

Identification de systèmes dynamiques microbiologiques complexes par filtrage non linéaire

Jean-Pierre Vila, Jean-Pierre Gauchi, Caroline Bidot

► **To cite this version:**

Jean-Pierre Vila, Jean-Pierre Gauchi, Caroline Bidot. Identification de systèmes dynamiques microbiologiques complexes par filtrage non linéaire. 41èmes Journées de Statistique, SFdS, Bordeaux, 2009, Bordeaux, France, France. inria-00386650

HAL Id: inria-00386650

<https://hal.inria.fr/inria-00386650>

Submitted on 22 May 2009

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.

Identification de systèmes dynamiques microbiologiques complexes par filtrage non linéaire

Jean-Pierre Vila

*UMR Analyse des Systèmes et Biométrie INRA/SupAgro
2, place P. Viala, 34060 Montpellier, France
vila@supagro.inra.fr*

Jean-Pierre Gauchi & Caroline Bidot

*UR341 Mathématiques et Informatique Appliquées INRA
Domaine de Vilvert, 78352 Jouy-en-Josas, France
Jean-Pierre.Gauchi@jouy.inra.fr ; Caroline.Bidot@jouy.inra.fr*

Résumé : L'étude statistique des systèmes dynamiques microbiens rencontrés dans la fabrication et la conservation des aliments est d'un grand intérêt pour les microbiologistes (accroissement des connaissances sur ces systèmes, analyse de risques alimentaires). La prédiction des évolutions de populations bactériennes passe par la modélisation et l'identification de tels systèmes par des techniques appropriées. Cependant, la spécificité de ces systèmes dynamiques (systèmes stochastiques à espace d'état non linéaire) dans lesquels les variables d'intérêt (effectifs) ne sont qu'estimées au travers d'un processus complexe de prélèvements, dilutions, mises en culture et comptages, compromet l'application des méthodes classiques (moindres carrés, maximum de vraisemblance). On doit alors avoir recours à des techniques plus spécialisées, par exemple de type estimations séquentielles. Dans cette communication, nous présentons les premiers résultats de l'application d'une telle approche à un système dynamique simulé bien connu des microbiologistes. Elle repose sur la mise en œuvre d'un nouveau filtre particulière aux propriétés théoriques récemment étudiées et d'un logiciel utilisateur convivial, spécialement développé.

Abstract : We consider microbiological dynamic systems met in food processing and conservation (e.g. *Listeria monocytogenes*). In order to predict the evolution of bacteria populations, for a better microbiological knowledge or other objectives as risk analysis, the modeling and identification of such stochastic dynamic systems has to be carried first, through adapted statistical approaches. However the specificities of such systems (non-linear state space systems), in which the variables of interest are only indirectly observed through a complex process of sampling, dilution series and counting, make difficult the use of classical methods such as least squares or maximum likelihood. One has then to resort to specialized approaches as sequential estimation techniques. In this paper are shown the first results of such a promising approach involving a new particular filter applied to a simulated microbial dynamics.

Key words : Nonlinear filters, convolution kernels particle filters, predictive modelling, microbiology.

1 Introduction

La complexité des systèmes microbiologiques de type alimentaire est à la fois de nature structurelle (écosystèmes impliquant plusieurs espèces bactériennes en interaction) et de nature fonctionnelle. Cette dernière résulte de la distribution de leur dynamique selon au moins trois niveaux hiérarchiques que l'on peut approximativement représenter par: a) un niveau externe, seul accessible aux mesures (comptages en milieu de culture après prélèvement et séries de dilutions), b) un niveau intermédiaire, correspondant aux dynamiques de croissances ou décroissances bactériennes proprement dites, non mesurables directement dans le substrat alimentaire considéré, mais a priori modélisables sous forme de modèles stochastiques dits primaires, c) un niveau plus profond, caractérisé par des variables cinétiques qui conditionnent les dynamiques précédentes, et elles-mêmes résultats d'interactions de différents facteurs biotiques et abiotiques (conditions de milieux telles le pH, la température, l'activité de l'eau, ..). Ces éléments cinétiques peuvent quelques fois être modélisables sous forme de modèles stochastiques dits secondaires.

