REVUE DE STATISTIQUE APPLIQUÉE

PHILIPPE GIRARD
ERIC PARENT

Analyse bayésienne du modèle linéaire avec erreur autocorrélée par échantillonnage de Gibbs : application à la modélisation d'un procédé agroalimentaire à partir de données recueillies sur ligne

Revue de statistique appliquée, tome 48, n° 2 (2000), p. 5-34
<http://www.numdam.org/item?id=RSA_2000__48_2_5_0>

© Société française de statistique, 2000, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (http://www.sfds.asso.fr/publicat/rsa.htm) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM
Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

Rev. Statistique Appliquée, 2000, **XLVIII** (2), 5-34

ANALYSE BAYÉSIENNE DU MODÈLE LINÉAIRE AVEC ERREUR AUTOCORRÉLÉE PAR ÉCHANTILLONNAGE DE GIBBS : APPLICATION À LA MODÉLISATION D'UN PROCÉDÉ AGROALIMENTAIRE À PARTIR DE DONNÉES RECUEILLIES SUR LIGNE

Philippe Girard*,** et Eric Parent*

** Laboratoire GRESE, ENGREF, 19 avenue du Maine 75015 Paris*

*** Nestlé, département Assurance Qualité,*
7, Bd P. Carle, BP 90, 77446 Noisiel Cedex 02

RÉSUMÉ

Le modèle linéaire à erreur autocorrélée (**MLEAR**) apparaît comme une extension naturelle de la régression lorsqu'on modélise un procédé à partir de données recueillies au cours du temps. Compte tenu de la structure de dépendance complexe entre les paramètres de ce modèle, l'analyse du **MLEAR** ne peut pas être menée sous une forme explicite, que l'on adopte le point de vue de la statistique classique ou celui de l'approche bayésienne. Toutefois, comme ce modèle résulte de l'association du modèle linéaire et du modèle autorégressif, il possède une structure conditionnelle adéquate qui permet d'en réaliser l'inférence bayésienne grâce à l'échantillonnage de Gibbs. L'ajout d'une structure conditionnelle prenant en compte une erreur dans les variables ou la modélisation multidimensionnelle du procédé ne complique pas l'inférence bayésienne, qui peut être aussi réalisée par cette même technique de simulation. A notre connaissance, en agroalimentaire comme dans d'autres domaines appliqués, cette méthodologie n'a jamais été mise en pratique. Pourtant, elle peut être facilement mise en œuvre sans gros efforts de programmation et le **MLEAR** est une formulation réaliste et suffisamment large pour modéliser de nombreux procédés industriels. En outre, la statistique bayésienne offre la possibilité d'intégrer des informations extérieures dans la démarche d'estimation, ce qui lui donne un atout supplémentaire vis-à-vis de préoccupations d'ingénierie.

Mots-clés : *modélisation, modèle linéaire à erreur autocorrélée, erreur dans les variables, statistique bayésienne, échantillonnage de Gibbs, informations a priori*

ABSTRACT

The linear model with autocorrelated disturbance (**MLEAR**) appears as a natural extension of regression technics in order to model a process when data are collected over time. Due to the intricate structure of dependence between parameters, no analytic solution

for parameter inference is available, had the bayesian or the conventional statistical framework been adopted. However, conditional structure resulting from the association of the linear model and the autoregressive model allows bayesian inference to be based on Gibbs sampling. We also show that adjunction of a supplementary conditional structure to take account of error-in-variable or multivariate modelling can be handled straightdforwardly by Gibbs sampling. To our knowledge, in food industry or in other applied areas, this methodology was never used. Indeed, it can be implemented very easily with no programmation effort and **MLEAR** is realistic enough to model a broad range of industrial processes. Moreover, bayesian statistics allow to integrate exogenous information which is valuable advantage in an engineering context.

Keywords *: modelisation, linear model with autocorrelated disturbance, error-in-variable, bayesian statistics, Gibbs sampling, a priori informations.*

1. Introduction

Disposant d'observations relatives à un phénomène, la tâche première du scientifique est de caractériser ce phénomène, c'est-à-dire de tenter de connaître les variations de certaines quantités en fonction d'une autre ou de plusieurs autres. Cette caractérisation est réalisée par la construction d'un modèle mathématique mettant en relation une ou plusieurs quantités observées, appelées variables réponses, avec d'autres quantités observées, appelées variables explicatives. En général, lorsque les variables explicatives sont réelles, on les nomme covariables alors que lorsqu'elles sont qualitatives, on parle de facteurs.

Un modèle statistique est constitué d'une fonction décrivant la relation entre la variable réponse et les variables explicatives. Puis, des paramètres $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$ règlent la forme de cette fonction. Enfin, des hypothèses sur la structure du terme aléatoire mesurant l'écart entre les observations et le modèle achèvent la spécification du modèle. Parmi les modèles statistiques, le modèle linéaire (lorsque la fonction est linéaire en θ) est sûrement le modèle statistique auquel les praticiens ont le plus recours. Ceci s'explique par le fait que le modèle linéaire offre une première approximation acceptable pour représenter une vaste classe de phénomènes. Une fois le modèle caractérisé (c'est-à-dire une fois que les paramètres θ sont estimés), le modèle peut être utilisé à des fins opérationnelles pour évaluer, par exemple, si l'influence de certaines variables explicatives est significative ou pour prédire de nouvelles valeurs du phénomène connaissant des nouvelles entrées.

Un procédé de fabrication agroalimentaire réalise une transformation physique ou chimique des matières premières pour générer un produit final. Pour piloter ses installations, l'industriel possède une connaissance empirique très fine de ses procédés de fabrication sans avoir besoin de recourir à un modèle statistique explicite. Néanmoins, cette maîtrise technologique peut diminuer en raison de modifications techniques importantes et un modèle statistique peut alors devenir un outil précieux. Tel est le cas, par exemple, du procédé de fabrication du Lait concentré Sucré (LCS). Le LCS est un produit phare de la société Nestlé car, non seulement, on le doit au fondateur Henri Nestlé de la société (1868), mais il véhicule aussi une image emblématique de qualité. Parmi les caractéristiques du LCS, la viscosité est sans

doute celle qui intéresse principalement le consommateur car elle participe de près à la facilité d'emploi et donc de consommation du LCS. Afin de maîtriser le procédé au niveau de la viscosité, Nestlé met en œuvre tous les moyens de mesure de la viscosité et enregistre tous les paramètres de fabrication (voir figure 1). Jusqu'il y a une dizaine d'années, la maîtrise de la viscosité était assurée par des opérateurs et des contremaîtres, spécialisés dans la maîtrise du procédé du LCS, qui, par leur expérience, connaissaient empiriquement tous les facteurs influençant la viscosité. Or, depuis quelques années, les évolutions technologiques du procédé et le renouvellement important de la main-d'œuvre entraînent une perte importante du savoir-faire et, par voie de conséquence, une perte de maîtrise du procédé.

Les mesures enregistrées sur la ligne de fabrication ainsi que la viscosité mesurée (voir figure 1) constituent un ensemble de données qui nous renseignent sur l'influence des paramètres de fabrication sur la viscosité. Comme nous ne possédons aucun modèle de connaissance sur la viscosité (Girard, [21]), le modèle linéaire peut sembler, dans un premier temps, une bonne approximation pour expliquer le phénomène.

Toutefois, l'exploitation par un modèle statistique des données recueillies sur la ligne de fabrication pose des problèmes qui sont courants en modélisation statistique : la non indépendance des observations, l'erreur dans les variables explicatives, l'aspect multidimensionnel. Nous proposons de les aborder au cours de cet article.

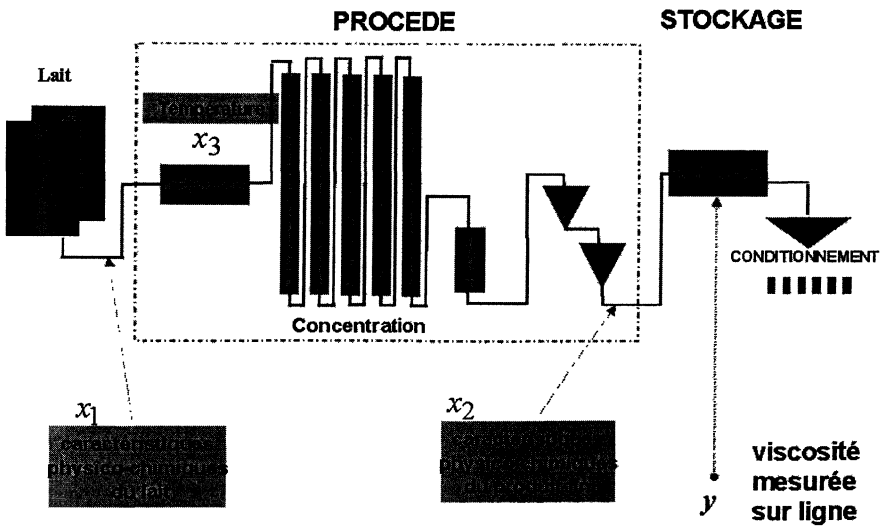

FIGURE 1
Schéma de fabrication du LCS et mesure de la viscosité.

En effet, le recueil des données sur la ligne de fabrication au cours du temps introduit une structure de dépendance temporelle dans les données. Or, Neter *et al.* [26], Draper et Smith [16], Doran [15] et Weisberg [31] indiquent que cette dépendance modifie notablement les résultats classiques du modèle linéaire. Ils remarquent ainsi la présence d'autocorrélation dans les erreurs et rapportent les principales conséquences de cette autocorrélation dans le modèle linéaire lorsqu'on utilise directement les résultats des moindres carrés ordinaires sans développements complémentaires. Elles concernent en grande partie la qualité des estimateurs en terme de variance : par exemple,

- les estimateurs classiques des moindres carrés sont toujours sans biais mais ils ne possèdent plus leur propriété de variance minimum et perdent en général leur caractère admissible;
- le carré moyen résiduel peut sérieusement sous-estimer l'amplitude des termes d'erreurs;
- cette sous-estimation de l'écart-type des résidus peut avoir de graves conséquences au niveau de l'inférence (par exemple, en concluant à tort qu'une variable est significative);
- on ne peut plus s'appuyer sur les distributions de Student et de Fisher-Snedecor pour construire les intervalles de confiance habituels.

Le modèle linéaire avec erreur autocorrélée (**MLEAR**) figure parmi les modèles les plus étudiés, notamment en économétrie. La littérature foisonne ainsi d'articles traitant de ce modèle selon une approche statistique conventionnelle (estimation itérative des paramètres : [10]; estimation des paramètres par maximum de vraisemblance : [1], [2], [21]; prévision pour des données futures : [14]). Mais, tous ces travaux se limitent à une estimation ponctuelle des paramètres (Cochrane et Orcutt [10]) dont ils étudient la loi par approximation asymptotique normale. Par ailleurs, des avancées concurrentes sur le **MLEAR** ont été menées en statistique bayésienne (Zellner et Tiao [33], Zellner [32], Fomby et Guilkey [17], Broemeling [6], Chalton et Troskie [7]). Malheureusement, pour les problèmes numériques usuels, tous ces auteurs se heurtent au problème d'identification des lois a posteriori et ont tous recours à des techniques numériques pour obtenir les lois a posteriori des paramètres. Ainsi, tous les auteurs cités ci-dessus se limitent à un modèle dont l'erreur suit un processus autorégressif d'ordre 1. Chib [8] est le premier à proposer une méthodologie originale d'estimation du **MLEAR** basée sur l'échantillonnage de Gibbs (Geman et Geman, [20]; Gelfand et Smith, [18]). En raison de la facilité d'emploi de l'échantillonnage de Gibbs, en particulier, dans le cas du **MLEAR**, et de la généralisation de l'emploi des ordinateurs personnels avec lesquels un tel algorithme peut être facilement programmé, il est assez surprenant que cette méthodologie n'ait pas été appliquée par la suite pour la modélisation de procédé et le contrôle statistique de la qualité.

Dans cet article, nous nous plaçons dans une perspective bayésienne et proposons des extensions méthodologiques aux travaux de Chib [8] avec d'une part, l'étude du **MLEAR** avec erreur dans les variables explicatives et, d'autre part, l'étude du **MLEAR** multidimensionnel. L'algorithme de Gibbs permet d'obtenir sans compétence numérique et avec un programme informatique rudimentaire un échantillon de la loi a posteriori des paramètres. Un élément important en faveur de cette technique de simulation est qu'elle ne nécessite pas le recours à des lois

instrumentales comme c'est le cas pour les méthodes habituelles d'intégration de Monte-Carlo. En effet, l'échantillonnage de Gibbs consiste uniquement en des tirages successifs dans certaines distributions conditionnelles des paramètres du modèle. Son emploi est d'autant plus facile que ces lois conditionnelles sont standard et faciles à obtenir à partir d'un générateur de nombre aléatoire. Nous restons ici dans le cadre de la famille conjuguée du modèle normal et ces distributions sont alors soit des lois Normales multidimensionnelles, soit des lois Gamma, soit des lois Wishart. Bien évidemment, la méthode offre la possibilité de considérer des informations a priori pour incorporer l'expertise d'un opérateur dans l'analyse.

