\frac{Ea_{DHA}^{chim}}{R} \left(\frac{1}{T} - \frac{1}{T_{ref}}\right)\right)$$

Des profils temps-températures décrivent la montée en température lors de la stérilisation par des pas (t_i, T_i) , $i=1, \dots, n$, où t_i et T_i sont le temps et la température à l'instant i . Le pas i a une durée de $t_i - t_{i-1}$ à une température moyenne de $(T_i + T_{i-1})/2$.

Lors du pas i :

$$C_{DHA}^i = C_{DHA}^{i-1} \cdot \exp(-k_2^i \cdot (t_i - t_{i-1}))$$

$$\text{Et } k_2^i = k_{DHA}^{ref} \cdot \exp\left(-\frac{Ea_{DHA}^{chim}}{R} \left(\frac{1}{(T_i + T_{i-1})/2} - \frac{1}{T_{ref}}\right)\right)$$

D'où :

$$C_{DHA}^n = C_{DHA}^{n-1} \cdot \exp(-k_2^n \cdot (t_n - t_{n-1}))$$

$$C_{DHA}^n = C_{DHA}^{n-2} \cdot \exp(-k_2^{n-1} \cdot (t_{n-1} - t_{n-2})) \exp(-k_2^n \cdot (t_n - t_{n-1}))$$

.....

$$C_{DHA}^n = C_{DHA}^0 \cdot \exp(-k_2^1 \cdot (t_1 - t_0)) \dots \exp(-k_2^{n-1} \cdot (t_{n-1} - t_{n-2})) \exp(-k_2^n \cdot (t_n - t_{n-1}))$$

$$C_{DHA}^n = C_{DHA}^0 \cdot \exp(-k_2^1 \cdot (t_1 - t_0) \dots - k_2^{n-1} \cdot (t_{n-1} - t_{n-2}) - k_2^n \cdot (t_n - t_{n-1}))$$

$$C_{DHA}^n = C_{DHA}^0 \cdot \exp\left(-\sum_{i=1}^n k_2^i \cdot (t_i - t_{i-1})\right)$$

$$C_{DHA}^n = C_{DHA}^0 \cdot \exp\left(-\sum_{i=1}^n k_{DHA}^{ref} \cdot \exp\left(-\frac{Ea_{DHA}^{chim}}{R} \left(\frac{1}{(T_i + T_{i-1})/2} - \frac{1}{T_{ref}}\right)\right) \cdot (t_i - t_{i-1})\right)$$

$$C_{DHA}^n = C_{DHA}^0 \cdot \exp\left(-k_{DHA}^{ref} \sum_{i=1}^n (t_i - t_{i-1}) \exp\left(-\frac{Ea_{DHA}^{chim}}{R} \left(\frac{1}{(T_i + T_{i-1})/2} - \frac{1}{T_{ref}}\right)\right)\right)$$

$$C_{DHA}^n = C_{DHA}^0 \cdot \exp(-k_{DHA}^{ref} \cdot t_{eq \ vit})$$

$$\text{Avec } t_{eq \ vit} = t_{eq \ vit}(E_{a,DHA}) = \sum_{i=1}^n (t_i - t_{i-1}) \exp\left(-\frac{Ea_{DHA}^{chim}}{R} \left(\frac{1}{(T_i + T_{i-1})/2} - \frac{1}{T_{ref}}\right)\right).$$

Cette dernière formule définit $t_{eq \ vit}$, le temps équivalent de chauffage (en min) à une température de référence T_{ref} pour la dégradation de DHA. Ce temps équivalent dépend d'une part du traitement thermique appliqué et d'autre part de l'énergie d'activation de DHA.

Transformation des temps équivalents vitamines calculés à partir des différents relevés temps-températures par la méthode de l'enveloppe

Une méthode de transformation similaire à celle utilisée pour le temps équivalent en microbiologie (Chapitre 3) a été utilisée. Cette transformation permet de prendre en compte une plus grande variété de gammes de montées en températures que les six profils temps-températures disponibles et de séparer l'effet du traitement thermique et de la « résistance thermique » de DHA (représentée par le paramètre Ea_{DHA}^{chim}). De plus, cette séparation permettra dans le Chapitre 5 relatif à l'analyse risque-bénéfice de simuler des temps équivalents « vitamine » et des temps équivalents « *G. stearothermophilus* » correspondant à un même traitement thermique, mais prenant en compte dans le premier cas la résistance « thermique » de DHA, et dans le second cas la résistance thermique de la bactérie *G. stearothermophilus*.

Le temps équivalent vitamine $t_{eq\ vit}$ dépend de l'énergie d'activation Ea_{DHA}^{chim} (Figure 4.6 (a)). Cette dépendance a été modélisée par la méthode de l'enveloppe (Vose, 2000). Une relation « globalement linéaire » a été observée entre $\ln(t_{eq\ vit})$ et Ea_{DHA}^{chim} (Figure 4.6 (b)). Une régression linéaire a été effectuée de $\ln(t_{eq\ vit})$ sur Ea_{DHA}^{chim} , modélisée par :

$$\ln(t_{eq\ vit}) = a' + b' \cdot Ea_{DHA}^{chim} + \epsilon'$$

où $\epsilon' \sim N(0, c')$ suit une loi normale de moyenne 0 et d'écart-type c' , représentant l'erreur du modèle qui ici représente principalement la variabilité des six profils de traitement thermique. Si on note $Ea_{DHA,ref}^{chim} = 80 \text{ kJ} \cdot \text{mol}^{-1}$ l'énergie d'activation de référence de DHA, l'expression se reformule en :

$$t_{eq\ vit} = \exp\left(b' \cdot (Ea_{DHA}^{chim} - Ea_{DHA,ref}^{chim})\right) \cdot \exp(X')$$

$$\text{où } X = a' + b' \cdot Ea_{DHA,ref}^{chim} + \epsilon' \sim N(a' + b' \cdot Ea_{DHA,ref}^{chim}, c').$$

En posant $F_1 = \exp(X')$, l'équation devient :

$$t_{eq\ vit} = \exp\left(b' \cdot (Ea_{DHA}^{chim} - Ea_{DHA,ref}^{chim})\right) \cdot F_1.$$

F_1 représente le temps équivalent à la température de référence T_{ref} et à l'énergie d'activation de DHA de référence. La distribution sur F_1 représente la variation des traitements thermiques. Dans notre application, la température de référence est $T_{ref} = 85^\circ\text{C}$, l'énergie d'activation de DHA de référence est $Ea_{DHA,ref}^{chim} = 80 \text{ kJ} \cdot \text{mol}^{-1}$; les estimations suivantes ont été obtenues : $a'=2,84$; $b'=3,07 \cdot 10^{-5}$ et $c'=0,05$. La distribution de F_1 est donc une loi lognormale de paramètres (5.30 min, 0.05 min), et de moyenne environ 200 min :

$$F_1 \sim \text{LnN}(5.30, 0.05) \text{ min}$$

Dans le Chapitre 5 (analyse risque-bénéfice), F_1 sera liée à la valeur stérilisatrice F_0 , très utilisée en microbiologie et par les industriels.

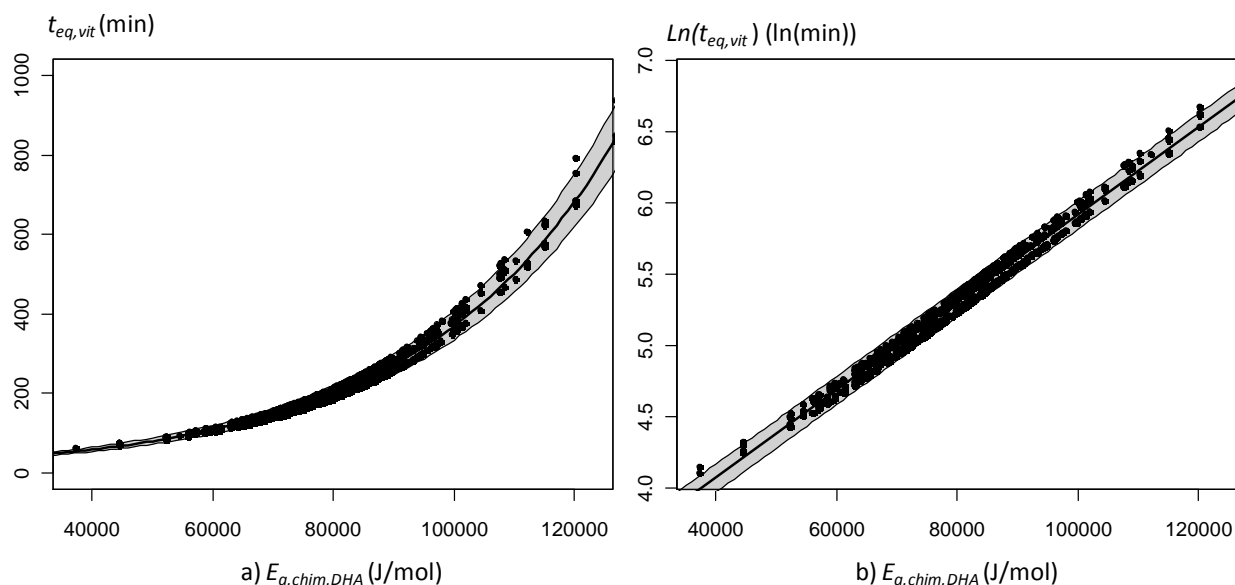


Figure 4.6. Temps équivalents « vitamine » $t_{eq, vit}$ à la température de référence 85°C, en fonction de l'énergie d'activation de DHA Ea_{DHA}^{chim} (a), et $\ln(t_{eq, vit})$ en fonction de Ea_{DHA}^{chim} (b). Points : échantillons de 1000 valeurs de Ea_{DHA}^{chim} et des temps équivalents vitamines correspondant $t_{eq, vit}$ (ou $\ln(t_{eq, vit})$). Lignes : modèles ajustés et intervalles de confiance à 95%, délimitant une enveloppe (partie grisée).

4.2.7 Modélisation de la diffusion des vitamines dans l'eau de blanchiment et dans le jus de couverture

4.2.7.a Le phénomène de diffusion

Une partie de la vitamine C est perdue par transfert de matière dans l'eau de cuisson au blanchiment, ou dans le jus de couverture au cours de la conservation. Ce phénomène de diffusion est la tendance naturelle d'un système à rendre homogènes les concentrations des espèces chimiques en son sein. Sous l'effet de l'agitation thermique, on observe un déplacement des constituants des zones de forte concentration vers celles de faible concentration, jusqu'à atteindre un équilibre. Des pertes proches de 50%, dont la grande majorité provenait du phénomène de diffusion, ont ainsi pu être mesurées lors du blanchiment de petits pois à 85°C (Selman, 1994). On va distinguer la modélisation de la diffusion pendant le blanchiment et à la conservation.

4.2.7.b Modélisation de la diffusion au blanchiment par la loi de Fick

La diffusion au blanchiment a été modélisée par la loi de Fick et plus particulièrement par la solution complète de la loi de Fick en assimilant un haricot vert à un cylindre (Selman, 1994; Arroqui et al., 2002). Les données disponibles sur AA ont été utilisées et les diffusions de AA et DHA ont été supposées similaires. En conséquence un même coefficient de diffusivité a été utilisé pour AA et pour DHA, ainsi qu'une même énergie d'activation de diffusion.

La loi de Fick :

Si on suppose une diffusion unidirectionnelle selon l'axe x , que Dif (en m^2s^{-1}) est le coefficient de diffusivité de la vitamine C, et que C est la concentration en vitamine C, la loi de Fick s'écrit :

$$\frac{\partial C}{\partial t} = Dif \cdot \frac{\partial^2 C}{\partial x^2}$$

La solution de la 2^{ème} loi de Fick (Crank, 1975; Selman, 1994; Arroqui et al., 2002) s'écrit :

$$\frac{C_{Av}^t - C^\infty}{C^0 - C^\infty} = 4 \sum_{i=1}^{\infty} \left(\frac{1}{R_i^2} \right) \exp \left(\frac{-R_i^2 \cdot Dif \cdot t}{r_{GB}^2} \right)$$

où :

- t (s) est la durée de la diffusion
- Dif (m^2s^{-1}) est le coefficient de diffusivité apparent de la vitamine C dans le haricot vert
- R_i est la $i^{ème}$ racine de la fonction de Bessel d'ordre zero J_0
- r_{GB} (m) est le rayon du haricot vert (assimilé à un cylindre)
- C_{Av}^t (mg/100g) est la concentration moyenne en vitamine C dans un (cylindre de) haricot vert au temps t
- C^0 (mg/100g) est la concentration initiale ($t=0$) uniforme en vitamine C dans un (cylindre de) haricot vert
- C^∞ (mg/100g) est la concentration en vitamine C à la surface du haricot vert, qui est égale à la concentration en AA dans l'eau de blanchiment.

La vitamine C se comportant comme un sucre, il est supposé que toute la vitamine C est peut être extraite du haricot vert et donc que $C_{A\infty} = 0$. Les 10 premiers termes de la solution suffisent à avoir une bonne approximation de la solution, qui s'écrit :

$$C_{Av}^t = C^0 \left(4 \sum_{i=1}^{10} \left(\frac{1}{R_i^2} \right) \exp \left(\frac{-R_i^2 \cdot Dif \cdot t}{r_{GB}^2} \right) \right)$$

CHAPITRE 4. Modèle d'évaluation du bénéfice

Le coefficient de diffusion Dif dépend de la température T . La loi d'Arrhénius permet de modéliser cet effet de la température, toujours à partir de la température de référence $T_{ref} = 85^\circ\text{C}$ correspondant à la température moyenne du blanchiment :

$$Dif(T) = Dif_{ref} \cdot \exp\left(-\frac{Ea_{Dif}}{R} \left(\frac{1}{T} - \frac{1}{T_{ref}}\right)\right)$$

Détermination du rayon rad_{GB} des haricots verts :

Comme le rayon des haricots verts est au carré dans la solution de la loi de Fick, ce paramètre peut avoir une influence assez forte : plus le rayon est petit, plus les pertes sont grandes. Différents calibres de haricot vert existent en boîte 4/4 : extra fins (diamètre < 6.5mm), très fins (diamètre < 8mm), fins (diamètre < 9mm). Dans nos prélèvements, le calibre des haricots verts était « fin » ou « très fin ». Du haricot vert fin a été modélisé ici, ce qui correspond à des diamètres entre 8 et 9mm, d'où des rayons entre 4 et 4.5 mm. La distribution du rayon des haricots verts dans le modèle est :

$$rad_{GB} \sim BP(\min = 0.004; ml = 0.00425; \max = 0.0045) m$$

Détermination du coefficient de diffusivité Dif_{ref} et de l'énergie d'activation Ea^{diff} liée à la diffusion de la vitamine C :

La diffusion de la vitamine C (ou AA) a été étudiée dans différentes publications dans de la pomme de terre, dont certains résultats sont présentés dans le Tableau 4.6 (Kozempel et al., 1982; Garrote et al., 1986; Garrote et al., 1988; Kincal and Kaymak, 1990; Arroqui et al., 2001; Varzakas et al., 2005). A défaut de données dans du haricot vert, la pomme de terre présente des caractéristiques assez similaires au haricot vert, car il s'agit également d'un « tissu cellulaire rempli ».

Les valeurs observées dans les différents travaux sont du même ordre de grandeur (coefficient de diffusivité autour de $1,5 \cdot 10^{-9} \text{m}^2 \text{sec}^{-1}$ et énergie d'activation autour de $30 \text{kJ} \cdot \text{mol}^{-1}$). L'étude sur la diffusion de AA lors du blanchiment de cylindres de pomme de terre dans de l'eau recyclée pour des températures comprises entre 65 et 93°C (Arroqui et al., 2002), proches de la température de référence vitamine $T_{refv} = 85^\circ\text{C}$, a été privilégiée. L'incertitude sur les paramètres est modélisée par les distributions suivantes :

$$Ea^{diff} \sim N(20; 5)T(0; 200) \text{kJ} \cdot \text{mol}^{-1}$$

$$Dif_{ref} \sim 10^{-9} \cdot BP(\min = 1.6; ml = 1.8; \max = 2.1) \text{m}^2 \text{s}^{-1}$$

Tableau 4.6. Coefficients de diffusivité de la vitamine C publiés dans la littérature.

Référence	Légume	T (°C)	Coefficient de diffusivité $D_{if} \times 10^9 \text{ m}^2 \text{ s}^{-1}$	Energie d'activation liée à la diffusion E_a^{dif} (kJ mol ⁻¹)
Garrote et al. 1986	Pomme de terre	65	0.93	49.5
		80	1.43	
		100	2.14	
Garrote et al. 1988	Pomme de terre	65	1.78	25
		80	2.03	
Kincal and Kaymak, 1990	Pomme de terre	80	1.24	9.5
		90	1.38	
Kozempel et al. 1982	Pomme de terre	77	9.56	
Arroqui et al. 2001	Pomme de terre (eau recyclée)	80	1.61	
Arroqui et al. 2002	Pomme de terre (eau recyclée)	65	1.3	17.1
		80	1.61	
		93	2.06	
	Pomme de terre (eau distillée)	65	1.25	31
		80	1.74	
		93	2.48	

4.2.7.c Modélisation de la diffusion au cours de l'entreposage des boîtes de haricots verts

L'équilibre de la concentration en vitamine C dans la boîte (c'est-à-dire une concentration homogène) a été supposé atteint au bout d'un ou deux mois, et donc atteint au moment de l'achat des boîtes. Les analyses confirment la rapidité d'atteinte de cet équilibre : des concentrations identiques en vitamine C dans les haricots et dans le jus ont été retrouvées après 2 mois de conservation, or l'entreposage dure de 10 jours à 3 ans. L'élimination du jus de couverture avant la consommation des haricots verts entraîne la perte des vitamines présentes dans le jus. L'effet de la diffusion des vitamines dans le jus est donc modélisé en supposant des pertes égales à la proportion de jus dans les boîtes, soit 56.5% (480mL/850mL) de pertes des vitamines restantes après les dégradations chimiques.

4.3 Résultats du modèle

4.3.1 Évolution des concentrations de vitamine C prédites par le modèle

Trois millions de boîtes ont été simulées par la méthode de simulation de Monte Carlo à 2 dimensions, en effectuant $N_u = 1500$ tirages dans la dimension incertitude et $N_v = 2000$ tirages dans la dimension variabilité, en utilisant le logiciel R (R Development Core Team, 2010).

CHAPITRE 4. Modèle d'évaluation du bénéfice

La Figure 4.7 représente l'évolution des concentrations en AA, DHA et vitamine C totale dans les boîtes le long de la chaîne de production, et le Tableau 4.7 représente l'évolution des concentrations en vitamine C totale le long de la chaîne (à incertitude et variabilité distinguées). Les teneurs initiales en vitamine C dans les haricots verts sont assez variables (entre 4.1 mg/100g et 30.3 mg/100g dans 95% des boîtes). Les pertes totales entre la matière première fraîche et les légumes appertisés après conservation sont de 86% en moyenne. Les pertes totales par rapport à la concentration initiale sont en moyenne de 15% après la préparation, de 56% après blanchiment, de 56% après transfert, et de 64% après stérilisation. Les pertes les plus importantes sont donc observées au blanchiment (environ 41%), mais les pertes sont aussi importantes pendant la conservation (environ 23%), le temps d'attente (environ 15%), et la stérilisation (environ 7%). Au niveau de la distinction entre AA et DHA, les pertes en AA sont observées principalement pendant le temps d'attente, au blanchiment et au cours de la conservation. Les pertes en DHA sont, elles, observées principalement pendant le blanchiment et la stérilisation. Globalement, l'incertitude sur les concentrations prédites en vitamine C totale est relativement faible (quelques mg/100g) (Figure 4.7 c), dans la mesure où elle ne remet pas en cause les principales conclusions de l'étude. L'incertitude sur les concentrations prédites en DHA est par contre assez importante (Figure 4.7 b).

La concentration finale en vitamine C est assez faible, puisqu'elle est en moyenne de 2.3 mg/100g, avec un intervalle de variation à 95% de [0.0 ; 5.6] mg/100g. Ces teneurs sont nettement inférieures au seuil d'allégation nutritionnelle qui est de 12 mg/100g (seuil permettant à l'industriel de mentionner que le produit contient une quantité significative de vitamine C). Au moins 95% des boîtes ont des concentrations en vitamine C sous ce seuil. Cependant il faut remarquer que 26% des portions de haricots verts frais, avant toute perte, possédaient des teneurs en vitamine C déjà inférieures à ce seuil. Par ailleurs, la fonction de répartition de la concentration finale en vitamine C (Figure 4.8) montre une assez forte variabilité des concentrations en vitamine C dans les boîtes après entreposage, ainsi qu'une incertitude importante représentée par les intervalles de confiance à 95% assez larges. Ainsi, l'intervalle d'incertitude à 95% sur le nombre de boîtes contenant des concentrations en vitamine C supérieures à 2 mg/100g est d'environ [20% ; 80%]. Par contre l'incertitude est négligeable sur le fait que toutes les boîtes ont des concentrations finales en vitamine C inférieures à 12 mg/100g.

Tableau 4.7. Concentrations en vitamine C totale après chaque étape du procédé. Des intervalles de confiance à 95% représentent l'incertitude sur les estimations.

Vitamine C (mg/100g)	Initial	Préparation	Blanchiment	Transfert	Stérilisation	Stockage
Quantile $Q_{0.025}$	4.1 [2.6; 6.1]	1.4 [0.3; 3.3]	0.6 [0.1; 1.6]	0.6 [0.1; 1.6]	0.3 [0.0; 1.3]	0.0 [0.0; 0.0]
Concentration moyenne	17.0 [15.5; 18.5]	14.4 [12.8; 16.0]	7.5 [6.3; 8.6]	7.4 [6.2; 8.6]	6.2 [4.3; 7.9]	2.3 [1.3; 3.3]
Quantile $Q_{0.975}$	30.3 [27.6; 33.2]	27.9 [25.0; 30.8]	14.9 [12.8; 17.1]	14.9 [12.8; 17.1]	12.5 [8.9; 15.7]	5.6 [3.6; 7.3]

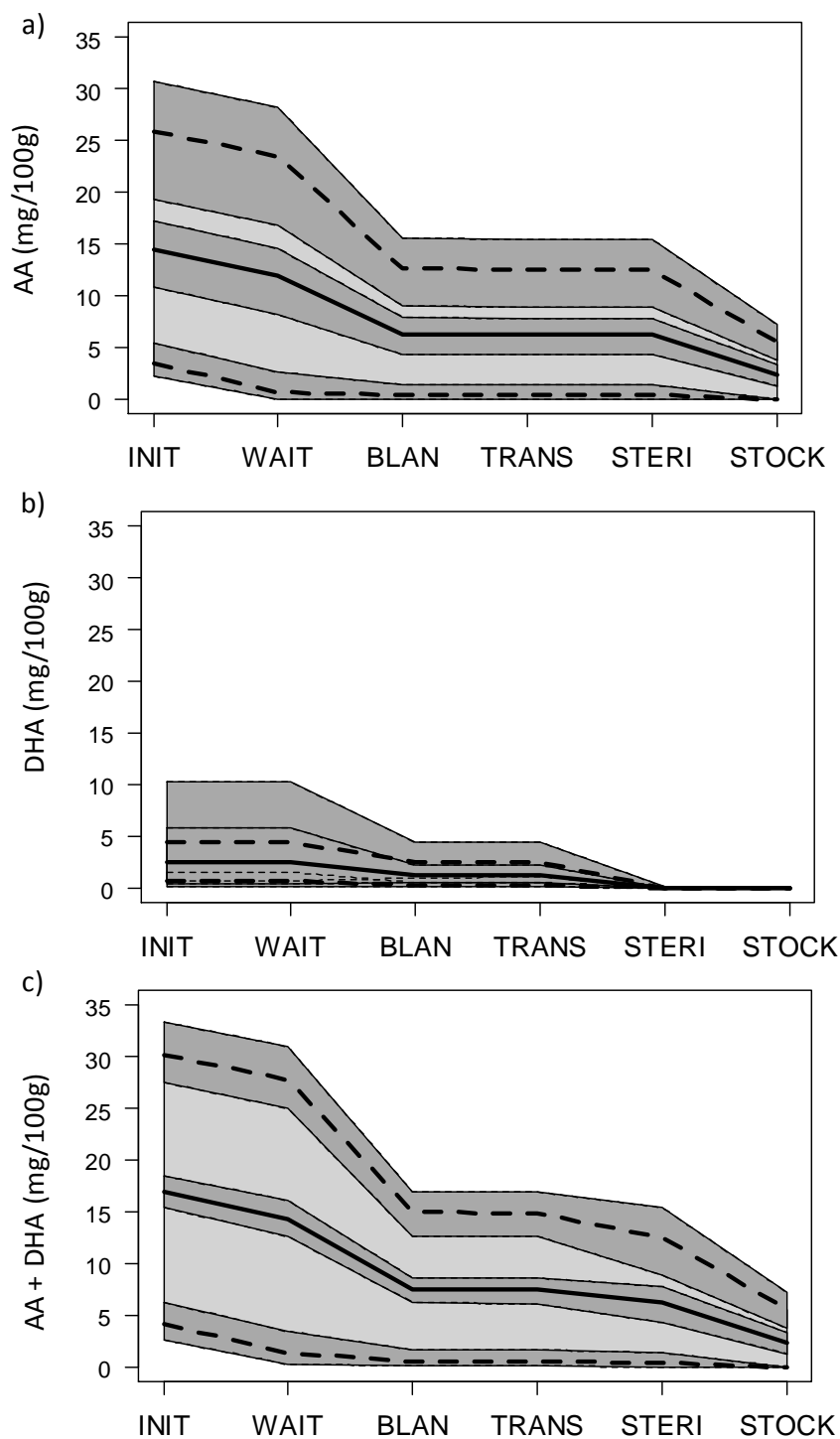


Figure 4.7. Evolution des concentrations en AA (a), en DHA (b) et en vitamine C totale (c) le long de la chaîne en mg/100g. Le trait plein épais représente la moyenne des concentrations, et les traits pointillés épais délimitent l'intervalle de variabilité à 95%. Les zones en gris foncé représentent les intervalles d'incertitude à 95% sur la moyenne et sur les limites de l'intervalle de variabilité à 95%.

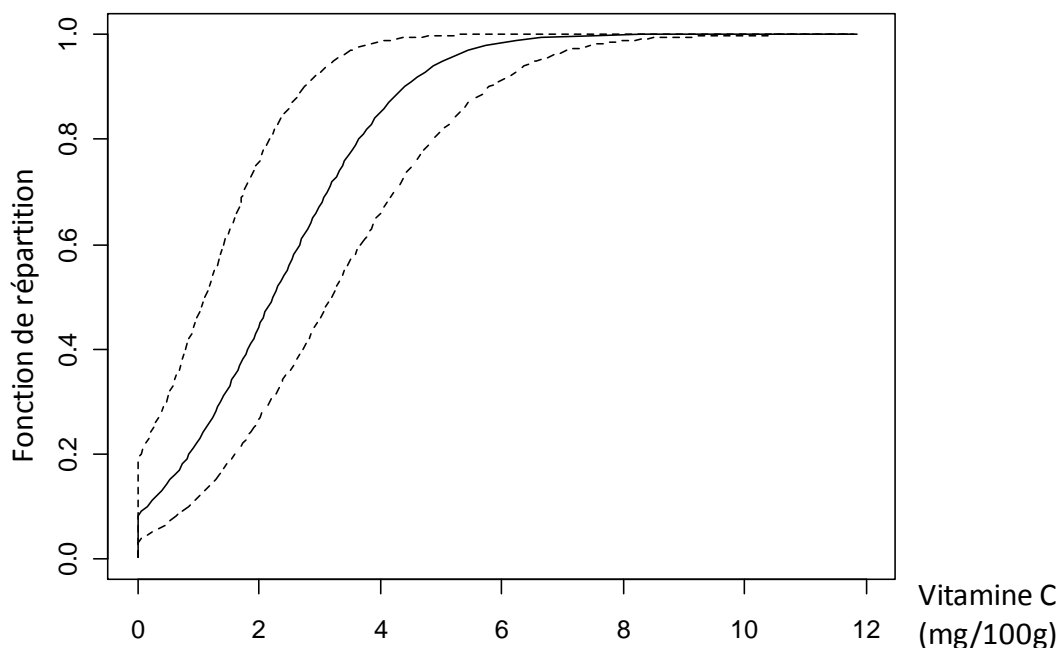


Figure 4.8. Fonction de répartition empirique (trait plein) de la variabilité de la concentration en vitamine C dans la boîte après conservation, avec intervalle d'incertitude à 95% (délimité par les traits pointillés). Pour chaque quantile de variabilité (en ordonnée), les concentrations (en abscisse) comprises entre les lignes pointillées correspondent à l'intervalle d'incertitude à 95% sur le quantile.

4.3.2 Confrontation avec des données indépendantes pour étudier la validité du modèle

Concentrations mesurées le long de la chaîne (Projet ANR Ribenut) :

Dans le cadre du projet ANR Ribenut, les concentrations de vitamine C ont été mesurées à différentes étapes de la chaîne industrielle, et sur une chaîne pilote. Les données ont été obtenues à partir de différentes variétés de haricot vert (Proton et Livorno), différents formats de boîtes (1/4, 1/2, et 4/4), différents types de blanchiments (immersion/aspersion), différents types de coupes des haricots verts, et sont présentées en Tableau 4.8. Certaines données sont à prendre avec précaution car une incertitude non négligeable est attachée à ces mesures, faites par trois laboratoires différents avec différents procédés expérimentaux, et présentant parfois des contradictions (voir l'incertitude liée au dosage de la vitamine C en partie 4.1).

CHAPITRE 4. Modèle d'évaluation du bénéfice

Tableau 4.8. Variations ou teneurs observées en vitamine C au cours du procédé industriel lors des analyses effectuées dans le cadre du projet ANR Ribenut. MP signifie « matière première ».

