



HAL
open science

Comment construire une base de données en microbiologie alimentaire? Projet national Sym'previus

B Leporq, P Buche, J M Membré, J P Guyonnet

► To cite this version:

B Leporq, P Buche, J M Membré, J P Guyonnet. Comment construire une base de données en microbiologie alimentaire? Projet national Sym'previus. COMPTES RENDUS-ACADEMIE D AGRICULTURE DE FRANCE, 2003, Paris, France. hal-03766404

HAL Id: hal-03766404

<https://hal.science/hal-03766404>

Submitted on 1 Sep 2022

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.

Comment construire une base de données en microbiologie alimentaire ?

Projet national Sym'previus

B Leporq¹, P. Buche², JM Membré¹ et JP Guyonnet^{3*}

1. LGPTA-INRA, BP 39, 59651 Villeneuve d'Ascq Cédex, France.
2. UMR INA PG/INRA BIA, 16, rue Claude Bernard, 75231 Paris Cédex 5, France
3. Arilait Recherches, 42 rue de Chateaudun, 75009 Paris, France.

1 Le projet Sym'previus

Le projet Sym'previus a pour objectif de bâtir un réseau d'expertise en microbiologie alimentaire, dont la réflexion et l'analyse seront établies à partir des outils disponibles en microbiologie prévisionnelle. Cet ambitieux projet, qui a débuté en 1999 suite à la demande des industriels de l'agroalimentaire auprès des ministères, rassemble un large panel de compétences (technologie et procédés alimentaires, microbiologie, statistiques, informatiques) et peut être considéré comme une première dans le domaine de l'application de la biomathématique aux problèmes concrets de l'industrie. La version 1 de l'outil logiciel sera mise sur le marché début 2004.

1.1 La place de la microbiologie prévisionnelle dans la sécurité alimentaire

La qualité et la sécurité des aliments sont devenues un sujet croissant de préoccupation. Afin de répondre à cette attente des consommateurs, les filières agroalimentaires soutenues par les Pouvoirs Publics, et avec l'aide de la recherche, de leurs qualitatifs et de fournisseurs, ont développé des méthodes et des moyens d'analyse, de contrôle, de garantie, et d'assurance. Ainsi sont apparus depuis des années, les « guides de bonnes pratiques hygiéniques », la certification ISO 9000, la méthode HACCP.

En sécurité des aliments, le développement de flores microbiennes pathogènes, en cas de contamination à un stade ou un autre de la fabrication ou de la vie du produit demeure un des problèmes majeurs à traiter. Ce développement conditionne bien évidemment la durée pendant laquelle un produit est consommable sans danger. Pour estimer cette durée, par rapport au danger microbien, on peut bien entendu procéder à des tests en laboratoire (appelés « tests d'épreuve » ou en anglais « challenge tests ») en pratiquant des contaminations artificielles, et en simulant les conditions supposées de l'environnement futur du produit. Les résultats de ces tests, comparés aux normes de populations microbiennes admissibles, assorties d'un coefficient de sécurité, permettent de fixer les DLC (Date limite de consommation) ou DLUO (date limite d'utilisation optimale) du produit – dans la mesure où le critère « pathogène » est bien le facteur limitant. Cette méthode empirique aboutit à de bons résultats. Mais elle est lourde et coûteuse, et le fait que l'on ne puisse pas multiplier à l'infini les expérimentations, pour recouvrir l'intégralité des situations possibles, en limite la portée.

La microbiologie prévisionnelle apporte à ce problème une solution moderne, puisqu'elle permet une exploitation *optimisée* et *élargie* des données recueillies au cours des tests d'épreuve, grâce à une *modélisation* mathématique de l'évolution microbienne que l'on peut en déduire.

* Pour plus d'information, contactez JP Guyonnet (jpguyonnet@arilait.com) ou visitez le site www.inra.fr/symprevius

Comme son nom l'indique, la microbiologie prévisionnelle consiste à prévoir le comportement (croissance, survie ou mort) des micro-organismes de façon quantitative, au moyen de *représentations* ou de modèles mathématiques. Cette discipline relativement jeune implique des équipes de chercheurs pluri disciplinaires réunissant biologistes et mathématiciens.