Dans cet article, nous allons utiliser les notations et définitions suivantes. La loi Normale p-dimensionnelle, Student p-dimensionnelle et la distribution gamma inverse sont notées respectivement $\mathbf{N}_p(\mu, \Sigma)$, $\mathbf{T}_p(\upsilon, \mu, \Sigma)$ et $\mathbf{GI}(\upsilon, s)$. Une variable matrice $W : p \times q$ qui suit une distribution Normale matricielle sera notée $\mathbf{NM}_{pq}=(W\,|\mu, A \otimes B)$ ce qui signifie que $\mathrm{vec}(W) \sim \mathbf{N}_{pq}(\mu, A \otimes B)$ où A est une matrice $p \times p$, B est une matrice $q \times q$ et $\mathrm{vec}(\cdot)$ est l'opérateur matriciel qui transforme une matrice en un vecteur colonne. Sa fonction de densité est

$$(2\pi)^{-pq/2}\,|A|^{-p/2}\,|B|^{-q/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\left(A^{-1}(W-\mu)B^{-1}(W-\mu)'\right)\right\}$$

Finalement, pour une matrice symétrique définie positive Ω qui suit une distribution de Wishart p-dimensionnelle avec υ degré de liberté et une matrice d'échelle A, nous utilisons la notation $\mathbf{W}_p\left(\upsilon, A^{-1}\right)$. Sa fonction de densité est

$$k\,|A|^{\upsilon/2}\,|\Omega|^{(\upsilon-p-1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\,(A\Omega)\right\}$$

où k est une constante de normalisation, A est une matrice d'hyperparamètres et $|A|$ est le déterminant de A (Gelman *et al.*, [19]).

L'article s'organise de la façon suivante :

- la section 2 rappelle l'analyse bayésienne du **MLEAR** ainsi que les principaux problèmes rencontrés lors de cette analyse (notamment, l'identification de la loi jointe a posteriori et l'expression des lois marginales a posteriori);
- les principaux résultats de l'échantillonnage de Gibbs et de son application au **MLEAR** sont expliqués dans la section 3;
- des extensions du **MLEAR** au **MLEAR** avec erreur dans les variables explicatives et au **MLEAR** multidimensionnel sont présentées dans la section 4;
- la section 5 est alors consacrée à l'étude de plusieurs applications.

2. Analyse bayésienne du modèle linéaire à erreur autocorrélée (MLEAR)

Nous considérerons, dans cette section, que l'erreur du modèle linéaire suit un processus autorégressif d'ordre 1 et nous noterons $\mathbf{MLEAR}_1$ ce modèle. Les calculs d'inférence ci-après peuvent être étendus sans difficulté en suivant la même ligne de raisonnement au cas où l'erreur suit un processus autorégressif d'ordre supérieur.

2.1. Formulation du modèle linéaire à erreur autocorrélée d'ordre 1 ($\mathbf{MLEAR}_1$)

Considérons le modèle :

$$\mathbf{MLEAR}_1 \ : \ y_t = x_t\beta + \varepsilon_t, \qquad \varepsilon_t = \rho\varepsilon_{t-1} + u_t \tag{1}$$
$$t = 1, \ldots, T$$

où une observation de la variable réponse, y_t, recueillie au temps t, est reliée à q variables explicatives $x_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,q})$ au moyen d'une combinaison linéaire avec $\beta = (\beta_1, \ldots, \beta_q)' \in \mathbf{R}^q$ et une erreur ε_t autocorrélée telle que $\rho \in \,]-1; 1[$ et où les innovations u_t, $t = 1, \ldots, T$, sont identiquement et indépendamment distribués selon une loi Normale, $u_t \sim \mathbf{N}\left(0, \sigma^2\right)$. La première variable explicative peut être prise égale à 1 pour considérer un terme constant.

Le modèle $\mathbf{MLEAR}_1$ s'écrit en prenant les notations matricielles suivantes $Y' = (y_1, y_2, \ldots, y_T)$, $X' = (x'_1, x'_2, \ldots, x'_T)$, $Y'_{-1} = (y_0, y_1, \ldots, y_{T-1})$, $X'_{-1} = \left(x'_0, x'_1, \ldots, x'_{T-1}\right)$, $Y_\rho = Y - \rho Y_{-1}$, $X_\rho = X - \rho X_{-1}$ et $u' = (u_1, u_2, \ldots, u_T)$, sous la forme classique d'une équation de régression :

$$Y_\rho = X_\rho\beta + u \tag{2}$$

avec $u \sim \mathbf{N}_T\left(0, \sigma^2\mathbf{I}_T\right)$ tel que $\mathbf{I}_T$ soit la matrice identité de dimension T. Notez que derrière les grandeurs Y_ρ et X_ρ se cache le paramètre additionnel inconnu de corrélation entre les résidus, ρ.

Lorsque $t = 1$, l'équation (1) devient $y_1 = \rho y_0 + x_1\beta - \rho x_0\beta + u_1$ où y_0 et x_0 apparaissent comme des quantités observées. Si nous supposons que l'équation (1) est représentative de ce qui s'est passé pour $t = 0, -1, -2, \ldots, -T_0$, où T_0 est inconnu, nous avons, par exemple, $y_0 = x_0\beta + \rho\left(y_{-1} - x_{-1}\beta\right) + u_0$ en supposant que u_0 est un bruit blanc tel que $u_0 \sim \mathbf{N}\left(0, \sigma^2\right)$ comme les $u_t, t = 1, \ldots, T$. Comme y_{-1} et x_{-1} ne sont pas des quantités connues, il est plus simple d'écrire que $y_0 = A + u_0$ où A est une fonction de quantités inobservées[1]. Parmi tous les auteurs cités précédemment, seuls Zellner et Tiao [33] considèrent explicitement A comme un paramètre du modèle sans résoudre complètement le problème. Dans cet article, A est considéré comme un paramètre à part entière du modèle, c'est lui qui règle la condition initiale de la trajectoire des y_t : le vecteur de paramètres du modèle $\mathbf{MLEAR}_1$ est donc $\theta = \left(\beta, \sigma^2, A, \rho\right)$.

Pour l'analyse de ce modèle, nous disposons d'un ensemble de variables réponses $\mathbf{y} = \{y_0, y_1, \ldots, y_T\}$ et de variables explicatives $\mathbf{X} = \{x_0, x_1, \ldots, x_T\}$. Considérant l'équation (1) et utilisant le caractère markovien de la série des y_t, la vraisemblance peut s'écrire sous la forme suivante :

$$p\left(\mathbf{y} \left| \beta, \sigma^2, A, \rho, \mathbf{X}\right.\right) = \left\{\prod_{t=1}^{T} p\left(y_t \left| x_t, x_{t-1}, y_{t-1}, \beta, \rho, \sigma^2, A\right.\right)\right\} \times p\left(y_0 \left| \sigma^2, A\right.\right)$$

[1] D'autres hypothèses sans doute plus appropriées dans d'autres situations peuvent considérer, par exemple, que $u_0 \sim \mathbf{N}\left(0, \sigma_0^2\right)$, σ_0^2 connu.

Sous les hypothèses du modèle $\mathbf{MLEAR}_1$, la vraisemblance est alors :

$$p\left(\mathbf{y}\,|\beta,\sigma^2,A,\rho,\mathbf{X}\right) = \left\{\prod_{t=1}^{T}\mathbf{N}\left(y_t-\rho y_{t-1}-x_t\beta+\rho x_{t-1}\beta,\sigma^2\right)\right\}\times\mathbf{N}\left(y_0-A,\sigma^2\right) \quad (3)$$

Avec les notations matricielles introduites ci-dessus, la vraisemblance s'écrit aussi :

$$p\left(\mathbf{y}\,|\beta,\sigma^2,A,\rho,\mathbf{X}\right) = \left(\frac{1}{\sqrt{2\pi}}\right)^{T+1}\times\frac{1}{\sigma^{2\left(\frac{T+1}{2}\right)}} \exp-\frac{1}{2\sigma^2}\left\{\left(Y_\rho-X_\rho\beta\right)'\left(Y_\rho-X_\rho\beta\right)+\left(y_0-A\right)^2\right\} \quad (4)$$

2.2. Rappels de l'analyse bayésienne d'un modèle statistique

La statistique bayésienne se distingue de la statistique classique (fréquentiste) par le fait qu'elle représente l'incertitude portant sur le paramètre θ du modèle par une distribution de probabilité conditionnellement à un état de connaissance de l'analyste (Berger, [3]). Cette incertitude doit être distinguée de l'aléa qui modélise l'écart du modèle à la réalité : en statistique classique, c'est uniquement ce dernier qui est pris en compte dans la caractérisation du modèle. La représentation de l'incertitude portant sur le paramètre θ du modèle est nommée loi a priori de θ et est notée ici $p\left(\theta\right)$. Elle s'interprète comme la représentation formelle sous forme probabiliste de la connaissance sur les paramètres détenue par l'analyste, par l'expert, ... avant une phase de recueil de données (expériences antérieures, expertise, connaissance subjective). Toutefois, si l'on souhaite uniquement traiter l'information apportée par les données, certains auteurs comme Jeffreys [23] et Box et Tiao [5] se sont préoccupés de la construction de lois a priori non informatives. Ces lois permettent de traduire autant que faire se peut l'ignorance éventuelle de l'analyste.

Finalement, le théorème de Bayes permet de mettre à jour la loi a priori du vecteur des paramètres θ du modèle par incorporation de l'information apportée par de nouvelles données $D=\{\mathbf{y},\mathbf{X}\}$ pour obtenir la probabilité a posteriori du vecteur des paramètres θ $p(\theta|D)$ selon la formule bien connue :

$$p\left(\theta\,|D\right) = \frac{p\left(D\,|\theta\right)p\left(\theta\right)}{m\left(D\right)} \quad (5)$$

où $p\left(D\,|\theta\right)$ est la vraisemblance de l'échantillon D. La loi prédictive a priori $m\left(D\right) = \int p\left(D\,|\theta\right)p\left(\theta\right)d\theta$ peut être interprétée comme une vraisemblance marginale intégrée sur les paramètres θ (Bernardo et Smith, [4]). Le choix des lois a priori dans une famille de lois dites conjuguées (Robert, [28]) permet un calcul commode de (5) car, la loi a posteriori $p\left(\theta\,|D\right)$ appartenant à cette même famille, la

formule de Bayes se résume à une simple mise à jour des hyperparamètres réglant les formes des lois $p(\theta)$ et $p(\theta|D)$.

La loi jointe a posteriori $p(\theta|D)$ est à la base de l'inférence bayésienne : en effet, elle représente les degrés de croyance que le modélisateur accorde à chaque valeur de θ après mise à jour de sa connaissance par incorporation de l'information apportée par les données. Elle peut être utilisée pour calculer l'espérance, le mode ou la médiane a posteriori de θ qui peuvent constituer des estimateurs ponctuels de θ, ou bien pour construire des zones de crédibilité (l'équivalent bayésien des ellipsoïdes de confiance) ou enfin pour obtenir des lois marginales $p(\theta_j|D)$, $j = 1, \ldots, k$, par intégration.

Sous certaines conditions (estimation stable) et si la loi a priori n'exclut rien du domaine de variation des paramètres, Berger ([3], p. 224) montre des résultats asymptotiques analogues à ceux des estimateurs du maximum de vraisemblance :

$p(\theta|D)$ est approximativement une loi normale $\mathbf{N}\left(\widehat{\theta}, \left[\mathbf{I}\left(\widehat{\theta}\right)\right]^{-1}\right)$ où $\widehat{\theta}$ est l'estimateur du maximum de vraisemblance et $\mathbf{I}\left(\widehat{\theta}\right)$ est la matrice d'information de Fisher définie comme $\mathbf{I}\left(\widehat{\theta}\right) = -\mathbf{E}\left[\dfrac{\partial^2}{\partial\theta_i\partial\theta_j}\log p(D|\theta)\right]_{\theta=\widehat{\theta}}$.

2.3. Caractérisation des lois a posteriori des paramètres du $\mathbf{MLEAR}_1$

Selon la formule de Bayes (5), la loi a posteriori du vecteur des paramètres $\theta = (\beta, \sigma^2, A, \rho)$ est proportionnelle au produit de la vraisemblance par la loi a priori, comme le dénominateur de la formule de Bayes n'est pas une fonction de θ. La loi a posteriori des paramètres est donc

$$\begin{aligned} p(\beta, \sigma^2, A, \rho|\mathbf{y}, \mathbf{X}) &\propto p(\mathbf{y}, \mathbf{X}|\beta, \sigma^2, A, \rho) \times p(\beta, \sigma^2, A, \rho) \\ &\propto p(\mathbf{y}|\beta, \sigma^2, A, \rho, \mathbf{X}) \times p(\beta, \sigma^2, A, \rho) \end{aligned} \tag{6}$$

sous l'hypothèse que les variables explicatives sont connues avec certitude.

2.3.1. Loi a priori des paramètres du $\mathbf{MLEAR}_1$

Préalablement au recueil des données expérimentales, l'homme d'étude possède une connaissance a priori sur les paramètres (β, σ^2) qui est indépendante de celle qu'il a pour le couple (A, ρ). Il semble donc raisonnable de supposer ici que

$$p(\beta, \sigma^2, A, \rho) = p(\beta, \sigma^2)\, p(A, \rho) = p(\beta, \sigma^2)\, p(\rho)\, p(A) \tag{7}$$

En procédant ainsi, ce sont les données qui établiront la covariation entre les paramètres relatifs à la partie «modèle linéaire» et ceux caractérisant la partie autorégressive : dans l'équation (6), la vraisemblance $p(\mathbf{y}|\theta, \mathbf{X})$ associe de façon complexe les deux types de paramètres selon la formule (4).