Procédé observé	Variation (ou teneur) en vitamine C dans les haricots verts
Entre MP et après blanchiment (immersion)	+14% ^{ci} ; -25% ^{si} ; -20% ^{ci} ; -19% ^{ap} ; -8% ^{ap} ; -35% ^{ap}
Entre MP et après blanchiment (aspersion)	-11% ^{ci} ; -60% ^{si} ; -33% ^{ci} ; -35% ^{ap} ; -28% ^{ap} ; -45% ^{ap}
Entre avant et après blanchiment (immersion)	-22% ^{ci} ; -53% ^{si}
Entre MP et après stérilisation	-71% ^{ci} ; -62% ^{ci} ; -69% ^{ci} ; -66% ^{ci} ; -85% ^{ap} ; -89% ^{ap} ; -80% ^{ap}
Entre avant blanchiment et après stérilisation	-60% ^{ci}
Teneur en vitamine C après stérilisation (mg/100g)	2.7 ^{ci} ; 2.9 ^{ci} ; 2.0 ^{ci} ; 2.9 ^{ci} ; 3.6 ^{ci} ; 2.4 ^{ap} ; 1.5 ^{ap} ; 3.0 ^{ap}

^{ci}Analyses effectuées par le CTCPA sur chaîne industrielle.

^{si}Analyses effectuées par l'UMR408 SQPOV (INRA) sur chaîne industrielle.

^{ap}Analyses effectuées par Aérial sur chaîne pilote.

Globalement, les pertes totales en vitamine C observées entre la matière première (MP) et les haricots après stérilisation sont assez similaires, en moyenne de 67% sur la chaîne industrielle et en moyenne de 85% sur la chaîne pilote. Les pertes observées entre la matière première et le blanchiment sont par contre assez variées, entre 5% et 60% (en supposant aberrante la donnée +14%). Ces pertes sont en moyenne de 15% au blanchiment par immersion et de 36% au blanchiment par aspersion. Comme des variations importantes des résultats sont observées selon la méthodologie d'analyse (par exemple pertes après blanchiment comprises entre 11% et 60% sur un même échantillon), il est difficile de savoir si toutes ces variations sont dues à des incertitudes d'analyses ou à de l'hétérogénéité naturelle.

Données de la littérature :

Une synthèse de quelques résultats de la littérature est rassemblée dans le Tableau 4.9.

CHAPITRE 4. Modèle d'évaluation du bénéfice

Tableau 4.9. Pertes observées en vitamine C d'après la littérature. * Résultats communiqués dans la revue bibliographique de Dewettinck et al. (2006).

Référence	Procédé observé	Pertes en vitamine C dans les haricots verts
Tosun et Yucecan, 2007	Préparation des légumes	Pertes de 18% au cours du lavage, ébouttage et coupe
	Préparation + blanchiment	Pertes totales de 25% après préparation des légumes et blanchiment
Oruna-Concha et al., 1998	Blanchiment	Pertes de 28% après blanchiment (4 min à 100°C)
Primo Yufera E., 1979	Blanchiment	Pertes de 26% après blanchiment
Howard et al., 1999	Blanchiment	Pertes de 17% au blanchiment à la vapeur (60 sec à 97°C)
Abou-Fadel et Miller, 1983*	Blanchiment	Pertes de 14% au blanchiment
	Appertisation	Pertes de 47% après appertisation
	Stockage des conserves	Pertes supplémentaires de 15% (resp. 21%) pour un stockage 4 mois à 25°C (resp. 31°C)
Hottenroth, B., 1976*	Appertisation	Pertes de 45%
	Stockage des conserves	Pertes de 6% (resp. 12%) pour le stockage pendant 1 an (resp. 2 ans)
Souci et al. 2000 (Food composition and Nutrition Table)	Taux final après appertisation	Taux de vitamine C : 4.3mg/100g Variation : entre 2 mg/100g et 8 mg/100g

Confrontation avec les prévisions du modèle :

Les données mesurées sur chaîne lors du projet ANR Ribenut et les données de la littérature (Tableau 4.9) peuvent être considérées indépendantes de celles utilisées pour construire le modèle. En effet seulement une partie de ces données a été utilisée, afin de modéliser la concentration initiale de vitamine C dans le modèle. On peut donc comparer les prévisions du modèle avec les pertes relevées lors de prélèvements du projet ANR Ribenut ou dans la littérature. Des incertitudes sur les mesures et des différences de procédé expérimentales ou conditions de traitements empêchent toutefois des comparaisons précises, il s'agit plutôt de comparer l'ordre de grandeur des pertes.

- Les pertes relevées dans la littérature pendant l'étape de préparation des haricots verts sont de 18%, ce qui est en accord avec les prévisions du modèle (15% en moyenne).
- Le modèle semble plutôt surestimer les pertes totales entre la matière première et le blanchiment (56% en moyenne pour le modèle ; entre 14% et 28% d'après la littérature ;

CHAPITRE 4. Modèle d'évaluation du bénéfice

environ 30% (entre 8% et 60%) d'après les mesures du projet Ribenut). L'effet du blanchiment semble donc un peu surestimé dans le modèle.

- Les pertes totales prédites par le modèle après stérilisation (64% en moyenne) sont en très bon accord avec les pertes observées sur la chaîne industrielle (67%), et sous-estiment légèrement les pertes observées sur la chaîne pilote (85%). Par contre elles surestiment un peu les pertes relevées dans la littérature (proches de 46%), qu'on considère les pertes prédites après stérilisation (64%) ou après conservation (86%).
- Les concentrations en vitamine C prédites dans les haricots verts juste après stérilisation sont globalement en accord avec les teneurs mesurées lors du projet ANR Ribenut même si elles les surestiment légèrement (prévisions de 6.2 mg/100g en moyenne, avec un intervalle de variation à 95% entre 0.3 mg/100g et 12.5 mg/100g ; mesures entre 1.5 et 3.6 mg/100g). (Il faut cependant remarquer que la diffusion des vitamines dans le jus n'est pas prise en compte à cet instant dans le modèle mais a peut-être déjà commencé dans haricots verts soumis aux mesures expérimentales, ce qui pourrait expliquer la légère surestimation.)
- Le taux final prédit de vitamine C dans les haricots verts appertisés est en assez bon accord avec les teneurs relevées dans la littérature (prévisions de 2.3 mg/100g en moyenne avec un intervalle de variation à 95% de [0 ; 5.6] mg/100g ; observations de 4.3 mg/100g en moyenne, variant entre 2 et 8 mg/100g).

Globalement, les prévisions du modèle semblent donc en assez bon accord avec les pertes et teneurs relevées dans la littérature et lors des mesures du projet ANR. Les pertes prédites au blanchiment semblent cependant un peu surestimées, peut-être aux dépens des pertes à la stérilisation qui pourraient être légèrement sous-estimées.

4.3.3 Analyse de sensibilité

4.3.3.a Impact de la variabilité des facteurs

Une analyse de sensibilité, basée sur les indices de Sobol (Saltelli, 2002) et les coefficients de corrélation de Spearman, a permis de déterminer quels étaient les paramètres variables qui influençaient le plus la variabilité de la concentration finale en vitamine C. Cette analyse de sensibilité a été faite à incertitude fixée dans la dimension variabilité, et répétée plusieurs fois pour différentes valeurs d'incertitudes, suivant la démarche de Mokhtari and Frey (2005) (initialement proposée à partir d'une ANOVA) . Des intervalles d'incertitude sur les indices de sensibilité sont ainsi obtenus. La gamme de variabilité des paramètres est celle définie dans la description du modèle (Tableau 4.2). Les indices d'ordre 1, notés S_i , prennent en compte l'influence du paramètre seul, tandis que les indices totaux, notés S_{ti} , prennent en compte l'influence du paramètre seul plus les interactions du paramètre avec les autres paramètres. Ces indices sont compris entre 0 et 1 et plus ils sont élevés, plus le paramètre est influent. Il est possible de les calculer car, conditionnellement aux

CHAPITRE 4. Modèle d'évaluation du bénéfice

variables incertaines, le modèle est déterministe et les facteurs variables sont indépendants. Par ailleurs, les coefficients de corrélation de Spearman entre les facteurs variables et la concentration en vitamine C après conservation ont également été calculés (notés $C_{Spearman,i}$). Compris entre -1 et 1, leur signe indique le sens de la corrélation, et plus ils sont proches de 1 en valeur absolue, plus la corrélation indiquée est forte.

Les différents indices ont été calculés en prenant $N_v = 20\ 000$ tirages dans la dimension variabilité, et $N_u = 1500$ tirages dans la dimension incertitude, et en utilisant le package R *sobolj2002* pour calculer les indices de Sobol.

Tableau 4.10. Indices de Sobol et coefficient de corrélation de Spearman mesurant l'influence de la variabilité des facteurs variables sur la variabilité du taux de vitamine C après conservation. Les intervalles de confiance à 95% représentent l'incertitude sur les estimations (incertitude issue de la dimension incertitude du modèle).

Facteurs	St_i	S_i	$C_{Spearman,i}$
C_{vitC}^{INIT}	0.97 [0.91; 1.04]	1.08 [0.95; 1.39]	0.96 [0.93; 0.97]
t_{blan}	0.04 [0.02; 0.09]	0.03 [0.02; 0.07]	-0.16 [-0.23; -0.12]
t_{wait}	0.02 [0.01; 0.04]	0.02 [0.01; 0.06]	-0.13 [-0.18; -0.09]
k_{ev}	0.01 [0.01; 0.03]	0.02 [0.01; 0.04]	-0.11 [-0.15; -0.08]
r_{GB}	0.00 [0.00; 0.00]	0.00 [0.00; 0.00]	0.03 [0.01; 0.04]
T_{blan}	0.00 [0.00; 0.00]	0.00 [0.00; 0.00]	-0.02 [-0.04; 0.00]
F_1	0.00 [0.00; 0.00]	0.00 [0.00; 0.00]	0.00 [-0.02; 0.01]
t_{trans}	0.00 [0.00; 0.00]	0.00 [0.00; 0.00]	-0.01 [-0.02; 0.01]
T_{trans}	0.00 [0.00; 0.00]	0.00 [0.00; 0.00]	0.00 [-0.01; 0.01]
t_{stock}	0.00 [0.00; 0.01]	0.00 [0.00; 0.01]	0.00 [-0.08; 0.01]
T_{stock}	0.00 [0.00; 0.01]	0.00 [0.00; 0.01]	0.00 [-0.07; 0.01]
air_{can}	0.00 [0.00; 0.00]	0.00 [0.00; 0.00]	-0.07 [-0.12; -0.04]

Les résultats sont présentés en Tableau 4.10. Malgré le grand nombre de tirages dans la dimension variabilité, l'estimation des indices de Sobol pose quelques problèmes car l'estimation de l'indice de Sobol d'ordre 1 de la concentration initiale en vitamine C (1.08 en moyenne) est supérieure à 1 et supérieure à l'indice de Sobol total correspondant (0.97 en moyenne). Ce problème d'estimation semble lié à l'influence très forte de ce paramètre. Cependant les résultats sont stables (indices estimés très proches quand on les recalcule) et sont en accord avec les coefficients de corrélation de Spearman, qui indiquent aussi l'influence majeure de la concentration initiale en vitamine C (coefficient de 0.96). Les différents indices montrent que la variabilité de la concentration finale en vitamine C dépend principalement de 4 paramètres : la concentration initiale totale en vitamine C, C_{vitC}^{INIT} , les durées de blanchiment et de préparation des légumes t_{blan} et t_{wait} , et le coefficient de vitesse de l'évolution physiologique k_{ev} . L'effet de la concentration initiale en vitamine

CHAPITRE 4. Modèle d'évaluation du bénéfice

C est largement prépondérante car elle explique à elle seule plus de 90% de la variabilité (intervalle de confiance de S_i au-delà de 90%). L'incertitude sur ces indices est assez faible et ne semble pas de nature à remettre en question l'influence importante de ces 4 paramètres. Il est intéressant de remarquer que le coefficient de corrélation de Spearman de air_{can} est significativement différent de 0 ($S_i = -0.07$; 95%CI=[-0.12 ; -0.04]), indiquant une faible corrélation négative entre la concentration finale en vitamine C et le volume de l'espace de tête, ce que n'indiquait pas les indices de Sobol. Par ailleurs, les variations du traitement thermique à la stérilisation (F_1) ne semblent pas avoir d'influence sur la variation de la concentration finale en vitamine C ($St_i = 0.00$ et $C_{Spearman,i} = 0.00$). La Figure 4.9 illustre la corrélation élevée entre les teneurs en vitamine C dans les haricots verts appertisés et dans les haricots verts frais.

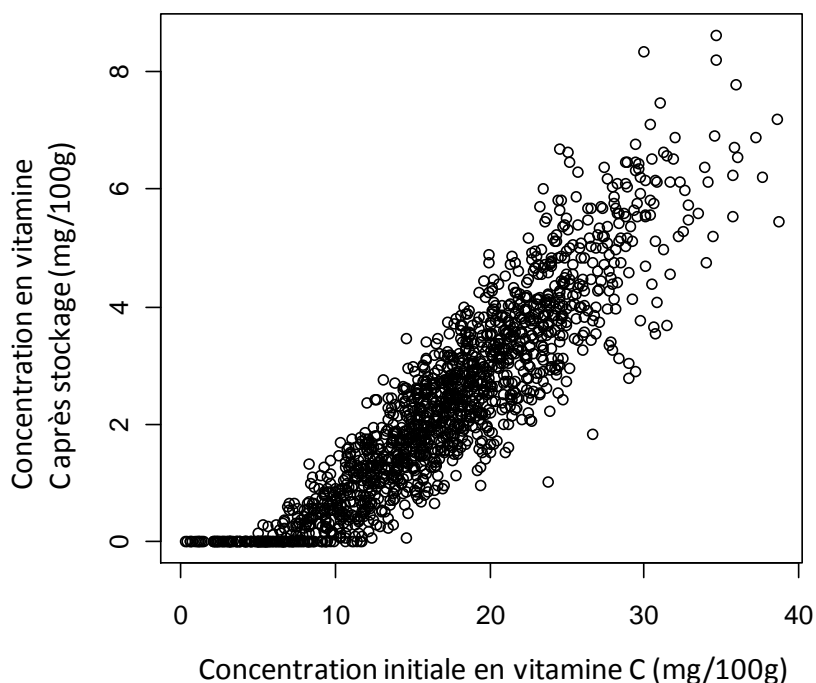


Figure 4.9. Concentration en vitamine C dans les haricots verts après entreposage en fonction de la concentration en vitamine C dans les haricots frais (échantillon de 1500 simulations).

4.3.3.b Impact de toutes les sources de variation des facteurs

Un autre type d'analyse de sensibilité a été réalisée, afin de répondre à une autre question que l'impact des variations des facteurs d'entrée variables sur la variation de la sortie. Cette fois il s'agit d'étudier l'influence de chaque facteur (variable et/ou incertain) sur la plage de variation totale qu'il prend dans le modèle. Cette analyse permet en particulier d'analyser l'impact de l'incertitude des facteurs incertains, qui sont nombreux dans ce modèle. Selon l'impact plus ou moins grand de cette incertitude, il sera plus ou moins important de rechercher à réduire cette incertitude.

CHAPITRE 4. Modèle d'évaluation du bénéfice

L'analyse de sensibilité a été appliquée à tous les facteurs d'entrée, sur un échantillon de 100 000 simulations, et la variable de sortie était toujours la concentration en vitamine C après entreposage. Afin d'estimer correctement tous les indices de sensibilité, le même nombre de simulations (100 000) a été utilisé pour tous les facteurs du modèle, qu'ils soient incertains, variables ou à la fois incertains et variables, et l'analyse de sensibilité a donc été faite à incertitude et variabilité mélangés. Les indices de Sobol d'ordre 1 et totaux, et le coefficient de corrélation de Spearman ont été calculés, ainsi que les intervalles d'incertitude estimés par bootstrap.

Tableau 4.11. Indices de Sobol et coefficient de corrélation de Spearman mesurant l'influence des variations de tous les facteurs d'entrée sur la variabilité du taux de vitamine C après conservation. Les intervalles de confiance à 95% représentent l'incertitude sur les estimations (incertitude issue de la méthode d'estimation et pas de la dimension incertitude du modèle).

Facteurs	St_i	S_i	$C_{Spearman,i}$
C_{vitC}^{INIT}	0.85 [0.83; 0.97]	0.76 [0.74; 0.78]	0.89 [0.89; 0.90]
$prop_{AA}$	0.10 [0.09; 0.10]	0.08 [0.07; 0.08]	0.26 [0.26; 0.26]
t_{blan}	0.04 [0.03; 0.04]	0.03 [0.03; 0.04]	-0.16 [-0.16; -0.15]
β	0.03 [0.03; 0.04]	0.03 [0.02; 0.03]	0.16 [0.15; 0.16]
t_{wait}	0.02 [0.01; 0.02]	0.02 [0.01; 0.02]	-0.12 [-0.13; -0.12]
k_{ev}	0.02 [0.01; 0.02]	0.01 [0.01; 0.02]	-0.11 [-0.12; -0.10]
k_{AA}^{ref}	0.01 [0.01; 0.01]	0.01 [0.01; 0.01]	-0.08 [-0.08; -0.07]
air_{can}	0.01 [0.01; 0.01]	0.00 [0.00; 0.00]	-0.07 [-0.08; -0.07]
Ea_{DHA}^{chim}	0.01 [0.01; 0.01]	0.00 [0.00; 0.01]	0.02 [0.02; 0.03]
Dif_{ref}	0.00 [0.00; 0.00]	0.00 [0.00; 0.00]	-0.03 [-0.04; -0.02]
t_{stock}	0.00 [0.00; 0.00]	0.00 [0.00; 0.00]	-0.01 [-0.01; 0.00]
T_{stock}	0.00 [0.00; 0.00]	0.00 [0.00; 0.00]	-0.01 [-0.01; 0.00]
Ea_{AA}^{chim}	0.00 [0.00; 0.00]	0.00 [0.00; 0.00]	-0.01 [-0.02; -0.00]
k_{DHA}^{ref}	0.00 [0.00; 0.00]	0.00 [0.00; 0.00]	-0.00 [-0.01; 0.00]
Ea^{diff}	0.00 [0.00; 0.00]	0.00 [0.00; 0.00]	-0.02 [-0.02; -0.02]
T_{blan}	0.00 [0.00; 0.00]	0.00 [0.00; 0.00]	-0.02 [-0.02; -0.01]
F_1	0.00 [0.00; 0.00]	0.00 [0.00; 0.00]	0.00 [-0.01; 0.00]
t_{trans}	0.00 [0.00; 0.00]	0.00 [0.00; 0.00]	-0.01 [-0.01; 0.00]
T_{trans}	0.00 [0.00; 0.00]	0.00 [0.00; 0.00]	0.00 [0.00; 0.01]
r_{GB}	0.00 [0.00; 0.00]	0.00 [0.00; 0.00]	0.02 [0.02; 0.03]

CHAPITRE 4. Modèle d'évaluation du bénéfice

Les résultats de cette analyse sont présentés en Tableau 4.11. Les facteurs variables les plus influents sont les mêmes que ceux déterminés dans l'analyse de sensibilité précédente, à savoir $C_{vit\ C}^{INIT}$, t_{blan} , t_{wait} et k_{ev} , avec toujours une influence majeure de la concentration initiale en vitamine C $C_{vit\ C}^{INIT}$ ($St_i = 0.85$ et $C_{Spearman,i} = 0.89$). De plus, des facteurs incertains ont une influence assez importante, à savoir la proportion $prop_{AA}$ de AA dans la concentration initiale en vitamine C ($St_i = 0.10$ et $C_{Spearman,i} = 0.26$), et le coefficient β d'ordre de la réaction d'oxydation de AA par rapport à l'oxygène ($St_i = 0.03$ et $C_{Spearman,i} = 0.16$). L'incertitude sur d'autres facteurs incertains comme les énergies d'activation semble par contre avoir peu d'impact. Cette fois on n'observe pas d'aberration sur l'estimation des indices de Sobol puisque tous les indices de Sobol et leurs intervalles de confiance sont bien compris entre 0 et 1 et les indices totaux sont supérieurs aux indices d'ordre 1. Cette amélioration des estimations s'explique peut-être par la présence de variables assez fortement influentes sur la sortie autres que $C_{vit\ C}^{INIT}$, comme $prop_{AA}$. Au vu de ces résultats, il semblerait donc qu'il serait judicieux de chercher à réduire l'incertitude prioritairement sur $prop_{AA}$ et β afin de réduire l'incertitude globale sur les résultats du modèle.

4.3.4 Tests de scénarios

Des scénarios de deux types ont été testés (Tableau 4.12), afin de tester soit des hypothèses incertaines du modèle (scénarios 1 à 6), soit des mesures permettant d'augmenter le bénéfice nutritionnel (scénarios 7 à 12). Le choix de ces scénarios s'inspire des résultats de l'analyse de sensibilité. Parmi les scénarios incertains, différentes valeurs ont été testées pour les paramètres incertains $prop_{AA}$ et β , dont les influences semblaient non négligeables sur la sortie du modèle d'après l'analyse de sensibilité. Les scénarios 5 et 6 visent en fait à tester l'hypothèse que seulement une partie de l'oxygène de l'espace de tête serait consommée par l'oxydation de AA lors de l'entreposage (ce qui peut se traduire par une réduction du volume de l'espace de tête air_{can}). Les scénarios 7 à 12 permettent de voir l'impact de réductions du temps standard de dégradation à 85°C de DHA F_1 et des durées d'attente et de blanchiment t_{wait} et t_{blan} .

Les résultats de ces scénarios montrent, au niveau des scénarios sur les hypothèses incertaines, que les valeurs de $prop_{AA}$ et β ont effectivement un impact assez important sur la concentration finale en vitamine C, puisque la concentration finale moyenne varie du simple au double selon les valeurs testées (1.2 mg/100g à 2.5 mg/100g pour $prop_{AA}$ et 1.0 mg/100g à 2.3 mg/100g pour β). Une diminution de moitié de la part d'oxygène de tête consommée entraînerait une légère augmentation de la concentration finale moyenne en vitamine C (de 2.3 mg/100g à 2.7 mg/100g). Il serait donc intéressant d'améliorer la connaissance de ces différents phénomènes, même si au final le taux de vitamine C reste dans tous les cas très bas.

Au niveau des possibles mesures pour augmenter la concentration finale en vitamine C, le test de réduction de F_1 va dans le même sens que les résultats de l'analyse de sensibilité puisque cette réduction n'a pas d'impact ; la diminution des durées de blanchiment et d'attente permet une augmentation des teneurs finales en vitamine C, de 65% en moyenne pour $t_{blan} = 1.5$ min et de 26% en moyenne pour $t_{wait} = 60$ min.

CHAPITRE 4. Modèle d'évaluation du bénéfice

Tableau 4.12. Tests de scénarios sur le modèle et résultats correspondants. Les intervalles de confiance à 95% représentent l'incertitude sur les estimations.

	What-if-scenario	Concentration moyenne en vitamine C (mg/100g) après entreposage
0	Modèle de référence	2.3 [1.2; 3.2]
1	$prop_{AA} = 90\%$	2.5 [1.8; 3.2]
2	$prop_{AA} = 60\%$	1.2 [0.7; 1.6]
3	$\beta = 0.9$	2.3 [1.4; 3.1]
4	$\beta = 0.5$	1.0 [0.3; 1.8]
5	$air_{can} = 2.25 \text{ mL}$	2.7 [1.7; 3.6]
6	$air_{can} = 1 \text{ mL}$	3.0 [2.0; 3.9]
7	$F_1 = 180 \text{ min}$	2.3 [1.3; 3.2]
8	$F_1 = 150 \text{ min}$	2.3 [1.3; 3.2]
9	$t_{wait} = 140 \text{ min}$	2.8 [1.8; 3.8]
10	$t_{wait} = 60 \text{ min}$	2.9 [1.8; 3.8]
11	$t_{blan} = 3.5 \text{ min}$	3.1 [1.7; 4.0]
12	$t_{blan} = 1.5 \text{ min}$	3.8 [2.3; 4.9]

4.4 Conclusion

Un modèle quantitatif d'analyse des bénéfices nutritionnels au cours de la transformation d'un aliment a été construit, en utilisant le cadre habituellement développé dans l'analyse quantitative des risques microbiologiques. Le développement de tels modèles est encore rare dans le domaine du bénéfice nutritionnel, d'autant plus que les sources d'incertitudes et de variabilité ont été prises en compte par simulation de Monte Carlo, et de manière séparée en faisant du Monte Carlo 2D.

La construction d'un tel modèle demande une bonne connaissance des phénomènes impliqués, l'existence de modèles mathématiques décrivant ces phénomènes, et la connaissance des valeurs possibles de toutes les variables chimiques, thermiques, physiques, environnementales intervenant dans ces modèles. La complexité des mécanismes de dégradation chimiques de la vitamine C est un obstacle au développement de modèles mécanistiques, qui sont moins développés qu'en analyse du risque microbiologiques, et constituent encore des thèmes de recherche prometteurs (Leskova et al., 2006; Pénicaud, 2009). Le manque de données pour estimer certains paramètres, tels que les paramètres chimiques relatifs au DHA, a été patent. De plus, les vitamines comme la vitamine C étant très fragiles, les méthodes de dosages des vitamines contenues dans des échantillons de légumes ont pu donner lieu à de nombreuses incertitudes. En conséquence, ce

CHAPITRE 4. Modèle d'évaluation du bénéfice

modèle de dégradation de la vitamine C dans la conserve de haricot vert n'a pas été simple à construire, et de nombreuses hypothèses, plus ou moins simplificatrices, ont été faites. Ainsi l'effet du pH a été négligé, les taux d'oxygène ont été approchés à partir de la table de Winkler, et la plupart des paramètres chimiques ont été estimés à partir d'analyses effectuées à partir de matrices différentes de celle du haricot vert, à des températures parfois un peu éloignées des procédés modélisés. Ce modèle a toutefois le mérite de proposer une première approche de modélisation de la dégradation de la vitamine C le long de la chaîne de fabrication de la conserve de haricot vert, et pourra être amélioré par la suite.

Les pertes globales prédites par le modèle entre la matière première fraîche et les légumes appertisés sont de 64% en moyenne juste après stérilisation et de 86% en moyenne après stérilisation et conservation. Ces pertes semblent en accord avec des résultats issues d'analyses effectuées sur chaîne industrielle et sur chaîne pilote, et semblent surestimer un peu les pertes reportées dans la littérature (partie 4.3.2). Le modèle semble donc globalement plutôt réaliste. Cependant les pertes prédites au blanchiment semblent un peu surestimées par rapport aux pertes observées, et en conséquence les pertes prédites à la stérilisation sont peut-être sous-estimées, ce qui nécessiterait des études et analyses supplémentaires pour comprendre ces écarts possibles et améliorer ces aspects du modèle.

Le taux final de vitamine C dans les haricots verts appertisés prédit par le modèle est assez faible (entre 0 et 5.7 mg/100g, en moyenne de 2.3 mg/100g), ce qui est en accord avec différentes mesures effectuées dans la littérature. Les concentrations en vitamine C semblent notamment nettement inférieures au seuil d'allégation nutritionnelle (12 mg/100g), montrant clairement que les haricots verts appertisés ne sont pas une source principale de vitamine C dans l'alimentation humaine. Les pertes en vitamine C au cours du procédé d'appertisation sont néanmoins élevées (64% en moyenne d'après le modèle ; proches de 46% dans la littérature), ce qui laisse entrevoir des pistes possibles pour diminuer ces pertes. Une possibilité intéressante du modèle est d'explorer des scénarios pouvant potentiellement limiter ces pertes (voir partie 4.3.4). L'analyse de sensibilité effectuée dans ce chapitre (partie 4.3.3) est une première approche permettant d'explorer les causes des variations des teneurs finales en vitamine C. Le paramètre le plus influent est la concentration initiale en vitamine C, non modifiable. Deux paramètres ayant une certaine influence sur la teneur finale en vitamine C sont les durées de blanchiment et de préparation des légumes, qui constituent des leviers potentiels d'optimisation. L'impact de la réduction de ces durées a été testé sous forme de scénarios, aboutissant à des augmentations allant jusqu'à 65% des teneurs finales en vitamine C, qui restent cependant globalement faibles. A contrario, les possibles variations de certains paramètres, comme la température de transfert ou la durée d'entreposage, semblent avoir une influence quasiment nulle sur la teneur finale en vitamine C, d'après le modèle. Pour simplifier le modèle sans perdre beaucoup d'information, ces paramètres pourraient donc être supposés fixes. De manière générale cependant il faut rester prudent avec les conclusions de l'analyse de sensibilité ou des scénarios testés tant que le modèle n'a pas été complètement validé.

Chapitre 5

Analyse risque-bénéfice dans la chaîne de fabrication des haricots verts appertisés

5.1 Introduction

Une analyse risque-bénéfice a été réalisée dans ce chapitre. Elle vise l'optimisation du processus de fabrication des haricots verts appertisés, dans le but d'augmenter le bénéfice nutritionnel des haricots verts appertisés tout en gardant à un niveau acceptable le risque de non-stabilité. Il s'agit de proposer des interventions pertinentes au cours du procédé de fabrication, en prenant en compte la faisabilité des modifications (contraintes industrielles). Ces interventions sont traduites sous la forme d'une modification des valeurs de certaines variables du modèle.

Le risque est représenté par le taux de non-stabilité dans la conserve de haricots verts (modèle présenté au Chapitre 3). Le bénéfice est représenté par la teneur en vitamine C dans les haricots verts appertisés (modèle présenté au Chapitre 4). Dans la suite de ce chapitre, le modèle représentant l'évolution des teneurs en vitamine C sera appelé « modèle bénéfice » et le modèle représentant l'évolution des concentrations en *G. stearothermophilus* sera appelé « modèle risque ». D'après ces modèles, la teneur en vitamine C dans les haricots verts appertisés après stockage est assez faible (2.3 mg/100g en moyenne), bien en dessous du seuil d'allégation nutritionnelle (12 mg/100g). Le risque de non-stabilité est également faible (0.5% en moyenne). Il s'agit donc de chercher à augmenter le bénéfice tout en gardant un risque faible.