Ces caractéristiques fonctionnelles rendent toute identification et modélisation prévisionnelle de ces dynamiques bactériennes par les approches classiques (*e.g.* moindres carrés non linéaires), particulièrement difficile sinon impossible. Des travaux récents [1, 2, 3] ont permis d'apporter une réponse méthodologique pertinente, aux deux problèmes de l'identification paramétrique de ces systèmes et de l'estimation des densités de probabilités prévisionnelles de l'évolution des concentrations bactériennes dans les substrats d'intérêt. Ils reposent sur la mise en œuvre d'une nouvelle technique de filtrage non linéaire particulière, à noyau de convolution.

L'objectif de cette présentation est de montrer les résultats obtenus en appliquant ce nouveau type de filtrage à un système dynamique microbiologique réel, de grand intérêt dans la communauté des microbiologistes alimentaires, à l'aide d'une nouvelle procédure conviviale écrite en langage Matlab [4]. On ne s'étendra pas ici sur les fondements théoriques de la méthode, publiée dans [1, 2, 3]. On se contentera d'en indiquer quelques éléments au paragraphe 2.2.

2 Méthodologies mises en oeuvre

2.1 Acquisition statistique des données

Les procédures actuelles d'acquisition des données sont séquentielles : à partir d'un tube primaire de solution, un prélèvement est effectué, dilué en cascade (dilutions en série de Fisher), suivi d'un étalement sur des boîtes de Petri. Après un certain temps les unités bactériennes formant colonies (UFC) sont dénombrées, sous l'hypothèse que chaque colonie correspond à une seule bactérie déposée. La pratique actuelle des microbiologistes est de déduire l'effectif des bactéries présentes dans la solution initiale du tube primaire par simple extrapolation à partir des facteurs de dilutions successives en ignorant les erreurs d'échantillonnage (lors du prélèvement), de pipetage, de dilution, et de comptages sur

les boîtes. Cette estimation déterministe très approximative est éminemment perfectible. Sous des hypothèses de répartitions spatiales à valider expérimentalement (distributions poissonniennes, distributions agrégatives) et de lois d'erreur de manipulation, nous avons effectué les caractérisations probabilistes de ces comptages (fonction des effectifs initiaux), prenant en compte des estimations des variances de toutes les erreurs successives.

2.2 Estimation de densités conditionnelles d'effectifs

Les systèmes dynamiques considérés ici s'apparentent aux systèmes à espace d'état non linéaires de forme générale

$$\begin{cases} x_{t+1} = f_t(x_t, \theta_x, \varepsilon_t) \\ y_{t+1} \sim \mathcal{L}(\cdot | x_{t+1}, \theta_y) \end{cases} \quad (1)$$

où $x_t \in \mathbb{R}^d$ est le vecteur des variables d'état, $y_t \in \mathbb{R}^q$ le vecteur des variables d'observation, $\theta = (\theta_x^T, \theta_y^T)^T \in \Theta \subset \mathbb{R}^p$ un vecteur de paramètres inconnus et constants, ε_t , un vecteur de bruits blancs et \mathcal{L} la loi de probabilité de y_{t+1} . Dans l'approche par filtrage à convolution qui est la nôtre, on introduit l'équation d'état supplémentaire $\theta_{t+1} = \theta_t$, pour associer l'estimation des densités conditionnelles des paramètres θ à celle des variables x_t . A partir de densités a priori à l'instant $t = 0$ pour le vecteur des variables et pour le vecteur des paramètres, ces estimations à l'instant t s'obtiennent par:

$$\begin{aligned} p_t^n(\theta | y_{1:t}) &= \frac{\sum_{i=1}^n K_{h_n}^y(\tilde{y}_t^i - y_t) \times K_{h_n}^\theta(\tilde{\theta}_t^i - \theta)}{\sum_{i=1}^n K_{h_n}^y(\tilde{y}_t^i - y_t)} \\ p_t^n(x | y_{1:t}) &= \frac{\sum_{i=1}^n K_{h_n}^y(\tilde{y}_t^i - y_t) \times K_{h_n}^x(\tilde{x}_t^i - x)}{\sum_{i=1}^n K_{h_n}^y(\tilde{y}_t^i - y_t)} \end{aligned}$$

où $(\tilde{x}_t^i, \tilde{\theta}_t^i, \tilde{y}_t^i), i = 1, \dots, n$, sont n particules générées à partir des densités correspondantes estimées à l'instant précédent et du modèle (1). $K_{h_n}^x, K_{h_n}^\theta$ et $K_{h_n}^y$ sont des noyaux de Parzen-Rosenblatt de dimension d, p et q respectivement. On se reportera à Rossi et Vila [2, 3] pour plus de détails sur cet algorithme de filtrage et ses conditions de convergence.

3 Applications

On présente ici le plus simple des deux systèmes microbiens étudiés. Le second sera évoqué lors de l'exposé.

- **Equation d'état** : modèle de croissance primaire de Baranyi et Roberts [5, 6]. La forme auto-régressive correspondant au système différentiel d'origine discrétisé est

$$N_{t+1} = \delta N_0 \exp(\mu_{max} A_t) \frac{1}{B_t} \left(\mu_{max} \frac{dA_t}{dt} - \frac{dB_t}{dt} \frac{1}{B_t} \right) + N_t + \varphi_t. \quad (2)$$

où :

$$A_t = t + \frac{1}{\mu_{\max}} \ln (\exp (-\mu_{\max} t) + \exp (\mu_{\max} \lambda) - \exp (-\mu_{\max} t - \mu_{\max} \lambda))$$

$$B_t = 1 + \frac{\exp (\mu_{\max} A_t) - 1}{\frac{N_{\max}}{N_0}}$$

avec: a) variable d'état: N_t , nombre de bactéries dans le milieu, à l'instant t . b) paramètres à estimer: N_0 , nombre de bactéries dans le milieu à l'instant initial; N_{\max} , nombre de bactéries maximum; λ , temps de latence des bactéries avant croissance; μ_{\max} , vitesse de croissance maximum. c) terme d'erreur: φ_t , variable de Poisson centrée $\mathcal{P}(N_t c) - (N_t c)$, où $c \in \mathbb{R}$ est une constante fixée ou à estimer et $N_t c$ le paramètre de la loi de Poisson. d) pas de discrétisation: δ .

Dans le second modèle évoqué plus haut, le paramètre μ_{\max} obéit lui-même à un modèle complexe dépendant de quatre paramètres et du facteur température [7].