La vraisemblance (3) et (4) indique que les lois des paramètres du modèle appartiennent à la famille exponentielle ce qui implique l'existence de lois a priori conjuguées (Robert, [28]). En fait, il est pratique, pour tirer partie de la structure du **MLEAR**, de prendre des lois a priori conjuguées par bloc de paramètres : ainsi, nous prenons pour (β, σ^2) la forme conjuguée du modèle linéaire et pour ρ la forme conjuguée de l'$\mathbf{AR}(1)$, soit

$$\beta \,|\sigma^2 \sim \mathbf{N}_q \left(\beta_0, \sigma^2 \Sigma_0^{-1}\right), \sigma^2 \sim \mathbf{GI}\left(\frac{\upsilon_0}{2}, \frac{\upsilon_0 s_0^2}{2}\right) \tag{8}$$

$$\text{et } A \sim \mathbf{N}\left(A_0, \delta_0^{-1}\right), \rho \sim \mathbf{N}\left(\rho_0, V_0^{-1}\right) 1_{|\rho|<1}$$

où $\left(\beta_0, \Sigma_0^{-1}, \upsilon_0, s_0^2, A_0, \delta_0^{-1}, \rho_0, V_0^{-1}\right)$ sont des hyperparamètres qui règlent la forme de chacune des lois a priori[2].

La loi a priori du vecteur de paramètres $\theta = (\beta, \sigma^2, \rho, A)$ est donc une combinaison d'une normale et d'une gamma inverse pour (β, σ^2), une normale pour A et une normale tronquée sur l'intervalle de définition $]-1; 1[$ pour ρ. La sélection de valeurs particulières dans (8) pour les hyperparamètres $\left(A_0, \delta_0^{-1}, \beta_0, \Sigma_0^{-1}, \upsilon_0, s_0, \rho_0, V_0^{-1}\right)$ relève de la responsabilité du modélisateur (voir section 5.1.1). Elles peuvent être choisies pour donner diverses formes générales à la distribution a priori des paramètres et prendre en compte des connaissances préalables du phénomène étudié. Les valeurs des hyperparamètres peuvent provenir d'information historique voire même de connaissances subjectives. Quand l'information a priori est inexistante, il est possible de considérer une loi a priori dite non informative ou vague (Box et Tiao, [5]). Des lois a priori non informatives pour $(\beta, \sigma^2, \rho, A)$ sont ici obtenues en fixant les hyperparamètres de (8 de telle sorte que $A_0 = y_0$, $\delta_0^{-1} = V_0^{-1} \to \infty$, $\Sigma_0^{-1} = c\mathbf{I}_q$ avec $c \to \infty$, $\upsilon_0 = -1$, $\beta_0 = \upsilon_0 s_0^2 = 0$.

2.3.2. Loi jointe a posteriori et lois marginales a posteriori des paramètres du $\mathbf{MLEAR}_1$

D'après (4), (6), (7) et (8), la loi a posteriori du vecteur de paramètres $\theta = (\beta, \sigma^2, A, \rho)$ est alors

$$p\left(\beta, \sigma^2, A, \rho \,|\mathbf{y}, \mathbf{X}\right) \propto \frac{1}{\sigma^{2\left(\frac{T+1+q+\upsilon_0}{2}+1\right)}} \exp\left[-\frac{\delta_0}{2}\left(A - A_0\right)^2 - \frac{1}{2\sigma^2}\right.$$
$$\left\{\upsilon_0 s_0^2 + \left(y_0 - A\right)^2 + \left(\beta - \beta_0\right)' \Sigma_0 \left(\beta - \beta_0\right) + \left(Y_\rho - X_\rho \beta\right)' \left(Y_\rho - X_\rho \beta\right)\right\} \tag{9}$$
$$\left. - \frac{V_0}{2}\left(\rho - \rho_0\right)^2\right] 1_{|\rho|<1}$$

[2] Bien évidemment, d'autres lois *a priori* peuvent être proposées. Notamment, l'indépendance *a priori* entre β et σ^2 peut être supposée. Il est facile de voir que cela ne modifie que très peu les résultats suivants. Toutefois, nous verrons que l'emploi de lois a priori conjuguées permet d'obtenir facilement des lois conditionnelles complètes, lois utilisées dans l'algorithme de Gibbs.

Quoique faisant partie de la famille exponentielle et malgré l'emploi de lois a priori partiellement conjuguées, la loi jointe a posteriori (9) n'est pas standard. On remarque notamment sous le terme exponentiel de l'équation (9) un polynôme dont le terme d'ordre le plus élevé est du type $\rho^2\beta'\beta$. A la connaissance des auteurs, $p\left(\beta,\sigma^2,A,\rho\,|\mathbf{y},\mathbf{X}\right)$ ne peut pas être obtenue sous une forme complétement explicite. Toutefois, si ρ et A sont supposés connus, c'est-à-dire conditionnellement à ρ et A, et compte tenu de nos hypothèses sur les lois a priori, nous reconnaissons l'expression classique de la loi jointe a posteriori des paramètres $\left(\beta,\sigma^2\right)$ du modèle linéaire $Y_\rho = X_\rho\beta + u$ avec $u \sim \mathbf{N}\left(0,\sigma^2\mathbf{I}_T\right)$ (Box et Tiao, [5]; Broemeling, [6]). Ainsi, nous déduisons que $p\left(\beta,\sigma^2\,|\rho,A,\mathbf{y},\mathbf{X}\right)$ est une Normale-Gamma inverse comme l'était sa loi a priori $p\left(\beta,\sigma^2\right)$. De façon symétrique, conditionnellement à β et σ^2, nous pouvons identifier un processus $\mathbf{AR(1)}$ sur l'erreur ε du modèle linéaire principal avec $\varepsilon = Y - X\beta$ (Zellner, [32]). Cette structure conditionnelle sera à la base du raisonnement et de la résolution par échantillonnage de Gibbs qui suit.

Le tableau ci-dessous regroupe les résultats principaux de l'inférence bayésienne du modèle $\mathbf{MLEAR}_1$ (voir Zellner et Tiao [33], Broemeling [6], Chib [8], Chalton et Troskie [7]). Ne sont présentées dans ce tableau que les lois conditionnelles complètes (un paramètre conditionnellement à tous les autres) et une loi conditionnelle partielle connue. Nous notons $\varepsilon = Y - X\beta$ et $\varepsilon_{-1} = Y_{-1} - X_{-1}\beta$.

TABLEAU 1

Récapitulatif de l'analyse du modèle $\mathbf{MLEAR}_1$

paramètre	loi a posteriori	hyperparamètres
$\rho\,\|\beta,\sigma^2,A,\mathbf{y},\mathbf{X}$	$\mathbf{N}\left(\widehat{\rho},V^{-1}\right)1_{\|\rho\|<1}$	$\begin{cases}\widehat{\rho} = V^{-1}\left(\dfrac{\varepsilon'_{-1}\varepsilon}{\sigma^2} + V_0\rho_0\right)\\ V = \dfrac{\varepsilon'_{-1}\varepsilon_{-1}}{\sigma^2} + V_0\end{cases}$
$\beta\,\|\sigma^2,A,\rho,\mathbf{y},\mathbf{X}$	$\mathbf{N}_q\left(\widehat{\beta},\sigma^2\Sigma^{-1}\right)$	$\begin{cases}\widehat{\beta} = \Sigma^{-1}\left(X'_\rho Y_\rho + \Sigma_0\beta_0\right)\\ \Sigma = X'_\rho X_\rho + \Sigma_0\end{cases}$
$\sigma^2\,\|\beta,A,\rho,\mathbf{y},\mathbf{X}$	$\mathbf{GI}\left(\dfrac{\upsilon}{2},\dfrac{\upsilon s^2}{2}\right)$	$\begin{cases}\upsilon = \upsilon_0 + T + q + 1\\ \upsilon s^2 = \upsilon_0 s_0^2 + (Y_\rho - X_\rho\beta)'(Y_\rho - X_\rho\beta)\\ \qquad + (\beta-\beta_0)'\Sigma_0(\beta-\beta_0)\\ \qquad + (y_0 - A)^2 + \delta_0(A - A_0)^2\end{cases}$
$A\,\|\beta,\sigma^2,\rho,\mathbf{y},\mathbf{X}$	$\mathbf{N}\left(\widehat{A},\delta^{-1}\sigma^2\right)$	$\begin{cases}\widehat{A} = \delta^{-1}\left(\dfrac{y_0}{\sigma^2} + \delta_0 A_0\right)\\ \delta = \dfrac{1}{\sigma^2} + \delta_0\end{cases}$
$\beta\,\|\rho,\mathbf{y},\mathbf{X}$	$\mathbf{T}_q\left(\lambda,\widehat{\beta},\Psi\right)$	$\begin{cases}\lambda = \upsilon_0 + T\\ \widehat{\beta} = \left(X'_\rho X_\rho + \Sigma_0\right)^{-1}\left(X'_\rho Y_\rho + \Sigma_0\beta_0\right)\\ \Psi = \lambda^{-1}\Sigma^{-1}\left(Y'_\rho Y_\rho + \upsilon_0 s_0^2\right.\\ \qquad \left. + \beta'_0\Sigma_0\beta_0 - \widehat{\beta}'\Sigma\widehat{\beta}\right)\end{cases}$

Nous notons que seule la loi conditionnelle de $\sigma^2 |\beta, \rho, A, \mathbf{y}, \mathbf{X}$ dépend explicitement de A.

Les lois marginales a posteriori $p(\theta_i |\mathbf{y}, \mathbf{X}), i = 1, \ldots, k$ sont obtenues grâce aux propriétés des lois conditionnelles ou par intégration. Par exemple, la loi marginale a posteriori de $\rho |\mathbf{y}, \mathbf{X}$ s'écrit, Σ étant défini dans le tableau 1 :

$$p(\rho |\mathbf{y}, \mathbf{X}) = \frac{p(\beta, \rho |\mathbf{y}, \mathbf{X})}{p(\beta |\rho, \mathbf{y}, \mathbf{X})} \propto |\Sigma|^{\frac{1}{2}} \left[Y_\rho' Y_\rho + \upsilon_0 s_0^2 + \beta_0' \Sigma_0 \beta_0 - \widehat{\beta}' \Sigma \widehat{\beta} \right]^{-\left(\frac{T+\upsilon_0}{2}\right)}$$

Cette loi marginale a posteriori n'est pas une loi standard car on ne connait pas d'expression analytique de sa constante de normalisation. De la même façon, les autres lois marginales a posteriori du modèle $\mathbf{MLEAR}_1$ ne s'expriment pas explicitement. Il est alors nécessaire de recourir soit à des méthodes numériques soit à des méthodes de simulation pour obtenir la loi conjointe et les lois marginales a posteriori (Robert, [28]; Robert, [29]).

2.4. Synthèse de l'analyse bayésienne du $\mathbf{MLEAR}_1$

L'analyse bayésienne du modèle $\mathbf{MLEAR}_1$ ne peut être menée à terme de façon explicite. Nous pouvons bien sûr obtenir numériquement les lois marginales a posteriori par intégration de la loi jointe a posteriori (voir ci-dessus). Mais, quelle que soit la méthode employée, l'expérience montre que les méthodes d'intégration numérique deviennent très instables (et donc peu fiables) lorsque la dimension de θ augmente, et qu'au delà de trois intégrales, il est souvent préférable d'utiliser des méthodes de simulation (Robert, [29]). De plus, les méthodes de simulation sont maintenant accessibles à tout praticien n'ayant qu'une connaissance rudimentaire de la programmation informatique.

La structure conditionnelle du modèle $\mathbf{MLEAR}_1$ (tableau 1, Chib [8]) conduit très naturellement à utiliser l'échantillonnage de Gibbs (fondé sur les travaux de Geman et Geman [20]). De plus, comme la technique d'échantillonnage de Gibbs utilise les structures conditionnelles et/ou hiérarchiques du modèle bayésien, des extensions au modèle $\mathbf{MLEAR}_1$ peuvent être développées facilement (voir section 4).

3. Implémentation de l'échantillonnage de Gibbs pour le MLEAR

L'échantillonnage de Gibbs appartient à la classe des méthodes de Monte-Carlo par Chaîne de Markov qui facilitent la résolution de problèmes pratiques d'inférence [29]. Il permet d'obtenir un échantillon de la loi jointe par tirage successif des lois conditionnelles complètes (un paramètre par rapport à tous les autres) : l'algorithme est rappelé en annexe 1. La première apparition de l'échantillonnage de Gibbs dans la littérature scientifique est due à Geman et Geman [20] dans un contexte non statistique (problème discret de reconnaissance d'image). Plus récemment, Gelfand et Smith [18] ont montré ses potentialités énormes d'application dans le cadre de l'analyse bayésienne. Le lecteur intéressé se reportera à cet article pour une discussion plus complète sur l'histoire de la méthode et la démonstration de ses propriétés. Par ailleurs, Gelfand et Smith [18], Tanner [30] et Robert [29] proposent toute une panoplie de

méthodes de simulation des lois marginales a posteriori en fonction de la structure du modèle.