L'analyse risque-bénéfice a été faite au moyen de critères probabilistes de dépassement de seuils de référence. Dans le cas de la vitamine C, le seuil d'allégation nutritionnelle a été choisi car il est concrètement utilisable du point de vue industriel : la présence de vitamine C peut être mentionnée sur l'étiquette du produit à condition que sa teneur soit supérieure au seuil d'allégation nutritionnelle. Dans le cas de *G. stearothermophilus*, le seuil de non-stabilité ($7 \log_{10}$ UFC/g) a été choisi car il correspondait aussi à une réalité industrielle. La probabilité que la teneur en vitamine C soit au-dessus du seuil d'allégation nutritionnelle (12 mg/100g) dans une boîte de conserve qui reste

CHAPITRE 5. Analyse risque-bénéfice

stable a été calculée ; elle constitue un critère englobant simultanément le risque et le bénéfice et peut donc être utilisée facilement pour comparer différents scénarios. Cette approche a été retenue pour sa simplicité. Des indicateurs plus complexes type DALY ou QALY, mesurant la qualité de la vie (voir partie 1.1.3), ne seraient pas adaptées à cette étude où la chaîne modélisée finit à la fin de la production et où le risque n'est pas sanitaire.

Dans ce chapitre, les variables communes entre le modèle risque et le modèle bénéfique sont explicitées. Un lien entre les deux types de temps équivalents de stérilisation (pour la vitamine et pour le risque) est notamment construit. Les résultats conjoints des deux modèles sont ensuite analysés, pour rechercher les corrélations possibles entre risque et bénéfice. Enfin, des mesures d'optimisation du compromis risque-bénéfice ont été recherchées au moyen de tests de scénarios sur le modèle global risque-bénéfice, en utilisant des critères probabilistes de dépassement simultané de seuils de référence.

5.2 Variables communes aux modèles risque et bénéfique

5.2.1 Lien entre le temps standard de dégradation à 85°C de DHA F_1 et la valeur stérilisatrice F_0

Dans le modèle risque comme dans le modèle bénéfique, la stérilisation est une étape clé de l'évolution des quantités d'intérêt. Des temps équivalents stérilisation pour la vitamine C et pour *G. stearothermophilus* ont été construits à partir des mêmes relevés de profils temps/températures de stérilisation fournis par l'industriel. Recourir à ces profils temps/températures et à des temps équivalents permet de prendre en compte de manière précise le traitement thermique, dont le rôle est particulièrement crucial dans le modèle risque. Le calcul des temps équivalents de stérilisation pour le risque et le bénéfique fait respectivement appel aux paramètres de résistance thermique de *G. stearothermophilus* et de la vitamine C (DHA). Les temps équivalents sont donc différents dans les deux modèles. La transformation appliquée aux temps équivalents (voir Chapitres 3 et 4) a permis de distinguer l'effet du traitement thermique de celui des paramètres de résistance thermique, ce qui va permettre d'appliquer un traitement thermique similaire à la vitamine C et à la bactérie *G. stearothermophilus* présents dans une même boîte de conserve.

Ainsi dans le modèle risque, on a montré que le temps équivalent stérilisation pouvait s'écrire :

$$t_{eq,Gbs}(z_T, F_0) = \exp\left(b\left(\frac{1}{z_T} - \frac{1}{z_{T,REF}}\right)\right) \cdot F_0$$

avec $F_0 \sim LnN(3.40, 0.11)$ min.

CHAPITRE 5. Analyse risque-bénéfice

F_0 correspond ici à la valeur stérilisatrice classiquement utilisée en microbiologie, correspondant à un z_T de référence $z_{T,REF} = 10^\circ\text{C}$ et à la température de référence de 121.1°C .

Pour le modèle bénéfice, on a montré que le temps équivalent de stérilisation pouvait s'écrire :

$$t_{eq\ vit}(Ea_{DHA}^{chim}, F_1) = \exp\left(b'(Ea_{DHA}^{chim} - Ea_{DHA,ref}^{chim})\right) \cdot F_1$$

$$\text{avec } F_1 \sim LnN(5.30, 0.05) \text{ min.}$$

F_1 représente ici le temps équivalent vitamine à la température de référence de 85°C et à l'énergie d'activation de DHA de référence $Ea_{DHA,ref}^{chim} = 80 \text{ kJ/mol}$, aussi appelé temps standard de dégradation à 85°C de DHA.

Les distributions sur F_0 et F_1 représentent toutes deux la variabilité des traitements thermiques. Une façon de lier ces deux paramètres est de tirer le même quantile dans chacune des deux distributions. Par exemple si pour une boîte de conserve on tire le quantile d'ordre 0.85 pour la valeur stérilisatrice F_0 (ce qui correspond à un F_0 de 33.6 min), en tirant le quantile 0.85 dans la distribution de F_1 ($F_1 = 210.2 \text{ min}$), un traitement thermique du même ordre sera appliqué aux vitamines (DHA). Dans le modèle global risque-bénéfice, l'intensité du traitement thermique sera expliquée à partir du paramètre F_0 uniquement, qui est le paramètre de référence des industriels. La valeur de F_1 sera déduite de celle de F_0 (en simulant le même quantile).

L'hypothèse faite pour lier F_1 à F_0 est donc de faire correspondre leurs quantiles. Une justification à cette hypothèse est obtenue par la comparaison des valeurs de F_0 et F_1 calculées directement et indépendamment à partir des temps équivalents des 6 profils temps-températures fournis par l'industriel (en utilisant $z_T=10^\circ\text{C}$ et $Ea_{DHA}^{chim} = 80 \text{ kJ/mol}$). Ces 6 valeurs de couples (F_0, F_1) sont représentées en Figure 5.1 (o) ; un coefficient de corrélation linéaire de Pearson de 0.84, très significativement différent de 0, leur est associé. Ainsi, même si la tendance observée n'est pas parfaitement croissante, en particulier sur les 3 profils les plus bas, F_1 est fortement corrélé positivement à F_0 , et donc globalement plus le traitement thermique appliqué est élevé du point de vue microbiologique, plus il est élevé du point de vue vitamine, ce qui justifie la correspondance effectuée entre F_0 et F_1 par le biais des quantiles. Un aléa aurait pu être rajouté mais cela n'a pas été fait par simplification et car d'un point de vue théorique nous voulions appliquer un traitement thermique « identique » dans les 2 modèles. Sur la Figure 5.1, les valeurs de F_1 associées aux mêmes quantiles que les valeurs de F_0 (pour les distributions déterminées pour F_1 et F_0 dans les modèles) sont représentées par des croix (+). Il serait toutefois intéressant d'approfondir l'étude du lien entre F_0 et F_1 à partir d'un nombre plus élevé de profils temps-températures (des vrais ou à défaut, des profils simulés).

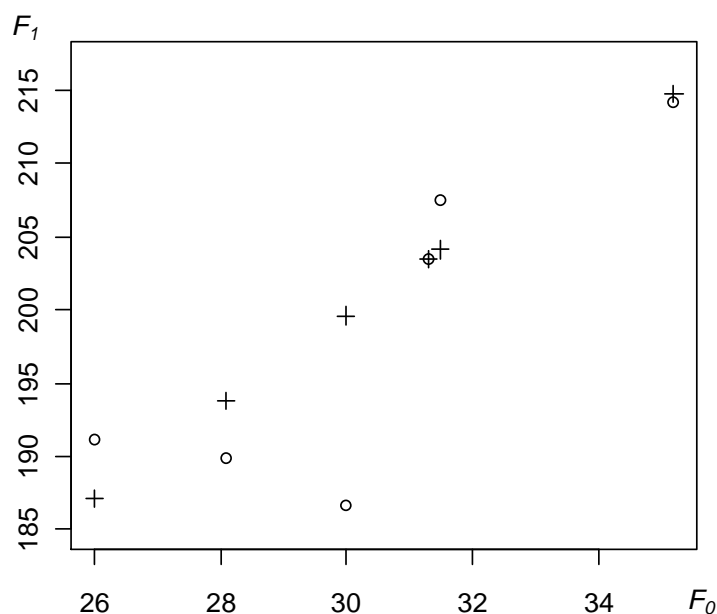


Figure 5.1. Temps standard de dégradation à 85°C de DHA F_1 en fonction de la valeur stérilisatrice F_0 pour les 6 profils temps-températures disponibles fournis par l'industriel. F_0 a été calculé à partir des profils et de la formule des temps équivalents bactériens (pour $z_T = 10^\circ\text{C}$). (o) F_1 a été calculé à partir des profils et de la formule du temps équivalent pour DHA (pour $Ea_{DHA}^{chim} = 80\text{kJ/mol}$). (+) F_1 a été calculé de manière à correspondre au même quantile que F_0 pour leurs distributions respectives dans le modèle.

5.2.2 Graphe du modèle global risque-bénéfice

La Figure 5.2 présente le graphe simplifié du modèle global « risque-bénéfice » (les paramètres non communs aux deux sous-modèles risque et bénéfique sont supposés marginalement indépendants). Les paramètres temps/températures du procédé de fabrication sont représentés par des ellipses vertes. Les seuls paramètres communs entre les deux sous-modèles sont la durée et la température de blanchiment (t_{blan} , T_{blan}), ainsi que la valeur stérilisatrice (F_0).

D'après l'analyse de sensibilité effectuée sur le modèle risque (Chapitre 3, partie 3.2.3.d), la variation des durées et températures de blanchiment t_{blan} et T_{blan} influence très peu le taux de non-stabilité, alors que la variation de la valeur stérilisatrice F_0 l'influence fortement. D'après l'analyse de sensibilité effectuée sur le modèle bénéfique (Chapitre 4, partie 4.3.3), la variation des durées et températures de blanchiment t_{blan} et T_{blan} influence légèrement la teneur en vitamine C après entreposage, et la variation du temps standard de dégradation à 85°C de DHA (F_1) ne semble pas avoir d'influence sur la teneur en vitamine C après entreposage. Les modèles risque et bénéfique

CHAPITRE 5. Analyse risque-bénéfice

partagent donc seulement trois facteurs en commun. Qui plus est les facteurs ayant une influence sur un modèle ont peu d'influence sur l'autre modèle. Une faible corrélation entre les concentrations finales en vitamine C et en *G. stearotherophilus* est donc attendue.

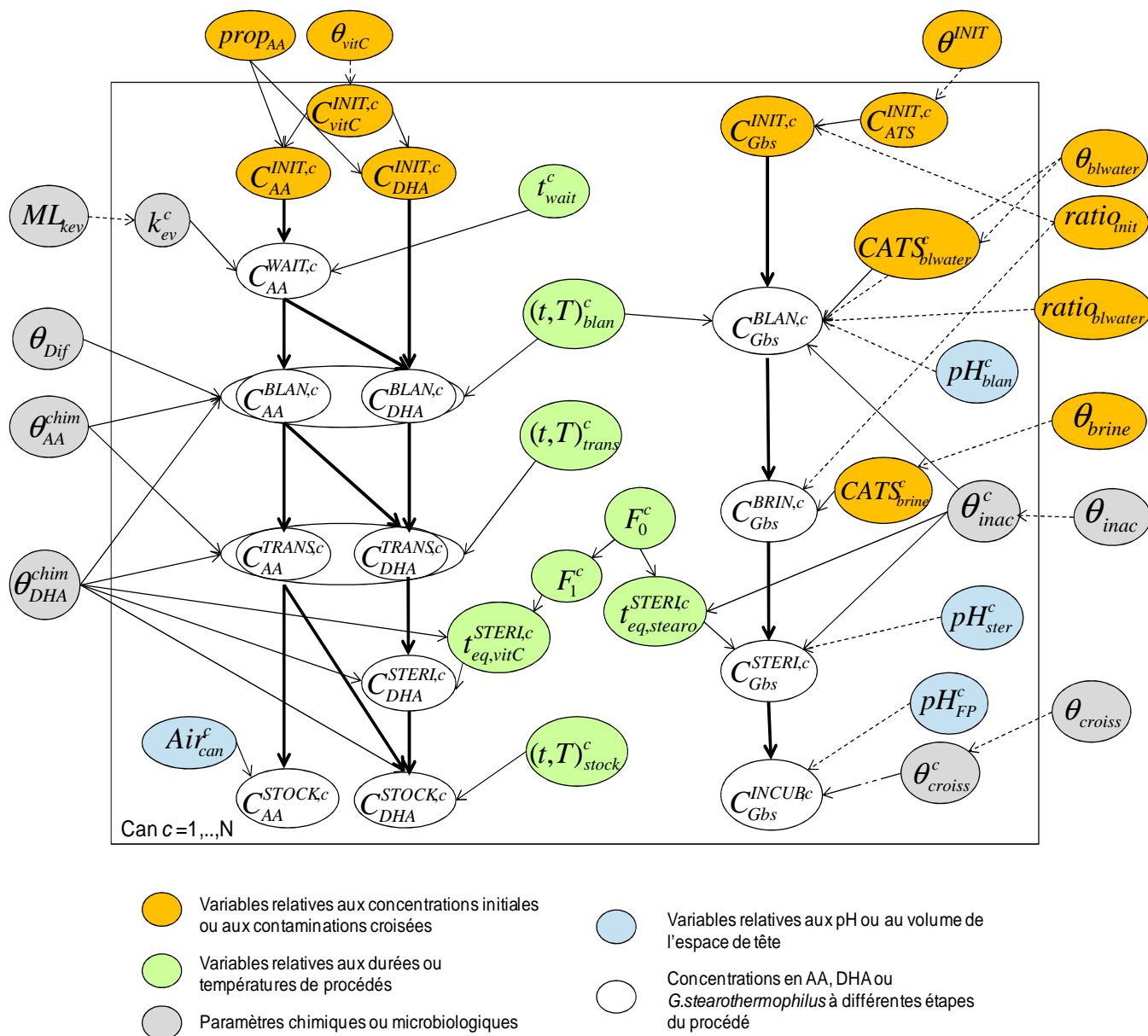


Figure 5.2. DAG simplifié du modèle global risque-bénéfice dans la chaîne de fabrication des haricots verts appertisés. Les variables sont définies dans les Chapitres 3 et 4, à l'exception des facteurs qui ont été regroupés pour plus de lisibilité (θ_X).

5.3 Résultats simultanés des modèles risque et bénéfice

Le modèle global risque-bénéfice a été étudié en simulant 1,5 millions de boîtes de haricots verts appertisés ($N_u = 1000$, $N_v = 1500$), en suivant simultanément l'évolution des concentrations en *G. stearothermophilus* et en vitamine C dans les mêmes boîtes (Figure 5.3).

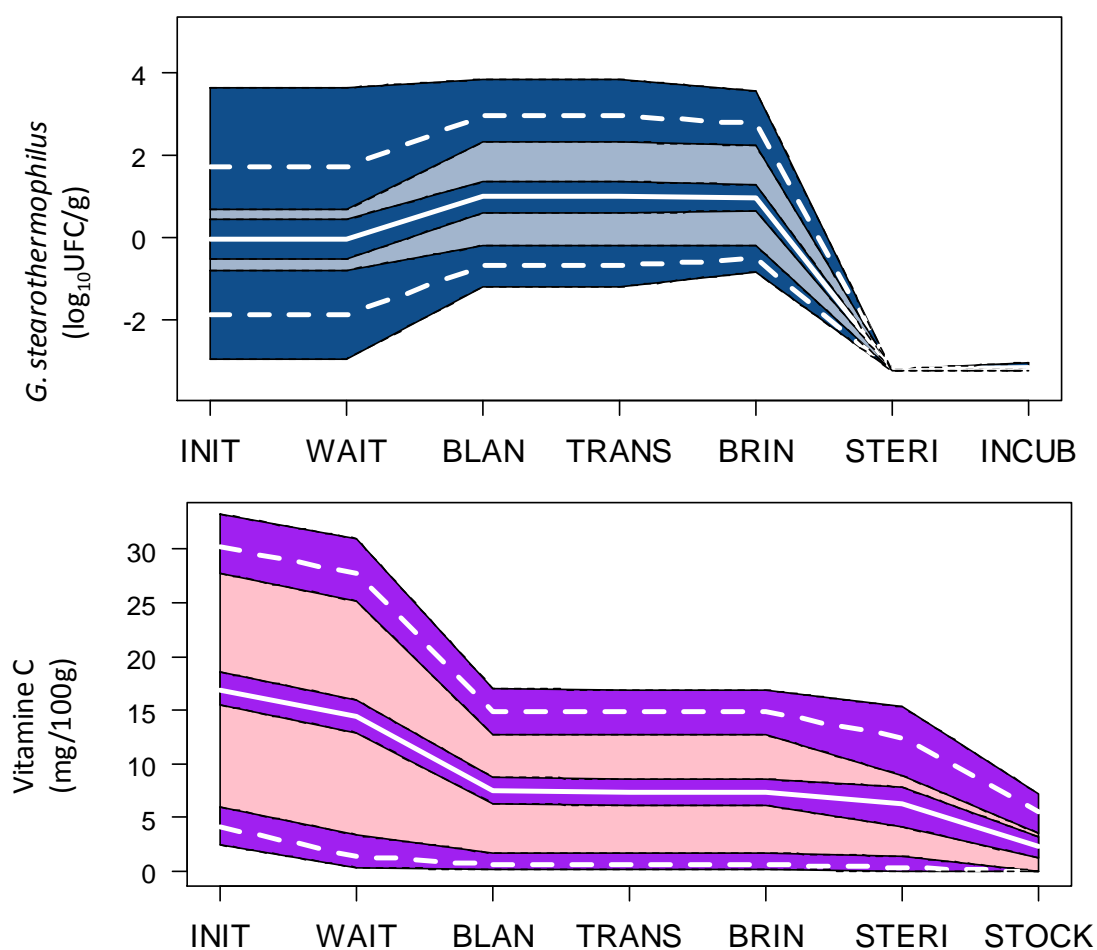


Figure 5.3. Evolution simultanée des concentrations en *G. stearothermophilus* et en vitamine C le long de la chaîne de transformation des haricots verts appertisés et aux étapes d'incubation ou d'entreposage (incertitude et variabilité distinguées). Les traits blancs pleins épais représentent les moyennes des concentrations, et les traits blancs pointillés épais délimitent les intervalles de variabilité à 95%. Les zones en bleu foncé et violet représentent les intervalles d'incertitude à 95% sur la moyenne et sur les limites de l'intervalle de variabilité à 95%. Les étapes sont celles utilisées dans les Chapitres 3 et 4 (INIT : initiale, WAIT : attente, BLAN : Blanchiment, TRANS : transfert, BRIN : jutage, STERI : stérilisation, INCUB : incubation, STOCK : stockage/entreposage). Pour faciliter la

CHAPITRE 5. Analyse risque-bénéfice

lecture du graphe, dans les boîtes contenant strictement moins de 1 spore, la concentration a été arbitrairement fixée à $-2.9 \log_{10}\text{CFU/g}$ aux étapes INIT, WAIT, BLAN et TRANS et à $-3.2 \log_{10}\text{UFC/g}$ aux étapes BRIN et STERI, INCUB, ce qui correspond à 0.5 spore par boîte.

La Figure 5.4 montre les concentrations simultanées en vitamine C et en *G. stearothermophilus* dans les haricots verts immédiatement après la stérilisation (concentrations dans l'ensemble « haricots verts + jus » pour *G. stearothermophilus*, et dans les haricots verts seulement pour la vitamine C). La stérilisation est la dernière étape commune aux deux modèles du procédé de fabrication. Il n'y a pas de corrélation visible entre les deux concentrations, ce qui est confirmé par la non-significativité du coefficient de corrélation de Spearman, estimé à 0.001. Ce résultat était attendu étant donné qu'il n'y a pas de facteur commun aux deux modèles influant fortement à la fois sur les deux concentrations (voir partie 5.2.2). Les concentrations en *G. stearothermophilus* et en vitamine C dans le produit fini sont donc relativement indépendantes.

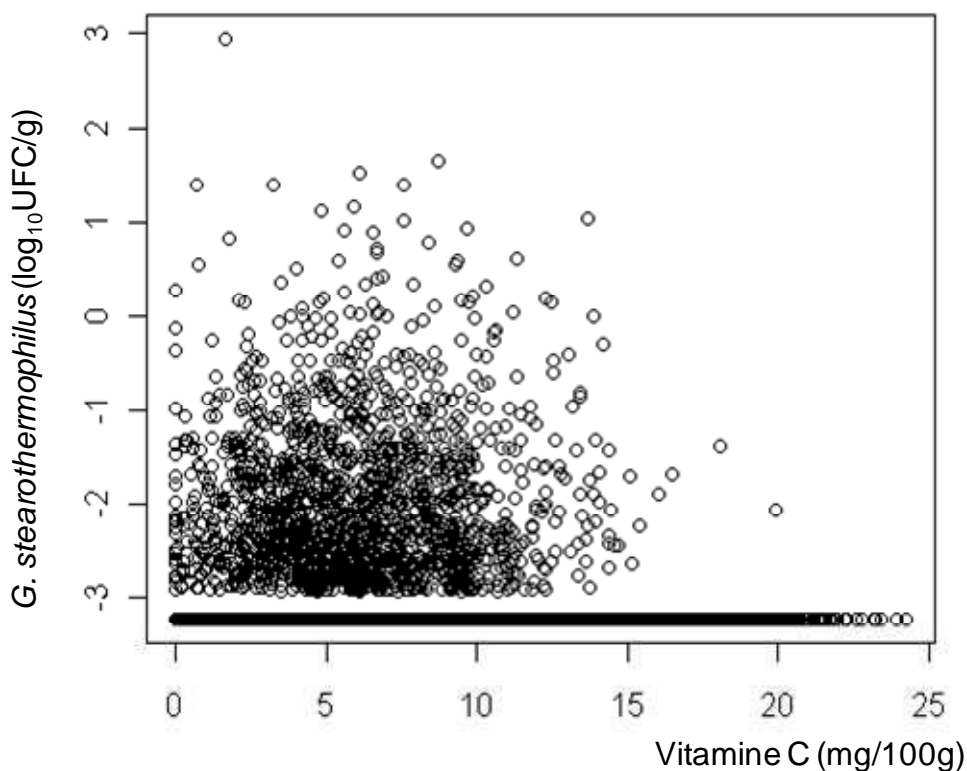


Figure 5.4. Concentrations simultanées en vitamine C et en *G. stearothermophilus* dans des haricots verts appertisés juste après stérilisation. Graphe obtenu sur un échantillon de 300 000 boîtes (incertitude et variabilité combinés). Dans les boîtes contenant strictement moins d'une spore, la concentration a été arbitrairement fixée à $-3.2 \log_{10}\text{UFC/g}$, ce qui correspond à 0.5 spore par boîte, pour faciliter la lecture du graphe.

5.4 Optimisation du compromis risque-bénéfice

5.4.1 Choix des facteurs modifiables

L'optimisation du compromis risque/bénéfice va être facilitée par la relative indépendance des teneurs en vitamine C et en *G. stearothermophilus*, puisque la modification d'un facteur (dans les plages de valeurs étudiées dans le modèle) dans le but d'augmenter le bénéfice ne modifiera globalement pas ou peu le risque, et inversement.

Les facteurs modifiables susceptibles de constituer des leviers d'actions sont les facteurs variables, puisque dans la réalité les facteurs seulement incertains ne peuvent pas être modifiés (leur incertitude peut seulement être réduite). De plus, les facteurs modifiables ne peuvent pas être des paramètres chimiques ou microbiologiques caractérisant la vitamine C ou *G. stearothermophilus*, comme les résistances thermiques ou l'énergie d'activation. Les facteurs modifiables sont donc à rechercher parmi les durées, températures ou caractéristiques physiques et environnementales des procédés (volume de l'espace de tête et pH du produit), et certaines sources de contaminations bactériennes. La contamination bactérienne de haricots verts frais semble difficilement modifiable, mais il est éventuellement possible d'agir sur certaines sources de contamination croisée. Les différentes modifications doivent néanmoins être techniquement réalisables par l'industriel, et doivent permettre de garantir la bonne qualité du produit fabriqué (sécurité alimentaire, cuisson, saveur, etc.).

L'industriel a été interrogé sur la faisabilité de modifier certains paramètres. Les paramètres qui pourraient éventuellement être modifiés sont :

- La durée d'attente des légumes avant passage sur la chaîne (t_{wait}). Elle pourrait éventuellement être réduite en travaillant en flux tendu plutôt qu'en stockant les légumes tout juste récoltés.
- La valeur stérilisatrice (F_0). Elle pourrait être augmentée ou faiblement réduite (valeurs déjà basses).
- La contamination de l'eau de blanchiment, qui pourrait être réduite par exemple par un renouvellement plus fréquent de l'eau.
- La contamination des éventuels débris, qui pourrait être réduite par augmentation des actions de nettoyage.
- Le pH du produit. L'abaissement du pH par ajout d'acides organiques ou l'utilisation de microorganismes de fermentation est parfois utilisé pour assurer la conservation.

Par contre, il semble plus difficile pour l'industriel d'envisager une modification des durées et températures de blanchiment (il est nécessaire que le haricot vert soit attendri par une cuisson légère pour être mis en boîte), des durées et températures de transfert, des durées et températures de l'entreposage, de la contamination du jus ou de la quantité d'air dans l'espace de tête (déjà optimisée).

5.4.2 Tests de scénarios et dépassements de seuils

De nouveaux scénarios ont été testés sur le modèle global risque-bénéfice, afin de rechercher des mesures possibles d'optimisation du compromis risque-bénéfice. A chaque fois 1,5 millions de boîtes de haricots verts appertisés ($N_u = 1000$, $N_v = 1500$) ont été simulées. Les scénarios testés portent sur une variation de la valeur stérilisatrice, une absence de contamination de l'eau de blanchiment, la réduction des durées d'attente ou de blanchiment, la baisse du pH à la stérilisation et du produit fini, et sur des mesures combinant des modifications de plusieurs de ces facteurs en même temps (Tableau 5.1).

L'étude simultanée du risque et du bénéfice a été faite au moyen des probabilités de dépassement de seuils (Van der Voet et al., 2007). Pour chaque scénario, le risque a été mesuré par la probabilité que la concentration en *G. stearothermophilus* après incubation dépasse le seuil de non-stabilité $S_r = 7 \log_{10} \text{UFC/g}$ (c'est-à-dire le taux de non-stabilité). Le bénéfice a été mesuré par la concentration moyenne en vitamine C après entreposage des boîtes, et par la probabilité que la concentration en vitamine C juste après stérilisation dépasse le seuil d'allégation nutritionnelle de la vitamine C, $S_b = 12 \text{ mg/100g}$. La concentration en vitamine C immédiatement après la stérilisation a été choisie comme critère car la concentration en vitamine C après l'entreposage des boîtes était systématiquement sous le seuil d'allégation nutritionnelle (pour tous les scénarios), ce qui limitait l'intérêt de l'étudier. Enfin, une mesure conjointe du risque et du bénéfice est donnée par la probabilité que simultanément dans une boîte de conserve, la concentration en vitamine C dépasse le seuil d'allégation nutritionnelle S_b et que la concentration en *G. stearothermophilus* après incubation soit sous le seuil de non-stabilité S_r , c'est-à-dire que la conserve soit stable à 55°C.

Avec le modèle de référence, la probabilité que la concentration en vitamine C après stérilisation dépasse le seuil est de 3.2% seulement en moyenne (intervalle d'incertitude à 95% assez large de [0.1% ; 13.2%]), le taux de non-stabilité est de 0.5% en moyenne (intervalle d'incertitude à 95% de [0.1% ; 1.3%]), et la probabilité jointe d'avoir un bénéfice supérieur au seuil d'allégation nutritionnelle dans des boîtes stables est de 3.2% en moyenne (intervalle d'incertitude à 95% de [0.1% ; 13.1%]). L'augmentation de la valeur stérilisatrice F_0 diminue le risque de non-stabilité et une forte réduction de F_0 augmente sévèrement le risque de non-stabilité, par contre l'effet de la variation de F_0 a très peu d'impact sur le bénéfice ; le meilleur compromis risque-bénéfice est donc obtenu pour $F_0 = 40 \text{ min}$ avec une probabilité jointe de 3.6% (scénarios 1 à 3). La diminution de la contamination de l'eau de blanchiment réduit le risque de non-stabilité (0.2% versus 0.5% en moyenne) et n'a pas d'impact sur le bénéfice, d'où une probabilité jointe de 3.7% (scénario 4). Une forte diminution du temps d'attente des légumes se traduit sous la forme d'une augmentation du bénéfice (probabilité de dépassement allant jusqu'à 7.1%) et n'a pas d'impact sur le risque (scénarios 5 et 6), d'où un compromis risque-bénéfice plus élevé (probabilité jointe jusqu'à 7.1%). La réduction de la durée du blanchiment à 2 min augmente fortement le bénéfice (probabilité moyenne de 21.2%) et n'impacte pas le risque, d'où une forte probabilité jointe de 21.1% (scénario 7). L'effet d'une baisse du pH à la stérilisation et dans le produit fini entraîne une forte diminution du risque et n'affecte pas le bénéfice (scénario 8). À partir de ces informations, des modifications simultanées de plusieurs paramètres ont été envisagées. Ainsi les scénarios 9 à 11 permettent d'atteindre des probabilités de dépasser le seuil d'allégation nutritionnelle après stérilisation dans des boîtes stables allant de 21.3% à 34.3%. Les meilleurs compromis risque-bénéfice sont ainsi obtenus avec une

CHAPITRE 5. Analyse risque-bénéfice

réduction des durées d'attente et de blanchiment, et soit une augmentation de la valeur stérilisatrice F_0 associée à une baisse de la contamination de l'eau du blancher, soit une baisse du pH à la stérilisation et dans le produit fini.

Tableau 5.1. Scénarios et mesures du risque et du bénéfice correspondant. De gauche à droite les mesures sont la concentration moyenne en vitamine C dans les haricots verts après entreposage, la probabilité que la concentration en vitamine C dans les haricots verts après stérilisation dépasse le seuil d'allégation nutritionnelle $S_b = 12$ mg/100g, la probabilité que la concentration en *G. stearothermophilus* après incubation dépasse le seuil $S_r = 7 \log_{10}$ UFC/g (= taux de non-stabilité), et la probabilité d'être simultanément sous le seuil de risque et au dessus du seuil de bénéfice. Les intervalles représentent l'incertitude sur l'estimation.

