
Critères BIC et AIC pour les chaînes de Markov cachées

Application aux communications numériques

Noura Dridi, Yves Delignon, Wadih Sawaya

Institut Mines Telecom/Telecom Lille, LAGIS UMR CNRS 8219
Cité Scientifique-Rue Guglielmo Marconi
59658 Villeneuve d'Ascq Cedex, France
prenom.nom@telecom-lille.fr

RÉSUMÉ. Dans cet article, nous proposons de développer les critères d'Information d'Akaike (AIC) et Bayésien (BIC) pour sélectionner l'ordre des chaînes de Markov cachées (CMC) et pour la sélection de modèles CMC. Les critères de sélection asymptotiques d'Akaike (AIC) ou Bayésien (BIC) devant respecter l'indépendance des données, nous proposons dans le cas des modèles de Markov cachés d'appliquer les critères sur les données observées conditionnelles aux données cachées préalablement estimées. Les algorithmes de sélection de modèles et d'estimation aveugle sont par la suite développés dans le cadre des communications numériques large bande. Les performances de ces critères sont évaluées au regard de la taille de l'échantillon et de la forme du canal en termes de taux de bonne sélection et de taux d'erreur binaire. En particulier, la pertinence de l'estimation aveugle avec sélection de modèles est comparée à l'estimation aveugle de modèles combinés.

ABSTRACT. This paper aims at developing both the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) for selecting the order of hidden Markov Chain (HMC) and for selecting HMC parametric models. Since AIC and BIC selection methods require the independence of the data, the proposed AIC and BIC methods are based on the observed data and the estimated hidden process. Order selection algorithms and model selection algorithms coupled to blind estimation methods are subsequently studied in the frame of wide band communication systems. Performances of algorithms are assessed with respect to both the size of the samples and the shape of the channel in terms of root mean square error of the channel and of bit error rate. Finally, the relevance of the blind estimation joint to a model selection is compared to the blind estimation of combined models.

MOTS-CLÉS : chaînes de Markov cachées, AIC, BIC, estimation du canal.

KEYWORDS: hidden Markov chain, AIC, BIC, channel estimation.

DOI:10.3166/TS.31.383-400 © 2014 Lavoisier

Extended abstract

Model selection methods have been used in various signal processing areas such as remote sensing image, image compression or digital communication. Amongst the numerous selection methods, the Akaike and the Bayesian information criteria are the most cited one. The first is based on the minimization of the Kullback-Leibler divergence, while the later maximizes the posterior probability. Both of them assume the independence of the observed data. In digital communication, the emitted signal is received through a channel modelled by a finite impulse response filter with additive Gaussian noise. So, blind receiver consists in estimating both the channel coefficients, the noise power and the transmitted symbols. When dealing with supervised estimation, a priori informations about the channel length and its shape are taken into account in order to guide the algorithm to the true channel and to estimate the transmitted symbols. In the absence of a priori information, in addition to symbols, both coefficients and the order of the channel are arbitrary. Most of usual blind estimation algorithms consist in estimating only the channel and the symbols, assuming known the channel length or its shape. In case of the joint channel estimation and symbol detection, the couple of hidden process and observed one are modelled as a hidden Markov channel (HMC) so that the observed data don't respect the independence hypothesis. We propose to develop AIC and BIC criteria for HMC model by exploiting the observed data and the estimated hidden process. We present results of both the order model selection method and parametric shape model selection method in the framework of blind channel estimation and symbol detection. Performance results attest the relevance of including a model selection method in terms of root mean square error of the channel and of bit error rate.

1. Introduction

Les méthodes de sélection de modèles ont fait l'objet d'un intérêt soutenu par la communauté scientifique depuis les années 1970. Parmi celles-ci, les critères entropiques pénalisés sélectionnent le modèle en fonction de la log-vraisemblance maximale pénalisée par le nombre de paramètres du modèle et la taille de l'échantillon. Les critères les plus cités sont ceux d'Akaike (AIC) (Akaike, 1974) et le critère d'information Bayésien (BIC) (Schwarz, 1978). Le critère AIC est connu pour privilégier les modèles surparamétrés par rapport au critère du BIC pour lequel la pénalité de surparamétrisation est plus grande. Parmi les extensions du critère AIC, le critère AICc est préféré lorsque le nombre de paramètres est comparable à la taille de l'échantillon (Sugiura, 1978 ; Hurvich, Chih Ling, 1989). Les critères de sélection de modèles sont largement utilisés en traitement du signal (Stoica, Selen, 2004), télédétection (Bouyahia *et al.*, 2008), compression d'image (Coq *et al.*, 2007), communication numérique (Liavas *et al.*, 1999), sociologie (Raftery, 1995)... Parce qu'ils permettent de prendre en compte des dépendances spatiales ou temporelles des processus, les modèles de Markov sont aussi très présents dans de nombreux secteurs du traitement des données tels que la reconnaissance de voix (Allen, 1971), le traitement du signal/image (Kaleh, Vallet, 1994 ; Anton-Haro, Fonollosa, 1997),

l'imagerie radar (Fjortoft *et al.*, 2003 ; Tison *et al.*, 2004)... Pour ce type de processus, l'ordre du modèle de Markov peut être obtenu par un critère de sélection d'ordre décrit dans les articles (Tong, 1975 ; Olivier *et al.*, 1997).