3.1. Mise en œuvre pratique de l'échantillonnage de Gibbs pour le MLEAR

Les lois conditionnelles complètes du vecteur des paramètres $\theta = (\beta, \sigma^2, \rho, A)$ du modèle $\mathbf{MLEAR}_1$ ont été récapitulées dans le tableau 1. L'échantillonnage de Gibbs appliqué au $\mathbf{MLEAR}_1$ consiste alors en des tirages successifs dans les lois de $\rho \,|\beta, \sigma^2, A, \mathbf{y}, \mathbf{X}$; $\beta \,|\sigma^2, A, \rho, \mathbf{y}, \mathbf{X}$; $\sigma^2 \,|\beta, A, \rho, \mathbf{y}, \mathbf{X}$ et $A \,|\beta, \sigma^2, \rho, \mathbf{y}, \mathbf{X}$ dans cet ordre en partant, par exemple, des estimateurs du maximum de vraisemblance $\left(\widehat{\beta}^{MV}, \widehat{\sigma}^{2MV}, \widehat{A}^{MV}, \widehat{\rho}^{MV}\right)$. Les M première itérations sont éliminées. Ensuite, après N itérations, nous obtenons un échantillon de la loi jointe $p\left(\beta, \sigma^2, A, \rho \,|\mathbf{y}, \mathbf{X}\right)$, soit $\left\{\left(\beta^{(g)}, \sigma^{2(g)}, A^{(g)}, \rho^{(g)}\right); g = 1, \ldots, N\right\}$ cet échantillon (voir annexe 1).

Nous pouvons remarquer que la contrainte imposée à ρ assure, entre autres, que la matrice X_ρ soit de rang q ce qui garantit l'existence de Σ^{-1} et permet le tirage de β. Notons, par ailleurs, que les espérances conditionnelles des paramètres sont exactement les moments conditionnels sur lesquels sont basés les méthodes d'estimation récursive décrites en statistique classique (procédure de Cochrane-Orcutt [10]) ou en statistique bayésienne (Lindley et Smith [25] cité par Broemeling [6]). Ici, au lieu d'une estimation ponctuelle, l'échantillonnage de Gibbs nous permet d'obtenir un échantillon simulé de la loi jointe des paramètres; nous pouvons ainsi représenter toutes sortes de distributions ainsi que la covariation des paramètres entre eux.

3.2. Estimation des lois marginales a posteriori par simulation

L'échantillon $\left\{\theta^{(g)}, g = 1, \ldots, N\right\}$ des paramètres du modèle obtenu par l'échantillonnage de Gibbs peut être utilisé pour estimer les densités marginales $p\left(\theta_i \,|\mathbf{y}, \mathbf{X}\right)$, $i = 1, \ldots, k$. L'histogramme de $\left(\theta_i^{(1)}, \ldots, \theta_i^{(N)}\right)$, $i = 1, \ldots, k$, pour un θ_i unidimensionnel, se révèle être alors un estimateur rudimentaire de $p\left(\theta_i \,|\mathbf{y}, \mathbf{X}\right)$. Compte tenu que pour chaque variable θ_i la loi conditionnelle complète $p\left(\theta_i \,|\theta_{(\neq i)}, \mathbf{y}, \mathbf{X}\right)$, c'est-à-dire $p\left(\theta_i \,|\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_k, \mathbf{y}, \mathbf{X}\right)$, est connue (disponible pour les besoins de l'échantillonnage de Gibbs), une estimation plus efficace au sens de la variance est obtenue en s'appuyant sur la loi conditionnelle complète. Par un argument de type Rao-Blackwell, Gelfand et Smith [18] et Robert ([28], p. 318) recommandent alors l'estimation de la densité $p\left(\theta_i\right)$ par la formule utilisant les N réplicats de θ_i (générés par l'échantillonnage de Gibbs) et les formules du tableau 1 avec $\theta_i \equiv \rho, \sigma^2, A$, ou $\beta_j, j = 1, \ldots, q$[3],

$$\widehat{p}\left(\theta_i\right) = \frac{1}{N} \sum_{g=1}^{N} p\left(\theta_i \,\middle|\, \theta_{(\neq i)}^{(g)}, \mathbf{y}, \mathbf{X}\right). \tag{10}$$

L'argument formel pour l'emploi de cet estimateur provient du théorème de Rao-Blackwell (Lehmann, [24], Chapitre 10). Gelfand et Smith [18] démontrent qu'il est meilleur que tout estimateur à noyau de la densité $p\left(\theta_i \,|\mathbf{y}, \mathbf{X}\right)$.

---

[3] D'après des résultats classiques (voir aussi Gelman *et al.*, [19]), la loi de β_j est une loi normale $\mathbf{N}\left(\widehat{\beta}_j, \sigma^2 s_{jj}\right)$ où s_{jj} est le $j^{\text{ième}}$ élément de la diagonale de la matrice Σ^{-1}, définie dans le tableau 1.

De même, l'espérance d'une fonction des paramètres $f(\theta)$ peut être estimée par la moyenne empirique $\widehat{E}(f(\theta)) = \frac{1}{N}\sum_{g=1}^{N} f\left(\theta^{(g)}\right)$. Par un argument analogue à (10) il est possible de réduire la variance numérique de l'estimateur en utilisant les lois conditionnelles. Supposons que pour tout i, l'espérance conditionnelle, $f_i\left(\theta_{(\neq i)}\right) \equiv E\left(f(\theta)\left|\theta_{(\neq i)}\right.\right)$ est disponible selon une forme simple, alors l'estimateur $\widehat{f}_i\left(\theta_{(\neq i)}\right) = \frac{1}{N}\sum_{g=1}^{N} f_i\left(\theta_{(\neq i)}^{(g)}\right)$ est meilleur en terme de variance que $\widehat{E}(f(\theta))$. Par exemple, pour calculer $E(\theta_i)$, on a intérêt à utiliser $\widehat{E}(\theta_i) = \frac{1}{N}\sum_{g=1}^{N} E\left(\theta_i\left|\theta_{(\neq i)}^{(g)}\right.\right)$.

4. Analyse bayésienne d'autres modèles linéaires avec erreur autocorrélée

L'étude de la généralisation du $\mathbf{MLEAR}_1$ en considérant que l'erreur du modèle linéaire suit un processus $\mathbf{AR}(p)$ est détaillée dans Chib [8]. Nous nous en distinguons par la considération explicite du niveau initial du procédé, paramétré par A dans notre modèle. Il faut donc rajouter aux lois explicitées par Chib ([8], pp. 280-281), la loi de $A\left|\beta, \sigma^2, \rho, \mathbf{y}, \mathbf{X}\right.$ donnée dans le tableau 1. Pour autant, l'approche est similaire : l'utilisation de l'échantillonnage de Gibbs est facilitée par l'emploi de lois a priori partiellement conjuguées (voir paragraphe 2.3.).

Le recours à l'échantillonnage de Gibbs repose sur la structure conditionnelle adéquate du modèle étudié. Celle-ci peut être facilement appréhendée si l'on considère que le $\mathbf{MLEAR}$ n'est autre que l'association d'un modèle linéaire ($\mathbf{ML}$) pur et d'un processus $\mathbf{AR}(p)$ et si l'on emploie des lois a priori conjuguées par bloc de paramètres (voir section 2.3.1). Si l'emploi de telles lois est inadapté pour rendre compte des connaissances a priori, le recours à d'autres techniques de simulation doit être envisagé. Toutefois, l'emploi de lois a priori conjuguées par bloc permet d'ajouter une structure hiérarchique supplémentaire comme l'illustre la section suivante : le modèle qui en résulte est alors tout aussi facilement estimé à l'aide de l'échantillonnage de Gibbs. En effet, comme l'indiquent de façon plus générale Gelfand et Smith [18], les modèles hiérarchiques possèdent intrinséquement une structure conditionnelle adéquate pour exhiber toutes les lois conditionnelles complètes.

4.1. $\mathbf{MLEAR}_1$ *avec erreur dans les variables explicatives (*$\mathbf{MLEAR}_1-\mathbf{EdV}$*)*

Le modèle linéaire repose sur l'hypothèse implicite que les variables explicatives sont connues avec certitude. Or, il est très fréquent que les variables explicatives soient entachées d'erreur. Une possibilité technique consiste à considérer des modèles robustes (voir par exemple, Crettaz de Rotten et Helbling, [11]). Comme nous nous trouvons dans une perspective bayésienne, nous prenons avantage de celle-ci pour tenir compte explicitement d'une erreur dans les variables explicatives dans la formulation du modèle.

Dans le cas du $\mathbf{MLEAR}_1$, Dagenais [12] souligne que la présence d'erreurs dans les variables explicatives est fortement préjudiciable à l'estimation des paramètres. Il indique, en effet, que les estimateurs obtenus sans prendre en compte l'autocorrélation sont parfois meilleurs en terme de biais et d'écart quadratique moyen que les estimateurs obtenus pour un modèle de type $\mathbf{MLEAR}_1$.

Afin de prendre en compte l'erreur dans les variables explicatives (**EdV**), nous modifions le modèle $\mathbf{MLEAR}_1$ de la façon suivante :

$$\mathbf{MLEAR}_1\text{–}\mathbf{EdV}: \; y_t = x_t\beta + \varepsilon_t, \qquad \varepsilon_t = \rho\varepsilon_{t-1} + u_t \quad \text{et} \quad x^a_{t,q} = x_{t,q} + \xi_t$$

où $x_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,q})$ est le vecteur ligne $(1 \times q)$ des variables explicatives tel que sa dernière composante $x_{t,q}$ est entachée d'erreur, $\beta \in \mathbf{R}^q$, $\rho \in]-\mathbf{1}; \mathbf{1}[$ et u_t est un bruit blanc $\mathbf{N}\left(0, \sigma^2\right)$. Nous considérons dans ce modèle que seule la quantité $x_{t,q}$ est disponible et que $x^a_{t,q}$ représente la variable explicative (non observée) qui a réellement une action sur la variable d'intérêt y_t. La quantité inobservée $x^a_{t,q}$ n'est connue qu'à une variable aléatoire près $\xi_t \sim \mathbf{N}\left(0, \tau^2\right)$ avec τ^2 connu[4] tel que ξ_t soit indépendant aussi bien de x_t que de $u_t, t = 1, \ldots, T$. On considère donc que $x^a_{t,q}$, $t = 1, \ldots, T$ est une variable normale d'espérance $x_{t,q}$.

Dans la formulation bayésienne du modèle, les $x^a_{t,q}$, $t = 1, \ldots, T$ doivent être considérés comme des paramètres additionnels du modèle (Dellaportas et Stephens, [12]). Pour faciliter les calculs, nous introduisons les notations suivantes :

- $X^0 = (x_{1,q}, x_{2,q}, \ldots, x_{T,q})'$ est le vecteur de la variable explicative entachée d'erreur et $X^a = \left(x^a_{1,q}, x^a_{2,q}, \ldots, x^a_{T,q}\right)'$ est le vecteur de la variable explicative non observée;
- la matrice X des variables explicatives est alors décomposée en deux sous-matrices $X = \left(\overline{X}, X^0\right)$ où $\overline{X}$ est le complément de la matrice X à X^0 et l'on note $\overline{\beta}' = (\beta_1, \beta_2, \ldots, \beta_{q-1})$, β_q étant le paramètre associé à la variable entachée d'erreur;
- enfin, Y_+, X_+, X^a_+, X^0_+, $\overline{X}_+$ sont les matrices augmentées de leurs valeurs initiales respectives, c'est-à-dire, par exemple, $Y_+ = (Y_0, Y)'$.

Considérant les $x^a_{t,q}$, $t = 1, \ldots, T$ comme des variables aléatoires, le vecteur des paramètres associés au modèle $\mathbf{MLEAR}_1\text{–}\mathbf{EdV}$ devient donc $\theta = \left(\beta, \sigma^2, \rho, A, X^a_+\right)$ de loi a priori $p(\theta) = p\left(\beta \left| \sigma^2 \right.\right) p\left(\sigma^2\right) p(\rho)\, p(A)\, p\left(X^a_+ \left| X^0_+ \right.\right)$. On suppose en plus des hypothèses de la section 2.3.1 que X^a_+ est indépendant a priori des autres paramètres : les lois a priori sont identiques à celles prises dans la section 2.3.1 et nous avons de plus $p\left(X^a_+\right) = \mathbf{N}\left(X^0_+, \tau^2 \mathbf{I}_{T+1}\right)$. La loi jointe a posteriori des paramètres du modèle $\mathbf{MLEAR}_1\text{–}\mathbf{EdV}$ est alors

$$p\left(\beta, \sigma^2, \rho, A, X^a_+ \left| \mathbf{y}, \mathbf{X} \right.\right) \propto p\left(\mathbf{y} \left| \beta, \sigma^2, \rho, A, X^a_+, \overline{X}_+ \right.\right) \\ p\left(X^a_+ \left| X^0_+ \right.\right) p\left(\beta \left| \sigma^2 \right.\right) p\left(\sigma^2\right) p(\rho)\, p(A)$$

[4] τ^2 a été calculé à partir d'expériences antérieures. Ce peut être, par exemple, l'erreur de reproductibilité pour une mesure analytique ou la variabilité d'un facteur de production autour d'une valeur cible générée par une régulation.