	Scénario	C_{VitC}^{STOCK} moyen (mg/100g)	$\Pr[C_{VitC}^{STERI} \geq S_b]$ (%)	$\Pr[C_{Gbs}^{INCUB} \geq S_r]$ (%)	$\Pr[(C_{Gbs}^{INCUB} < S_r)$ et $(C_{VitC}^{STERI} \geq S_b)]$ %
0	Modèle de référence	2.3 [1.2; 3.2]	3.2 [0.1; 13.2]	0.5 [0.1; 1.3]	3.2 [0.1; 13.1]
1	$F_0 = 40$ min	2.3 [1.3; 3.2]	3.6 [0.1; 13.0]	0.1 [0.0; 0.5]	3.6 [0.1; 13.0]
2	$F_0 = 20$ min	2.3 [1.2; 3.2]	3.5 [0.1; 13.3]	2.0 [0.8; 4.2]	3.5 [0.1; 13.1]
3	$F_0 = 10$ min	2.3 [1.3; 3.2]	3.3 [0.1; 12.8]	13.4 [7.7; 19.6]	2.9 [0.1; 11.2]
4	$N_{BLANCC} = 0$ CFU/g	2.3 [1.2; 3.2]	3.7 [0.1; 13.5]	0.2 [0.0; 0.8]	3.7 [0.1; 13.5]
5	$t_{wait} = 140$ min	2.8 [1.7; 3.8]	6.3 [0.3; 19.0]	0.5 [0.1; 1.3]	6.3 [0.3; 18.9]
6	$t_{wait} = 45$ min	2.9 [1.8; 3.8]	7.1 [0.3; 20.6]	0.5 [0.1; 1.3]	7.1 [0.3; 20.3]
7	$t_{blan} = 2$ min	3.6 [2.3; 4.7]	21.2 [3.4; 38.9]	0.5 [0.1; 1.3]	21.1 [3.4; 38.7]
8	$pH_{steri} = 5.45$ et $pH_{FP} = 5.15$	2.3 [1.2; 3.1]	3.6 [0.1; 12.5]	0.1 [0.0; 0.7]	3.6 [0.1; 12.5]
9	$F_0 = 40$ min et $t_{blan} = 2$ min	3.5 [2.3; 4.6]	21.4 [4.2; 37.4]	0.1 [0.0; 0.5]	21.3 [4.2; 37.3]
10	$F_0 = 40$ min et $t_{blan} = 2$ min et $t_{wait} = 45$ min et $N_{BLANCC} = 0$ CFU/g	4.4 [3.1; 5.5]	34.4 [10.1; 51.9]	0.1 [0.0; 0.3]	34.3 [10.1; 51.8]
11	$pH_{steri} = 5.45$ et $pH_{FP} = 5.15$ et $t_{blan} = 2$ min et $t_{wait} = 45$ min	4.4 [3.0; 5.6]	33.5 [7.8; 52.5]	0.1 [0.0; 0.7]	33.5 [7.8; 52.5]

5.5 Conclusion

Cette analyse risque-bénéfice propose des solutions d'optimisation du compromis risque-bénéfice dans les haricots verts appertisés. Un résultat est que dans le modèle global, il n'y a pas de facteur influençant significativement à la fois le risque et le bénéfice (dans la plage de variation étudiée des facteurs). Des mesures peuvent donc être prises pour diminuer le risque (exemple : augmenter la valeur stérilisatrice ou diminuer le pH du produit) et d'autres mesures peuvent être prises pour augmenter le bénéfice (exemple : réduire les durées d'attente ou de blanchiment). En combinant plusieurs mesures pertinentes, la probabilité d'avoir simultanément une concentration en vitamine C dans les haricots verts après stérilisation supérieure au seuil d'allégation nutritionnelle de la vitamine C (12 mg/100g) et une concentration en *G. stearothermophilus* après incubation sous le seuil de non-stabilité passe de 3.2% à 35.6% en moyenne (voir partie 5.4.2).

Des pistes d'optimisation sont donc possibles. Il serait néanmoins préférable de travailler avec des modèles complètement validés, car les résultats de l'analyse risque-bénéfice dépendent étroitement de la qualité des modèles. Il serait aussi intéressant de rajouter à cette analyse risque-bénéfice une étude plus approfondie de la faisabilité et du coût industriel de telles mesures. Des ratios coût-utilité sont par exemple déterminés pour différentes interventions de réduction du risque dans Havelaar et al. (2007). Leur construction demande toutefois une très grande quantité d'information, difficilement disponible.

Chapitre 6

Confrontation d'un modèle d'analyse du risque microbiologique à des données de contamination dans des produits finis

Ce chapitre présente un travail d'inférence bayésienne réalisée sur un modèle global d'évaluation du risque dû à *Bacillus cereus* dans une chaîne de production de purée de courgettes, afin d'améliorer les estimations des paramètres de ce modèle à partir de données de contamination sur produits finis. Ce travail a fait l'objet d'un article reproduit dans la partie 6.2, accepté en août 2012 dans la revue *Risk Analysis*.

6.1 Considérer un modèle d'évaluation du risque comme un réseau bayésien

Les chapitres 3 et 4 ont présenté la construction de modèles d'évaluation du risque et du bénéfice, à partir de nombreuses données qui chacune participait à la construction d'un sous-module. Une fois les modèles construits, les méthodes de simulation de Monte Carlo ont été utilisées pour propager l'incertitude et la variabilité dans les modèles et analyser leurs résultats. La simulation de Monte Carlo fait parcourir un modèle depuis les facteurs d'entrée jusqu'à la sortie du modèle, dans un sens « descendant », c'est-à-dire des variables d'entrée vers les variables de sortie. Dans une autre optique, un modèle d'évaluation des risques (ou du bénéfice) peut être considéré comme un réseau bayésien (Albert et al., 2008; Smid et al., 2010). Si des données relatives à la sortie ou une variable intermédiaire du modèle sont disponibles, une façon différente d'intégrer de l'information provenant de données est d'utiliser une méthode d'inférence bayésienne pour « confronter » le modèle aux données. La méthode d'inférence va permettre de faire « remonter » l'information issue des données dans le modèle, notamment jusqu'aux facteurs d'entrée. Ici, l'information issue des données met à jour (plus ou moins fortement) l'ensemble des variables du modèle. Pour relier les paramètres du modèle aux données disponibles, il est nécessaire d'« augmenter » le modèle, c'est-

CHAPITRE 6. Confrontation d'un modèle à des données

à-dire de rajouter une ou plusieurs étapes dans le modèle, en particulier la vraisemblance des données. Cette méthode complémentaire aux précédentes a été mise en place dans ce chapitre.

La simulation de Monte Carlo à 2 dimensions et l'inférence bayésienne se combinent bien ensemble car elles permettent toutes deux de prendre en compte les sources d'incertitude et/ou de variabilité dans les modèles. L'inférence bayésienne est bien adaptée aux modèles hiérarchiques qui modélisent séparément l'incertitude et la variabilité, et qui peuvent être associés à une simulation de Monte Carlo 2D (cf partie 1.2.2). En inférence bayésienne, les lois a priori expriment en effet de l'incertitude sur la vraie valeur du paramètre, et le but de l'inférence bayésienne est d'obtenir des lois a posteriori sur les paramètres où cette incertitude est diminuée. Les lois a priori doivent donc bien traduire de l'incertitude et pas de la variabilité, qui elle n'est pas réductible. Typiquement, la méthode d'inférence bayésienne peut servir à estimer les paramètres en modélisant la variabilité et l'incertitude d'un phénomène, puis les estimations obtenues pourront être utilisées pour réaliser une simulation de Monte Carlo 2D, comme cela a été fait dans les chapitres 2 et 3. Dans ce chapitre, la méthode d'inférence bayésienne est utilisée pour mettre à jour les estimations de tous les paramètres du modèle complet d'évaluation des risques à partir de données de sortie de modèle.

Ce chapitre porte sur une analyse du risque due à la bactérie *B. cereus* dans de la purée de courgettes. Un modèle d'analyse des risques pour ce couple de produit et de pathogène a été construit par Afchain et al. (2008) ; il s'agissait d'un modèle « d'ordre 1 », dont les distributions des facteurs d'entrée traduisaient alternativement de l'incertitude, de la variabilité ou un mélange des deux (DAG représenté en Figure 6.1). Ce modèle avait été étudié par simulation de Monte Carlo. Le but de l'étude présentée ici était de confronter ce modèle avec des données de contaminations dans des paquets de purée de courgettes, par le biais d'une inférence bayésienne. Il a d'abord été nécessaire de transformer le modèle initial d'analyse du risque en un modèle « d'ordre 2 » modélisant séparément l'incertitude et la variabilité, afin de pouvoir notamment prendre en compte les différentes sources de variabilité entre les données. Les données étaient en effet constituées de mesures de contamination en *B. cereus* dans 64 paquets de purée de courgettes provenant de différents batchs de légumes produits sur une période de production étendue (2 ans). De nombreux facteurs, comme la contamination des légumes ou la durée des étapes du procédé, étaient ainsi susceptibles de varier lors de la production des différents produits (variabilité inter-batch). La transformation d'un modèle d'ordre 1 en modèle d'ordre 2 a été délicate et a nécessité de nombreuses hypothèses (sur la nature des distributions, sur la façon de modéliser l'incertitude sur certains paramètres des distributions de variabilité), parfois discutables. Ce travail s'est fait en étroite collaboration avec des experts en microbiologie, pour une meilleure interprétation des variables du modèle. Une fois le modèle d'ordre 2 construit, l'inférence bayésienne a été réalisée à l'aide du logiciel Jags (Plummer, 2007). La complexité du modèle, tenant compte en particulier des 6 groupes génétiques de *B. cereus*, d'un partitionnement de batchs de légumes en petits paquets, de phénomènes non-linéaires, d'interactions et de censures, explique en partie la longueur des calculs effectués afin d'obtenir une bonne convergence des chaînes de Markov et une auto-corrélation limitée (1,4 millions d'itérations en tout, ce qui a pris environ 6,5 jours de calculs). Enfin, une fois l'inférence bayésienne réalisée, des simulations de Monte Carlo 2D ont été effectuées à partir des lois a priori puis à partir des lois a posteriori des paramètres du modèle. Les deux résultats du modèle d'évaluation du risque ont été comparés, afin d'étudier les conséquences sur le risque des informations apportées par les données.

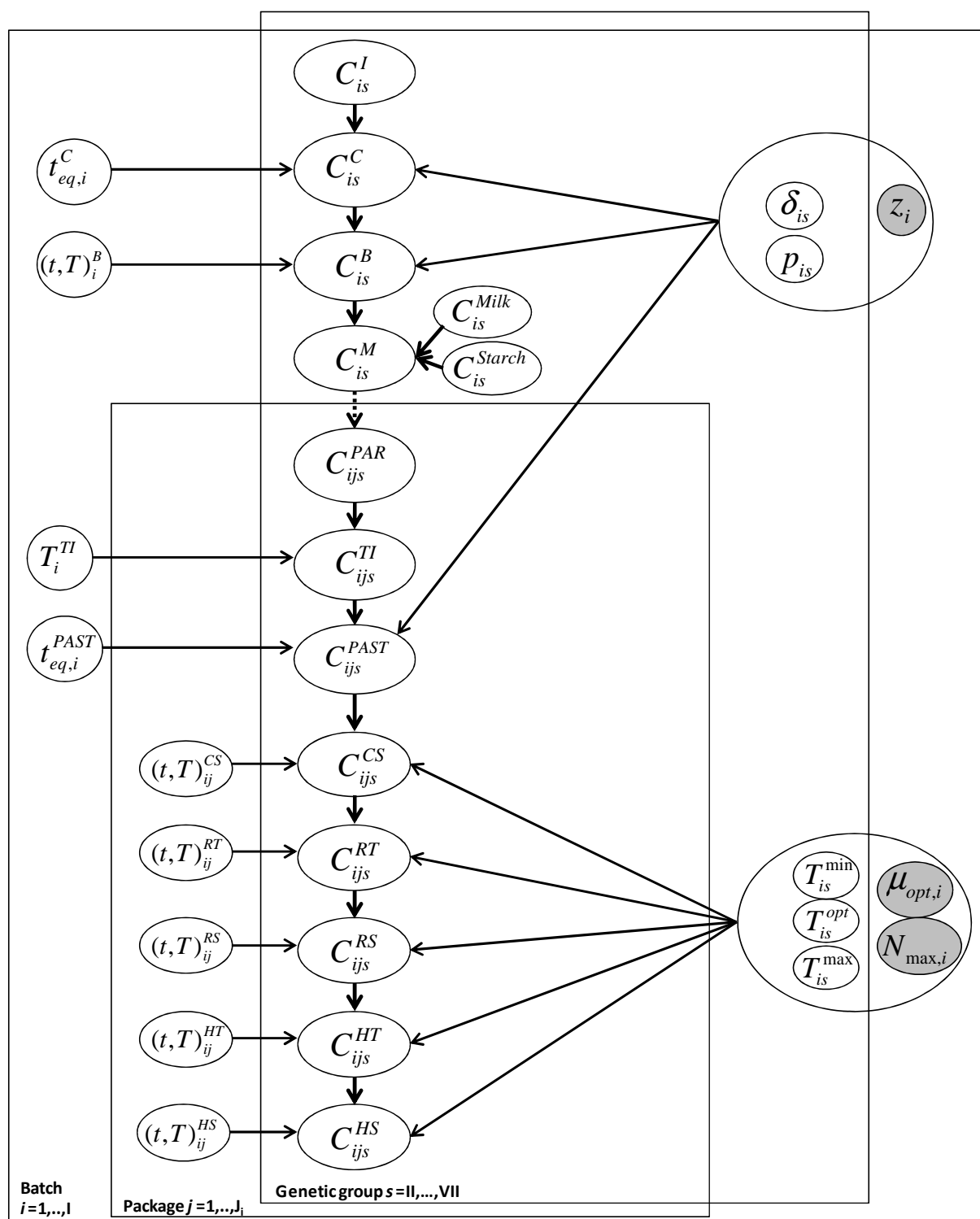


Figure 6.1. Le DAG du modèle d'analyse des risques d'ordre 1 (Afchain et al., 2008). Voir le Tableau 6.3 pour la description des variables. Dans l'article d'Afchain et al. (2008), les distributions étaient alternativement désignées comme distributions de « variabilité » ou « d'incertitude » selon les variables d'entrées, mais la méthode de simulation était identique pour toutes les distributions.

6.2 Article: Inferring an augmented Bayesian network to confront a complex quantitative microbial risk assessment model with durability studies: Application to *Bacillus cereus* on a courgette purée production chain

Cette partie est la reproduction d'un article écrit en collaboration entre Clémence Rigaux^a, Sophie Ancelet^a, Frédéric Carlin^{b,c}, Christophe Nguyen-thé^{b,c} et Isabelle Albert^a. Il a été accepté dans la revue Risk Analysis en août 2012.

^aINRA, UR 1204, Met@risk, Food Risk Analysis Methodologies, F-75005 Paris, France

^bINRA, UMR 408, Sécurité et Qualité des Produits d'Origine Végétale, F-84000 Avignon, France

^cUniversité d'Avignon et des Pays de Vaucluse, UMR 408, Sécurité et Qualité des Produits d'Origine Végétale, F-84000 Avignon, France

Abstract

The Monte-Carlo (MC) simulation approach is traditionally used in food safety risk assessment to study Quantitative Microbial Risk Assessment (QMRA) models. When experimental data are available, performing Bayesian inference is a good alternative approach that allows backward calculation in a stochastic QMRA model to update the experts' knowledge about the microbial dynamics of a given food-born pathogen. In this paper, we propose a complex example where Bayesian inference is applied to a high-dimensional second-order QMRA model. The case study is a farm-to-fork QMRA model considering genetic diversity of *Bacillus cereus* in a cooked, pasteurized and chilled courgette purée (Afchain et al., 2008). Experimental data are *Bacillus cereus* concentrations measured in packages of courgette purées stored at different time-temperature profiles after pasteurization. To perform a Bayesian inference, we first built an augmented Bayesian network by linking a second-order QMRA model to the available contamination data. We then ran a Markov Chain Monte Carlo (MCMC) algorithm to update all the unknown concentrations and unknown quantities of the augmented model. About 25% of the prior beliefs are strongly updated, leading to a reduction in uncertainty. Some updates interestingly question the QMRA model.

Keywords

Quantitative Microbial Risk Assessment (QMRA), Bayesian inference, Second-order models, *Bacillus cereus*, Markov Chain Monte Carlo (MCMC)

6.2.1 Introduction

Quantitative Microbial Risk Assessment (QMRA) models describe the dynamics of food-borne pathogens in a food chain, in order to assess microbial risk in a “farm-to-fork” approach. Following the Modular Process Risk Modelling (MPRM) approach (Nauta, 2002), these models are typically based on a logic and sequential combination of conditionally independent modules, each module modelling one stage of the food chain. In food safety risk assessment, a stochastic QMRA model consists of a set of variables defining, for instance, biological and/or process parameters, bacteria concentrations and a set of mathematical relations defining their dependencies (e.g., basic microbial processes impacting microbial contamination). As defined, such a model may be considered as a Bayesian Network (BN) (Pearl, 1988; Jensen, 2001), i.e. a framework that defines the conditional dependencies between variables (also called nodes in this context) both qualitatively through a Directed Acyclic Graph (DAG) and quantitatively with local transition probabilities or deterministic equations. The conditional dependencies may be deterministic or stochastic, and are represented by arrows (also called directed edges) on the DAG (Pearl, 1988; Jensen, 2001; Albert et al., 2008). Formally, if there is a directed edge from variable A to variable B in a given DAG, we say that B is a child of A and A is parent of B. In the following, the inputs of a given BN denote all the variables without parents, the outputs are all the variables without children and the intermediate variables denote all the remaining variables included in the model. The parameters are the unknown inputs i.e., the ones which are not fixed.

QMRA models are often studied using the Monte Carlo (MC) simulation approach, which consists of drawing values for all the inputs of the model from their assigned probability distributions and then generating samples of all the intermediate variables and outputs by propagating such random initial values through the system. The assigned distributions, which reflect some intrinsic variability and/or epistemic uncertainty (Nauta, 2000; Vose, 2000) of the inputs, are marginally defined on the basis of external knowledge (e.g., experimental or historical data, expert opinions). Intrinsic variability, reduced to “variability” hereafter, denotes the natural heterogeneity of a random quantity irreducible by additional measurements whereas epistemic uncertainty, reduced to “uncertainty” hereafter, denotes some lack of knowledge about a given variable, reducible by additional measurements. At the end, MC estimates may be computed, from the simulated samples, for summary statistics of interest. Additionally, the MC simulation approach can be used as a powerful tool to perform sensitivity analyses. Nevertheless, one major weakness of the MC simulation approach is that it is unidirectional by nature, meaning that the dependency relationships between the variables cannot be inverted. As a consequence, in such an approach, strength cannot be borrowed from raw data obtained “downstream” in the food chain, i.e., data that could bring information about parameters, intermediate variables or outputs, the distributions of which are already induced by the MC simulation algorithm.

An alternative approach consists in applying Bayesian inference to a QMRA model (Albert et al., 2008). Bayes' rule (Robert, 2007) allows the updating of prior beliefs on all the unknown variables (i.e., parameters, intermediate variables and outputs) of the model given new raw data. In such “inverse problems,” the initial BN representing the QMRA model must be augmented, i.e. extended by the specification of the conditional distribution, typically known as the “observational model” or “likelihood model,” of the data given some variables of the QMRA model plus, if necessary, some

CHAPITRE 6. Confrontation d'un modèle à des données

additional unobserved variables and the conditional dependencies between them. In the case of a complex BN, Bayesian inference can then be implemented through Markov Chain Monte Carlo (MCMC) algorithms (Robert et al., 2004). Once convergence is reached, such iterative algorithms generate a sample from the joint posterior distribution of the model's unknown parameters. This distribution reflects the remaining uncertainty about the unknown variables after borrowing strength across the available data.

Until now, applying Bayesian inference to an augmented QMRA model has been less common than MC simulations. Smid *et al.* (2010) have recently pointed out the strengths of Bayesian inference compared to MC simulations but underlined the over-simplified structure of the augmented QMRA models so far proposed in the literature. Hence, some authors have proposed some Bayesian inference on QMRA models using continuous variables but limited to only a few modules of the QMRA model (Pouillot et al., 2003; Delignette-Muller et al., 2006; Spor et al., 2010) or to simplified "farm-to-fork" QMRA models (Albert et al., 2008), or on QMRA models using discretized continuous variables (Barker et al., 2005; Smid et al., 2010; Barker and Gómez-Tomé, 2011; Smid et al., 2011; Smid et al., 2012). Even if these latter models do not necessarily perform worse (or better) than models based on continuous variables when discretization is thin enough, discretizing continuous variables generally may lead to a less realistic model. Indeed, the real nature of the variables is then lost: such a model only supposes the exploration of a finite and a priori fixed number of values among the possible ones. In particular the distributions tails are less explored which is problematic when they are critical (Smid et al., 2010).

In this paper, we propose a complex example where Bayesian inference is applied to a high-dimensional augmented BN implying many continuous variables and using experimental data indirectly linked to some given variables of the initial "farm-to-fork" QMRA model. Such augmented BN is based on a second-order QMRA model (Cullen et al., 1999; Frey and Burmaster, 1999; Vose, 2000; Pouillot et al., 2007), making it possible to account for the whole variability of the data (durability studies elapsed over 2 years and coming from several batch units). Our case study clearly shows that such an approach can improve our prior beliefs about the dynamics of food-borne pathogens in a food chain and that it can be a powerful tool for questioning a QMRA model in light of posterior distributions.

We focus on a published first-order QMRA model previously used through a standard MC simulation approach (Afchain et al., 2008). This QMRA model formalizes the dynamics of a heterogeneous population of *Bacillus cereus* in a cooked, pasteurized and chilled courgette purée as an example of REfrigerated Processed Food of Extended Durability (REPFED). The model represents a complex "farm-to-fork" process with inactivation, germination and growth of *B. cereus* at each stage of the food pathway, and takes into account the genetic diversity of *B. cereus*. First, we build a second-order QMRA model (Cullen et al., 1999; Frey et al., 1999; Vose, 2000; Pouillot et al., 2007) as in a well-established approach in food safety risk assessment to separate uncertainty and variability when modelling unknown quantities based on the initial first-order model. Then, following Albert *et al.* (2008), we propose a possible way of augmenting this BN by using raw experimental data from durability studies not previously used to learn about the multiple variables implied in the courgette purée food pathway. Finally, we run a Markov Chain Monte Carlo (MCMC) algorithm to update the prior beliefs of the whole unknown quantities implied in the proposed augmented BN.

6.2.2 Materials and methods

6.2.2.a The QMRA model

Afchain *et al.* (2008) have proposed a first-order stochastic QMRA model that describes the transmission of a heterogeneous population of *B. cereus* along a courgette purée's food pathway, from the initial contamination in raw courgettes to home consumption. Figure 6.2 and Table 6.1 give an overview of such a food pathway based on the combination of twelve conditionally independent modules describing each stage of the processing and chill chain. From one module to another, the implied variables, which mainly formalize the concentration of the food-borne pathogen, are linked either by deterministic relations modelling *B. cereus* heat inactivation, germination or growth, or a stochastic model describing the stage of partitioning a batch of courgettes purée into packages. Such models are controlled by various inputs (e.g., time-temperature profiles, biological parameters such as cardinal growth temperatures).

The *B. cereus* population may be considered as a mixture of six genetic groups, ranging from mesophilic and highly heat resistant to psychrotrophic and moderately heat resistant (Guinebretiere *et al.*, 2008). Note that the proposed QMRA model accounts for this genetic diversity by predicting the *B. cereus* concentration for each genetic group at each stage of the food pathway.

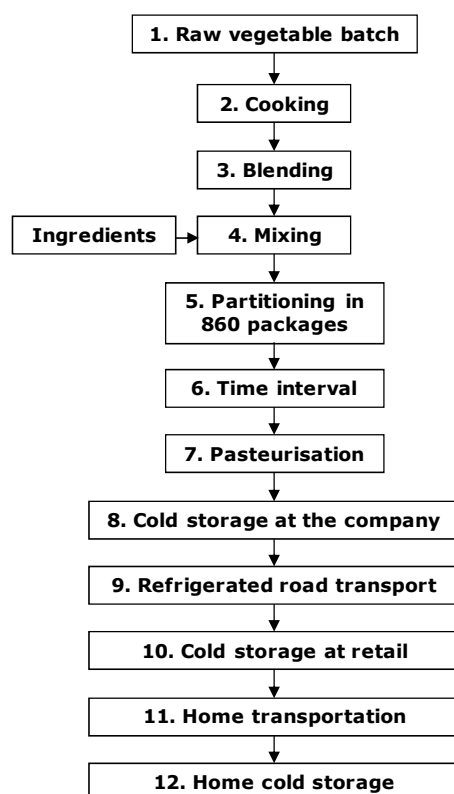


Figure 6.2. Overview of the food pathway (processing and chill chain) of cooked, pasteurized and chilled courgette purées.

Table 6.1. Description of the deterministic and stochastic models associated to each stage of the food pathway of the QMRA model, depending on the involved mechanisms.

Impacting mechanisms	Step in the food pathway	Deterministic model or probability distribution
Contamination	1, 4	BetaPert distribution
Inactivation	2, 3, 7	Primary model: Weibull model (Peleg et Cole, 1998) Secondary model: Bigelow model (Bigelow, 1921)
Partitioning	5	Multinomial distribution
Germination	6, 8	Empirical model (see Afchain <i>et al.</i> 2008)
Growth	8, 9, 10, 11, 12	Primary model: Logistic model with delay (Rosso <i>et al.</i> 1995) Secondary model: cardinal temperature model with inflection point (Rosso <i>et al.</i> 1993) Competition model (see Afchain <i>et al.</i> 2008)

As in a well-established framework in food safety risk assessment to separate uncertainty and variability when modelling unknown quantities (Cullen *et al.*, 1999; Frey *et al.*, 1999; Vose, 2000; Pouillot *et al.*, 2007), we have built a second-order version of the initial QMRA model proposed by Afchain *et al.* (2008). Figure 6.3 provides the simplified DAG associated with the second-order QMRA model (represented by the white or light grey nodes only). The variables description is given in Table 6.2. Let I be the number of simulated batches and let J_i be the number of packages coming from batch i ($i=1, \dots, I$). The total number of simulated packages is $P = \sum_{i=1}^I J_i$. The second-order QMRA model holds 53 parameters (light grey ellipses) and $6P$ outputs globally (i.e., the concentrations of the 6 genetic groups in P packages after home cold storage).

To construct the second-order model, we first decided whether or not each input of the initial first-order QMRA model could be affected by batch (or package) specific variability. Such variability may be due, for instance, to heterogeneous environmental conditions affecting the batches (or packages) during the cooking and chill chain. Then, we decided whether the variability distributions should be considered as known (i.e., defined by fixed inputs) or uncertain (i.e., defined by uncertain inputs). These two main modelling choices were based on microbiological considerations or process knowledge, and made to propose a second-order model as parsimonious as possible.

Therefore, the growth parameters N_{max} and μ_{opt} and the resistance parameter z were not assumed to be affected by batch (or package)-specific variability. Indeed, previous works (Carlin *et al.*, 2013) indicate that growth rates of *B. cereus* strains in optimal conditions (μ_{opt}) are very similar. It seems also reasonable according to the few available data to assume a constant N_{max} in a given growth media or food substrate. Finally, in the absence of microbiological counter-arguments and due to a limited number of z data available for *B. cereus*, we opted for the parsimonious hypothesis that z was not variable among batches and packages. In contrast, all the other inputs of the first-order model were supposed to be affected by batch (or package)-specific variability. It was the case for all the process durations and temperatures, based on process knowledge. A variability of

CHAPITRE 6. Confrontation d'un modèle à des données

minimum and maximal temperatures allowing growth among several strains within each genetic group was also observed (Guinebretiere et al., 2008). Finally, as shown by results of spore heat resistance experimental measures (Carlin et al., 2006), the resistance parameters δ_{is} and p_{is} differed among strains from the same genetic group.

Following Pouillot *et al.* (2007), the probability distributions describing a food item-specific variability were assumed to be perfectly known for all the duration and temperature variables, except for the equivalent heating times (calculated from durations, temperatures and z (Afchain *et al.*, 2008)). These modelling choices were made to be parsimonious and given that the industrial had a rather good knowledge of the process (except for the equivalent heating times). In contrast, the probability distributions describing the inter-batch variability for the *B. cereus* concentrations (C_{is}^I , C_{is}^{Milk} , C_{is}^{Starch}), the resistance parameters δ_{is} and p_{is} , the growth cardinal temperatures T_{is}^{\min} , T_{is}^{opt} , T_{is}^{\max} and the equivalent cooking and pasteurization times $t_{eq,i}^C$ and $t_{eq,i}^{PAS}$ were assumed to be defined by some uncertain inputs. Indeed, until now, these distributions have only been fitted from a conversion model (Afchain et al., 2008) using an uncertain z -value (for $t_{eq,i}^C$ and $t_{eq,i}^{PAS}$), or from a limited number of data which may be not representative of all the sources of variability (e.g: only two strains per genetic groups for the cardinal temperatures T_{is}^{\min} , T_{is}^{opt} , T_{is}^{\max}).

For this last type of variables, uncertainty and variability were modelled separately by considering, as proposed in Pouillot *et al.* (2003), the uncertain parameters of such a parametric distribution as random variables. An uncertainty distribution was then assigned to these random variables. Table 6.3 summarizes the uncertainty and variability distributions separately defined for these variables of the initial first-order model of Afchain *et al.* (2008). These distributions were mainly chosen to be biologically interpretable, to keep the marginal distributions of the main variables relatively close to their distributions in the first-order model, and to satisfy some sound assumptions on the dispersion of uncertainty around the marginal distributions in the second-order model. These modelling assumptions were also motivated by parsimony and prior beliefs. Thus, for instance, we assumed that the means of the normal and exponential distributions and the most likely values of the BetaPert distributions were uncertain, and that some parameters related to dispersion were fixed (standard deviation for the normal distributions and minimum and maximum values of the support of the BetaPert distributions). These choices were not equivalent for the three types of distributions. For instance these choices led to a fixed support of the marginal distribution in the BetaPert case but to an uncertain support of the marginal distribution in the exponential and normal cases. But biological interpretations motivated these choices. For instance, for the genetic groups that were not detected in the food samples, the maximum value of the support of the BetaPert distributions was fixed to the limit of detection by enrichment, and the minimum one was fixed assuming the support size to be equal for all the genetic groups (one range rather well known in one group). Moreover, it can be shown that the decomposition we have proposed for the normal distribution (see Table 6.3) allows the marginal distribution on X in the second-order model to be equal to the distribution on X in the first-order model, and allows an easy interpretation of the part of total randomness on X coming from uncertainty (variance parameter s_1^2) or coming from variability (variance parameter s_2^2). Without further information, we simply assumed that $s_1^2 = s_2^2 = s^2 / 2$. The uncertainty distributions chosen for the most likely values of the BetaPert distributions

CHAPITRE 6. Confrontation d'un modèle à des données

and for the mean of the exponential distributions were chosen to make a trade-off between the degree of uncertainty and a small distance between the marginal distribution on X in the second-order model (integrating the uncertainty) and the distribution on X in the first-order model. Such a small distance was checked graphically by comparing the plots of these two distributions, and the spaghetti plots of the possible variability distributions corresponding to different uncertain values (results not shown).