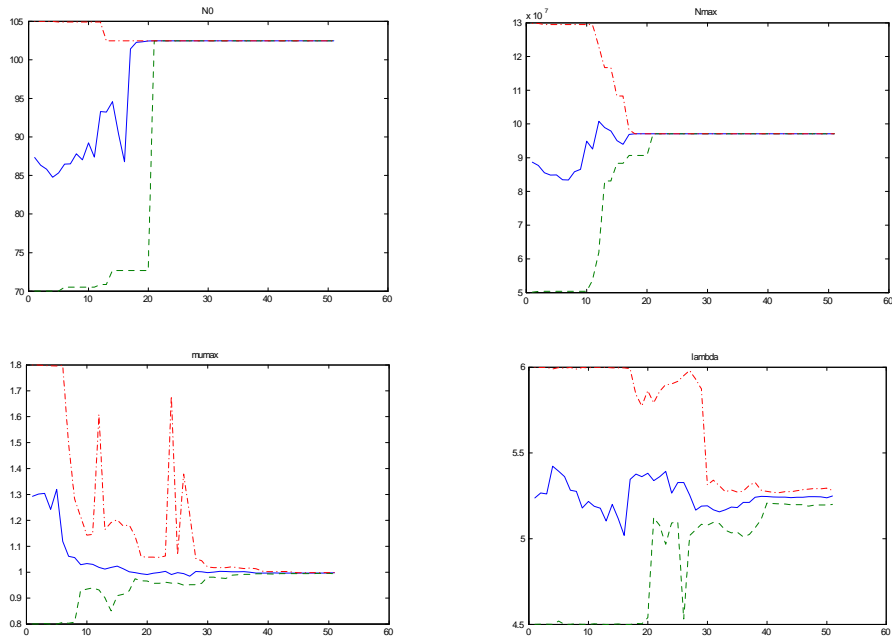
- **Modèle d'observation:** Il correspond à la loi résultant du processus de prélèvement-dilution (typiquement lois de poisson) et comptage des UFC sur les boîtes de Petri, notée $\mathcal{L}(\cdot | x_{t+1}, \theta_y)$ dans (1). Cette loi n'est pas caractérisable analytiquement mais peut être simulée, ce qui suffit à la mise en oeuvre du filtre.

4 Etude de cas par simulation

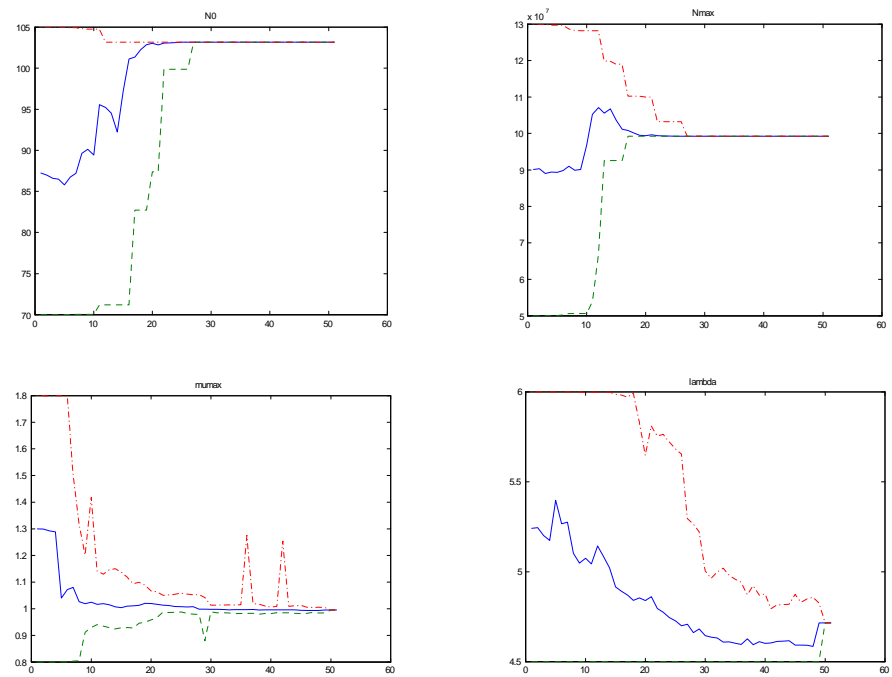
Deux filtrages du système microbien de Baranyi considéré ont été réalisés sur données simulées, le premier avec 1000 particules, le second avec 5000 (nombres volontairement réduits).