De façon analogue au $\mathbf{MLEAR}_1$, le calcul explicite des lois marginales a posteriori n'est pas possible et nous devons recourir à des techniques de simulation. En outre, l'ajout de cette structure hiérarchique ne modifie pas la structure conditionnelle du $\mathbf{MLEAR}_1$. Nous obtenons, en effet, des lois conditionnelles complètes identiques à celles récapitulées dans le tableau 1 à ceci près qu'elles sont conditionnées en plus par rapport à X_+^a. Ainsi, pour utiliser l'échantillonnage de Gibbs, nous avons seulement besoin d'expliciter la loi conditionnelle complète de X_+^a. Un calcul immédiat montre que

$$X_+^a \left| \beta, \sigma^2, \rho, A, \mathbf{y}, \overline{X}_+ \right. \sim N\left(\widehat{X}_+^a, \Sigma_{X^a}\right) \tag{11}$$

avec

$$\widehat{X}_+^a = \Sigma_{X_a}^{-1}\left(\frac{\beta_q}{\sigma^2} S'S\left(Y^+ - \overline{X}_+\overline{\beta}\right) + \frac{1}{\tau^2} X_+^0\right)^5$$

et $\Sigma_{X^a} = \dfrac{\beta_q^2}{\sigma^2} S'S + \dfrac{1}{\tau^2}\mathbf{I}_{T+1}$ en considérant la matrice S de taille $T \times (T+1)$ définie telle que $Y_\rho - X_\rho\beta = S\left(Y_+ - X_+\beta\right)$.

Si τ^2 n'est pas connu, nous pouvons ajouter un niveau de conditionnement supplémentaire par rapport à τ^2 et la loi a priori devient $p(\theta) = p\left(\beta \left| \sigma^2 \right.\right) p\left(\sigma^2\right) p(\rho)\, p(A)\, p\left(X_+^a \left| X_+^0, \tau^2 \right.\right) p\left(\tau^2\right)$. En faisant l'hypothèse que $\tau^2 \sim \mathbf{GI}\left(\dfrac{\vartheta_0}{2}, \dfrac{\vartheta_0\varphi_0^2}{2}\right)$, nous ajoutons alors aux lois conditionnelles complètes précédemment citées la loi suivante :

$$\tau^2 \left| \beta, \sigma^2, \rho, A, X_+^a, \mathbf{y}, \overline{X}_+ \right. \sim \mathbf{GI}\left(\frac{\vartheta}{2}, \frac{\vartheta\varphi^2}{2}\right) \tag{12}$$

avec $\vartheta = \vartheta_0 + 1$ et $\vartheta\varphi^2 = \vartheta_0\varphi_0^2 + \left(X_+^a - X_+^0\right)'\left(X_+^a - X_+^0\right)$.

L'algorithme de Gibbs est alors facilement mis en œuvre en générant dans l'ordre ρ, β, A, σ^2, X_a^+ et τ^2 en utilisant les lois conditionnelles du tableau 1 et les lois conditionnelles précédentes (11) et (12).

4.2. MLEAR *multidimensionnel (*MLEARm*)*

Par ailleurs, il est rare que le procédé que l'on souhaite modéliser soit unidimensionnel. L'analyse bayésienne du modèle linéaire multidimensionnel est assez directe (voir Broemeling [6] et Box et Tiao [5]). L'extension multidimensionnelle du **MLEAR** a un fort potentiel tant les applications sont multiples dans le domaine de la maîtrise de la qualité où l'on suit couramment des caractéristiques multidimensionnelles de qualité du produit. Chib et Greenberg [9] proposent l'analyse bayésienne du modèle classique des régressions en apparence non reliées (**SUR** : Seemingly Unrelated Regression) avec une extension avec des erreurs autocorrélées. Toutefois, dans

---

5 $\widehat{X}_+^a$ s'interprète ainsi comme une moyenne pondérée entre la valeur connue X_+^0 et une valeur déduite de la régression de Y sur les autres composantes $\overline{X}_+$ de la matrice X.

le modèle **SUR**, on considère que les variables explicatives peuvent être différentes d'une composante de la variable réponse à une autre autre. En fait, nous considérons ici que la variable réponse $y_t, t = 1, \ldots, T$, est de dimension $r \times 1$ et que l'on observe un unique vecteur de variables explicatives $x_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,q})$ de dimension $1 \times q$ et les résultats de Chib et Greenberg [9] ne sont pas applicables. Dans notre cas de figure, le **MLEAR** multidimensionnel s'écrit alors

$$Y = X\mathbf{B} + \mathcal{E} \qquad \text{tel que} \quad \mathcal{E} = \mathcal{E}_{-1}\Phi + U$$

avec

- $Y = (y_1, y_2, \ldots, y_T)'$ une matrice $T \times r$;
- $X = (x_1', x_2', \ldots, x_T')'$ une matrice $T \times q$;
- $\mathbf{B} = (\beta_j)\,, j = 1, \ldots, r$, une matrice $q \times r$;
- $\mathcal{E} = (e_1, e_2, \ldots, e_T)'$ une matrice $T \times r$ où $e_{tr\times 1} = y_t - \mathbf{B}'x_t'$
- Φ une matrice $r \times r$ des paramètres du vecteur autorégressif d'ordre 1;
- et U une matrice de variables aléatoires normales telle que

$$U \sim \mathbf{NM}_{rT}\left(0, (\Sigma \otimes \mathbf{I}_T)^{-1}\right)$$

Nous avons considéré, d'une part, pour des facilités de calcul et, d'autre part, compte tenu que les variables décrivent un phénomène identique, que Φ se réduit à un unique paramètre autorégressif commun à toutes les variables réponses, c'est-à-dire que $\Phi = \rho\mathbf{I}_r$.

Considérant une loi a priori pour le paramètre $\theta = (\mathbf{B}, \Sigma, \rho)$ analogue à celle de la section 2.3.1, nous avons $p(\mathbf{B}, \Sigma, \rho) = p(\mathbf{B}\,|\Sigma)\, p(\Sigma)\, p(\rho)$ avec $p\left(\mathbf{B}\,\middle|\Sigma^{-1}\right) = \mathbf{NM}_{rq}\left(\mathbf{B}_0, (\Sigma \otimes A_0)^{-1}\right)$, $p(\Sigma) = \mathbf{W}_q\left(\upsilon, S_0^{-1}\right)$ et $p(\rho) = \mathbf{N}\left(\rho_0, V_0^{-1}\right)$.

Une nouvelle fois, la loi jointe a posteriori des paramètres n'a pas une forme standard. Toutefois, nous pouvons obtenir un échantillon de la loi jointe a posteriori grâce à l'échantillonnage de Gibbs (voir section 3). Pour ce faire, il suffit de spécificer les lois conditionnelles complètes :

- $\mathbf{B}\,|\Sigma, \rho, Y, X \sim \mathbf{MN}_{rq}\left(\widehat{\mathbf{B}}, \Sigma^{-1} \otimes \left(X_\rho' X_\rho + A_0\right)\right)$

avec $\widehat{\mathbf{B}} = \left(X_\rho' X_\rho + A_0\right)^{-1}\left(X_\rho' Y_\rho + A_0 B_0\right)$;

- $\Sigma\,|B, \rho, Y, X \sim \mathbf{W}_q\left(\upsilon + q, (\upsilon + q)\left(S_0 + (\mathbf{B} - \mathbf{B}_0)'\right.\right.$
$\left.\left. A_0(\mathbf{B} - \mathbf{B}_0) + (Y - X\mathbf{B})'(Y - X\mathbf{B})\right)\Sigma^{-1}\right)$

- $\rho\,|\Sigma, B, Y, X \sim \mathbf{N}\left(\widehat{\rho}, V^{-1}\right)$

avec $V = V_0 + \text{vec}\,(E_{-1})'\,(\Sigma \otimes \mathbf{I}_T)\,\text{vec}\,(E_{-1})$

et $\widehat{\rho} = V^{-1}\left(\text{vec}\,(E_{-1})'\,(\Sigma \otimes \mathbf{I}_T)\,\text{vec}\,(E) + V_0\rho_0\right)$.

et de tirer successivement $\mathbf{B}, \Sigma$ et ρ dans les lois conditionnelles complètes précédentes.

5. Applications à la modélisation du procédé de fabrication du L.C.S.

5.1. MLEAR *simple*

La viscosité du LCS est un phénomène très complexe (Girard, [21]). Depuis 1868, Les fabricants de LCS possédaient une connaissance empirique de l'influence du procédé sur la viscosité. Or, depuis quelques années, en raison de modifications techniques, cette maîtrise a été perdue. Nous proposons dans cette section une modélisation de la viscosité en fonction de paramètres du procédé de fabrication et de caractéristiques physico-chimiques du lait.

5.1.1. Spécification des lois a priori

Un des avantages de la statistique bayésienne est de pouvoir incorporer des connaissances avant l'estimation des paramètres à travers la spécification des lois a priori. Alors que certains statisticiens bayésiens se préoccupent de proposer des lois a priori non-informatives (Box et Tiao, [5]), nous souhaitons pour cette application exploiter le savoir détenu par les fabricants. En effet, les opérateurs qui sont actuellement sur la ligne possèdent une expérience. Berger [3] propose à cette fin une dizaine de techniques opérationnelles pour obtenir des lois a priori à partir d'informations subjectives.

L'observation de la maîtrise opérationnelle de la viscosité sur la ligne de fabrication montre que les opérateurs fixent une nouvelle valeur d'une variable explicative en fonction de

1. la précédente valeur de la viscosité observée y_{t-1} et
2. d'une valeur empirique qui mesure l'influence de l'incrémentation d'une unité de la variable de contrôle considérée soit $x_{t,3} - x_{t-1,3}$.

En terme mathématique, les opérateurs ont construit empiriquement le modèle suivant :

$$y_t = y_{t-1} + \gamma \left(x_{t,3} - x_{t-1,3} \right) \tag{13}$$

Nous avons considéré que le **MLEAR** consitue une approximation raisonnable du modèle empirique précédent (13). Nous avons ainsi proposé un **MLEAR** avec 3 variables explicatives (voir figure 1) et une constante :

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \beta_3 x_{t,3} + \varepsilon_t, \\ \varepsilon_t &= \rho \varepsilon_{t-1} + u_t \quad \text{tel que} \quad u_t \sim \mathbf{N}\left(0, \sigma^2\right) \\ t &= 1, \ldots, T = 304 \end{aligned} \tag{14}$$

Restant dans le cadre développé dans l'article, nous supposons que les lois a priori sont de la forme proposée dans la section 2.3.1, c'est-à-dire $p\left(\beta, \sigma^2, A, \rho\right) = p\left(\beta \left| \sigma^2 \right.\right) p\left(\sigma^2\right) p\left(A\right) p\left(\rho\right)$. La spécification des hyperparamètres, qui est de la responsabilité du modélisateur, a été réalisée à partir des connaissances détenues par les opérateurs de la ligne :

- $p\left(\rho\right) = \mathbf{N}\left(\rho_0, V_0\right) 1_{|\rho|<1}$

Les hyperparamètres ρ_0 et V_0 sont déduits du modèle empirique précédent (13). Un poids important est mis autour des valeurs proches de 0.9 pour s'approcher au plus du modèle empirique précédent, avec $\rho_0 = 0.9$ et $V_0 = 1$.

$$- \beta \left| \sigma^2 \right. \sim \mathbf{N}_q \left(\beta_0, \sigma^2 \Sigma_0^{-1} \right)$$

Les hyperparamètres β_0 et Σ_0^{-1} ont été définis à partir d'une enquête réalisée auprès de la production. Compte tenu de l'état de notre connaissance sur le phénomène modélisé, il est raisonnable de considérer que la matrice de corrélation Σ_0^{-1} est diagonale, c'est-à-dire que les variables explicatives n'ont pas d'interaction entre elles pour le phénomène considéré. Les valeurs β_0 et les termes de la diagonale de la matrice Σ_0^{-1} sont ensuite évaluées à partir de l'approximation normale réalisée sur l'histogramme obtenu pour chacune des variables considérées indépendamment.

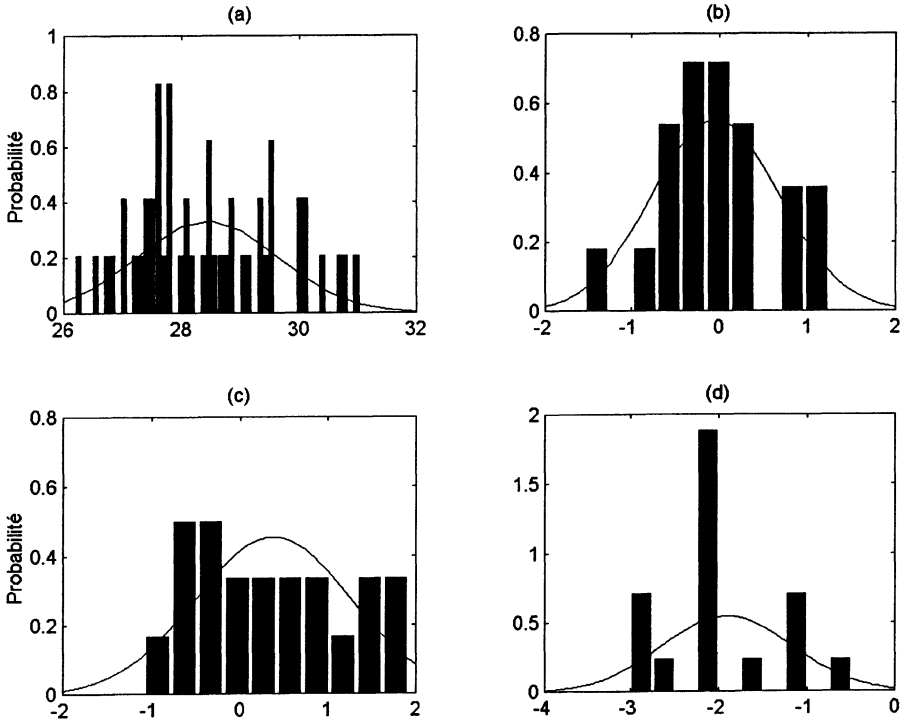

FIGURE 2
Encodage de la loi a priori du paramètre β :
(a) - terme constant β_0; (b) - β_1; (c) - β_2; (d) - β_3.