Actuellement les modèles mathématiques développés, à partir de travaux de laboratoire, permettent de calculer les taux de destruction, de croissance, ou de survie, en prenant en compte principalement 3 paramètres : température, pH, et activité de l'eau (aw). Cependant de nombreuses équipes travaillent aujourd'hui à l'amélioration des modèles, et en particulier à la prise en compte de paramètres supplémentaires comme les inhibiteurs (chimiques, organiques...), mais aussi les interactions entre les paramètres principaux, l'état des microorganismes (stress), la compétition entre les différentes flores, et plus récemment les effets « matrices » liés à la nature propre des aliments.

Ainsi ce qui était il y a quinze ans une curiosité de laboratoire, est en train de devenir un outil utilisable sur le terrain. Et cet outil peut aussi servir dans le processus d'analyse des points critiques en HACCP, ainsi que dans la démarche d'analyse quantitative des risques (AQR). Toutes ces possibilités sont à la base du projet Sym'previus.

1.2 Les expériences analogues dans le monde

Plusieurs logiciels de microbiologie prévisionnelle sont à la disposition des professionnels de l'agroalimentaire, à divers titres. Le logiciel Food MicroModel, mis au point au Royaume Uni est assez largement commercialisé dans ce pays. Le logiciel américain Pathogen Modelling Programme (P.M.P), diffusé par la F.D.A., est quant à lui mis à disposition gratuitement sur son site (www.arserrc.gov/mfs/pathogen.htm). Enfin, actuellement en cours de développement, le projet anglais ComBase, auquel est associée l'USDA, a pour objectif de mettre en ligne gratuitement sur Internet une base de données sur la microbiologie des aliments.

Signalons aussi d'autres logiciels plus spécialisés, notamment en Europe du nord, pour la sécurisation des produits de la pêche ().

En France, le système Dyn@card, développé et utilisé aujourd'hui par Danone, est basé sur le modèle « cardinal » de prévision, étudié notamment par Laurent Rosso (1995). Ce type de modèle est considéré actuellement par les spécialistes comme donnant de bons résultats. Dyn@card présente l'intérêt d'être réellement utilisé par les praticiens pour le suivi des fabrications de l'entreprise et préfigure, pour un type de produits, ce à quoi est destiné le projet Sym'previus : donner des réponses concrètes aux questions que posent les professionnels concernant le développement des germes dans les aliments, en ajoutant à la connaissance des germes celle des procédés et des produits. Car les modèles ne suffisent pas, il faut aussi des données, les plus nombreuses, pertinentes et complètes possible, pour mettre au point ces modèles, les valider, puis les utiliser à bon escient.

1.3 La place de la base de données dans l'outil Sym'previus

La base de données dans le projet Sym'previus a donc pris une place majeure. Elle permet de rassembler des connaissances (issues de la bibliographie ou des professionnels de l'agroalimentaire) sur le couple microorganisme – aliment, indépendantes mais complémentaires aux modèles.

La construction de la base de données a entraîné trois séries de questions : 1) quel type de données est-il nécessaire de stocker ? à partir de quelle source ? sous quelle forme ?...

2) comment donner un accès rapide aux données saisies dans une base? sous quelle forme faut-il les restituer à l'utilisateur ?...

3) comment valider, analyser les données avant de les stocker ? faut-il recouper les informations ? quelle est la masse critique de données avant leur exploitation à des fins d'expertise ?

Ces trois séries de questions ont été traitées au travers d'un groupe de travail rassemblant des professionnels de l'agro-alimentaire, des chercheurs en microbiologie, en informatique et en statistique (huit équipes différentes).

2 La base de données

L'objectif a été de rassembler des connaissances sur le couple microorganisme – aliment, indépendantes des simulations et permettant d'obtenir une information complémentaire au modèle.