The probability distributions assigned to the other variables ($z, N_{max}, \mu_{opt}, t_i^X, t_{ij}^X, T_i^X, T_{ij}^X, X=B, TI, CS, RT, RS, HT, HS, i=1,..,I, j=1,..,J_i$), remained unchanged as described in Afchain *et al.* (2008). They were assumed to be uncertainty distributions for z, N_{max} and μ_{opt} , and to be inter-batch (or package) variability distributions for the durations and temperatures $t_i^X, t_{ij}^X, T_i^X, T_{ij}^X$. Finally, all the parameters (light grey ellipses) and the batch or package-specific durations and temperatures (i.e., $t_i^X, t_{ij}^X, T_i^X, T_{ij}^X$ for $X=B, TI, CS, RT, RS, HT, HS, i=1,..,I, j=1,..,J_i$) of the second-order model presented in Figure 6.3 were supposed to be independent.

Table 6.2. Uncertainty and variability distributions redefining the initial modelling of some biological variables of the courgette purée' food pathway. See Table 6.3 for the definition of the variables. $N(a,b)$ denotes the normal distribution of mean a and standard deviation b , eventually truncated on the interval $[\cdot]$. $BP(a, b, c)$ denotes the BetaPert distribution, where a represents the minimal value, b the most likely value, and c the maximal value of the Pert distribution. $Exp(1/a)$ denotes the exponential distribution of mean a . Bold letters $\mathbf{X}, \boldsymbol{\mu}, \mathbf{o}$ represent random variables, whereas normal lower case letters m, s, s_1, s_2, n, ml, x represent different fixed values.

Variable X	Distribution in the first order model ^{(1)a}	Distributions in the second order model	
		Variability distribution	Uncertainty distribution
$T_{is}^{\min}, T_{is}^{opt}, T_{is}^{\max}$	$\mathbf{X} \sim N(m, s)$	$\mathbf{X} \sim N(\boldsymbol{\mu}, s_2)$ $s_2^2 = s^2/2$	$\boldsymbol{\mu} \sim N(m, s_1)$ $s_1^2 = s^2/2$
$C_{is}^I, C_{is}^{Milk}, C_{is}^{Starch}$ δ_{is}, p_{is}	$\mathbf{X} \sim BP(n, ml, x)$	$\mathbf{X} \sim BP(n, \mathbf{o}, x)$	$\mathbf{o} \sim BP((n+ml)/2, ml, (ml+x)/2)$
$t_{eq,i}^C$	$\mathbf{X} \sim 20 + Exp(1/m)$	$\mathbf{X} \sim 20 + Exp(1/\boldsymbol{\mu})$	$\boldsymbol{\mu} \sim N(m, s)$ on $[0; 1000]$
$t_{eq,i}^{PAS}$	$\mathbf{X} \sim Exp(1/m)$	$\mathbf{X} \sim Exp(1/\boldsymbol{\mu})$	$s = m/4$

^aCompared to the probability distributions described in Afchain *et al.* (Afchain et al., 2008), two main amendments to note are, first, the marginal distributions of the cardinal temperatures $T_{is}^{\min}, T_{is}^{opt}, T_{is}^{\max}$ used in the microbial growth model which have been updated from new experimental data (Carlin et al., 2013), and, secondly, the equivalent cooking time $t_{eq,i}^C$ that has been assumed to be a translated exponential distribution of mean 145min, with a lower support boundary fixed at 20 min instead of 0 min (Afchain et al., 2008). Such a distribution better fits the industrial data.

CHAPITRE 6. Confrontation d'un modèle à des données

Table 6.3. Description of the variables involved in the augmented second-order QMRA model.

Type	Class ^a	Nodes	Unit	Description
<i>B.cereus</i> concentrations	iv	C_{is}^X (resp. C_{ijs}^X)	log ₁₀ CFU/g	Concentration of genetic group <i>s</i> in <i>B. cereus</i> in a batch <i>i</i> (resp. in package <i>j</i> coming from batch <i>i</i>) at step <i>X</i> ^b of the food pathway
	par	O_s^I (resp. O_s^{Milk} , O_s^{Starch})	log ₁₀ CFU/g	Most likely <i>B.cereus</i> concentration in genetic group <i>s</i> in fresh courgettes (resp. in milk, starch)
	fn	n, x^I (resp. n, x^{Milk} , n, x^{Starch})	log ₁₀ CFU/g	Minimum and maximum <i>B.cereus</i> concentration in genetic group <i>s</i> in fresh courgettes (resp. in milk, starch)
Durations and temperatures of the process	iv	t_i^X (resp. t_{ij}^X)	min	Duration of step <i>X</i> in batch <i>i</i> (resp. in package <i>j</i> coming from batch <i>i</i>)
	iv	T_i^X (resp. T_{ij}^X)	°C	Temperature of step <i>X</i> in batch <i>i</i> (resp. in package <i>j</i> coming from batch <i>i</i>)
	fn	D^X	various	Parameters of the variability distributions of t_i^X , t_{ij}^X , T_i^X and T_{ij}^X at step <i>X</i> (see Afchain <i>et al.</i> ⁽¹⁾)
	iv	$t_{eq,i}^X$	min	Equivalent time at the target temperature of 90°C of step <i>X</i> in batch <i>i</i> (calculated from durations, temperatures and z ⁽¹⁾)
	par	$\mu_{t_{eq}}^X$	min	Mean equivalent time at the target temperature of 90°C at step <i>X</i>
Microbiological resistance variables	iv	δ_{is}	min	Time to the first decimal reduction time of the spore population for genetic groups <i>s</i> in batch <i>i</i> under an isothermal regime at T=90°C
	par	O_s^δ	min	Most likely value of the time to the first decimal reduction time of the spore population for genetic groups <i>s</i> under an isothermal regime at T=90°C
	fn	n, x^δ	min	Minimum and maximum values of the time to the first decimal reduction time of the spore population for genetic groups <i>s</i> under an isothermal regime at T=90°C
	par	z	°C	Increase in temperature resulting in a 10-fold reduction in the time to first decimal reduction
	iv	P_{is}	-	Shape parameter of the inactivation Weibull model, for genetic group <i>s</i> in batch <i>i</i>
	par	O_s^P	-	Most likely value of the shape parameter of the Weibull model for genetic group <i>s</i>
	fn	n, x^P	-	Minimum and maximum values of the shape parameter of the Weibull model for genetic group <i>s</i>
Microbiological growth variables	iv	T_{is}^{\min} (resp. T_{is}^{opt} , T_{is}^{\max})	°C	Minimal (resp. most likely, maximal) cardinal temperature for genetic group <i>s</i> in batch <i>i</i>
	par	$\mu_s^{T\min}$ (resp. μ_s^{Topt} , $\mu_s^{T\max}$)	°C	Mean minimum (resp. optimum, maximum) cardinal temperature for genetic group <i>s</i>
	fn	$S_{2,s}^{T\min}$ (resp. $S_{2,s}^{Topt}$, $S_{2,s}^{T\max}$)	°C	Standard deviation of the minimum (resp. optimum, maximum) cardinal temperature for genetic group <i>s</i>
	par	N_{\max}	log ₁₀ CFU/g	Maximum bacterial concentration in the product
	par	μ_{opt}	day ⁻¹	Most likely specific optimal growth rate in the product
Data and variables of the augmented model	fn	C_{ijs}^{obs}	log ₁₀ CFU/g	Data: observed concentration of the genetic group <i>s</i> in <i>B. cereus</i> in the package <i>j</i> coming from batch <i>i</i> after storage in the durability studies
	fn	α_{ijs}	log ₁₀ CFU/g	Censoring threshold for genetic group <i>s</i> in the observed package <i>j</i> coming from batch <i>i</i>
	par	σ	log ₁₀ CFU/g	Standard deviation of the lognormal distribution that quantifies the model error

^aDescription of classes: “iv” stands for intermediate variable, “par” stands for parameter, and “fn” stands for fixed node. ^bDescription of the steps *X* of the food pathway: *I*: initial, *C*: cooking, *B*:

CHAPITRE 6. Confrontation d'un modèle à des données

blending, *Milk*, *Starch*: ingredients, *M*: mixing, *PAR*: partitioning, *TI*: time interval, *PAST*: pasteurization, *CS*: cold storage at the company, *RT*: refrigerated transport, *RS*: refrigerated storage at retail, *HT*: home transportation, *HS*: home storage, *A*: storage in the durability studies.

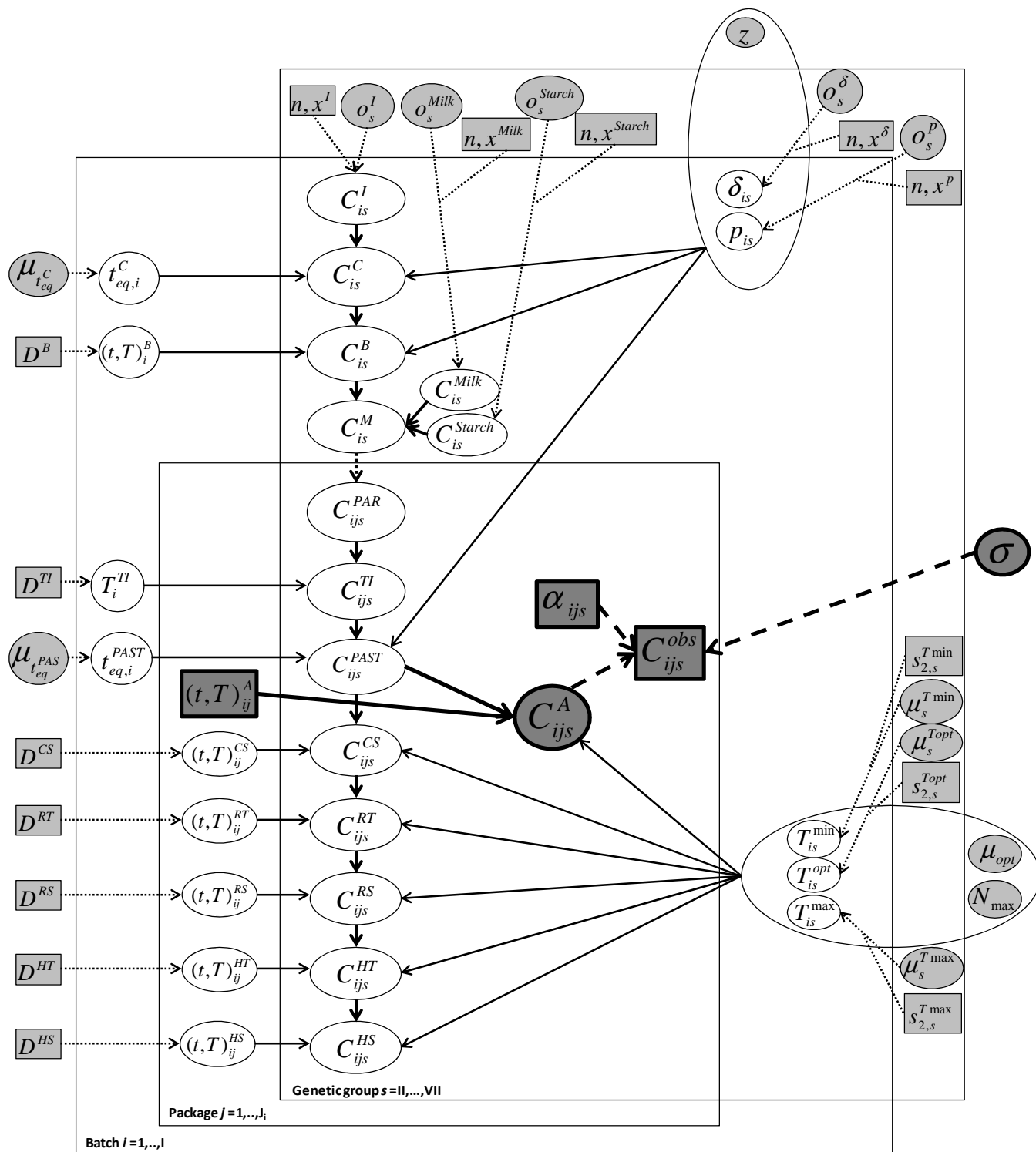


Figure 6.3. The DAG of the augmented BN as the combination of the second-order QMRA model and the likelihood model. Ellipses indicate random variables and rectangles indicate data or fixed nodes.

CHAPITRE 6. Confrontation d'un modèle à des données

Solid edges indicate deterministic links and dashed edges indicate stochastic links. The inputs of the second-order QMRA model are in light grey (rectangles pointing out the fixed inputs and ellipses pointing out the parameters) and the intermediate variables of the second-order QMRA model are in white. The variables involved in the likelihood model are pointed out in dark grey. See Table 6.2 for the definition of all the nodes.

6.2.2.b Data from durability studies

Data are observed concentrations of *B. cereus* in packages of courgette purée collected from the INRA-University of Avignon research laboratory, Avignon, France (called INRA source hereafter) and from the company manufacturing the courgette purée (called company source hereafter). Table 6.4 gives details about such durability studies and the collected experimental data. For both data sources, the packages analyzed followed the food pathway modelled by Afchain *et al.* (2008) up to the end of the pasteurization step and were then stored under different and potentially extreme time-temperature profiles (compared to industrial and home environments). 14 individual 800g packages came from one initial batch for the INRA source and 50 individual 400g packages came from different batches analyzed at different dates over two years for the company source. Thus the whole data set was obtained from quite various conditions. The global concentration of *B. cereus* was measured (in \log_{10} CFU/g) for each package by plating appropriate dilutions of the purée on MYP agar media as described in Guinebretière *et al.* (2003) For the positive packages, whose global concentration is above the limit of detection (LoD), *B. cereus* isolates were typed using both M13-PCR and DAF methods (Guinebretière and Nguyen-the, 2003) to determine their genetic groups (Guinebretiere et al., 2008). Only one dominant genetic group was found per package, except for one package where two genetic groups were identified. Following simple microbiological considerations based on specific decimal dilutions spread on agar plates, we suppose that, even if only one (or two) dominant genetic group(s) is (are) detected per package, the non-detected genetic groups might be present but they represent less than 10 % of the whole population. Hence, we define the specific concentrations of the non-detected genetic group as inferior to the global concentration in CFU/g divided by 10.

CHAPITRE 6. Confrontation d'un modèle à des données

Table 6.4. Description of the durability studies and the *B. cereus* concentration data measured in packages of courgette purée after different storage scenarios. The genetic groups refer to the groups defined by Guinebretière *et al.* (2008).

Data (source)	Product	Time and temperature of storage	Concentration in log ₁₀ CFU/g (detected genetic groups)
14 packages coming from the same batch (INRA)	5 packages of 800g	21 days at 4°C	< 1,7
			5,5 (VI)
	4 packages of 800g	21 days at 10°C	3 (VI)
			3,8 (II)
			3,1 (II)
			6 (IV)
			7,2 (II and IV)
50 packages coming from different batches (company)	5 packages of 800g	5 days at 20-25°C	6,4 (IV)
			6,3 (IV)
			6,3 (IV)
			3,7 (II)
			2,7 (II)
	27 packages of 400g	20 days at 4°C then 10 days at 8°C	3,7 (II)
			5 (II)
			< 2 for the 23 other packages
			5,2 (VI)
	23 packages of 400g	10 days at 4°C then 20 days at 8°C	4,4 (VI)
		4,7 (VI)	
		< 2 for the 20 other packages	

6.2.2.c The likelihood model

The experimental data, described in section 6.2.2.b, do not directly match the above second-order QMRA model (see section 6.2.2.a). Indeed, this model describes the fate of *B. cereus* in a standard “farm-to-fork” food chain, not in pasteurized packages of courgette purée submitted to durability studies (and therefore stored in potentially extreme conditions). To match with the data features, we built an augmented model that describes the dynamics of *B. cereus* in 64 packages coming from 51 different batches. Consequently, as illustrated in Figure 6.3 (dark grey nodes), we added new stages to the model to link the data to the QMRA model. The growth of *B. cereus* concentration during the durability studies was modeled with the same growth and competition models as the ones previously defined (see Table 6.1). To model the partitioning into packages (Afchain *et al.*, 2008), to describe the concentration of *B. cereus* in each of the 800g INRA packages all coming from the same batch, we used a multinomial distribution with probability parameters $p_1, \dots, p_{430} = 1/430$, partially generated by recursive sampling from 14 binomial distributions (Nauta, 2005) (only 14 packages were needed). On the other hand, a binomial distribution with probability parameter $p = 1/860$ was simply used for the 50 independent 400g company packages as each came from a different batch. Probabilities of 1/430 and 1/860 were used because a batch of 344kg was divided into 860 packages of 400g, or into 430 packages of 800g.

CHAPITRE 6. Confrontation d'un modèle à des données

We now describe the likelihood model which takes into account the model error and the data censoring (i.e., when the *B. cereus* concentration is lower than the limit of detection) that imply discrepancies between the observed concentration data and those induced by our second-order QMRA model. Let LoD_j be the limit of detection of the total concentration in package j (see Table 6.4: $LoD_j = 1.7$ or $2 \log_{10}$ CFU/g). Let $C_{ijTotal}^{obs}$ be the observed total concentration of *B. cereus* (in CFU/g) in package j coming from batch i . For each genetic group s , let C_{ijs}^{obs} be the observed concentration of *B. cereus* (in CFU/g) in package j coming from batch i and C_{ijs}^A the predicted concentration of *B. cereus* in package j coming from batch i after storage in accordance with durability conditions. These two random variables are linked through the following model:

- For the uncensored concentration i.e., if $\log(C_{ijs}^{obs}) > LoD_j$:

$$\log(C_{ijs}^{obs}) \sim N(\log(C_{ijs}^A), \sigma^2)$$

- For the censored concentrations i.e., for the concentrations of all the genetic groups s if $\log(C_{ijTotal}^{obs}) \leq LoD_j$ (case (i)), or for the concentrations of the non-detected genetic groups s if $\log(C_{ijTotal}^{obs}) > LoD_j$ (case (ii)):

$$\log(C_{ijs}^{obs}) \sim N(\log(C_{ijs}^A), \sigma^2)C(; \alpha_{ijs})$$

where $N(,)$ is the normal distribution, $N(,)C(; \alpha_{ijs})$ is the right censored normal distribution at the limit of $\alpha_{ijs} = LoD_j$ in case (i), and $\alpha_{ijs} = \log(C_{ijTotal}^{obs} / 10)$ in case (ii), and σ^2 is an unknown variance parameter that may be interpreted as the level of model error. Without a priori knowledge, we assigned the rather vague uniform prior distribution $U[0, 10]$ (in \log_{10} CFU/g) to σ . Figure 6.3 gives the DAG of the proposed augmented BN (composed of white, light grey and dark grey nodes).

6.2.2.d Bayesian inference of the augmented BN

MCMC sampling algorithms were applied as implemented in the software Jags 2.1.0. (Plummer, 2007) to obtain a sample from the posterior distribution of all the unknown variables of our augmented BN. In particular, all the quantities called parameters were inferred (see Table 6.2 column Class). The uncertainty prior distributions of z , N_{max} , μ_{opt} are described in Afchain *et al.* (2008) (because no inter-batch variability has been assumed for these parameters) whereas the uncertainty prior distributions assigned to all the remaining parameters are given in Table 6.3. Note that these priors are informative, based on expert's opinion of the company staff manufacturing the courgette purée, industrial and literature data. The fixed nodes (see Table 6.2 and rectangles) are not inferred.

We ran two independent MCMC chains (using different initial values for the parameters) of 1,400,000 simulations with a burn-in period of 1,000,000 and kept every 200th to reduce autocorrelation in the MCMC samples. Our results are therefore based on thinned samples of size 4000. Convergence of the MCMC run was assessed by graphical inspection of the chains and by

computing the Gelman-Rubin statistics as modified by Brooks and Gelman (Brooks et al., 1998) and autocorrelations.

6.2.2.e Prior and posterior predictive concentrations in the QMRA model

To point out the impact of the experimental data on *B. cereus* concentrations assessment through the standard "farm-to-fork" food chain (see Figure 6.2), we also performed prior and posterior predictive simulations for our "second-order" QMRA model (see Figure 6.3). By sampling in the prior and then in the posterior parameters distributions, we simulated the *B. cereus* prior and then posterior concentrations at each step of the standard "farm-to-fork" food chain. Only 10 packages per batch were followed to reduce computation time and computational memory space. The prediction procedure is well known in food safety risk assessment as a second-order Monte Carlo simulation method (Vose, 2000; Pouillot et al., 2007), and is as follows:

For $u=1,\dots,N_u\{$

Simulate a vector ϑ_u of possible values for all the model parameters from their uncertainty distribution.

For $i=1,\dots,N_v\{$

- Given the vector ϑ_u , simulate a vector $\varphi_i(\vartheta_u)$ of possible values for all the batch-specific intermediate variables and 10 vectors $\psi_{i,1}(\vartheta_u), \dots, \psi_{i,10}(\vartheta_u)$ of possible values for all the package-specific intermediate variables from their respective variability distribution.

- Using ϑ_u , $\varphi_i(\vartheta_u)$ and $\psi_{i,1}(\vartheta_u), \dots, \psi_{i,10}(\vartheta_u)$, for each genetic group s generate predictive concentrations at the different steps of the food chain in a batch i and in 10 packages coming from this batch i .

}

}

The 2D simulations were performed using $N_u = N_v = 200$, to avoid too much computation time and computational memory problems. This procedure leads to the simulation of $N_u \cdot N_v$ batches and $10 \cdot N_u \cdot N_v$ packages of courgette purée, i.e. 40 000 batches and 400 000 packages. We applied exactly the same procedure to make prior predictive simulations and posterior predictive simulations, except that all the parameters in the first case were simulated from their uncertainty prior distribution and in the second case from their joint uncertainty posterior distribution. Typically, we used posterior values from the MCMC inference algorithm previously applied to the full experimental data to simulate values in the posterior distributions. Note that the software JAGS 2.1.0 (Plummer, 2007) makes it possible to perform such a posterior prediction procedure in parallel with the MCMC algorithm.

For the main key stages of the food pathway, we compared:

CHAPITRE 6. Confrontation d'un modèle à des données

- The prior and posterior predictive prevalences (in %) of packages (or batches, depending on the stage) containing at least one *B. cereus* spore whatever the genetic group and per genetic group (see Table 6.5),

- The average prior and posterior predictive proportions (in %) of each genetic group relative to the whole *B. cereus* population in the positive packages (or batches), i.e. the units of food item containing at least one spore of *B. cereus* (see Table 6.6),

- The 95% prior and posterior predictive credible intervals (CI) of the *B. cereus* concentration whatever the genetic group in the positive packages (or batches) and per genetic group in the packages (or batches) containing at least one CFU of this group (see Table 6.7).

6.2.3 Results

6.2.3.a Update of prior beliefs

Convergence was fast for all but a few parameters, such as σ_{III}^{Milk} , which showed slower convergence due to a high level of intra-chain autocorrelation. Results show that the prior and posterior marginal distributions match very well for 40 of the 54 parameters of our augmented second-order QMRA model (i.e., the 53 parameters plus the error σ). This means that the data does not modify the associated prior beliefs.

Nevertheless, the prior marginal distributions were strongly updated for 14 of these 54 parameters, showing clear visual differences between the prior and posterior distributions in terms of central location, dispersion or shape. Figure 6.4 compares the prior and posterior marginal distributions of the 12 parameters whose prior beliefs were most strongly amended by the experimental data. Mainly two kinds of prior beliefs update can be distinguished:

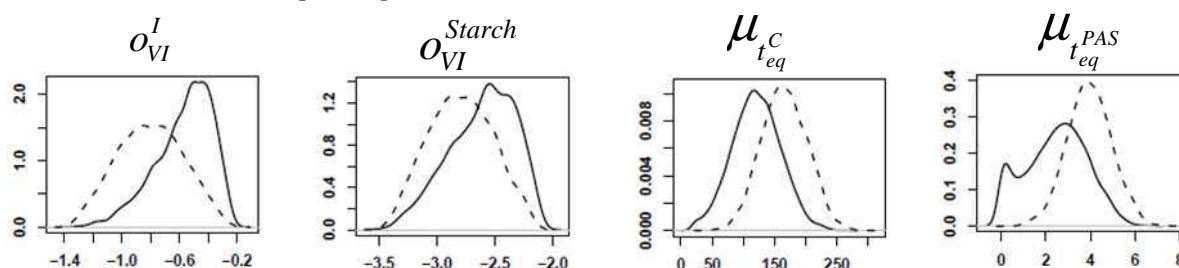
1. The range of the posterior marginal distribution is smaller than the prior one. This is the case, for instance, for $\mu_{II}^{T_{min}}$ i.e., the mean minimum cardinal temperature for genetic group II (prior standard deviation $sd= 1.5^{\circ}C$ vs posterior $sd= 0.8^{\circ}C$). We can also notice a very strong reduction in uncertainty for the error parameter σ whose posterior distribution is significantly peaked around $1.4 \log_{10} CFU/g$ (prior $sd= 2.8 \log_{10} CFU/g$ vs posterior $sd= 0.3 \log_{10} CFU/g$).
2. The posterior mean is substantially different from the prior one. This is the case for instance for $\mu_{t_{eq}^C}$ i.e., the mean equivalent cooking time at the target temperature of $90^{\circ}C$ (prior mean= 165min vs posterior mean= 121min) or for $\log_{10} N_{max}$ i.e., the log maximum concentration of *B. cereus* in the final product (prior mean= $8.5 \log_{10} CFU/g$ vs posterior mean= $6.6 \log_{10} CFU/g$). The higher prior mean modifications also concern the mean equivalent pasteurization time $\mu_{t_{eq}^{PAS}}$, the courgette most likely contamination in genetic group VI σ_{VI}^I , and the heat resistance parameters z and σ_{VI}^{δ} . The decrease in $\log_{10} N_{max}$

CHAPITRE 6. Confrontation d'un modèle à des données

indicates levels of final concentrations lower than expected. Smaller values in $\mu_{t_{eq}^C}$ and $\mu_{t_{eq}^{PAS}}$ lead to smaller heating treatments. A higher value in o_{VI}^I means a higher contamination in group VI in the initial vegetables. A higher o_{VI}^δ means more heat resistance for genetic group VI, and the decrease of z means that the *B. cereus* spores seem to be more sensitive to a temperature increase than expected.

Some batch- (or package)-specific intermediate variables were also updated from the experimental data. The differences between the updates of intermediate variables specific to different batches (or packages) are interesting because they reflect some biological or processing variability. For instance, the posterior mean of the equivalent cooking time at the target temperature of 90°C in batch n°3, i.e. $t_{eq,3}^C$, is estimated at 101 min, whereas $t_{eq,4}^C$ is estimated at 40 min in batch n°4 (prior mean is defined at 165min). Such a difference is due mainly to the strong heterogeneity observed between the data from these two batches. Indeed, the *B. cereus* concentration is lower than the LoD in the company package coming from batch n°3, whereas it is equal to 3.7 log₁₀ CFU/g in the company package coming from batch n°4 (see Table 6.4).

(a) Initial concentrations and process parameters



(b) Growth biological parameters

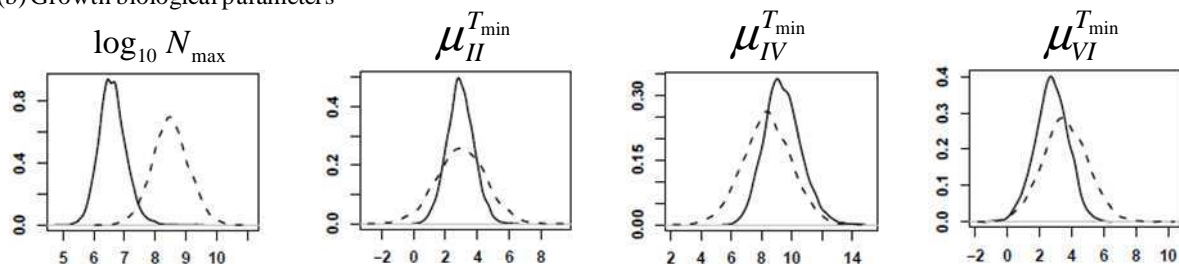
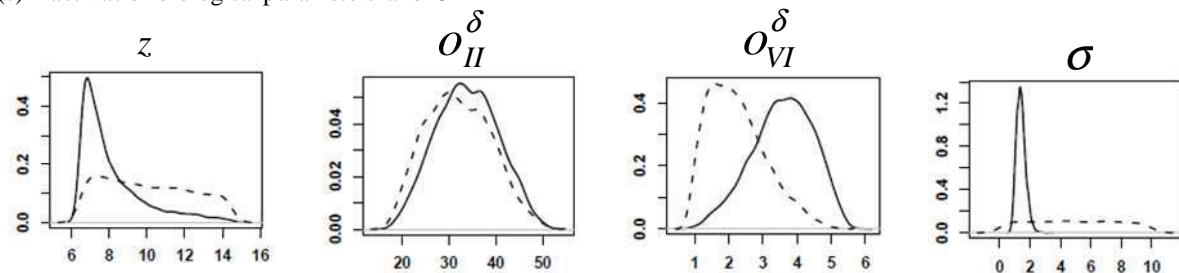
(c) Inactivation biological parameters and σ 

Figure 6.4. Prior (dashed line) versus posterior (solid line) marginal distributions of some parameters of the augmented BN. See Table 6.2 for definition of parameters.

6.2.3.b Comparison of the prior and posterior predicted *B. cereus* concentrations through the “farm-to-fork” food chain

Following the procedure explained in 2.5, we compare the prior and posterior predictive *B. cereus* concentrations through the food chain, obtained from second-order Monte Carlo simulations.