Les graphiques ci-dessous correspondent aux conditions suivantes: a) valeurs (pour les simulations) et plages de variation a priori pour les paramètres: $N_0 = [70 \leftarrow 100 \rightarrow 105]$; $N_{\max} = [5 \times 10^7 \leftarrow 10^8 \rightarrow 1.3 \times 10^8]$; $\lambda = [4.5 \leftarrow 5 \rightarrow 6]$; $\mu_{\max} = [0.8 \leftarrow 1 \rightarrow 1.8]$; les lois initiales sur ces paramètres sont des uniformes. b) nombre d'instantants d'observation correspondant à un prélèvement: 50 sur 100 heures au total correspondant à un pas de temps de 2 heures (ici une observation est faite à chaque pas de temps, mais la procédure permet de n'observer qu'à certains temps choisis a priori et non pas à chaque pas de temps), c) nombre de dilutions: 0 dilution de 0 à 10 heures, 1 dilution de 10 h à 20 h, 2 dilutions de 20 à 25 h, et 3 dilutions de 25 à 100 h, d) nombre de boîtes de Petri par prélèvement: 5, e) nombre de particules: 1000 pour le premier filtrage (première série de graphiques), et 5000 pour le second (deuxième série de graphiques). Cette saisie des informations est facilitée par une interface conviviale (qui sera présentée lors de l'exposé).

On peut noter la bonne qualité globale de ces estimations et notamment celles des paramètres μ_{\max} et N_{\max} , ainsi que l'amélioration apportée par l'augmentation du nombre des particules entre les deux filtrages.



Evolution de l'estimation de l'espérance des paramètres et de leurs intervalles de confiance, avec 1000 particules à chaque pas de temps (environ 45 secondes de calcul sur un PC-pentium IV).



Evolution avec 5000 particules (environ 250 secondes de calcul).

5 Conclusion

Cette approche d'identification par filtrage particulaire d'un système dynamique micro-biologique, constitue une démarche innovante pour la communauté des microbiologistes. Elle repose sur une prise en compte exhaustive de toutes les erreurs de manipulations et de comptage et permet une estimation rigoureuse des paramètres des modèles (ce qui n'est pas le cas avec les approches actuelles par pseudo-moindres carrés double passe). Elle devrait permettre d'améliorer de façon sensible les démarches de microbiologie prévisionnelle particulièrement utile pour une maîtrise des risques alimentaires. L'interface du logiciel développé, a été conçue pour faciliter sa mise en oeuvre par le praticien. La validation de cette approche sur données réelles est en cours actuellement.

L'étape méthodologique suivante consistera en l'optimisation des instants de prélèvements pour un coût expérimental donné, afin d'augmenter les vitesses de convergence des lois a posteriori estimées par filtrage.

Références

- [1] Rossi, V., Vila, J.P. (2003) Filtrage non linéaire en temps discret par convolution de particules. *Actes des XXXVèmes Journées de Statistique*, 823-826.
- [2] Rossi, V., Vila, J.P. (2005) Approche non paramétrique du filtrage de système non linéaire à temps discret et à paramètres inconnus. *C.R. Acad. Sci. Paris. Ser I* 340, 759-764.
- [3] Rossi, V., Vila, J.P. (2006) Nonlinear filtering in discrete time : a particle convolution approach. *Inst. Stat. Univ. Paris*, 3, 71-102.
- [4] Bidot, C., Gauchi, J.P., Vila, J.P. (2009) Programmation Matlab du filtrage non linéaire par convolution de particules pour l'identification et l'estimation d'un système dynamique micro-biologique. Rapport technique INRA/Jouy-en-Josas/MIA/ n°2009 – 1.
- [5] Baranyi, J., Roberts, T.A. (1994). A dynamic approach to predicting bacterial growth in food. *International Journal of Food Microbiology*, 23,277-294.
- [6] Baranyi, J., Roberts, T.A. (1995). Mathematics of predictive food microbiology *International Journal of Food Microbiology*, 26, 199-218.
- [7] Rosso, L., Lobry, J.R., Flandrois, J.P. (1993) An unexpected correlation between cardinal temperatures of microbial growth highlighted by a new model. *J. of Theoretical Biology*, 162, 447-463.