2.1 Les types d'information retenues dans la base

┌ Données de type **épidémiologique** (prévalence ou niveau de contamination sur aliment...), informations pour lesquelles la qualité de la description de l'aliment et des méthodes de détection/dénombrement des micro-organismes, a été retenue comme critère de sélection de l'information dans la base.

┌ **Effets des facteurs** (liés à la décontamination, la conservation, la formulation...), informations pour lesquelles la qualité de la description de l'aliment sur lequel l'étude a porté et de la méthodologie de l'étude elle-même (microbiologie, plan d'expérience, analyse de l'effet du facteur) a été retenue comme critère de sélection.

┌ **Challenge-tests** (inoculation du produit alimentaire puis suivi dans le temps du comportement du micro-organisme sur l'aliment), informations pour lesquelles la qualité de la description de l'aliment sur lequel est réalisé le challenge-test et de la méthodologie microbiologique (état physiologique du micro-organisme, choix de la souche, conditions de pré-culture...) a été retenue comme critère de sélection de l'information dans la base.

2.2 La segmentation de l'information

Les informations stockées dans la base ont été réparties en quatre niveaux :

┌ l'aliment et ses caractéristiques,

┌ le micro-organisme susceptible de contaminer l'aliment,

┌ les facteurs de variation du comportement du micro-organisme sur l'aliment,

┌ les réponses observées en fonction des facteurs étudiés.

Pour chacun de ces quatre grands niveaux, la base de données a été conçue de façon à intégrer des informations générales permettant, avec les connaissances actuelles (par ex. effet de la température sur le taux de croissance) et selon les connaissances futures (par ex. effet des faibles inoculum sur les mesures de temps de latence), de comparer les données aux modèles. De plus des informations purement qualitatives sont aussi intégrées à la base, afin de faciliter le travail d'analyse du comportement microbien.

A titre d'exemple, en ce qui concerne l'aliment, la base contient :

1- des données qualitatives :

- le nom du produit, le nom générique,
- selon la filière, le mode cultural, le mode d'élevage,
- le pays d'origine,
- la provenance du produit (champ, arrivée usine, sortie usine, grandes surfaces),
- la traçabilité du produit selon la date des opérations, géographique, de la nature de l'analyse effectuée,
- la structure du produit (solide, liquide, homogène, hétérogène...),
- le procédé : liste des opérations unitaires de fabrication-distribution,
- les caractéristiques de conditionnement, de stockage avant étude.

2- Des données quantitatives:

- les caractéristiques produit (pH, aw, conditionnement),
- les matières premières, la formulation, les ingrédients.

Les données qualitatives recueillies permettent d'affiner les comparaisons entre les comportements de micro-organismes sur matrice alimentaire, et ainsi d'ancrer notre expertise en microbiologie prévisionnelle au contexte agro-alimentaire (technologie, process).

2.3 *Autres spécificités*

D'autres contraintes liées aux sources d'informations sont venues s'ajouter aux éléments à prendre en compte dans la construction de la base de données. En effet, les informations peuvent provenir de trois origines différentes :

- La bibliographie internationale. La base de données a été conçue de façon à intégrer les références complètes des articles scientifiques.
- Les données issues des partenaires de Sym'previus, acquises avant le programme (littérature grise, auto-contrôle...). L'environnement informatique garantit l'anonymat des sources et la confidentialité des échanges d'informations.
- Les données acquises pendant le programme Sym'previus : connaissances scientifiques confortées par des travaux spécifiques sur tel ou tel couple micro-organisme – aliment ; travaux pour étendre les connaissances scientifiques dans le domaine de la sécurité microbiologie alimentaire. Dans les deux cas, l'accent est mis sur la méthodologie d'acquisition de données, afin de garantir une assise scientifique aux données stockées dans la base.

Globalement pour intégrer toutes ces informations, la base de données relationnelle créée contient 90 tables différentes.