Table 6.5 indicates that the posterior predictive prevalence of *B. cereus* is higher than the prior one for all the genetic groups at the cooking stage. At the other stages, the posterior predictive prevalence is smaller than the prior one for genetic group III, whereas it is higher than the prior one for genetic groups IV and VI, particularly at the end of the chain (1.1 times and 6.7 times respectively). These differences are due to the influence of the observed data in which genetic group VI is present at much higher levels than those predicted by the second-order QMRA model, whereas genetic group III is not detected (see Tables 6.4, 6.5 and discussion). Note that these changes are consistent with the posterior reduction of the equivalent cooking time as well as with the posterior higher values of o_{VI}^{δ} (see Section 3.1. and Figure 6.4).

Table 6.6 indicates that compared to the average prior predictive proportions, the average posterior predictive proportions of genetic groups IV and VI globally increase in the positive packages (or batches) whereas the average posterior predictive proportions of genetic group III decrease.

Table 6.7 shows that from prior to posterior predictive simulations the distribution of the *B. cereus* concentration in the contaminated packages remains globally unchanged at the main key stages of the food pathway. The main exception appears at the end of the chain. The upper boundaries of the 95% posterior predictive CI are higher than the prior ones for genetic groups II and VI and smaller for genetic groups III, IV and V.

Finally, both the prior and posterior predictive simulations indicate that genetic groups II, III and IV prevail in prevalence and proportion at the end of the food chain. The upper boundaries of the 95% posterior predictive CI are reached for psychrotrophic genetic groups II and VI in the prior and in the posterior results.

CHAPITRE 6. Confrontation d'un modèle à des données

Table 6.5. Prior and posterior predictive prevalences of unit^a containing at least one *B. cereus* spore, for each genetic group at main key stages of the food pathway.

Genetic Group	Initial	After cooking	After mixing with ingredients and partitioning	After pasteurization	After home cold storage
Prior prevalence (%)					
II	100	62	29	13	13
III	100	47	77	38	38
IV	100	48	100	82	82
V	100	24	4	0.1	0.1
VI	100	47	5	0.03	0.03
VII	100	65	9	0.6	0.6
Total	100	84	100	87	87
Posterior prevalence (%)					
II	100	64	29	13	13
III	100	48	72	33	33
IV	100	50	100	89	89
V	100	27	4	0.1	0.1
VI	100	57	9	0.20	0.20
VII	100	67	9	0.5	0.5
Total	100	86	100	91	91

^aUnits are batches before partitioning and packages after partitioning.

Table 6.6. Prior and posterior predictive average proportions of each genetic group relative to the whole *B. cereus* population in positive units at main key stages of the food pathway.

Genetic Group	Initial	After cooking	After mixing with ingredients and partitioning	After pasteurization	After home cold storage
Prior average proportions (%)					
II	49	52	5	6	9
III	0.4	9	18	17	17
IV	0.4	7	76	78	74
V	0.4	0.2	0.2	0.01	0.03
VI	49	16	0.3	0.00	0.01
VII	0.4	14	0.4	0.07	0.05
Posterior average proportions (%)					
II	43	48	4	5	10
III	0.3	9	13	10	11
IV	0.4	7	83	85	79
V	0.4	0.2	0.2	0.01	0.03
VI	56	22	0.4	0.04	0.11
VII	0.3	13	0.3	0.05	0.04

Table 6.7. 95% prior and posterior predictive credible intervals (CI) of the *B. cereus* concentrations (in log₁₀ CFU/g) per genetic group and per unit of food item^a at main key stages of the food pathway.

Genetic Group	Initial	After Cooking	After mixing with ingredients and partitioning	After pasteurisation	After home cold storage
95% predictive prior CI					
II	[-1.7; 0.1]	[-5.3; -0.9]	[-2.6; -0.9]	[-2.6; -1.0]	[-2.4; 4.8]
III	[-4.1; -2.3]	[-5.4; -2.8]	[-2.6; -0.9]	[-2.6; -0.9]	[-2.5; -0.2]
IV	[-4.1; -2.3]	[-5.4; -3.0]	[-2.3; -0.3]	[-2.5; -0.5]	[-2.5; 0.1]
V	[-4.1; -2.3]	[-5.4; -3.6]	[-2.6; -2.3]	[-2.6; -2.2]	[-2.6; 3.1]
VI	[-1.7; 0.1]	[-5.4; -2.2]	[-2.6; -2.3]	[-2.6; -2.3]	[-2.6; 4.3]
VII	[-4.1; -2.3]	[-5.4; -2.8]	[-2.6; -2.3]	[-2.4; -2.0]	[-2.4; -2.0]
Total	[-1.1; 0.3]	[-5.3; -1.0]	[-1.9; -0.3]	[-2.5; -0.4]	[-2.5; 1.6]
95% predictive posterior CI					
II	[-1.7; 0.1]	[-5.3; -1.0]	[-2.6; -1.0]	[-2.5; -1.0]	[-2.3; 5.2]
III	[-4.1; -2.3]	[-5.4; -2.8]	[-2.6; -1.0]	[-2.6; -1.0]	[-2.5; -0.5]
IV	[-4.1; -2.3]	[-5.4; -3.0]	[-2.1; -0.2]	[-2.5; -0.4]	[-2.5; -0.2]
V	[-4.1; -2.3]	[-5.4; -3.6]	[-2.6; -2.3]	[-2.6; -2.2]	[-2.5; 1.8]
VI	[-1.6; 0.2]	[-5.4; -2.0]	[-2.6; -2.1]	[-2.6; -2.1]	[-2.5; 6.3]
VII	[-4.1; -2.3]	[-5.4; -2.8]	[-2.6; -2.3]	[-2.4; -2.0]	[-2.3; -2.0]
Total	[-1.0; 0.3]	[-5.2; -1.1]	[-1.9; -0.2]	[-2.5; -0.4]	[-2.5; 1.6]

^aOnly units containing at least one spore of the genetic group were considered in calculating the 95% CI.

6.2.4 Discussion

In this paper, we deal with a complex case study in food safety risk assessment and highlight the benefits of Bayesian inference by confronting an observed reality - here *B. cereus* concentration data from durability studies – with a QMRA model – here describing the fate of *B. cereus* along a courgette purée's food pathway, from an augmented BN specification.

In the proposed case study, the available experimental data contribute to strongly update the prior beliefs of 25% of the parameters and of some intermediate variables. Particularly, the ranges of some posterior marginal distributions shrink, meaning a reduction in uncertainty. Such substantial changes indicate that, for the associated variables, the information conveyed by the data from durability studies has been able to spread throughout the whole chain during the Bayesian inference procedure, up to the biological and process inactivation parameters which are defined upstream in the augmented BN. These prior updates may be explained by a relative discrepancy between the observed *B. cereus* concentrations and the ones predicted by the augmented BN in a package coming from durability studies. For instance, the prior predictive prevalence of genetic

CHAPITRE 6. Confrontation d'un modèle à des données

group VI in a package after pasteurization is only 0.03% according to the QMRA model (leading to a predictive prevalence in packages after durability studies of 0.03% as well), whereas this genetic group is present in at least 7.8% of the observed packages. Moreover, genetic group III is not detected in the observed packages, although it is the genetic group with the second higher prior prevalence and proportion at the end of the chain as calculated by the prior model. Such a discrepancy is partly picked up by the prior updates and by the standard deviation parameter σ that quantifies the model error and whose posterior mean is $1.4 \log_{10}$ CFU/g, (such a value is rather reasonable considering usual microbiological standards).

Note that in our case study 75% of the prior beliefs about the parameters of the model were not updated from the experimental data and that the posterior distribution of the *B. cereus* concentration in the contaminated packages remained globally close to the prior one through the “farm-to-fork” food chain. This may suggest either a weakness in the available information conveyed by the data (including 96% left-censored data), or that, for these unknown quantities, the prior distribution corresponds approximately to the information conveyed by the data and does not need to be updated, or that the prior distribution already contains too many information to be updated by the new available data.

In this paper, we propose an extended version of the initial first-order model proposed by Afchain *et al.* (2008) to describe the transmission of *B. cereus* along the courgette purée's food pathway. In other words, we build a second-order QMRA model in which batch- (or package)- specific variability is modeled explicitly and separately from epistemic uncertainty about some parameters. The case study shows that, compared to a first-order model, a second-order framework, based on hierarchical modelling (Gelman *et al.*, 2003), makes it possible to properly describe the different sources of heterogeneity of the experimental data (see Table 6.4) and provides a better interpretation of Bayesian inference results. Indeed, provided the experimental data convey enough information, Bayesian inference should highlight uncertainty reduction for all the parameters of a QMRA model (contrary to intrinsic variability that is irreducible by additional measurements). Previous works (Delignette-Muller *et al.*, 2006; Pouillot *et al.*, 2007) have already underlined the interest of combining Bayesian inference and second-order QMRA models, but not in a complete “farm-to-fork” QMRA model. Note that this use of second-order modelling is also an adequate answer to respect the recommendations from the international organizations to take variability and uncertainty separately into account in food risk analysis (Codex Alimentarius Commission, 1999; FAO/WHO, 2002; European Commission, 2003). However, the transformation of a first-order model into a second-order model is not simple, and has led in this work to some (debatable) modelling choices (see section 2.1), mainly based on easy biological interpretation, microbiological and parsimonious considerations.

Some strong prior belief updates require discussion because they deeply question experts' prior beliefs. They need to be analyzed and interpreted from experts' feedback: how realistic are such prior belief updates? If the experts globally agree with the new distribution of a given parameter, then this approach may have contributed to reducing the prior uncertainty about this unknown quantity. If the experts disagree, because the modification cannot be realistic, then a scientific debate is open: either the modelling has to be improved, or the expert opinions should be changed or improved, depending on the strength of the beliefs. In the case study, the posterior distributions of all the parameters of the second-order QMRA model seem coherent enough to

validate or amend some conclusions of the risk assessment in the courgette puree's food pathway. The modifications of the concentrations of genetic group VI in the initial vegetables and in starch is clearly interpretable, as such genetic groups are observed in higher quantities in the observed data than predicted a priori. For the same reason, the increase in the most likely heat resistance parameter σ_{VI}^{δ} for group VI is also consistent, as it may lead to a reduction in destruction of group VI during heat treatments, and consequently to a higher prevalence. However, according to the expert, the fact that the a priori heat resistance estimation of spores from group VI can be underestimated to an extent greater than those of the other genetic groups is surprising, because the prior δ values have been defined from spore heat resistance values on a large set of group VI strains (Carlin et al., 2006; Afchain et al., 2008), as for the other genetic groups. This could be explained by insufficient representivity of group VI strains used in heat resistance tests. The diversity of spore heat resistance within group VI can also be assumed to be wider than for the other groups. It would be interesting to do more experiments to validate or invalidate these assumptions. In any case, however, the posterior distribution of δ values for group VI is still much lower than for other genetic groups, in agreement with the previous assumption that group VI was the most heat sensitive genetic group (Carlin et al., 2006; Afchain et al., 2008). Moreover, the reduction of the maximum concentration of *B. cereus* in the final product, $\log_{10}N_{max}$, seems consistent with the rather low observed concentrations in the packages. To produce a conservative simulation (i.e. a fail-safe model) and because $\log_{10}N_{max}$ is usually considered independent of the temperature of growth, Afchain et al. (2008) used a high $\log_{10}N_{max}$ value in the QMRA model. However, the $\log_{10}N_{max}$ posterior values are closer to the concentrations observed after storage at 10°C by Choma et al. (2000), suggesting that $\log_{10}N_{max}$ of *B. cereus* indeed depends on the incubation temperatures in these products. This would be consistent with the recent finding that without oxygen (which is the case of the vacuum packaged courgette purées), maximal growth of *B. cereus* is reduced at cold temperatures (De Sarrau et al., 2012). As for the means of the equivalent cooking and pasteurization times at the target temperature of 90°C $\mu_{t_{eq}}^C$ and $\mu_{t_{eq}}^{PAS}$, the posterior distributions are considerably lower than the prior ones. These process parameter modifications show an overestimation of the prior equivalent heating times and suggest that reconsideration of the computation of the equivalent time is necessary.

To conclude, a difficulty of the approach is linked to its computational feasibility. Basically, Bayesian inference through MCMC algorithms on high-dimensional BN including continuous variables is much more complex than MC simulations or Bayesian inference in BN using discrete variables. This approach is also more interesting because it is closer to reality. Very efficient software, e.g. Jags (Plummer, 2007), are powerful tools for performing complex Bayesian inference. However, in high-dimensional models as in our example, the convergence of the MCMC algorithms may be very slow, particularly due to high intra-chain autocorrelation. This may be due partly to the global scarcity of our data, compared to the complexity of the model (only 64 data, with many censored data, and only one package per batch for the company data). Such difficulties have to be compared with the advantages of having a powerful tool to improve and validate a QMRA model through discussion with the experts of the domains of interest.

Acknowledgments

This project is supported by grants from the Agence Nationale de la Recherche (ANR, France) within the frameworks of the Quant'HACCP and Ribenut projects. Thanks are due to Jean-Baptiste Denis (INRA, Unité MIA, 78352 Jouy-en-Josas Cedex - FRANCE) for his precious advice. Thanks are also due to Marie-Hélène Guinebretière (INRA, UMR 408, Sécurité et Qualité des Produits d'Origine Végétale, F-84000 Avignon, France) for kind communication of data.

Chapitre 7

Conclusion et perspectives

Dans ce chapitre, un bilan des différents travaux effectués dans cette thèse est proposé, puis les limites et les perspectives de ces travaux sont évoquées.

7.1 Bilan

7.1.1 Bilan des travaux

Le principal travail de cette thèse a porté sur une modélisation risque-bénéfice dans une chaîne de transformation d'un légume, la conserve de haricots verts. Ce travail était au cœur du projet ANR Ribenut ; il a demandé une grande collaboration avec les microbiologistes et les chimistes impliqués dans le projet. Le travail de modélisation a commencé avec la visite de l'usine de fabrication, dont le but était d'une part de recueillir des données et d'autre part d'avoir une vision de la réalité industrielle, qui est impressionnante par la dimension de l'usine et des machines. Cette visite était très intéressante car elle m'a permis, en tant que modélisatrice, d'appréhender la complexité de la réalité industrielle et de ses dimensions économiques, logistiques, sociales, environnementales. Ainsi, il peut sembler à première vue peu hygiénique d'utiliser de l'eau et du jus recyclés pour les procédés de blanchiment ou de jutage, car le recyclage entraîne des risques de contamination croisée. Mais d'un autre côté, ces recyclages permettent d'importantes économies d'eau, alors même que la consommation d'eau d'une telle usine de fabrication est énorme, équivalente à celle d'une ville entière de plus de 200 000 habitants... Une vision d'ensemble des chaînes de fabrication permet de mieux comprendre leur complexité et leurs différents enjeux. Par ailleurs, à l'issue d'une telle visite, le modélisateur ressort convaincu du principe bien connu de la modélisation : « un modèle est toujours faux », et mesure le challenge représenté par la construction d'un modèle représentant cette chaîne complexe de fabrication. Mais le but est finalement d'obtenir des « modèles utiles », ce qui reste compliqué mais est plus facilement réalisable car le modèle est construit selon les objectifs de la modélisation et ne cherche pas à englober toute la réalité. Une vision réaliste du phénomène à modéliser permet aussi de mieux appréhender le travail de modélisation et facilite la compréhension avec les experts du terrain, les gestionnaires du risque, les

CHAPITRE 7. Conclusion

microbiologistes et les chimistes, qui fournissent les données et les informations utiles à la modélisation.

Ce travail de modélisation risque-bénéfice a demandé le rassemblement de nombreuses données, sur les procédés industriels, sur les niveaux de contaminations bactériennes et les teneurs en vitamines, ou encore sur les paramètres microbiologiques et chimiques régissant le comportement des bactéries et des vitamines dans certaines conditions environnementales (température, pH, teneur en oxygène etc.). Les données proviennent d'analyses expérimentales spécifiquement menées lors du projet ANR, de la littérature ou d'avis d'experts. En particulier, la détermination des paramètres microbiologiques et chimiques a été faite essentiellement à partir de données de la littérature. Une méta-analyse a ainsi été réalisée pour déterminer les paramètres de résistance thermique de *G. stearothermophilus* (Chapitre 2). Cette méta-analyse a été faite par le biais d'une approche hiérarchique bayésienne, où l'importance de prendre en compte les sources de variabilité entre les données dans la méthode d'estimation a été mise en évidence par la comparaison de plusieurs modèles. Cette méta-analyse a permis de synthétiser l'information issue des nombreuses études sur la résistance thermique de *G. stearothermophilus*. Dans le cas de la vitamine C, beaucoup moins d'information était disponible, et les paramètres de la réaction chimique de dégradation de la vitamine C ont chacun été déterminés à partir d'une étude unique (partie 4.2.6.d). Des sources d'incertitudes étaient attachées à l'estimation des différents paramètres et ont été prises en compte dans les modèles.

Une fois les modèles construits, leurs résultats ont été analysés afin de déterminer le risque et le bénéfice dans les haricots verts appertisés (Chapitres 3 et 4). La simulation de Monte Carlo à deux dimensions a été utilisée, afin de propager conditionnellement incertitude et variabilité, comme recommandé par les organisations internationales (Codex Alimentarius Commission, 1999; FAO/WHO, 2006; European Commission, 2003). Cette méthode facilite l'interprétation des résultats et par conséquent la gestion des risques. Puis, ces modèles ont été confrontés à des données indépendantes afin d'étudier leur validité. Des méthodes d'analyses de sensibilité et des tests de scénarios ont ensuite été mis en place. Ces méthodes sont en effet cruciales car elles permettent de mieux comprendre le fonctionnement des modèles, en déterminant les facteurs clés, les principales sources de variation des sorties finales, en identifiant les principales sources d'incertitudes dans le modèle, et en donnant des pistes de simplification et de priorisation pour la réduction de l'incertitude. Parmi les scénarios testés, certains correspondaient à des tests d'hypothèses incertaines des modèles, sur l'impact de la germination des spores par exemple, et d'autres correspondaient à des mesures possibles de maîtrise du risque ou d'augmentation du bénéfice. Une fois le modèle risque et le modèle bénéfice construits et analysés, le modèle global risque-bénéfice a été étudié, avec recherche d'optimisation du compromis risque-bénéfice (Chapitre 5). Un critère probabiliste incluant à la fois le risque et le bénéfice a été introduit.

Les résultats principaux des modèles risque et bénéfice sont le faible risque de non-stabilité des haricots verts appertisés (intervalle de variabilité à 95% des concentrations en *G. stearothermophilus* après incubation de $[-3.2; -3.2]$ \log_{10} UFC/g, et taux de non-stabilité correspondant de 0.5% avec un intervalle d'incertitude à 95% de $[0.1\%; 1.2\%]$), et le faible bénéfice associé à la concentration en vitamine C dans les haricots verts appertisés (intervalle de variabilité à 95% de $[0; 5.6]$ mg/100g). Cette teneur en vitamine C est en dessous du seuil d'allégation nutritionnelle de 12 mg/100g, mais les haricots verts contribuent néanmoins à l'apport en vitamine C

CHAPITRE 7. Conclusion

dans l'alimentation. Par ailleurs, ce risque et ce bénéfice sont relativement indépendants étant donné les facteurs et leur plage de variation pris en compte dans les modèles ; il n'y a pas de facteur commun qui influence significativement à la fois le risque et le bénéfice (Chapitre 5). Cette relative indépendance est en fait un avantage dans la recherche d'optimisation du compromis risque-bénéfice, puisque les mesures prises pour augmenter le bénéfice ne devraient pas ou peu influencer le risque, et inversement les mesures prises pour diminuer le risque ne devraient pas ou peu influencer le bénéfice. Des légères modifications des procédés, comme la réduction de la durée d'attente des légumes frais après récolte, de la durée du blanchiment, de la contamination bactérienne de l'eau de blanchiment, ou encore la baisse du pH du produit fini, sont des pistes intéressantes pour optimiser le compromis risque-bénéfice.

Un travail indépendant et complémentaire à la modélisation risque-bénéfice a été mené sur une chaîne de fabrication de purée de courgettes, avec la bactérie *B. cereus*. Il s'agissait de confronter un modèle d'évaluation du risque préalablement construit à des données de vieillissement, par la méthode d'inférence bayésienne (Chapitre 6). Le modèle d'analyse de risque sur cette chaîne est assez semblable à celui construit dans les haricots verts appertisés, sauf qu'il prenait en compte 6 groupes génétiques différents de *B. cereus*, ne considérait pas de sources de contamination croisée et modélisait la concentration bactérienne jusque dans le réfrigérateur du consommateur. De plus, le modèle initialement existant a été transformé dans cette thèse en un modèle à deux dimensions, permettant l'ajustement de données provenant de différents batches de purée. La méthode développée dans ce travail cherche à améliorer les modèles existants en introduisant de l'information issue de nouvelles données. Ce travail fait appel à des méthodes et des problématiques similaires à celles utilisées dans la modélisation risque-bénéfice, comme le traitement de l'incertitude et de la variabilité ou l'inférence bayésienne. Une différence notable est que dans ce cas, l'inférence bayésienne a été appliquée sur le modèle global d'évaluation du risque et pas seulement sur une partie de celui-ci, et donc a visé à mettre à jour l'ensemble des paramètres du modèle et pas seulement à estimer des distributions sur certains paramètres d'entrées du modèle.

7.1.2 Pluridisciplinarité et collaboration industrielle

Cette thèse a été marquée par une très forte pluridisciplinarité, puisque c'est un travail de statistique appliqué à la microbiologie et à la chimie. Ce travail a requis une collaboration étroite avec les microbiologistes et les chimistes impliqués dans le projet ANR, qui ont fourni des données et les explications associées à ces données, ont déterminé les processus biologiques impactant les bactéries et les vitamines aux différentes étapes des chaînes de transformation des légumes, et ont aidé à l'interprétation des résultats des modèles.

Les mondes de « la paillasse » et du statisticien « accroché à son ordinateur » sont très différents, et les collaborations nécessitent écoute et compréhension. Le statisticien doit notamment comprendre et prendre en compte les contraintes des expérimentations : par exemple, les très grandes incertitudes sur certaines données (différences surprenantes de résultats entre différents laboratoires sur les teneurs en vitamine C), ou l'attente durant 6 mois ou 1 an pour recevoir les données de prélèvements effectués lors des campagnes de récolte des haricots verts, qui ont lieu aux

mois d'août et septembre seulement... La collaboration entre plusieurs domaines est cependant le propre du travail du statisticien appliqué, et constitue une part très enrichissante de son travail.

Par ailleurs, ce travail a aussi demandé une collaboration avec les industriels dont les chaînes de production ont été modélisées. Même s'il est parfois nécessaire de composer avec les exigences des entreprises sur certaines données touchant au secret industriel, de nombreuses et précieuses informations sur les procédés industriels ont été obtenues, notamment grâce à des analyses expérimentales réalisées directement sur les chaînes de production dans le cadre du projet ANR Ribenut ou dans le cadre d'audits du CTCPA (Centre technique de la conservation des produits agricoles).

7.2 Limites des travaux

7.2.1 Limites sur les données

Le manque de données ou la mauvaise qualité des données disponibles sont des problématiques fréquentes en statistique appliquée. Ce travail de thèse y a aussi été confronté. L'estimation de certains paramètres a été difficile, notamment lors de la construction du modèle vitamine, où peu d'information était disponible pour déterminer les énergies d'activation de la vitamine C, le taux d'oxygène dans le haricot vert aux différentes étapes de la chaîne, ou même la concentration initiale en vitamine C oxydée et non oxydée. Une difficulté vient notamment du fait que le comportement des vitamines peut varier assez fortement selon la matrice alimentaire, par exemple d'un légume à un autre. Des approximations ont donc été parfois nécessaires ; il a notamment été supposé que les paramètres chimiques de la vitamine C étaient semblables dans la pomme de terre et le haricot vert, ou encore que le taux d'oxygène dans les tissus du haricot vert blanchi était proche des taux d'oxygène dans son milieu environnant. Dans le modèle d'analyse des risques, la modélisation de la contamination croisée par les débris de haricots verts a aussi posé problème, car il n'y avait aucune information disponible sur la fréquence d'introduction de ces débris dans les conserves. La mesure des concentrations en bactéries aérobies thermophiles dans ces débris révélait néanmoins qu'ils pouvaient être très contaminés, et donc représenter un risque. Des hypothèses théoriques ont donc été faites, mais demanderaient à être étayées de données.

La construction d'un modèle peut d'ailleurs permettre d'orienter les efforts à effectuer en matière d'acquisition de données et d'analyses. Des paramètres mal connus peuvent apparaître importants dans le modèle, et nécessitent alors d'être mieux renseignés. Ainsi au cours de la construction du modèle d'évaluation du risque dans les haricots verts, une analyse de sensibilité a révélé le fort impact potentiel du pH du produit à la stérilisation. En absence de données sur ce pH, des analyses expérimentales ont été mises en place afin de le mesurer (merci au CTCPA d'Avignon), ce qui a permis de consolider la modélisation de l'effet du traitement thermique à la stérilisation.

Par ailleurs, dans le modèle d'analyse des risques sur la chaîne haricots verts appertisés, le risque de non-stabilité est faible, 0.5% en moyenne avec un intervalle d'incertitude de [0.1%; 1.2%]. Ce faible taux correspond au fait que le modèle a été construit à l'aide de données de concentrations bactériennes prélevées sur des chaînes de production avec des faibles taux de non-stabilité ; les

résultats du modèle sont très cohérents avec les taux de non-stabilité observés sur ces chaînes. Dans certaines usines et certaines années, ce taux peut atteindre 3%, 5% ou 10%, ce qui devient plus problématique. Il aurait été sans doute plus intéressant de travailler sur des chaînes de production avec des taux de non-stabilité plus élevés, où les mesures de maîtrise du risque proposées auraient eu un enjeu plus élevé et un impact plus fort. Cependant, la non-stabilité n'est pas connue à l'avance et le hasard a voulu que les prélèvements aient lieu sur des chaînes avec de faibles taux de non-stabilité ces années-là. Par ailleurs, il aurait aussi pu être intéressant de travailler avec des données provenant de plusieurs chaînes de production sur un plus grand nombre d'usines et d'années afin de modéliser une plus grande variété de situations.

7.2.2 Limites des méthodes

Une limite de la méthode de simulation de Monte Carlo à deux dimensions est le nombre relativement faible de simulations effectuées dans la dimension incertitude d'une part, et dans la dimension variabilité d'autre part. Les limites viennent des capacités de stockage et des temps de calcul encore longs des logiciels (R dans ces travaux). L'utilisation de matrices de grande dimension est en particulier difficile ; il est recommandé d'optimiser les programmes. Ainsi dans les simulations utilisées pour étudier les modèles d'analyse du risque et du bénéfice dans la chaîne haricots verts appertisés, $N_u = 1500$ tirages ont été réalisés dans la dimension incertitude, et $N_v = 2000$ ou $N_v = 3000$ tirages ont été réalisés dans la dimension variabilité, d'où la simulation d'un nombre total de boîtes supérieur à 3 millions. Un nombre de 1500 ou 2000 simulations est assez faible pour explorer en détail une distribution, en particulier pour aller tirer dans la queue de distribution. Cependant dans une simulation de Monte Carlo 2D réalisée sur le modèle d'évaluation du risque dans la purée de courgettes, un nombre plus faible de tirages a été réalisé à partir du logiciel Jags et des lois a posteriori issues de l'inférence bayésienne ($N_u = 200$ et $N_v = 200$, avec 10 paquets modélisés par batch et 6 groupes génétiques de *B. cereus*, d'où 40 000 batchs et 400 000 paquets simulés). Dans certains travaux de recherche, des valeurs de $N_u = 100$ et $N_v = 650$ (Mokhtari et al., 2005), de $N_u = 500$ et $N_v = 500$ (Delignette-Muller et al., 2006) ou de $N_u = 100$ et $N_v = 10\ 000$ (Mataragas et al., 2010) ont été utilisées. L'enjeu est d'autant plus fort lorsqu'il s'agit d'étudier des situations extrêmes, comme les risques alimentaires. Le nombre de tirages dans la dimension variabilité est alors parfois privilégié. Les capacités des logiciels et des ordinateurs ont cependant vocation à augmenter dans le futur, ce qui atténuera cette limite actuelle des outils de calculs.

Une autre limite de la simulation de Monte Carlo à deux dimensions, plus conceptuelle, concerne les difficultés qui apparaissent parfois lors de l'émission d'hypothèses sur la variabilité ou l'incertitude attachées aux entrées du modèle (puis lors de l'ajustement des distributions de variabilité ou d'incertitude correspondantes). Pour certaines entrées du modèle, les hypothèses sont assez simples à émettre, par exemple de nombreuses données microbiologiques montrent que la résistance thermique d'une bactérie (temps de réduction décimale D) peut varier en fonction de sa souche. Mais le caractère variable ou non de certaines entrées est parfois difficile à connaître, souvent par manque de connaissances assez précises. Différentes hypothèses de variabilité des paramètres de résistance thermique z_T et z_{pH} de *G.stearothermophilus* ont par exemple été émises dans le cadre de la méta-analyse en Chapitre 2. Il s'agit par ailleurs de préciser clairement quelles sont les sources de variabilité possibles (différentes souches, différents milieux de traitement,

CHAPITRE 7. Conclusion

variabilité en fonction du temps, etc.). Ces notions délicates de distinction entre la variabilité et l'incertitude sont ainsi apparues lors de la méta-analyse, lors de la construction des modèles risque et bénéfice sur la chaîne haricots verts, ainsi que sur le modèle d'évaluation du risque sur la purée de courgettes lors du passage du modèle 1D au modèle 2D, qui a nécessité de nombreuses hypothèses parfois discutables. Un compromis est à rechercher entre le niveau de connaissance actuelle sur une entrée/un phénomène, la quantité et la qualité des données disponibles, et le besoin de parcimonie du modèle. Malgré ces limites, la simulation de Monte Carlo à deux dimensions présente de nombreux avantages car son exigence de distinction variabilité/incertitude facilite la bonne interprétation des résultats du modèle (« que représente l'intervalle de confiance obtenu sur le risque ? »), ce qui est primordial pour son usage en gestion du risque.