Après approximation normale, nous obtenons :

$$\beta = (28.43, -0.11, 0.2, -1.91)' \text{ et } \Sigma_0^{-1} = \begin{pmatrix} 1.19 & 0 & 0 & 0 \\ 0 & 0.94 & 0 & 0 \\ 0 & 0 & 0.86 & 0 \\ 0 & 0 & 0 & 0.64 \end{pmatrix}$$

– $\sigma^2 \sim \mathbf{GI}\left(\frac{\upsilon_0}{2}, \frac{\upsilon_0 s_0^2}{2}\right)$

Les hyperparamètres υ_0 et s_0^2 se déduisent de la connaissance de l'erreur de reproductibilité de la mesure en égalant l'erreur de reproductibilité avec $\mathbf{E}\left(\sigma^2\right) = \frac{\upsilon_0}{\upsilon_0 - 2} s_0^2$ et la variance de l'erreur de reproductibilité avec $\mathbf{V}\left(\sigma^2\right) = 2\frac{\upsilon_0^2 s_0^4}{\left(\upsilon_0 - 2\right)^2 \left(\upsilon_0 - 4\right)}$. Des calculs simples montrent que $\upsilon_0 = 2\frac{\left(\mathbf{E}\left(\sigma^2\right)\right)^2}{\mathbf{V}\left(\sigma^2\right)} + 4$ et $s_0^2 = \frac{\upsilon_0 - 2}{\upsilon_0}\mathbf{E}\left(\sigma^2\right)$. Etant donné que l'erreur de reproductibilité est de 3 et que sa variance est de 1, nous avons $\upsilon_0 = 22$ et $s_0^2 = \frac{30}{11}$.

– $A \sim \mathbf{N}\left(A_0, \delta_0^{-1}\right)$

La dernière loi a priori est prise non informative en prenant A_0 quelconque (par exemple, $A_0 = y_0$) et $\delta_0 \to 0$.

5.1.2. Applications

La caractérisation du modèle (14) a été réalisée à partir de données récoltées dans l'usine de Nestlé ($T = 304$) correspondant à l'année 1997. L'algorithme de Gibbs (annexe 1) a été utilisé à partir des lois conditionnelles complètes répertoriées dans le tableau pour obtenir un échantillon de la loi jointe a posteriori des paramètres : après quelques expérimentations, la taille de l'échantillon a été fixé à 2000, après avoir éliminé les 100 premiers tirages. La formule (10) a été employée pour obtenir une estimation des lois marginales a posteriori.

La figure 3 présente ainsi les lois marginales a posteriori (trait pointillé) des paramètres du modèle linéaire pur (**ML**) pur obtenues par application des résultats classiques versus les lois marginales a posteriori des paramètres d'un **MLEAR** obtenues par simulation. Les deux premières variables explicatives sont des caractéristiques de la matière première tandis que la dernière est une température du procédé (voir figure 1).

La figure 4 représente la covariation des paramètres de la régression avec ρ.

On voit que c'est le coefficient de la température de procédé qui est le plus lié aux valeurs possibles de ρ. Ceci illustre les problèmes d'inférence rencontrés pour le modèle linéaire en cas de dépendance entre les erreurs (voir paragraphe 1). Ce résultat montre que notre connaissance a priori d'indépendance entre β et ρ a évolué au vue des données $\mathbf{y}$ et $\mathbf{X}$. En fait, l'hypothèse d'indépendance a priori des paramètres de

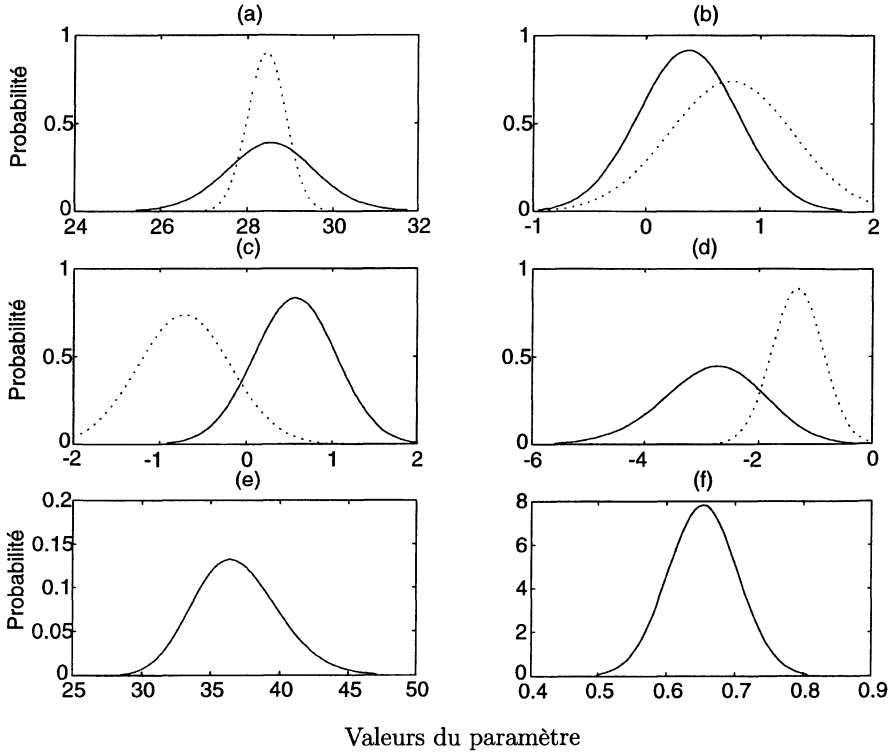

FIGURE 3
*Lois marginales a posteriori des paramètres du modèle linéaire pur (*ML*) en trait pointillé et du* MLEAR *en trait continu : (a) - terme constant β_0; (b) - β_1; (c) - β_2; (d)- β_3; (e) - σ^2; (f) - paramètre d'autocorrélation ρ.*

la régression β et du paramètre autorégressif ρ a été faite afin de faciliter les calculs et parce qu'elle correspondait à l'état de connaissances avant la collecte de données.

La figure 5 présente les lois a posteriori des paramètres versus les lois a priori. Remarquons que les lois marginales des paramètres ont été modifiées au vu des données. Notons particulièrement que les lois a posteriori associées aux variables explicatives sont généralement moins diffuses que les lois a priori, le mode a posteriori étant un peu modifié.

5.2. MLEAR – EdV

Les facteurs de production sont fixés par un opérateur à une valeur de consigne. Or, tous les facteurs de production (température, pression, par exemple) sont soumis à des systèmes de régulation qui permettent d'obtenir en moyenne la valeur de consigne sur un certain pas de temps. Il est possible aussi que, compte tenu de la construction des appareils de fabrication, certains facteurs de production fluctuent de façon non

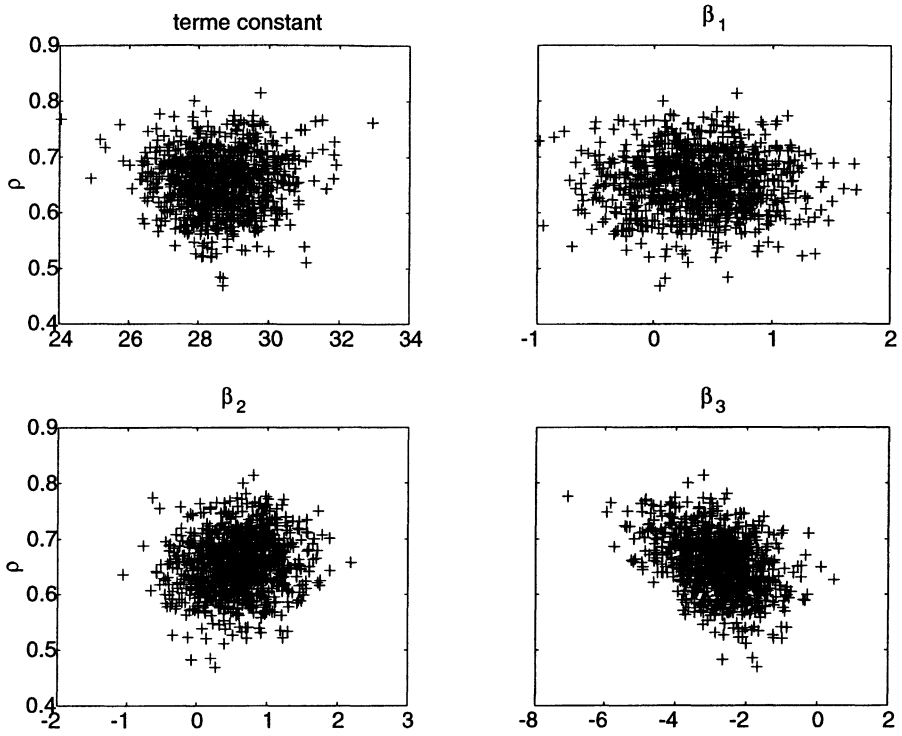

FIGURE 4
Covariation des paramètres β de la régression avec ρ.

volontaire en fonction d'autres. Tous ces éléments amènent à penser que le facteur de production considéré n'est pas exactement la valeur de consigne. Dans ce paragraphe, nous supposons que la dernière variable a été entachée d'une erreur normale.

5.2.1. Spécification des lois a priori

Les lois a priori pour les paramètres $(\beta, \sigma^2, A, \rho)$ ont été prise identiques à la section 5.1.1. Pour le **MLEAR – EdV**, nous avons le paramètre supplémentaire τ^2 qui modélise l'erreur qui entache une des variables explicatives. Une campagne de mesure a été réalisée pour avoir une première évaluation de τ^2. Si nous n'avons pas une bonne connaissance de τ^2 et si l'on veut donner plus de souplesse au modèle, il est possible de considérer que τ^2 est inconnu et d'estimer alors de façon simultanée tous les paramètres à l'aide de l'échantillonnage de Gibbs. Alors, sous l'hypothèse que ce nouveau modèle proposé représente bien les relations, l'échantillon $\left\{X_+^{a(g)}, g = 1, \ldots, N\right\}$ obtenu par échantillonnage de Gibbs permet de préciser la valeur de ce paramètre : par exemple, une estimation bayésienne empirique a posteriori de τ^2 peut être $\widehat{\tau^2} = \dfrac{1}{N-1}\displaystyle\sum_{g=1}^{N}\left\{\left(X_+^{a(g)} - X_+^0\right)'\left(X_+^{a(g)} - X_+^0\right)\right\}$.

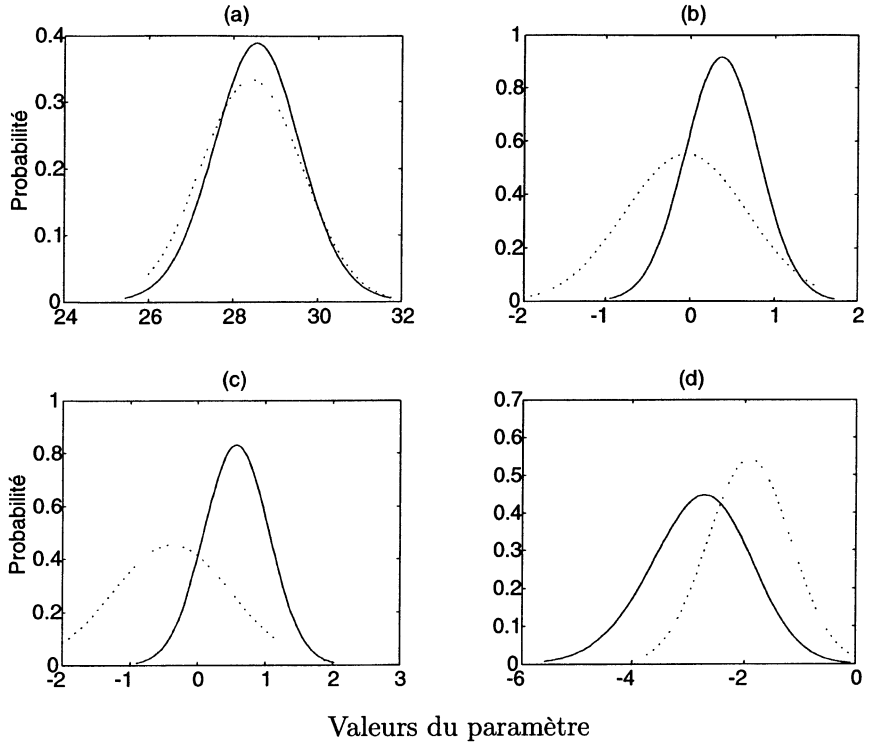

FIGURE 5
Loi a priori (trait pointillé) et loi a posteriori (trait continu) des paramètres de la régression pour un **MLEAR** *: (a) - terme constant β_0; (b) - β_1; (c) - β_2; (d) - β_3.*

5.2.2. Applications

La figure 6 présente les lois marginales des paramètres standard du **MLEAR** sans erreur dans les variables (trait continu) et avec erreur dans la dernière variable (trait discontinu, τ^2 connu).