3 Accès rapide aux données et analyse des informations

La base de données ainsi constituée, avec ses 90 tables recoupe des informations variées mais incomplètes (les informations obtenues pour chaque couple micro-organisme – aliment ne sont pas de même nature : toutes les rubriques listées dans le § 2 "la base de données" ne sont pas systématiquement complétées. Par ailleurs, l'information peut être plus ou moins précise (par ex. un même aliment peut avoir été décrit comme "viande de bœuf" ou comme "steak haché de viande de bœuf").

3.1 *L'interrogation de la base de données*

Sym'previus a donc opté pour un système d'interrogation multi-critères adapté spécifiquement à la base de données conçue et réalisée dans le cadre du programme (Buche et al. 2003). Ce système d'interrogation, appelé MIEL, est basé sur les travaux scientifiques de chercheurs en informatique, entrepris en amont du programme Sym'previus (Buche et Loiseau 1999, Buche et al. 2002, Thomopoulos et al. 2003). Il restitue à l'utilisateur les informations les plus pertinentes, vis à vis de ses critères d'interrogation. MIEL s'appuie, pour cela, sur une base de connaissances dans laquelle sont mémorisées les relations de proximité sémantique existant entre les produits (respectivement les germes et les facteurs analysés).

Le projet Sym'previus réunit des professionnels de l'agro-alimentaire, des chercheurs en microbiologie, en informatique et en statistique. Cette diversité est une richesse qui a été exploitée au sein du groupe de travail mis en place pour l'analyse des données.

La validation des données avant de les stocker dans la base, a été réalisée par lecture croisée de deux experts du domaine.

Une fois les données saisies, deux questions essentielles se sont posées :

- Quelle est la masse critique de données avant de pouvoir les exploiter à des fins d'expertise ?
- Faut-il recouper les informations pour faciliter le travail d'expertise ?

3.2 *La masse critique de données*

Nous avons jugé qu'il n'y avait pas de masse critique, une information unique pouvant être pertinente en tant que telle. Le crédit que l'expert portera à cette information unique sera relatif à la description de la méthodologie ayant abouti à l'information recherchée, d'où l'importance d'une base de données intégrant un grand nombre d'informations méthodologiques (cf. § 2 "la base de données").

3.3 *Le recoupement d'informations*

Faut-il recouper les informations pour faciliter le travail d'expertise ? **Sym'previus se veut être un logiciel d'aide à l'expertise et non pas un système expert.** Cependant, pour faciliter le travail de l'expert (le professionnel de l'agro-alimentaire, le scientifique...), nous avons décidé de rassembler les données traitant du même domaine (par ex. l'effet de la température sur le comportement d'une même espèce bactérienne en cas de contamination d'un même aliment) et de les analyser ensemble, à l'aide d'outils statistiques et plus particulièrement à l'aide des modèles prévisionnels. Les Fig. 1 et 2 illustrent ce type d'analyse.

Le graphe des données brutes (Fig. 1) permet de visualiser l'ensemble des données collectées et de mettre en évidence des divergences dans le comportement des différentes souches d'*E. coli* (ex. comportement à 4°C où la croissance est décrite par certains auteurs mais non par d'autres).

Pour chacune des courbes exploitables de la Fig. 1 (nombre de points >5), un modèle prévisionnel primaire (Rosso, 1995) est ajusté, ce qui permet d'estimer les latences et les taux de croissance d'*Escherichia coli* O157:H7 sur viande de bœuf, à des températures comprises entre 4 et 30°C.

Les taux de croissance ainsi obtenus peuvent être ensuite prédit en fonction du facteur de variation (ici la température), à l'aide de modèles prévisionnels secondaires (Zwietering et al. 1992 ; McMeekin et al. 1993 ; Rosso et al. 1995 ; Rosso and Robinson, 2001). Une marge d'erreur autour du taux de croissance est calculée (Huet et al. 1996) ; elle représente, selon les données, la variation due aux différents essais, aux souches bactériennes, aux laboratoires (en

France, à l'étranger) et aux modes opératoires pour obtenir ces cinétiques. Cette marge d'erreur autour des simulations n'est pas identique selon la température car elle reflète la variabilité des informations expérimentales observée à chaque température.