La « validation » ou l'« évaluation » des modèles est un sujet délicat. En général, un modèle est dit valide s'il est capable de reproduire correctement les données observées. Une façon de valider un modèle consisterait donc à appliquer en parallèle un ensemble de paramètres d'entrée à la fois au mécanisme réel et à son modèle, puis à comparer l'ensemble des paramètres de sortie produits par chacun d'eux. La répétition de cette opération permettrait de couvrir toute la plage de fonctionnement prévue, et le modèle serait validé s'il produit approximativement les mêmes valeurs que le mécanisme réel. Malheureusement, dans le cas de systèmes réels complexes comme les modèles d'analyse du risque ou du bénéfice construits dans cette thèse, il est très difficile, voire impossible, de recueillir des données aussi exhaustives (ensemble des valeurs des paramètres d'entrée, plus sortie), et encore plus dans de nombreuses conditions. La validité des modèles peut néanmoins être étudiée à partir des différentes sources de données indépendantes disponibles. Dans le cas du modèle d'analyse des risques dus à *G. stearothermophilus* dans les haricots verts appertisés, les résultats du modèle étaient globalement en bonne adéquation avec les concentrations mesurées indépendamment en un point de la chaîne de production et avec les taux de non-stabilité observés sur les chaînes. Même s'ils ne sont pas suffisants pour assurer la validité du modèle, ces éléments sont rassurants. Dans le cas du modèle d'analyse de la vitamine C dans les haricots verts appertisés, les résultats semblent globalement assez en accord avec les données observées, mais sont un peu discordants au niveau de l'effet de la stérilisation ou du blanchiment qui sont peut-être un peu sous-estimés ou surestimés. Il serait nécessaire d'approfondir ces vérifications en collaboration avec des experts chimistes et éventuellement de préciser certaines parties du modèle.

Dans le cadre de l'optimisation ou de l'analyse risque-bénéfique, une analyse rigoureuse et complète devrait prendre en compte tout un ensemble de critères, et pas seulement un risque et un bénéfice. La thèse de V. Sirot (2010) consacrée à une analyse risque-bénéfique de la consommation de poisson souligne que les recommandations découlant d'une analyse risque-bénéfique ne sont pas les mêmes selon que l'ensemble des principaux risques et bénéfices sont pris en compte, ou seulement un petit nombre d'entre eux. Ainsi les études comparant les risques liés au méthylmercure seul versus les bénéfices liés à certains acides gras (AGPI-LC n-3) dans la consommation de poisson concluent toujours à un bénéfice supérieur au risque (Rissanen et al., 2000; Cohen et al., 2005), alors que de nombreuses études montrent l'impossibilité d'atteindre les recommandations nutritionnelles tout en restant sous les seuils toxicologiques de référence lors de l'ajout dans l'analyse des effets néfastes des dioxines et des PCB (Foran et al., 2005; Antonijevic et al., 2007; Domingo et al., 2007). La pertinence des recommandations issues des analyses risques-bénéfices va dépendre du choix des risques et bénéfices pris en compte dans l'analyse. Dans l'exemple de la chaîne haricots verts appertisés, une analyse risque-bénéfique complète aurait également considéré le risque sanitaire

d'altération biologique (*Clostridium botulinum*), les bénéfices apportés par d'autres nutriments (folates, fibres, etc.), la qualité gustative du produit, la faisabilité et le coût des interventions. Il aurait été particulièrement intéressant de s'intéresser à l'acide folique et à ses dérivés, présents en plus grande quantité (relativement aux besoins nutritionnels) que la vitamine C dans les haricots verts. Cette étude avait été envisagée dans le projet ANR Ribenut, mais n'a pas été mise en place par manque de temps et par manque de connaissance sur les phénomènes à modéliser au moment de la construction des modèles. Prendre en compte un large ensemble de critères représente un défi considérable et un important travail nécessitant l'acquisition et la mise à disposition des modélisateurs d'une multitude d'informations. Un premier pas est accompli dans cette thèse par la prise en compte d'un risque microbiologique et d'un bénéfice nutritionnel sur la même chaîne de production et par une description unique des paramètres de commande du procédé agissant différemment sur deux phénomènes, ce qui est encore peu courant dans le domaine des transformations alimentaires.

7.3 Perspectives

Dans cette thèse, une méthode d'inférence bayésienne a été utilisée pour ajuster un modèle d'évaluation quantitative des risques microbiologiques à des données de vieillissement (Chapitre 6). Le but était d'améliorer la connaissance sur les paramètres du modèle à partir de l'information contenue dans des données en aval de la chaîne, qui étaient des concentrations bactériennes dans des produits prélevés à la fin de la chaîne de production et incubés. De telles données de sorties de modèles (données de contamination ou teneurs en vitamine d'un produit fini, éventuellement prévalence d'une maladie, etc.) sont finalement parfois plus simples à obtenir que certaines données d'entrées du modèle. L'intérêt de tels outils est donc renforcé. Dans ce travail, la comparaison des lois a priori et des lois a posteriori a montré que la connaissance sur certains paramètres avait été mise à jour, comme la concentration maximale atteignable ou la résistance thermique de certains groupes génétiques de *B. cereus*. Des méthodes de Monte Carlo par Chaîne de Markov (MCMC) ont permis de générer des échantillons des lois a posteriori. Ces méthodes étant assez lourdes à mettre en place sur un modèle complexe (difficulté de programmation, difficulté de convergence et auto-corrélation des chaînes, programmation impossible de certaines lois comme la loi multinomiale à taille aléatoire en Winbugs et Jags, etc.), il serait possible de chercher à développer une méthode de calcul bayésien approché (ABC) avec le même objectif. Les méthodes ABC ont été développées à partir de l'année 1997 initialement dans des disciplines comme la biologie évolutive et la dynamique des populations pour des modèles complexes pour lesquels la vraisemblance était incalculable en pratique (Tavaré et al., 1997; Beaumont et al., 2002). Ces modèles étaient cependant utilisables pour simuler des données; l'ABC se base sur cette faculté. Le principe de l'ABC est la méthode d'acceptation-rejet : des jeux de paramètres sont simulés sous la loi a priori, et des jeux de données correspondantes sont simulées; puis les jeux de paramètres ayant permis de simuler des données proches des données réelles sont acceptés, et les autres sont rejetés. Un échantillon de la loi a posteriori est alors constitué à partir des jeux de paramètres acceptés. Un des enjeux de cette méthode est la définition d'une « bonne » distance permettant de comparer les données simulées

CHAPITRE 7. Conclusion

aux données réelles ; pour cela des statistiques résumées des données sont souvent utilisées. Par ailleurs, si les résultats du modèle sont très éloignés des données, le taux d'acceptation est très faible ; des méthodes peuvent être utilisées pour améliorer le taux d'acceptation (Beaumont et al., 2002; Marjoram et al., 2003; Csilléry et al., 2010). Au cours de cette thèse, des programmes ont été implémentés pour appliquer la méthode ABC au modèle d'analyse des risques *B. cereus*, dans le but de comparer les résultats de l'ABC et du MCMC. La distance était basée sur le type de groupes génétiques détectés et les niveaux de contamination observés dans les 64 paquets de purée de courgettes. Cependant, les essais se sont avérés infructueux car les taux d'acceptations étaient toujours nuls. Une explication pourrait être que les prévisions du modèle étaient trop éloignées des données. Il était en effet impossible de simuler un ensemble de paquets dont les concentrations dans chaque condition de vieillissement et pour chaque groupe génétique étaient proches de celles observées dans les données ; en particulier, le groupe génétique VI était très peu prédit mais était détecté dans les données. Une autre explication pourrait être que les tirages indépendants réalisés dans les priors ne permettaient pas en un nombre raisonnable d'itérations (tout de même des millions !) de trouver l'espace a posteriori des paramètres. La méthode d'inférence bayésienne par méthode MCMC s'est révélée nettement plus efficace que l'ABC, puisqu'elle a fourni des lois a posteriori pertinentes. Il serait néanmoins intéressant de chercher à développer de nouveau la méthode ABC sur des modèles d'évaluation du risque, en commençant peut-être par des modèles plus simples que le modèle AQR *B. cereus*.

Au niveau des modèles risque et bénéfice construits, il serait intéressant d'approfondir la validation de ces modèles, en particulier le modèle vitamine qui demande à être amélioré sur certains points. De plus, une fois les modèles d'évaluation du risque et du bénéfice validés, ils pourraient être partiellement simplifiés, en se basant sur les résultats des analyses de sensibilité. Par exemple il a été montré que l'impact du blanchiment était négligeable sur l'évolution de la contamination en *G.stearothermophilus*, ou encore que l'altération après incubation était quasiment systématique dès qu'au moins une spore avait germé. Dès lors, ces parties de modèle pourraient être simplifiées, ce qui faciliterait de futurs travaux sur ces modèles en réduisant notamment le nombre de paramètres d'entrée.

Par ailleurs au niveau de l'analyse risque-bénéfice, il serait intéressant de l'approfondir en testant des méthodes complémentaires aux tests de scénarios, que nous n'avons pas eu le temps de mettre en place. Une optimisation des paramètres pourrait être menée au moyen d'algorithmes d'optimisation maximisant un critère combinant le risque et le bénéfice. Des procédures d'optimisation adaptées aux procédés alimentaires sont par exemple mises en place au sein de l'Unité Mixte de Recherche Génie des Procédés Alimentaires (INRA UMR-GENIAL) (Olmos Perez, 2003). Enfin, il serait pertinent de prendre en compte d'autres types de risques et bénéfices associés aux haricots verts appertisés, comme l'étude des folates ou la prise en compte du coût des interventions envisageables.

Références bibliographiques

Afchain, A.L., Carlin, F., Nguyen-The, C., Albert, I. 2008. Improving quantitative exposure assessment by considering genetic diversity of *B. cereus* in cooked, pasteurised and chilled foods. *International Journal of Food Microbiology* 128, 165-173.

AFNOR. 1988. Norme NF T72-231. Antiseptiques et désinfectants utilisés à l'état liquide, miscibles à l'eau - Détermination de l'activité sporicide - Méthode par filtration sur membranes., <http://www.boutique-normes.afnor.org/norme/nf-t72-231/antiseptiques-et-desinfectants-utilises-a-l-etat-liquide-miscibles-a-l-eau-determination-de-l-activite-sporicide-methode-pa/article/702102/fa001862>.

AFNOR. 1997. Norme NFV08-408 Microbiologie des aliments - Contrôle de la stabilité des produits appertisés et assimilés - Méthode de routine, <http://www.boutique.afnor.org/norme/nf-v08-408/microbiologie-des-aliments-contrôle-de-la-stabilité-des-produits-appertisés-et-assimilés-méthode-de-routine/article/757270/fa045341>.

AFNOR. 2011. NF V08-602 - Microbiologie des aliments - Dénombrement des spores dans les produits alimentaires avant traitement d'appertisation par comptage des colonies, <http://www.boutique.afnor.org/norme/nf-v08-602/microbiologie-des-aliments-denombrement-des-spores-dans-les-produits-alimentaires-avant-traitement-d-appertisation-par-comptag/article/697591/fa169301>.

AFSSA. 2003. Evaluation des risques et bénéfices nutritionnels et sanitaires des aliments issus de l'agriculture biologique, Maisons-Alfort, France.

AFSSA. 2010. Avis de l'Agence française de sécurité sanitaire des aliments relatif aux bénéfices / risques liés à la consommation de poissons. In: 2008-SA-0123, A.S.n., (Ed.), Maisons-Alfort, France.

Albert, I., Grenier, E., Denis, J.B., Rousseau, J. 2008. Quantitative risk assessment from farm to fork and beyond: a global Bayesian approach concerning food-borne diseases. *Risk Analysis* 28, 557-571.

Anderson, E.L., Hattis, D. 1999. Uncertainty and Variability. *Risk Analysis* 19, 47-49.

Anderson, R.A., Friesen, W.T. 1974. The thermal resistance of *Bacillus stearothermophilus* spores. The effect of temperature and pH of the heating medium. *Pharmaceutica Acta Helvetiae* 49, 295-298.

Ando, T. 2007. Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. *Biometrika* 94, 443-458.

André, S. 2011. Unpublished data obtained in the framework of the Ribenut project. CTCPA, Avignon, France.

REFERENCES BIBLIOGRAPHIQUES

- André, S., Hédin, S., Remize, F., Zuber, F. 2012a. Evaluation of peracetic acid sanitizers efficiency against spores isolated from spoiled cans in suspension and on stainless steel surfaces. *Journal of Food Protection* 75, 371-375.
- André, S., LeBa, D., Zuber, F. 2009. Thermal resistance of *Geobacillus stearothermophilus* spores in different food products: green peas, green beans and sweet corn. Poster at the International Congress Spore-Forming Bacteria in Food, June 15-17, Quimper, France.
- André, S., Zuber, F., Remize, F. 2012b. Thermophilic spore-forming bacteria isolated from spoiled canned food and their heat resistance. Results of a French ten-year survey. *International Journal of Food Microbiology*. Submitted.
- ANIA. 2009. Recommandations pratiques sur l'étiquetage nutritionnel. In: Association Nationale des Industries Alimentaires, (Ed.).
- ANSES. 2011. Fiche de description de danger biologique transmissible par les aliments. *Bacillus cereus*., Disponible à <http://www.anses.fr/Documents/MIC-Fi-FicheBacillusCereus.pdf>.
- Antonijevic, B., Matthys, C., Sioen, I., Bilau, M., Van Camp, J., Willems, J.L., De Henauw, S. 2007. Simulated impact of a fish based shift in the population n-3 fatty acids intake on exposure to dioxins and dioxin-like compounds. *Food and Chemical Toxicology* 45, 2279-2286.
- Arroqui, C., Rumsey, T.R., Lopez, A., Virseda, P. 2002. Losses by diffusion of ascorbic acid during recycled water blanching of potato tissue. *Journal of Food Engineering* 52, 25-30.
- Arroqui, C., Rumsey, T.R., Virseda, P., Lopez, A. 2001. Effect of different soluble solids in the water on the ascorbic acid losses during water blanching of potato tissue. *Journal of Food Engineering* 47, 123-126.
- Augustin, J.-C., Rosso, L., Carlier, V. 1999. Estimation of temperature dependent growth rate and lag time of *Listeria monocytogenes* by optical density measurements. *Journal of Microbiological Methods* 38, 137-146.
- Barker, G.C., Gómez-Tomé, N. 2011. A Risk Assessment Model for enterotoxigenic *Staphylococcus aureus* in pasteurized milk: A potential route to source-level inference. *Risk Analysis* doi: 10.1111/j.1539-6924.2011.01667.x.
- Barker, G.C., Malakar, P.K., Peck, M.W. 2005. Germination and growth from spores: variability and uncertainty in the assessment of food borne hazards. *International Journal of Food Microbiology* 100, 67-76.
- Beaumont, M.A., Zhang, W., Balding, D.J. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162, 2025-2035.
- Bigelow, W.D. 1921. The logarithmic nature of thermal death time curves. *Journal of Infectious Diseases* 29, 528-536.
- Bos, P.M., Boon, P.E., van der Voet, H., Janer, G., Piersma, A.H., Bruschweiler, B.J., Nielsen, E., Slob, W. 2009. A semi-quantitative model for risk appreciation and risk weighing. *Food and Chemical Toxicology* 47, 2941-2950.
- Brooks, S.P., Gelman, A. 1998. Alternative methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7, 434-455.

REFERENCES BIBLIOGRAPHIQUES

- Burgess, S.A., Lindsay, D., Flint, S.H. 2010. Thermophilic bacilli and their importance in dairy processing. *International Journal of Food Microbiology* 144, 215-225.
- Busschaert, P., Geeraerd, A.H., Uyttendaele, M., Van Impe, J.F. 2011. Sensitivity analysis of a two-dimensional quantitative microbiological risk assessment: keeping variability and uncertainty separated. *Risk Analysis* 31, 1295-1307.
- Carlin, F., Albagnac, C., Rida, A., Guinebretiere, M.H., Couvert, O., Nguyen-The, C. 2013. Variation of cardinal growth parameters according to phylogenetic affiliation in the *Bacillus cereus* Group and consequences for risk assessment. *Food Microbiology* 33, 69-76.
- Carlin, F., Pielaat, A.M., Heisterkamp, S., Shaheen, R., Salonen, M., Svensson, B., Nguyen-the, C., Ehling-Schulz, M. 2006. Emetic toxin-producing strains of *Bacillus cereus* show distinct characteristics within the *Bacillus cereus* group. *International Journal of Food Microbiology* 109, 132–138.
- Choma, C., Guinebretiere, M.H., Carlin, F., Schmitt, P., Velge, P., Granum, P.E., Nguyen-The, C. 2000. Prevalence, characterization and growth of *Bacillus cereus* in commercial cooked chilled foods containing vegetables. *Journal of Applied Microbiology* 88, 617-625.
- Codex Alimentarius Commission. 1999. Principles and Guidelines for the Conduct of Microbial Risk Assessment, CAC/GL-30. FAO edition, Rome, Available at www.codexalimentarius.org/input/download/standards/357/CXG_030e.pdf.
- Codex alimentarius Commission. 2003. Principles and guidelines for the conduct of microbiological risk assessment, Roma : FAO.
- Cohen, J.T., Bellinger, D.C., Connor, W.E., Kris-Etherton, P.M., Lawrence, R.S., Savitz, D.A., Shaywitz, B.A., Teutsch, S.M., Gray, G.M. 2005. A quantitative risk-benefit analysis of changes in population fish consumption. *American Journal of Preventive Medicine* 29, 325-334.
- Commission des Communautés Européennes. 2008. Directive 2008/100/CE de la Commission du 28 Octobre 2008 relative à l'étiquetage nutritionnel des denrées alimentaires, Journal officiel de l'Union européenne, Available at <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:285:0009:0012:FR:PDF>.
- Cook, A.M., Gilbert, R.J. 1968a. Factors affecting the heat resistance of *Bacillus stearothermophilus* spores. I. The effect of recovery conditions on colony count of unheated and heated spores. *International Journal of Food Science & Technology* 3, 285-293.
- Cook, A.M., Gilbert, R.J. 1968b. Factors affecting the heat resistance of *Bacillus stearothermophilus* spores. II. The effect of sporulation conditions and nature of the heating medium. *International Journal of Food Science & Technology* 3, 295-302.
- Corradini, M.G., Peleg, M. 2006. Prediction of vitamins loss during non-isothermal heat processes and storage with non-linear kinetic models. *Trends in Food Science & Technology* 17, 24-34.
- Couvert, O. 2002. Unpublished data. Université de Bretagne occidentale, Quimper, France.
- Couvert, O., Gaillard, S., Savy, N., Mafart, P., Leguerinel, I. 2005. Survival curves of heated bacterial spores: effect of environmental factors on Weibull parameters. *International Journal of Food Microbiology* 101, 73-81.

REFERENCES BIBLIOGRAPHIQUES

- Cox Jr., L.A. 2005. Potential human health benefits of antibiotics used in food animals: a case study of virginiamycin. *Environment International* 31, 549–563.
- Cox Jr., L.A., Popken, D.A. 2004. Quantifying human health risks from virginiamycin used in chickens. *Risk Analysis* 24, 271–288.
- Crank, J. 1975. *The mathematics of diffusion*, Second edition, New York.
- Crépet, A., Stahl, V., Carlin, F. 2009. Development of a hierarchical Bayesian model to estimate the growth parameters of *Listeria monocytogenes* in minimally processed fresh leafy salads. *International Journal of Food Microbiology* 131, 112-119.
- Csilléry, K., Blum, M.G., Gaggiotti, O.E., François, O. 2010. Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution* 25, 410-418.
- CTCPA. 2010. Fiche technique sur *Geobacillus stearothermophilus*, CTCPA, France. Disponible à http://ogi.ctcpa.eu/media/fck/file/ctcpa_fiche_Geobacillus_stearothermophilus.pdf.
- Cullen, A.C., Frey, H.C. 1999. *Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs* Plenum Press, New York.
- Davies, F.L., Underwood, H.M., Perkin, A.G., Burton, H. 1977. Thermal death kinetics of *Bacillus stearothermophilus* spores at ultra high temperatures. I. Laboratory determination of temperature coefficients. *Journal of Food Technology* 12, 115-129.
- Dawid, A.P., Lauritzen, S.L. 2000. Compatible prior distributions. *ISBA Proceedings*.
- De Sarrau, B., Clavel, T., Clerté, C., Carlin, F., Giniès, C., Nguyen-the, C. 2012. Influence of anaerobiosis and low temperature on *Bacillus cereus* growth, metabolism, and membrane properties. *Applied and Environmental Microbiology* 78, 1715-1723.
- 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. *International Journal of Food Microbiology* 106, 195–208.
- Delignette-Muller, M.L., Rosso, L. 2000. Biological variability and exposure assessment. *International Journal of Food Microbiology* 58, 203-212.
- Den Besten, H., Zwietering, M. 2012. Meta-analysis for quantitative microbiological risk assessments and benchmarking data. *Trends in Food Science & Technology* 25, 34-39.
- Dewettinck, K., Anthierens, K., Verbeken, D., Van Camp, J., Huyghebaert, A. 2006. Etude bibliographique sur les bénéfices nutritionnels des légumes prêts à l'emploi. , Université de Gand. Available at www.legumes-infos.com.
- Domingo, J.L., Bocio, A., Falcó, G., Llobet, J.M. 2007. Benefits and risks of fish consumption Part I. A quantitative analysis of the intake of omega-3 fatty acids and chemical contaminants. *Toxicology* 230, 219-226.
- Dragsted, L.O., Strube, M., Larsen, J.C. 1993. Cancer-protective factors in fruits and vegetables: biochemical and biological background. *Pharmacology & Toxicology* 72 (Suppl. 1), 116-135.

REFERENCES BIBLIOGRAPHIQUES

- EFSA. 2006. Risk-benefit analysis of foods. Methods and approaches. In: Report., E.S.C.S., (Ed.), Tabiano, Italy.
- EFSA. 2010. Guidance on human health risk-benefit assessment of foods. In: EFSA Scientific Committee, (Ed.), vol. 8. EFSA Journal, Parma, Italy. 1673.
- Eison-Perchonok, M.H., Downes, T.W. 1982. Kinetics of Ascorbic Acid autoxidation as a function of dissolved oxygen concentration and temperature. *Journal of Food Science* 47, 765-767.
- Ellouze, M., Gauchi, J.-P., Augustin, J.-C. 2010. Global sensitivity analysis applied to a contamination assessment model of *Listeria monocytogenes* in cold smoked salmon at consumption. *Risk Analysis* 30, 841-852.
- Etoa, F.X., Michiels, L. 1988. Heat-induced resistance of *Bacillus stearothermophilus* spores. *Letters in Applied Microbiology* 6, 43-45.
- European Commission. 2003. Risk Assessment of Food-Borne Bacterial Pathogens: Quantitative Methodology Relevant for Human Exposure Assessment, Available at: http://ec.europa.eu/food/fs/sc/ssc/out308_en.pdf.
- FAO/WHO. 2002. Principles and guidelines for incorporating microbiological risk assessment in the development of food safety standards, guidelines and related texts, Report of a Joint FAO/WHO Consultation Kiel. Available at <ftp://ftp.fao.org/docrep/fao/006/y4302e/y4302e00.pdf>, Germany, 18–22 March 2002.
- FAO/WHO. 2006. Food safety risk analysis. A guide for national food safety authorities, FAO Food and Nutrition, FAO/WHO, Roma. Available at http://www.fao.org/icatalog/search/dett.asp?aries_id=108051.
- Favell, J.D. 1998. A comparison of the vitamin C content of fresh and frozen vegetables. *Food Chemistry* 62, 59-64.
- Feeherry, F., Munsey, D.T., Rowley, D.B. 1987. Thermal inactivation and injury of *Bacillus stearothermophilus* spores. *Applied and Environmental Microbiology* 53, 365-370.
- Fernandez, P.S., Gomez, F.J., Ocio, M.J., Rodrigo, M., Sanchez, T., Martinez, A. 1995. D Values of *Bacillus stearothermophilus* spores as a function of pH and recovery medium acidulant. *Journal of food protection* 58, 628-632.
- Fernandez, P.S., Ocio, M.J., Sanchez, T., Martinez, A. 1994. Thermal resistance of *Bacillus stearothermophilus* spores heated in acidified mushroom extract. *Journal of Food Protection* 57, 37-41.
- Fernandez, P.S., Peck, M.W. 1997. Predictive model describing the effect of prolonged heating at 70 to 80 °C and incubation at refrigeration temperatures on growth and toxigenesis by nonproteolytic *Clostridium botulinum*. *Journal of Food Protection* 60, 1064-1071.
- Food and Nutrition Board, I.o.M. 2000. Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium, and Carotenoids National Academy Press, Washington, D.C.
- Foran, J.A., Good, D.H., Carpenter, D.O., Hamilton, M.C., Knuth, B.A., Schwager, S.J. 2005. Quantitative analysis of the benefits and risks of consuming farmed and wild salmon. *Journal of Nutrition* 135, 2639-2643.

REFERENCES BIBLIOGRAPHIQUES

- Frey, H.C. 1992. Quantitative Analysis of Uncertainty and Variability in Environmental Policy Making. In: EPA, A.U.S., (Ed.), Washington, DC. Available at http://www4.ncsu.edu/~frey/reports/frey_92.pdf.
- Frey, H.C., Burmaster, D.E. 1999. Methods for characterizing variability and uncertainty: Comparison of bootstrap simulation and likelihood-based approaches. *Risk Analysis* 19, 109–130.
- Frey, H.C., Mokhtari, A., Danish, T. 2003. Evaluation of selected sensitivity analysis methods based upon applications to two food safety process risk models. In: Office of Risk Assessment and Cost-Benefit Analysis, U.S.D.A., (Ed.), Washington, DC.
- Frey, H.C., Patil, S.R. 2002. Identification and review of sensitivity analysis methods. *Risk Analysis* 22, 553-578.
- Garrote, R., Silva, E.R., Bertone, R.A. 1986. Losses by diffusion of ascorbic acid during water blanching of potato tissue. *Lebensm-Wiss Technol* 19, 263-265.
- Garrote, R.L., Silva, E.R., Bertone, R.A. 1988. Effect of freezing on diffusion of ascorbic acid during water heating of potato tissue. *Journal of Food Science* 53, 473-487.
- Gauthier, C.A., Smith, G.M., Pflug, I.J. 1978. Effect of phosphate buffer concentration on heat-resistance of *Bacillus stearothermophilus* suspended in parenteral solutions. *Applied and Environmental Microbiology* 36, 457-464.
- Gelman, A., Carlin, J., Stern, H., Rubin, D. 2003. *Bayesian Data Analysis* Chapman & Hall/CRC, Boca Raton.
- Gil, M.I., Aguayo, E., Kader, A.A. 2006. Quality changes and nutrient retention in fresh-cut versus whole fruits during storage. *Journal of Agricultural and Food Chemistry* 54, 4284-4296.
- Gneiting, T., Raftery, A. 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* 102, 359-378.
- Gold, M.R., Stevenson, D., Fryback, D.G. 2002. HALYS and QALYS and DALYS, Oh My: similarities and differences in summary measures of population Health. *Annual Review of Public Health* 23, 115–134.
- Gonzales Barron, U., Soumpasis, I., Butler, F., S., D., Prendergast, D., Duffy, G. 2009. Estimation of prevalence of *Salmonella* on pig carcasses and pork joints, using a quantitative risk assessment model aided by meta-analysis. *Journal of Food Protection* 72, 274-285.
- Green, M.J., Medley, G.F., Browne, W.J. 2009. Use of posterior predictive assessments to evaluate model fit in multilevel logistic regression. *Veterinary Research* 40.
- Grothendieck, G. 2012. Package R nls2 : Non-linear regression with brute force, Available at <http://cran.r-project.org/web/packages/nls2/index.html>.
- Gschlößl, S., Czado, C. 2005. Spatial modelling of claim frequency and claim size in non-life insurance. Discussion Paper 461, Ludwig-Maximilians-Universität, Munich, Germany, Sonderforschungsbereich 368.
- Guevel, M.R., Sirot, V., Volatier, J.L., Leblanc, J.C. 2008. A risk-benefit analysis of French high fish consumption: a QALY approach. *Risk Analysis* 28, 37-48.

REFERENCES BIBLIOGRAPHIQUES

Guinebretiere, M.H., Girardin, H., Dargaignaratz, C., Carlin, F., Nguyen-The, C. 2003. Contamination flows of *Bacillus cereus* and spore-forming aerobic bacteria in a cooked, pasteurized and chilled zucchini puree processing line. *International Journal of Food Microbiology* 82, 223-232.

Guinebretière, M.H., Nguyen-the, C. 2003. Sources of *Bacillus cereus* contamination in a pasteurized courgette purée processing line, differentiated by two PCR-based methods. *FEMS Microbiology Ecology* 43, 207–215.

Guinebretiere, M.H., Thompson, F.L., Sorokin, A., Normand, P., Dawyndt, P., Ehling-Schulz, M., Svensson, B., Sanchis, V., Nguyen-The, C., Heyndrickx, M., De Vos, P. 2008. Ecological diversification in the *Bacillus cereus* Group. *Environmental Microbiology* 10, 851-865.

Haas, J., Behnsilian, D., Schubert, H. 1996. Determination of the heat resistance of bacterial spores by the capillary tube method. II — Kinetic parameters of *Bacillus stearothermophilus* spores. *Lebensmittel-Wissenschaft & Technologie* 29, 299-303.

Halder, A., Black, G., Davidson, M., Datta, A. 2010. Development of associations and kinetic models for microbiological data to be used in comprehensive food safety prediction software. *Journal of Food Science* 75, 107-120.

Havelaar, A.H., De Hollander, A.E., Teunis, P.F., Evers, E.G., Van Kranen, H.J., Versteegh, J.F., Van Koten, J.E., Slob, W. 2000. Balancing the risks and benefits of drinking water disinfection: disability adjusted life-years on the scale. *Environmental Health Perspectives* 108, 315–321.

Havelaar, A.H., Mangen, M.J., de Koeijer, A.A., Bogaardt, M.J., Evers, E.G., Jacobs-Reitsma, W.F., van Pelt, W., Wagenaar, J.A., de Wit, G.A., van der Zee, H., Nauta, M.J. 2007. Effectiveness and efficiency of controlling *Campylobacter* on broiler chicken meat. *Risk Analysis* 27, 831-844.