De façon générale, le fait de considérer que la dernière variable est entachée d'erreur permet de préciser l'estimation de son paramètre : ici, la distribution a posteriori de la dernière variable est moins diffuse. Par ailleurs, nous pouvons observer

- tout d'abord, que l'aléa du modèle paramétré par σ^2 de la loi Normale a fortement diminué. La différence a été «absorbée» par l'incertitude τ^2 modélisant l'erreur dans la dernière variable;
- ensuite, les deux autres variables que nous avons supposé connues avec certitude ont vu leur influence peu modifiée; la légère différence que l'on observe peut être imputée à la précision de l'estimation de chacune des densités, précision qui dépend de la taille des chaînes simulées;
- enfin, le paramètre représentant la mémoire du phénomène semble prendre une valeur plus forte.

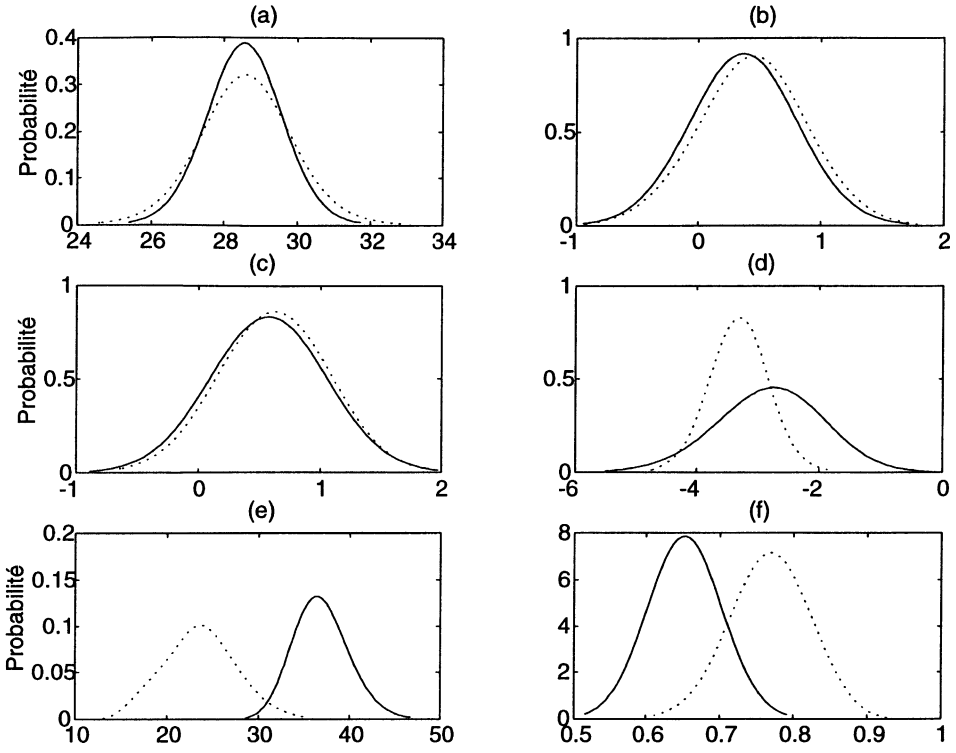

FIGURE 6
Lois marginales a posteriori des paramètres d'un **MLEAR** *(trait continu) et d'un* **MLEAR – EdV** *(trait pointillé) : (a) - terme constant β_0; (b) - β_1; (c) - β_2; (d) - β_3; (e) - σ^2; (f) - paramètre d'autocorrélation ρ.*

La figure 7 présente, quant à elle, les lois marginales des paramètres standard du premier **MLEAR – EdV** (trait continu) avec τ^2 connu et fixé à $4/5$ et du second **MLEAR – EdV** où le paramètre de précision τ^2 de l'erreur est inconnu (trait discontinu).

Dans le second cas où τ^2 est inconnu, un calcul à partir de l'échantillon simulé montre que la moyenne a posteriori calculée de τ^2 est significativement différente de sa première approximation : 5.8, calculée selon la formule donnée dans le paragraphe 5.2.1, contre $4/5$. Ce résultat nous amène à ouvrir la discussion sur la précision de la mesure des variables explicatives. D'une part, la dernière variable n'est sans doute pas la seule variable entachée d'erreur : par exemple, les deux premières variables sont des mesures de laboratoire dont on sait qu'elles sont entachées d'erreur (erreur de reproductibilité). Une modélisation plus complète considérant l'erreur dans chaque variable permettrait sans doute d'aboutir à un modèle plus représentatif de la réalité. L'objectif ici n'étant que d'illustrer la possibilité de prendre en compte une erreur

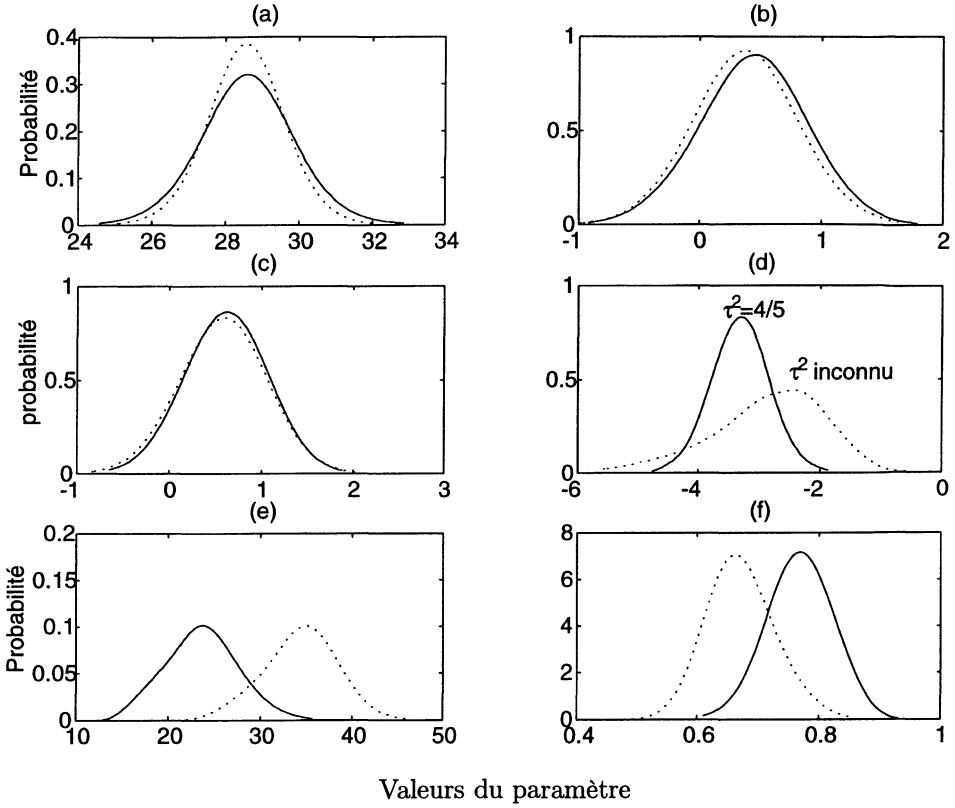

FIGURE 7
Lois marginales a posteriori des paramètres d'un **MLEAR – EdV**
dont le paramètre τ^2 *est connu (trait continu) ou inconnu (trait pointillé) :*
(a) - terme constant β_0*; (b) -* β_1*; (c) -* β_2*; (d) -* β_3*; (e) -* σ^2*;*
(f) - paramètre d'autocorrélation ρ*.*

dans les variables, nous ne poursuivons pas cette complexification. D'autre part, le fait de considérer que τ^2 n'est pas connu avec certitude (figure 7) donne peut-être trop de souplesse au modèle. Il est sans doute préférable dans notre application de considérer que τ^2 est fixé à la valeur préalablement estimée.

6. Conclusions et discussion

Même si le modèle linéaire avec erreur autocorrélée est un des modèles les plus étudiés en littérature statistique, la méthodologie bayésienne développée ici pour son étude est novatrice et prometteuse. D'une part, son emploi est accessible à n'importe quel statisticien praticien à partir du moment où il dispose d'un générateur de lois standard (disponibles même dans un tableur classique). Elle permet en outre d'accéder sans gros efforts à un échantillon simulé provenant des lois exactes des

paramètres. D'autre part, cette méthodologie présente une souplesse intéressante qui permet quelques extensions très réalistes comme nous l'avons présenté sur le modèle avec erreur dans les variables. L'extension multidimensionnelle peut être menée aussi aisément moyennant quelques hypothèses commodes mais réalistes qui facilitent les calculs. De plus, le recours aux hypothèses asymptotiques requises en statistique conventionnelle (Harvey et Phillips, [22]) n'est pas utile, et par conséquent l'analyse bayésienne du **MLEAR** peut être facilement réalisée même sur des échantillons de taille modeste.

Les conclusions suivantes ont été atteintes :

– Comme nous l'avons montré pour le **MLEAR**, les méthodes de simulation stochastique (Robert, [29]) apportent des solutions à des problèmes jusqu'alors traités par des méthodes numériques lourdes à mettre en œuvre. Parmi ces techniques de Monte Carlo, l'échantillonnage de Gibbs est notre favori tant par sa facilité de mise en œuvre (il ne requiert pas nécessairement la spécification de loi a priori informative) que par son implémentation informatique. De plus, l'emploi de la formule (10) (Gelfand et Smith, [18]) permet d'obtenir d'excellents estimateurs en terme de variance des lois marginales a posteriori. D'autres techniques peuvent être plus adaptées pour d'autres modèles comme le souligne Robert [29] lorsque nous ne disposons pas de toutes les lois conditionnelles complètes (par exemple, échantillonnage de substitution, Gelfand et Smith [18]) ou lorsque la paramétrisation du modèle est telle que nous ne pouvons obtenir aucune loi conditionnelle (par exemple, algorithme de Métropolis-Hastings [29]).

– L'algorithme de Gibbs nécessite la spécification des lois conditionnelles complètes des paramètres du modèle. Dans le cas d'un modèle possèdant une structure conditionnelle adapté à l'algorithme de Gibbs, l'ajout d'une structure hiérarchique supplémentaire, par exemple, sur les variables explicatives du modèle (voir section 4.1), ou sur les paramètres du modèle (voir Chib et Greenberg, [9]) ne modifie pas la structure conditionnelle du modèle : l'analyse du modèle résultant de cette modification peut être tout aussi facilement menée à l'aide de l'algorithme de Gibbs. De façon plus générale, Gelfand et Smith [18] montrent que les modèles hiérarchiques possèdent intrinséquement une structure conditionnelle adaptée à l'échantillonnage de Gibbs.

– La statistique bayésienne permet l'exploitation d'information exogène aux données. De façon distincte de la plupart des articles publiés en statistique bayésienne, le problème de l'encodage des informations a priori n'a pas été éludé en utilisant des lois a priori non informatives ou en spécifiant les hyperparamètres à partir d'un sous-ensemble de la base de données. Ici, nous avons tiré partie de l'expérience acquise par les opérateurs en réalisant une enquête et en calant les hyperparamètres à partir des données de l'enquête.

La méthodologie développée dans cet article s'appuie sur le fait que le modèle linéaire à erreur autocorrélée peut être interprété comme résultant de l'association de deux modèles simples : le modèle linéaire pur et le modèle $\mathbf{AR(p)}$. Conditionnellement à l'un des blocs de paramètres, toute l'analyse bayésienne de l'autre bloc de paramètres est connue et répertoriée dans les manuels classiques de statistique. La structure conditionnelle ainsi mise en évidence permet le recours à l'échantillonnage de Gibbs. Il est ainsi possible d'envisager d'étendre ce type de raisonnement pour la

caractéristisation de modèles «mixtes» issus de l'association de deux modèles classiques : ainsi, l'analyse bayésienne de modèles mixtes tels que le modèle linéaire à erreur MA, les modèles ARMA pourrait trouver une résolution aisée en suivant une méthodologie analogue. A contrario, la facilité de mise en œuvre des méthodes de simulation est à double tranchant car elle ouvre aussi la porte à la surparamétrisation des modèles (modèles hiérarchiques à plusieurs niveaux, voir Gelfand et Smith [18]) et très souvent la complexité du modèle ne trouve de limite que dans l'imagination du modélisateur. Pour l'utilisateur final (un opérateur, un fabricant) les hyperparamètres doivent avoir une interprétation (à peu près) tangibles et suffisamment explicites pour être facilement encodés. Ils ne doivent pas être seulement introduits comme un artifice technique par le modélisateur. Quand faut-il s'arrêter?