In fine, une simulation de l'effet de la température sur la croissance d'*Escherichia coli* O157 :H7 sur viande de bœuf peut être proposée (Fig. 2).

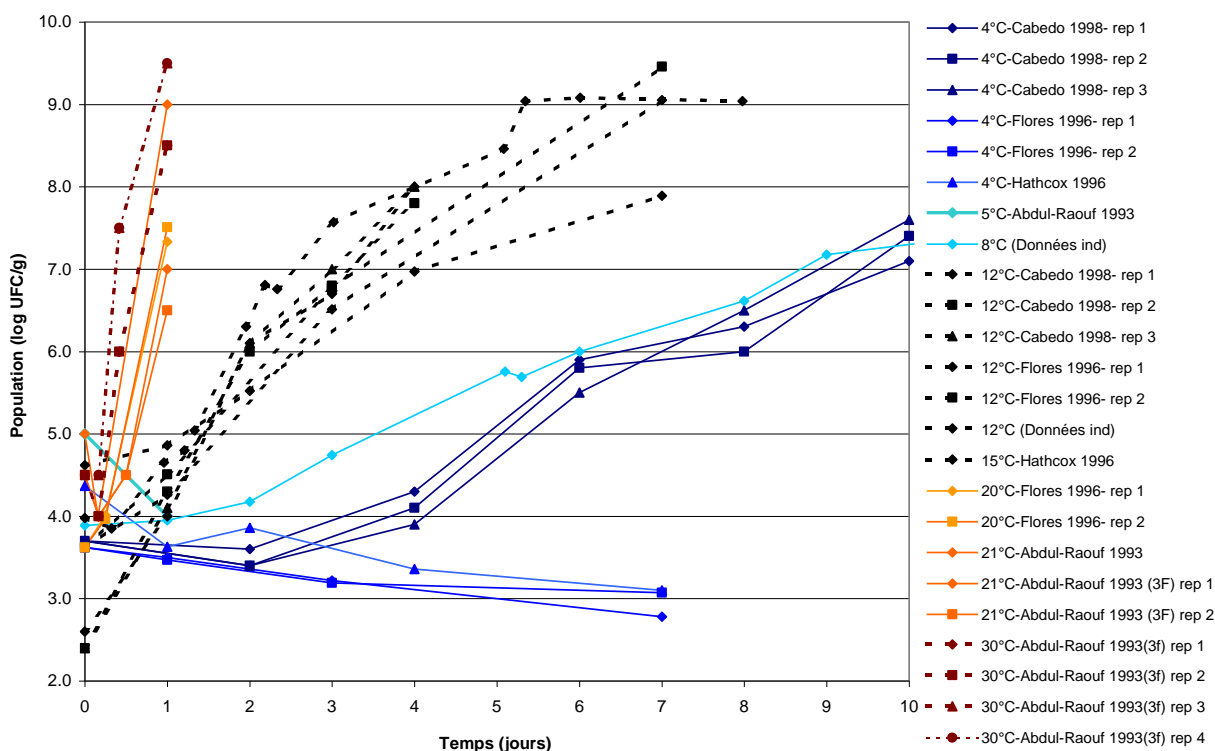


Fig. 1 : Données brutes, cinétiques de *Escherichia coli* O157 :H7 sur viande de bœuf, à des températures de 4 à 30°C, les références des sources sont portées en légende.