Heinrich, H.T.M., Bremer, P.J., McQuillan, A.J., Daughney, C.J. 2008. Modelling of the acid-base properties of two thermophilic bacteria at different growth times. *Geochimica Et Cosmochimica Acta* 72, 4185-4200.

Heldman, D., Lund, D. 2007. *Handbook of Food Engineering*, second edition, Boca Raton.

Howard, L.A., Wong, A.D., Perry, A.K., Klein, B.P. 1999. β -Carotene and Ascorbic Acid retention in fresh and processed vegetables. *Journal of Food Science* 64, 929-936.

Huemer, I.A., Klijn, N., Vogelsang, H.W.J., Langeveld, L.P.M. 1998. Thermal death kinetics of spores of *Bacillus sporothermodurans* isolated from UHT milk. *Int. Dairy J* 8, 851-855.

Iman, R.L., Conover, W.J. 1982. A distribution-free approach to inducing rank correlation among input variables. *Communication in Statistics B11*, 311-334.

Jaloustre, S., Cornu, M., Morelli, E., Noël, V., Delignette-Muller, M.L. 2011. Bayesian modeling of *Clostridium perfringens* growth in beef-in-sauce products. *Food Microbiology* 28.

Jaloustre, S., Guillier, L., Morelli, E., Noël, V., Delignette-Muller, M.-L. 2012. Modeling of *Clostridium perfringens* vegetative cell inactivation in beef-in-sauce products: A meta-analysis using mixed linear models. *International Journal of Food Microbiology* 154, 44-51.

Jensen, F.V. 2001. *Bayesian networks and decision graphs* Springer-Verlag, New York.

REFERENCES BIBLIOGRAPHIQUES

- Karhan, M., Aksu, M., Tetik, N., Turhan, I. 2004. Kinetic modeling of anaerobic thermal degradation of ascorbic acid in rose hip (*Rosa Canina* L.) pulp. *Journal of food Quality* 27, 311-319.
- Kincal, N.S., Kaymak, F. 1990. Simultaneous diffusion and degradation of ascorbic acid in potato blanching, London: Elsevier Applied Science.
- Kozempel, M.F., Sullivan, J.F., Monica, E.S.D., Egoville, E.A., Talley, M.J., Jones, W.J. 1982. Application of leaching model to describe potato nutrient losses in hot water blanching. *Journal of Food Science* 47, 1519-1523.
- Le Jean, G., Abraham, G., Debray, E., Candau, Y., Piar, G. 1994. Kinetics of thermal destruction of *Bacillus stearothermophilus* spores using a two reaction model. *Food Microbiology* 11, 229-241.
- Leguerinel, I., Spegagne, I., Couvert, O., Gaillard, S., Mafart, P. 2005. Validation of an overall model describing the effect of three environmental factors on the apparent D-value of *Bacillus cereus* spores. *International Journal of Food Microbiology* 100, 223-229.
- Leskova, Kubikova, Kovacikova, Kosicka, Porubska, Holcikova. 2006. Vitamin losses: Retention during heat treatment and continual changes expressed by mathematical models. *Journal of Food Composition and Analysis* 19, 252-276.
- Lin, S.H., Agalloco, J. 1979. Degradation kinetics of Ascorbic Acid. *Process Biochemistry* 14, 22-26.
- Llaudes, M.K., Zhao, L., Duffy, S., Schaffner, D.W. 2001. Simulation and modelling of the effect of small inoculum size on time to spoilage by *Bacillus stearothermophilus*. *Food Microbiology* 18, 395-405.
- Logan, N.A., De Vos, P. 2009. Genus VII. *Geobacillus* Nazina et al. 2001, 442AL. In: De Vos, P., Garrity, G.M., Jones, D., Krieg, N.R., Ludwig, W., Rainey, F.A., Schleifer, K.H., Whitman, W.B., (Ed.), *Bergey's Manual of Systematic Bacteriology*, vol. Three. The Firmicutes, Second ed. Springer, Dordrecht. 21-128.
- Lopez, M., Gonzalez, I., Condon, S., A., B. 1996a. Effect of pH heating medium on the thermal resistance of *Bacillus stearothermophilus* spores. *International Journal of Food Microbiology* 28, 405-410.
- Lopez, M., Gonzalez, I., Mazas, M., Gonzalez, J., Martin, R., Bernardo, A. 1997. Influence of recovery conditions on apparent heat resistance of *Bacillus stearothermophilus* spores. *International Journal of Food Science and Technology* 32, 305-311.
- Lopez, M., Martinez, S., Gonzalez, J., Martin, R., Bernardo, A. 1998. Sensitization of thermally injured spores of *Bacillus stearothermophilus* to sodium benzoate and potassium sorbate. *Letters in Applied Microbiology* 27, 331-335.
- Lopez, M., Mazas, M., Gonzalez, I., Gonzalez, J., Bernardo, A. 1996b. Thermal resistance of *Bacillus stearothermophilus* spores in different heating systems containing some approved food additives. *Letters in Applied Microbiology* 23, 187-191.
- Lunn, D.J., Spiegelhalter, D., Thomas, A., Best, N. 2009. The BUGS project: Evolution, critique, and future directions. *Statistics in Medicine* 28, 3049-3067.
- Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D. 2000. WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 10, 325-337.

REFERENCES BIBLIOGRAPHIQUES

- Magnússon, S.H., Gunnlaugsdóttir, H., Loveren, H., Holm, F., Kalogeras, N., Leino, O., Luteijn, J.M., Odekerken, G., Pohjola, M.V., Tijhuis, M.J., Tuomisto, J.T., Ueland, Ø., White, B.C., Verhagen, H. 2012. State of the art in benefit-risk analysis: food microbiology. *Food and Chemical Toxicology* 50, 33-39.
- Mallidis, C.G., Scholefield, J. 1986. Evaluation of recovery media for heated spores of *Bacillus stearothermophilus*. *Journal of Applied Bacteriology* 61, 517-523.
- Manso, M., Oliveira, F., Oliveira, J., Frias, J. 2001. Modelling ascorbic acid thermal degradation and browning in orange juice under aerobic conditions. *International Journal of Food Science & Technology* 36, 303-312.
- Marckmann, P., Grønbaek, M. 1999. Fish consumption and coronary heart disease mortality. A systematic review of prospective cohort studies. *European Journal of Clinical Nutrition* 53, 585-590.
- Marjoram, P., Molitor, J., Plagnol, V., Tavaré, S. 2003. Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences* 100, 15324-15328.
- Marshall, E.C., Spiegelhalter, D.J. 2003. Approximate cross-validated predictive checks in disease mapping models. *Statistics in Medicine* 22, 1649-1660.
- Mataragas, M., Zwietering, M.H., Skandamis, P.N., Drosinos, E.H. 2010. Quantitative microbiological risk assessment as a tool to obtain useful information for risk managers — Specific application to *Listeria monocytogenes* and ready-to-eat meat products. *International Journal of Food Microbiology* 141, 170-179.
- Membré, J.-M., Kan-King-Yu, D., Blackburn, C.W. 2008. Use of sensitivity analysis to aid interpretation of a probabilistic *Bacillus cereus* spore lag time model applied to heat-treated chilled foods (REPFEDs). *International Journal of Food Microbiology* 128, 28-33.
- Mertz, C., Brat, P., Caris-Veyrat, C., Gunata, Z. 2010. Characterization and thermal lability of carotenoids and vitamin C of tamarillo fruit (*Solanum betaceum* Cav.). *Food Chemistry* 119, 653-659.
- Mokhtari, A., Frey, H.C. 2005. Sensitivity analysis of a two-dimensional probabilistic risk assessment model using analysis of variance. *Risk Analysis* 25, 1511-1529.
- Muri, S.D., van der Voet, H., Boon, P.E., van Klaveren, J.D., Bruschweiler, B.J. 2009. Comparison of human health risks resulting from exposure to fungicides and mycotoxins via food. *Food and Chemical Toxicology* 47, 2963-2974.
- Murray, C. 1994. Quantifying the burden of disease: the technical basis for disability-adjusted life years. *Bulletin of World Health Organization* 72, 429-445.
- Nauta, M.J. 2000. Separation of uncertainty and variability in quantitative microbial risk assessment models. *International Journal of Food Microbiology* 57, 9-18.
- Nauta, M.J. 2001. A modular process risk model structure for quantitative microbiological risk assessment and its application in an exposure assessment of *Bacillus cereus* in a REPFED. RIVM, Bilthoven.
- Nauta, M.J. 2002. Modelling bacterial growth in quantitative microbiological risk assessment: is it possible? *International Journal of Food Microbiology* 73(2-3), 297-304.

REFERENCES BIBLIOGRAPHIQUES

- Nauta, M.J. 2005. Microbiological risk assessment models for partitioning and mixing during food handling. *International Journal of Food Microbiology* 100, 311-322.
- Ng, T.M., Schaffner, D.W. 1997. Mathematical models for the effects of pH, temperature, and sodium chloride on the growth of *Bacillus stearothermophilus* in salty carrots. *Applied and Environmental Microbiology* 63, 1237-1243.
- Ng, T.M., Viard, E., Caipo, M.L., Duffy, S., Schaffner, D.W. 2002. Expansion and validation of a predictive model for the growth of *Bacillus stearothermophilus* in military rations. *Journal of Food Science* 67, 1872-1878.
- NOR: AGRG9700991A. 1997. Arrêté du 28 mai 1997 relatif aux règles d'hygiène applicables à certains aliments et préparations alimentaires destinés à la consommation humaine. . Available on www.legifrance.gouv.fr, JORF n°126 du 1 juin 1997 page 8785
- Ocio, M.J., Fernández, P., Rodrigo, F., Martínez, A. 1996. Heat resistance of *Bacillus stearothermophilus* spores in alginate-mushroom puree mixture. *International Journal of Food Microbiology* 29, 391-395.
- Oey, I., Verlinde, P., Hendrickx, M., Loey, V. 2006. Temperature and pressure stability of L-ascorbic acid and/or 6s,5-methyltetrahydrofolic acid: A kinetic study. *European Food Research and Technology* 223, 71-77.
- Olmos Perez, A. 2003. Contribution à l'optimisation de la conduite des procédés alimentaires, Thèse de doctorat, sous la direction de Gilles Trystam. ENSIA (AgroParisTech).
- Oruña-Concha, M.J., Gonzalez-Castro, M.J., Lopez-Hernandez, J., Simal-Lozano, J. 1998. Monitoring of the vitamin C content of frozen green beans and padron peppers by HPLC. *Journal of the Science of Food and Agriculture* 76, 477-480.
- Parent, E., Bernier, J. 2007. Le raisonnement bayésien, Springer-Verlag, France.
- Patazca, E., Koutchma, T., Ramaswamy, H.S. 2006. Inactivation kinetics of *Geobacillus stearothermophilus* spores in water using high-pressure processing at elevated temperatures. *Journal of Food Science* 71, M110-M116.
- Pearl, J. 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible inference Morgan Kaufman Publishers, Inc, San Mateo, CA.
- Peleg, M., Cole, M.B. 1998. Reinterpretation of microbial survival curves. *Critical Reviews in Food Science and Nutrition* 38, 353-380.
- Pénicaud, C. 2009. Etude et modélisation du couplage entre le transfert d'oxygène et les réactions d'oxydation dans les aliments au cours de leur conservation., Thèse de Doctorat, Université Montpellier II.
- Periago, P.M., Fernandez, P.S., Ocio, M.J., Martinez, A. 1998a. A predictive model to describe sensitization of heat-treated *Bacillus stearothermophilus* spores to NaCl. *Zeitschrift fur Lebensmittel-Untersuchung und -Forschung. A* 206, 58-62.
- Periago, P.M., Fernandez, P.S., Salmeron, M.C., Martinez, A. 1998b. Predictive model to describe the combined effect of pH and NaCl on apparent heat resistance of *Bacillus stearothermophilus*. *International Journal of Food Microbiology* 44, 21-30.

REFERENCES BIBLIOGRAPHIQUES

- Periago, P.M., Leontidis, S., Fernandez, P.S., Rodrigo, C., Martinez, A. 1998c. Note. Kinetic parameters of *Bacillus stearothermophilus* spores under isothermal and non-isothermal heating conditions. *Food Science and Technology International* 4, 443-447.
- Plummer, M. 2007. JAGS (Just Another Gibbs Sampler) 2.1.0 ed, Available at <http://mcmc-jags.sourceforge.net>.
- Plummer, M. 2008. Penalized loss functions for Bayesian model comparison. *Biostatistics* 9, 523–539.
- 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*. *International Journal of Food Microbiology* 81, 87–104.
- Pouillot, R., Delignette-Muller, M.L. 2010. Evaluating variability and uncertainty separately in microbial quantitative risk assessment using two R packages. *International Journal of Food Microbiology* 142, 330-340.
- 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 Analysis* 27, 683-700.
- Prevost, S., Andre, S., Remize, F. 2010. PCR detection of thermophilic spore-forming bacteria involved in canned food spoilage. *Current Microbiology* 61, 525-533.
- R Development Core Team. 2010. R: a Language and Environment for Statistical Computing (Version 2.12.1), R Foundation for Statistical Computing, Vienna, Austria. Available at <http://cran.r-project.org/>.
- Renwick, A.G., Flynn, A., Fletcher, R.J., Müller, D.J., Tuijelaars, S., Verhagen, H. 2004. Risk-benefit analysis of micronutrients. *Food and Chemical Toxicology* 42, 1903-1922.
- Rigaux, C., Denis, J.B., Albert, I., Carlin, F. 2013. A meta-analysis accounting for sources of variability to estimate heat resistance reference parameters of bacteria using hierarchical Bayesian modeling: estimation of D at 121.1°C and pH 7, z_T and z_{pH} of *Geobacillus stearothermophilus*. *International Journal of Food Microbiology* 161, 112-120.
- Rimbaud, L., Heraud, F., La Vieille, S., Leblanc, J.C., Crepet, A. 2010. Quantitative risk assessment relating to adventitious presence of allergens in food: a probabilistic model applied to peanut in chocolate. *Risk Analysis* 30, 7-19.
- Rissanen, T., Voutilainen, S., Nyssonen, K., Lakka, T.A., Salonen, J.T. 2000. Fish oil-derived fatty acids, docosahexaenoic acid and docosapentaenoic acid, and the risk of acute coronary events: the Kuopio ischaemic heart disease risk factor study. *Circulation* 102, 2677-2679.
- Robert, C. 2007. *The Bayesian Choice* Springer, New York.
- Robert, C., Casella, G. 2004. *Monte Carlo Statistical Methods* Springer-Verlag, New-York.
- Rodrigo, F., Fernandez, P.S., Rodrigo, M., Ocio, M.J., Martinez, A. 1997. Thermal resistance of *Bacillus stearothermophilus* heated at high temperatures in different substrates. *Journal of Food Protection* 60, 144-147.

REFERENCES BIBLIOGRAPHIQUES

- Rodrigo, F., Rodrigo, C., Fernandez, P.S., Rodrigo, M., Martinez, A. 1999. Effect of acidification and oil on the thermal resistance of *Bacillus stearothermophilus* spores heated in food substrate. *International Journal of Food Microbiology* 52, 197-201.
- Rosso, L., Lobry, J.R., Bajard, S., Flandrois, J.P. 1995. Convenient model to describe the combined effects of temperature and pH on microbial growth. *Applied and Environmental Microbiology* 61, 610-616.
- Rosso, L., Lobry, J.R., Flandrois, J.P. 1993. An unexpected correlation between cardinal temperature of microbial growth highlighted by a new model. *Journal of Theoretical Biology* 162, 447-463.
- Sablani, S.S., Opara, L.U., Al-Balushi, K. 2006. Influence of bruising and storage temperature on vitamin C content of tomato fruit. *Journal of Food, Agriculture & Environment* 4, 54-56.
- Saltelli, A. 2002. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications* 145, 280-297.
- Sanchez-Mata, M.C., Camara, M., Diez-Marques, C. 2003. Extending shelf-life and nutritive value of green beans (*Phaseolus vulgaris* L.), by controlled atmosphere storage: micronutrients. *Food Chemistry* 80, 317-322.
- Santillana Farakos, S., Zwietering, M. 2011. Data analysis of the inactivation of foodborne microorganisms under high hydrostatic pressure to establish global kinetic parameters and influencing factors *Journal of Food Protection* 74, 2097-2106.
- Selman, J.D. 1994. Vitamin retention during blanching of vegetables. *Food Chemistry* 49, 137-147.
- Sioen, I., De Henauw, S., Van Camp, J. 2007. Evaluation of benefits and risks related to seafood consumption. *Verh K Acad Geneesk Belg* 69, 249-289.
- Sirot, V. 2010. Une approche d'analyse risque/bénéfice de la consommation de poissons et produits de la mer, Thèse. AgroParisTech, Paris, France.
- Smid, J.H., Heres, L., Havelaar, A.H., Pielaat, A. 2012. A biotracing model of *Salmonella* in the pork production chain. *Journal of food protection* 75, 270-280.
- Smid, J.H., Swart, A.N., Havelaar, A.H., Pielaat, A. 2011. A practical framework for the construction of a biotracing model: Application to *Salmonella* in the pork slaughter chain. *Risk Analysis* 31, 1434-1450.
- Smid, J.H., Verloo, D., Barker, G.C., Havelaar, A.H. 2010. Strengths and weaknesses of Monte Carlo simulation models and Bayesian belief networks in microbial risk assessment. *International Journal of Food Microbiology* 139 Suppl 1, S57-63.
- Souci, Fachman, Kraut. 2000. *Food composition and Nutrition Table*, 6th Revised edition, Garching b.Munhen.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van der Linde, A. 2002. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* 64, 583-616.
- Spor, A., Dillmann, C., Wang, S., de Vienne, D., Sicard, D., Parent, E. 2010. Hierarchical Bayesian Modelling for *Saccharomyces cerevisiae* population dynamics. *International Journal of Food Microbiology* 142, 25-35.

REFERENCES BIBLIOGRAPHIQUES

- Tavaré, S., Balding, D.J., Griffiths, R.C., Donnelly, P. 1997. Inferring coalescence times from DNA sequence data. *Genetics* 145, 505-518.
- Tee, E.S., Young, S.I., HO, S.K., Mizura, S.S. 1988. Determination of vitamin C in fresh fruits and vegetables using the dye-titration and microfluorometric methods. *Pertanika* 11, 39-44.
- Tejedor, W., Rodrigo, M., Martinez, A. 2001. Modeling the combined effect of pH and temperature on the heat resistance of *Bacillus stearothermophilus* spores heated in a multicomponent food extract. *Journal of Food Protection* 64, 1631-1635.
- Thompson, P.J., Thames, O.A. 1967. Sporulation of *Bacillus stearothermophilus*. *Applied Microbiology* 15, 975-979.
- Tijhuis, M.J., de Jong, N., Pohjola, M.V., Gunnlaugsdóttir, H., Hendriksen, M., Hoekstra, J., Holm, F., Kalogeras, N., Leino, O., van Leeuwen, F.X., Luteijn, J.M., Magnússon, S.H., Odekerken, G., Rompelberg, C., Tuomisto, J.T., Ueland, Ø., White, B.C., Verhagen, H. 2012. State of the art in benefit-risk analysis: food and nutrition. *Food and Chemical Toxicology* 50, 2-25.
- Toledo, M., Ueda, Y., Imahori, Y., Ayaki, M. 2003. L-ascorbic acid metabolism in spinach (*Spinacia oleracea* L.) during postharvest storage in light and dark. *Postharvest Biology and Technology* 28, 47-57.
- Tosun, B.N., Yucesan, S. 2007. Influence of Home Freezing and Storage on Vitamin C contents of some vegetables. *Pakistan Journal of Nutrition* 6, 472-477.
- Tuffin, B. 2010. *La simulation de Monte Carlo*, Editions Lavoisier. Paris, France.
- Van Asselt, E.D., Zwietering, M.H. 2006. A systematic approach to determine global thermal inactivation parameters for various food pathogens. *International Journal of Food Microbiology* 107, 73-82.
- Van Boekel, M.A.J.S. 1996. Statistical aspects of kinetic modeling for food science problems. *Journal of Food Science* 61, 477-489.
- Van Boekel, M.A.J.S. 2008. Kinetic modeling of food quality: a critical review. *Comprehensive reviews in Food Science and Food Safety* 7, 144-158.
- Van der Voet, H., De Mul, A., Van Klaveren, J.D. 2007. A probabilistic model for simultaneous exposure to multiple compounds from food and its use for risk-benefit assessment. *Food and Chemical Toxicology* 45, 1496-1506.
- Van Kreijl, C.F., Knaap, A.G.A.C., Van Raaij, J.M.A. 2006. Our food, our health – Healthy diet and safe food in the Netherlands. RIVM report 270555009. In: National Institute of Public Health and the Environment, (Ed.), Bilthoven, The Netherlands.
- Van Lieverloo, J.H.M., De Roode, B.M., Fox, M.B., Zwietering, M.H., Wells-Bennik, M.H.J. 2012. Multiple regression model for thermal inactivation of *Listeria monocytogenes* in liquid food products. *Food Control* 29, 394-400.
- Van Raaij, J., Hendriksen, M., Verhagen, H. 2009. Potential for improvement of population diet through reformulation of commonly eaten foods. *Public Health Nutrition* 12, 325-330.

REFERENCES BIBLIOGRAPHIQUES

- Varzakas, T., Leach, G., Israilides, C., Arapoglou, D. 2005. Theoretical and experimental approaches towards the determination of solute effective diffusivities in foods. *Enzyme and Microbial Technology* 37, 29-41.
- Verger, P., Khalfi, N., Roy, C., Blanchemanche, S., Marette, S., Roosen, J. 2008. Balancing the risk of dioxins and polychlorinated biphenyls (PCBs) and the benefit of long-chain polyunsaturated fatty acids of the n-3 variety for French fish consumers in western coastal areas. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess* 25, 765-771.
- Verhagen, H., Vos, E., Francl, S., Heinonen, M., van Loveren, H. 2010. Status of nutrition and health claims in Europe. *Archives of Biochemistry and Biophysics* 501, 6–15.
- Vose, D. 2000. *Risk analysis: a quantitative guide*. John Wiley & Sons, Chichester, NewYork, Weinheim, Brisbane, Singapore, Toronto.
- Wallace, M.J., Nordsiden, K.L., Wolf, I.D., Thompson, D.R., Zottola, E.A. 1978. Thermal inactivation of *Clostridium-Sporogenes* PA-3679 and *Bacillus-Stearothermophilus* 1518 in low-acid home canned foods. *Journal of Food Science* 43, 1738-1740.
- Wang, H., Cao, G., Prior, R.L. 1996. Total antioxidant capacity of fruits. *Journal of Agricultural and Food Chemistry* 44, 701-705.
- Wescott, G.G., Fairchild, T.M., Foegeding, P.M. 1995. *Bacillus cereus* and *Bacillus stearothermophilus* spore inactivation in batch and continuous-flow systems. *Journal of Food Science* 60, 446-450.
- Winkler. 1888. Die bestimmung im waser gelôsten sauertoffes. . *Berichte der deutschen chemischen gesellschaft* 21, 2843-2854.
- Yoo, J.A., Hardin, M.T., Chen, X.D. 2006. The influence of milk composition on the growth of *Bacillus stearothermophilus*. *Journal of Food Engineering* 77, 96-102.
- Yuan, J., Chen, F. 1998. Degradation of Ascorbic Acid in Aqueous Solution. *Journal of Agricultural and Food Chemistry* 46, 5078-5082.
- Zhao, Y., Singh, M.K., Ogino, A., Nagatsu, M. 2010. Effects of VUV/UV radiation and oxygen radicals on low-temperature sterilization in surface-wave excited O₂ plasma. *Thin Solid Films* 518, 3590–3594.
- Zwietering, M.H., Van Gerwen, S.J.C. 2000. Sensitivity analysis in quantitative microbial risk assessment. *International Journal of Food Microbiology* 58, 213-221.
- Zwietering, M.H., Wiltzes, T., De Wit, J.C., K., V.t.R. 1992. A decision support system for prediction of the microbial spoilage in foods. *Journal of Food Protection* 55, 973-979.

Communications orales et posters

Poster, 22nd International ICFMH symposium, Food Micro 2010, August 30 to September 3, 2010, Copenhagen, Danemark. *An augmented Bayesian network to infer the fate of Bacillus cereus in a cooked, pasteurized and chilled food*. Clémence Rigaux, Sophie Ancelet, Frédéric Carlin, Christophe Nguyen-the, Isabelle Albert.

Poster, « Euro-mediterranean symposium on fruit and vegetable processing », 18-20 Avril 2011, Avignon. *Modeling risk-benefit in a food chain: nutritional benefit versus microbial spoilage risk in canned green beans*. Clémence Rigaux, Stéphane André, Catherine Renard, Isabelle Albert, Christophe Nguyen-the, Frédéric Carlin

Communication orale, séminaire Applibugs, 17 juin 2011, Paris. *MCMC ou ABC ? Bonheurs et tourments de mise en œuvre sur un cas de risque sanitaire*.

Communication orale, Quatrièmes rencontres des jeunes statisticiens, du 5 au 9 septembre 2011, Aussois. *Mise en place d'une inférence bayésienne sur un modèle complexe de risque alimentaire*

Poster, AgroStat2012, 28 février–2 mars 2012, Paris. *Modeling risk-benefit in a food chain: nutritional benefit versus microbial spoilage risk in canned green beans*. Clémence Rigaux, Stéphane André, Catherine Renard, Isabelle Albert, Christophe Nguyen-the, Frédéric Carlin

Communication orale, 7èmes Rencontres Statistiques de Rochebrune (UMR518 AgroParisTech / INRA), 3 avril 2012, Rochebrune. *Inférence d'un réseau bayésien augmenté visant à confronter un modèle complexe d'analyse quantitative du risque microbiologique et des données de vieillissement*.

Communication orale, Journée modélisation de l'UMR Genial de l'INRA, GMPA GPAN-Irstea, 24 mai 2012, Massy. *Prise en compte des incertitudes et de la variabilité par modélisation à 2 dimensions, dans des modèles d'analyse du risque microbiologique*.

Communication orale, Comité de programmation légumes du CTCPA, 11 septembre 2012, Paris. *Modélisation risque-bénéfice : risque microbiologique de non stabilité due à Geobacillus stearothermophilus et taux de vitamine C dans la conserve de haricots verts*.

Communication orale, Journées 2012 du GDR « Statistique et Santé », 20 et 21 septembre 2012, Rennes. *Méta-analyse statistique bayésienne de paramètres de résistance d'une bactérie de référence en microbiologie des aliments*

Communication orale, Journées de la Société Française de Biométrie, 8 et 9 novembre 2012, Paris. *Evaluation du risque microbiologique de non stabilité à 55°C causée par Geobacillus stearothermophilus dans la conserve de haricots verts*.

Résumé

L'objectif du travail est de mettre en place des méthodes d'évaluation des risques microbiologiques et des bénéfices nutritionnels lors de la transformation des légumes, en vue d'une analyse risque-bénéfice. Le risque (industriel) considéré est l'altération des conserves de haricots verts due à la bactérie thermophile *Geobacillus stearothermophilus*, et le bénéfice nutritionnel est la teneur en vitamine C des haricots verts appertisés. Des paramètres de référence ont d'abord été acquis, par une méta-analyse par approche bayésienne pour la partie risque. Les paramètres de résistance thermique D à 121.1°C et pH 7, z_T et z_{pH} de *G. stearothermophilus* ont ainsi été estimés respectivement à 3.3 min, 9.1°C et 4.3 unités pH en moyenne en milieu aqueux. Les modèles risque et bénéfice ont ensuite été analysés par la méthode de simulation de Monte Carlo à deux dimensions, permettant une propagation séparée de la variabilité et de l'incertitude. Les pertes en vitamine C prédites par le modèle entre la matière première et les légumes appertisés sont de 86% en moyenne, et le taux de non-stabilité à 55°C prédit est de 0.5% en moyenne, globalement en accord avec la réalité. Une analyse risque-bénéfice a ensuite visé à optimiser le bénéfice tout en gardant le risque à un niveau acceptable, par exploration de scénarios d'interventions possibles basés sur les résultats d'analyses de sensibilité. Enfin, un modèle d'évaluation des risques dus à la bactérie pathogène *Bacillus cereus* dans une purée de courgettes a été confronté à des données de contamination dans des produits incubés, par le biais d'une inférence bayésienne.

Mots clés : Statistique bayésienne, simulation de Monte-Carlo du second ordre, analyse risque-bénéfice, évaluation des risques alimentaires, évaluation du bénéfice nutritionnel, *Geobacillus stearothermophilus*, *Bacillus cereus*, Vitamine C, légumes, conserves.

Abstract

The aim of this work is to set up microbiological risk and nutritional benefit assessment methods in the transformation of vegetables, in view of a risk-benefit analysis. The considered (industrial) risk is the alteration of green bean cans due to thermophilic bacteria *Geobacillus stearothermophilus*, and the nutritional benefit is the vitamin C content in appertized green beans. Reference parameters have first been acquired, by a meta-analysis using Bayesian inference for the risk part. Thermal resistance parameters D at 121.1°C and pH 7, z_T and z_{pH} of *G. stearothermophilus* have been respectively estimated at 3.3 min, 9.1°C and 4.3 pH units on average in aqueous media. The risk and benefit models have then been analyzed by a two-dimensional Monte Carlo simulation method, allowing a separated propagation of uncertainty and variability. The vitamin C losses between fresh and appertized green beans predicted by the model are of 86% on average, and the predicted non-stability at 55°C rate is of 0.5% on average, in good accordance with reality. A risk-benefit analysis has then been carried out to optimize benefit while keeping risk at an acceptable level, by exploring possible intervention scenarios based on some sensibility analysis results. Finally, a risk analysis model involving pathogenic bacteria *Bacillus cereus* in a courgette puree has been confronted to incubated product contamination data, by means of a Bayesian inference.

Keywords: Bayesian statistics, second order Monte Carlo simulation, risk-benefit analysis, food risk assessment, nutritional benefit assessment, *Geobacillus stearothermophilus*, *Bacillus cereus*, Vitamin C, vegetables, canned food.