Par ailleurs, afin d'utiliser l'échantillonnage de Gibbs, nous avons besoin de simuler les lois conditionnelles complètes a posteriori. Pour ce faire, il faut qu'elles aient une forme commode, c'est-à-dire en pratique que l'on ait choisi des lois a priori dans une classe de conjuguées. Une telle approche est évidemment restrictive et s'oppose aux considérations purement subjectivistes, suivant lesquelles les lois a priori sont déterminées en fonction du problème, des connaissances a priori du phénomène et du décideur et non en fonction de leur régularité et de leur souplesse. L'emploi de lois conjuguées est-elle toujours compatible avec un codage adéquat du savoir initial de l'homme d'étude? La spécification des lois a priori reste un problème central dans l'analyse bayésienne : si l'on ne sait pas quelle loi a priori choisir, il est possible d'une part d'utiliser des lois a priori non informatives (Jeffreys [23], Box et Tiao [5]) et d'autre part de réaliser une analyse de sensibilité en la modifiant. Dans cet article, nous nous sommes limités à l'utilisation de lois conjuguées car elles correspondaient à la structure conjuguée par bloc de notre modèle mais une recherche plus poussée devrait être réalisée pour exploiter de façon plus approfondie le cadre bayésien.

Enfin, en suivi opérationnel de la qualité, les données sont récoltées quotidiennement, ce qui permet d'envisager une mise à jour séquentielle des lois des paramètres. Le cadre bayésien correspond tout à fait à cette demande si l'on interprète la formule de Bayes comme un générateur d'information. Or, compte tenu de la structure du **MLEAR**, il n'existe pas de solution explicite pour la mise à jour séquentielle des paramètres. Il est cependant possible au moyen de l'échantillonnage de Gibbs de proposer un schéma séquentiel basé sur l'utilisation duale des lois a posteriori et lois a priori. Un problème technique apparaît lorsque nous voulons utiliser les lois a posteriori comme loi a priori au temps suivant en raison de la dépendance a posteriori entre β et ρ et de l'hypothèse d'indépendance a priori des lois a priori. Une recherche supplémentaire est nécessaire sur l'approximation des lois a priori pour une implémentation de la mise à jour des lois. Enfin, en statistique, la caractérisation d'un modèle n'a de sens que si elle est utilisée par la suite dans un contexte décisionnel. Tout d'abord, compte tenu des informations dont on dispose, nous devons procéder à une sélection des variables explicatives. Cette sélection peut être réalisée assez facilement dans un cadre bayésien au moyen de la sélection de modèles (Bernardo et Smith, [4]). Par ailleurs, le fabricant souhaite avoir à disposition un outil de maîtrise de son procédé. L'échantillonnage de Gibbs permet d'obtenir directement un échantillon de la loi prédictive du modèle. Cet échantillon peut être utilisé pour déterminer dans un contexte décisionnel, par exemple sous contrainte économique, la valeur optimale d'une variable de contrôle.

Remerciements

Nous remercions vivement M. Cazes pour ses remarques qui ont permis d'éclaircir les points insuffisamment précis d'une première version de ce papier.

Annexe 1. Présentation de l'échantillonnage de Gibbs

Pour utiliser l'échantillonnage de Gibbs, nous devons, relativement à une collection de k variables aléatoires, disposer de l'ensemble des lois conditionnelles complètes. De façon plus précise, pour une telle collection de variables aléatoires, $U = (U_1, \ldots, U_k)$, la loi jointe complète $p(U_1, \ldots, U_k)$ doit être déterminée de façon unique par la donnée des densités conditionnelles $p\left(U_i \left| U_{(\neq i)}\right.\right)$, $i = 1, \ldots, k$, où $U_{(j \neq i)}$ représente la collection complémentaire des $(k-1)$ variables aléatoires $(U_1, \ldots, U_{i-1}, U_{i+1}, \ldots, U_k)$. L'échantillonnage requiert que les densités conditionnelles complètes soient «disponibles», c'est-à-dire qu'elles doivent être générées facilement, connaissant les valeurs des variables $U_{(\neq i)}$.

L'échantillonnage de Gibbs est un schéma de mise à jour markovienne qui fonctionne de la manière suivante. Partant d'un ensemble de valeurs arbitraires de départ $U^{(0)} = \left(U_1^{(0)}, \ldots, U_k^{(0)}\right)$, nous générons $U_1^{(1)}$ suivant $p\left(U_1 \left| U_2^{(0)}, \ldots, U_k^{(0)}\right.\right)$ puis $U_2^{(1)}$ suivant $p\left(U_2 \left| U_1^{(1)}, U_3^{(0)}, \ldots, U_k^{(0)}\right.\right)$, ainsi de suite jusqu'à $U_k^{(1)}$ suivant $p\left(U_k \left| U_1^{(1)}, \ldots, U_{k-1}^{(1)}\right.\right)$ pour compléter la première itération du schéma. Ainsi, chaque variable est générée selon un ordre naturel et une itération de ce schéma demande k générations aléatoires. Après M itérations, nous obtenons $U^{(M)} = \left(U_1^{(M)}, \ldots, U_k^{(M)}\right)$. Sous certaines conditions faibles de régularité sur la loi jointe et sur les lois conditionnelles, il peut être montré que le vecteur généré $U^{(M)}$ tend en loi vers la distribution jointe $p(U_1, \ldots, U_k)$ quand M tend vers l'infini (Geman and Geman, [Geman84]), c'est-à-dire $\left(U_1^{(M)}, \ldots, U_k^{(M)}\right) \xrightarrow{d} p(U_1, \ldots, U_k)$. Répétant de façon indépendante ce procédé N fois, nous obtenons N réplicats indépendants de U. L'échantillon $\left\{U^{(M)_g},\ g = 1, \ldots, N\right\}$ ainsi obtenu constitue un échantillon simulé de la loi jointe a posteriori de U . Alternativement, il est possible de générer une seule chaîne de longueur $M + N$, d'éliminer les M premiers tirages et de conserver les N derniers tirages pour constituer l'échantillon $\left\{U^{(g)},\ j = 1, \ldots, N\right\}$ (Gelfand and Smith [Gelfand90b]). C'est cette pratique que nous avons employée dans notre étude.

Nous pouvons remarquer que l'utilisation pratique de l'échantillonnage de Gibbs requiert la détermination de M et N. De façon générale, ces valeurs sont fixées après plusieurs expérimentations. Nous ne considérerons pas ce problème comme un obstacle dès lors que la génération de variable aléatoire n'est pas coûteuse en temps. Dans cette étude, les valeurs M et N ont été fixées respectivement à 100 et 2000 pour les exemples présentés en section 5.

Notons finalement que les estimations ci-dessous sont basées sur des tirages non indépendants. Il est alors nécessaire afin de calculer numériquement les écarts-types des estimateurs d'emprunter des méthodes du domaine des séries temporelles (Ripley [26]).

Le fonctionnement de l'échantillonnage de Gibbs est résumé dans la figure 8.

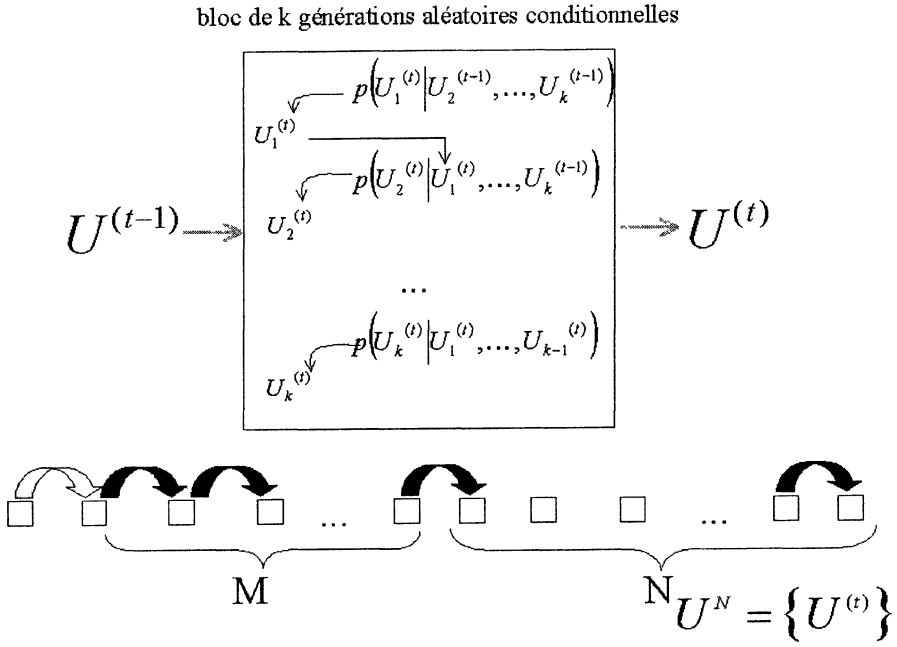

FIGURE 8
Echantillonnage de Gibbs.

Références

[1] I.V. BASAWA and L. BILLARD. Large-sample inference for a regression model with autocorrelated errors. *Biometrika*, 76 : 283–288, 1989.

[2] C.M. BEACH and J. G. MACKINNON. A maximum likelihood procedure for regression with autocorrelated errors. *Econometrica*, 46 : 51–58, 1978.

[3] J. O. BERGER. *Statistical decision theory and bayesian analysis*. Springer Verlag, New York, 2nd edition, 1980.

[4] J. M. BERNARDO and A. F. M. SMITH. *Bayesian theory*. Wiley, Londres, 1 edition 1994.

[5] G. E. P. BOX and G. C. TIAO. *Bayesian inference in statistical analysis*, Wiley, New York, 1 st edition, 1973.

[6] L. D. BROEMELING. Bayesian analysis of linear models. Marcel Dekker, Inc., New York and Basel, 1 edition, 1984.

[7] D. CHALTON and C. TROSKIE. Multiple regression with autocorrelated errors : bayesian analysis with different prior. *South African Statistical Journal*, 27 : 51–62, 1993.

[8] S. CHIB. Bayes regression with autoregressive errors. *Journal of econometrics*, 58 : 275–294, 1993.

[9] S. CHIB and E. GREENBERG. Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameters models. *Journal of econometrics*, 68 : 339–360, 1995.

[10] D. COCHRANE and G. ORCUTT. Application of least square regressions to relationships containing auto-correlated error terms. *Journal of the American Statistical Association*, 44 : 32–61, 1949.

[11] F. CRETTAZ DE ROTEN and J.-M. HELBLING. Données manquantes et aberrantes : le quotidien du statisticien analyste de données. *Revue de Statistiques Appliquées*, 44 : 105–115.

[12] M.G. DAGENAIS. Parameter estimation in regression models with errors in the variables and autocorrelated disturbances. *Journal of econometrics*, 64 : 145–163, 1994.

[13] P. DELLAPORTAS and D. A. STEPHENS. Bayesian analysis of error-in-variables regression models, *Biometrics*, 51 : 1085-1095, 1995.

[14] T.E. DIELMAN. Regression forecasts when disturbances are autocorrelated. *Journal of forecasting*, 4 : 263–271, 1985.

[15] H. DORAN. *Applied regression analysis in econometrics*. New York Marcel Dekker Inc. edition, 1989.

[16] N. R. DRAPER and H. SMITH. *Applied Regression Analysis*. Wiley, New York, 2 nd edition, 1981.

[17] T.B. FOMBY and D. K. GUILKEY. On choosing the optimal level of significance for the Durbin-Watson test and the bayesian alternative. *Journal of econometrics*, 8 : 203–213, 1978.

[18] A. GELFAND and A. SMITH. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85 : 398–409, 1990.

[19] A. GELMAN and J. B. CARLIN and H. S. STERN and D. B. RUBIN. *Bayesian data analysis*, Londres, Chapman and Hall edition, 1995.

[20] S. GEMAN and D. GEMAN. Stochastic relaxation, Gibbs distributions and the bayesian restoration of image. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 6 : 721–741, 1984.

[21] P. GIRARD. Les mécanismes physico-chimiques responsables de la viscosité du lait concentré sucré - Synthèse bibliographique. Rapport Nestlé, 1997.

[22] A.C. HARVEY and G. D. A. PHILLIPS. Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. *Biometrika*, 66 : 49-58, 1979.

[23] H. JEFFREYS. Theory of probability, Oxford University Press, New York, 3 rd edition, 1961.

[24] E. L. LEHMAN. *Theory of point estimation*, John Wiley and Sons, New York, 1 st edition, 1983.

[25] D.V. LINDLEY and A. F. M. SMITH. Bayes estimates for the linear model - with discussion, *Journal of the Royal Statistical Society B*, 34 : 1–41, 1972.

[26] J. NETER and W. WASSERMAN and M. H. KUTNER. *Applied linear statistical model - Regression, analysis of variance and experimental design.* IRWIN, Honewood, 2 nd edition, 1985.

[27] B. D. RIPLEY. *Stochastic simulation.* Wiley, New York, 1 edition, 1987.

[28] C. ROBERT. *Analyse statistique bayesienne.* Economica, Paris, 1 st edition, 1992.

[29] C. ROBERT. *Méthode de simulation de Monte-Carlo par chaînes de Markov.* Economica, paris, economica edition, 1996.

[30] M. A. TANNER. *Tools for statistical inference : methods for the exploration of posterior distribution and likelihood functions.* Springer Verlag, New York, 3rd edition, 1996.

[31] S. WEISBERG. *Applied linear regression.* John Wiley, New York, 2 nd edition, 1985.

[32] A. ZELLNER, *An introduction to Bayesian Inference in econometrics.* Krieger Robert E., Londres, 1 st edition, 1971.

[33] A. ZELLNER and G. C. TIAO. Bayesian analysis of the regression model with autocorrelated errors. *Journal of the American Statistical Association*, 59 : 763–778, 1964.