D'une façon générale, la valeur ajoutée de tels traitements d'informations se partage sur différents aspects :

- plus le nombre de données brutes est important (Fig. 1), plus on multiplie les possibilités de comparaison entre les cinétiques obtenues sous différentes conditions. Indépendamment même des modèles, ce recoupement d'informations permet d'avoir instantanément l'état de l'art dans le domaine (synthèse bibliographique "en ligne").
- En se fondant sur les simulations (Fig. 2), on peut se simplifier le travail de réalisation d'un challenge-test (préparation et conduite expérimentale facilitées : répartition des prélèvements, temps total de manips...).
- En se fondant sur les simulations également, on peut éventuellement s'épargner du travail expérimental : association de tests d'épreuve réels (quelques expériences "clés") et virtuels (pour couvrir toute la gamme de température par ex.).
- enfin, l'interprétation des tests d'épreuve sur le même germe et le même produit alimentaire se trouve facilitée ou relativisée : si par exemple le résultat présente un écart important par rapport aux simulations, il y a lieu de s'interroger sur sa propre expérimentation avant de prendre des décisions, car les simulations proposées sont fondées sur une réflexion scientifique (microbiologique et statistique) collective (réseau

Sym'previus) et objective (issues de données expérimentales ou de la bibliographie internationale, avec transparence des méthodes utilisées).

Cet exemple d'exploitation montre combien une base de données convenablement documentée et exploitée facilite le travail d'expertise en microbiologie alimentaire.

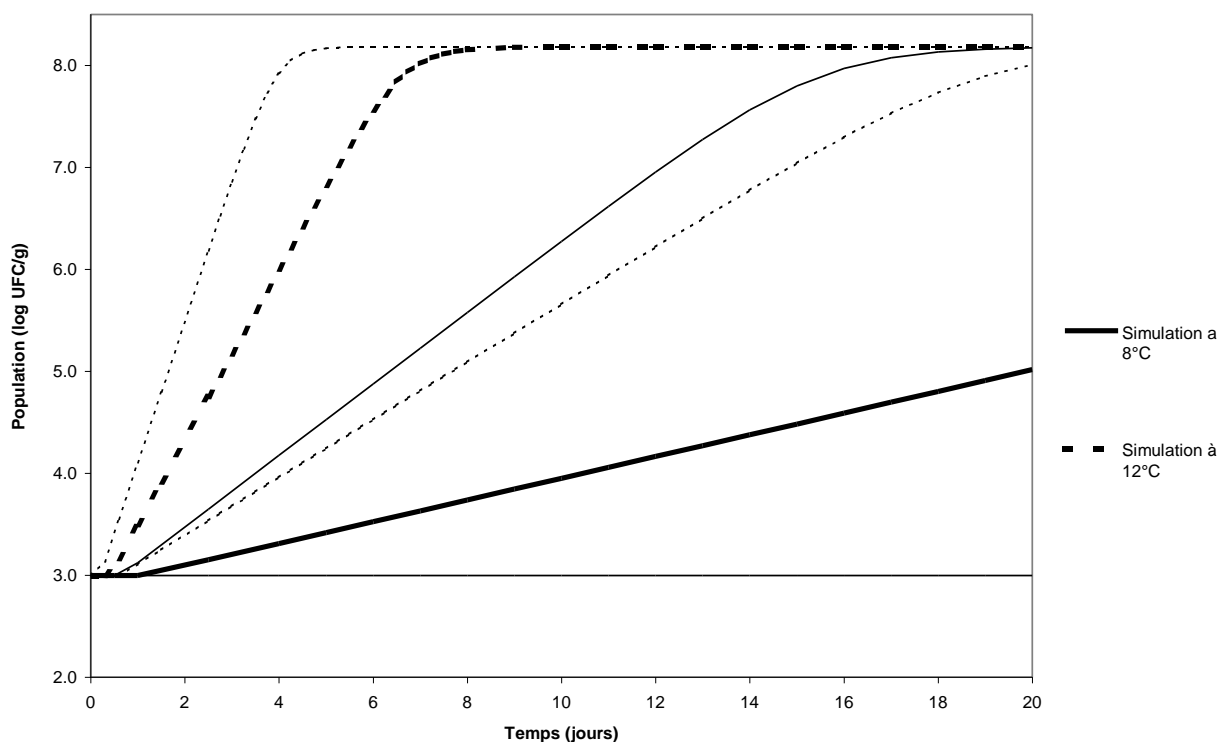


Fig. 2 : Simulation de l'effet de la température sur le comportement de Escherichia coli O157:H7 sur viande de bœuf : valeur moyenne et marge d'erreur.

4 Conclusion et perspectives

Sym'previus est un outil destiné aux industriels de l'agroalimentaire et à tous leurs partenaires, pour les aider à répondre à des questions concrètes de sécurité des produits.

Il est cependant difficile d'inclure dans une base de données tous les aliments, toutes leurs caractéristiques ainsi que tous les scénarii de contamination possibles. Les simulations quant à elles, même si elles donnent des indications précieuses, ne peuvent pas encore couvrir tous les champs de la microbiologie alimentaire. Il faut donc bien rester conscient du fait que l'utilisation d'un tel système ne peut se faire à la légère : Sym'previus donnera des indications pertinentes, certes, mais il faudra les interpréter dans leur contexte.

L'approche multi-partenaires de Sym'previus, permet une consultation d'experts en cas de besoin, ce qui est pour cela un avantage certain.

Bien entendu, tel qu'il a été pensé par ses créateurs, Sym'previus sera un outil évolutif. Il bénéficiera en effet des apports des données (anonymées) en provenance de ses utilisateurs, afin de fournir des prévisions de plus en plus « documentées ». Simultanément une réflexion sera conduite en vue de l'amélioration continue des modèles de prévision, et la prise en compte de nouveaux paramètres, et aussi de microorganismes supplémentaires. Des travaux de recherche associés au projet, spécifiquement orientés vers des besoins de données particuliers, ou une meilleure connaissance de phénomènes (notamment stress bactériens et temps de latence), contribueront à enrichir et affiner le système.

5 Références.

- Buche P. and Loiseau S, 1999, Using Contextual Fuzzy Views to Query Imprecise Data. Lecture Notes in Computer Science 1677. Proceedings DEXA'99 (Database and EXpert system Application) , pp 460-472. Florence, August 99.
- Buche P., Dervin C., Brouillaud-Delattre A. and Gnanou-Besse N. 2002. Combining fuzzy querying of imprecise data and predictive microbiology using category-based reasoning for prediction of the possible microbial spoilage in foods : application to *Listeria monocytogenes*. *International Journal of Food Microbiology* 73, 171-185.
- Buche P., Dervin C., Haemmerlé O., Surleau E. and Thomopoulos R. 2003. Carrying out the microbial risk analysis of food product using the MIEL software: a new tool to query incomplete, imprecise and heterogeneously structured experimental data stored in a relational database. Communication acceptée dans 4^{ème} *International Conference, Predictive modelling in foods*. 15-19 Juin. Quimper, France.
- Huet S., Bouvier A ., Gruet M.A. and Jolivet E. 1996. Statistical Tools for Nonlinear Regression. Springer-Verlag, New-York, USA.
- McMeekin T.A., Olley J.N., Ross T., and Ratkowsky D.A.. 1993. Predictive Microbiology: Theory and Application. Research Studies Press, Taunton, UK.
- Rosso L. 1995. Modélisation et Microbiologie Prévisionnelle : élaboration d'un nouvel outil pour l'agro-alimentaire, Thèse de doctorat, présentée devant l'Université Claude Bernard - Lyon (France)
- Rosso L., Lobry J.R., Bajard S. and 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. and Robinson T.P. 2001. A cardinal model to describe the effect of water activity on the growth of moulds. *International Journal of Food Microbiology* 63, 265-273.
- Thomopoulos R., Buche P. and Haemmerle O. 2003. Representation of weakly structured imprecise data for fuzzy querying. A paraître dans *Fuzzy sets and System*.
- Zwietering M. H., Wiltjes T., De Wit J.C. and Van't Riet K. 1992. A decision support system for prediction of the microbial spoilage in foods. *Journal of Food Protection*. 55 : 973-979.