Dans cet article, nous proposons d'étudier les méthodes de sélection de modèles de chaînes de Markov cachées (CMC) en considérant les observations et les données manquantes estimées. Nous montrons que les critères AIC et BIC gardent les mêmes expressions asymptotiques pour les CMC. Les méthodes de sélection sont ensuite appliquées aux communications numériques à large bande, en particulier pour sélectionner l'ordre du canal à travers lequel est transmis le signal d'entrée. Les différents critères sont étudiés en fonction des paramètres du modèle tels que la taille de l'observation, l'ordre du filtre et la forme de la réponse impulsionnelle (symétrique, exponentielle, uniforme).

Le reste du document est organisé comme suit : la deuxième section est consacré aux critères de sélection de modèles CMC par log-vraisemblance pénalisée. Dans la section 3, nous détaillons le problème d'estimation aveugle de symboles issus d'un canal à large bande. Les algorithmes de sélection de l'ordre du canal sont développés dans la section 4 alors que les méthodes de sélection de modèles paramétriques CMC sont décrits dans la section suivante. Les performances des algorithmes sont analysées dans la section 6 à la fois pour les méthodes de sélection d'ordre des CMC et pour les méthodes de sélection de modèles paramétriques CMC. Ces résultats sont illustrés sur la problématique de lecture de codes graphiques dont les flous de mouvement, de défocalisation ou combiné sont connus. En particulier, la comparaison entre sélection de modèles et combinaison de modèles (ici convolution de réponse impulsionnelle) pour l'estimation aveugle de symboles est étudiée.

Notation

x^T est la transposée de la matrice x . I_{L-1} est la matrice d'identité de taille $L - 1 \times L - 1$. Les lettres majuscules désignent des données aléatoires, les caractères gras sont utilisés pour les vecteurs ou les matrices.

2. Sélection de modèles pour les chaînes de Markov cachées

Considérons N observations $y = (y_1, y_2, \dots, y_N)$ indépendantes et distribuées suivant la même densité g tel que $\int |\ln g(y)| g(y) dy < \infty$. Considérons de plus un ensemble de K modèles probabilistes $\mathcal{M} = \{M_i, i \in \{1, \dots, K\}\}$ contenant g et où chaque modèle M_i est défini par :

$$M_i = \{f_{Y|\theta_i}(y)\}, i \in \{1, \dots, K\}$$

où θ_i est un vecteur de paramètres de taille K_i , et $f_{Y|\theta_i}(y)$ est la vraisemblance des observations.

Les méthodes de sélection de modèles consistent à choisir parmi K modèles celui qui optimise un critère spécifique. Le critère d'information Bayésien (BIC) consiste à maximiser la probabilité a posteriori :

$$M_{BIC}^* = \arg \max_i P(M_i | y) \quad (1)$$

(Schwarz, 1978) démontra qu'asymptotiquement, la maximisation de la probabilité a posteriori revenait à pénaliser la log-vraisemblance par un terme fonction de la taille de l'échantillon et du nombre de paramètres du modèle :

$$M_{BIC}^* = \arg \min_i [-2l_{M_i}(\tilde{\theta}_i) + K_i \ln N] \quad (2)$$

où $l_{M_i}(\cdot)$ est la log-vraisemblance du modèle M_i , $l_{M_i}(\tilde{\theta}_i) = \ln f_{Y|\tilde{\theta}_i}(y)$. $\tilde{\theta}_i$ est l'estimateur par maximum de vraisemblance de θ_i . Le critère d'Akaike (AIC) est basé sur la minimisation de la divergence de Kullback-Leibler entre la distribution d'un modèle de \mathcal{M} et la vraie distribution des observations Y notée g :

$$M_{AIC}^* = \arg \min_i KL(g | f_{Y|\tilde{\theta}_i}) \quad (3)$$

avec $KL(g | f_{Y|\theta}) = \int g(y) \ln \frac{g(y)}{f_{Y|\theta}(y)} dy$ la divergence de Kullback-Leibler. (Akaike, 1974) montre qu'asymptotiquement, le critère est équivalent à :

$$M_{AIC}^* = \arg \min_i [-2l_{M_i}(\tilde{\theta}_i) + 2K_i] \quad (4)$$

Le critère d'Akaike corrigé AICc (Hurvich, Chih Ling, 1989 ; Sugiura, 1978) est valable pour des échantillons de taille réduite par rapport au nombre des paramètres. La formule asymptotique du critère est donnée par :

$$M_{AICc}^* = \arg \min_i [-2l_{M_i}(\tilde{\theta}_i) + 2 \frac{NK_i}{N - K_i - 1}] \quad (5)$$

avec $N > K_i - 1$. L'utilisation de ces critères pour les chaînes de Markov cachées, présente une difficulté liée à la dépendance des observations et au calcul de la log-vraisemblance. Dans ce travail nous proposons de construire les critères de sélection de modèles en se basant sur les données complètes, c'est à dire les données observées Y et les données cachées S qui sont au préalable estimées. Dans ce cas les lois conditionnelles $f_{Y|S,\theta_i}$ composent l'ensemble des modèles \mathcal{M} . Les observations étant indépendantes entre elles conditionnellement au processus caché, les démonstrations d'Akaike et Schwarz s'appliquent directement, les versions asymptotiques sont données par les équations (2, 4, 5), avec $l_{M_i}(\theta_i) = \ln f_{Y|S,\theta_i}(y)$ la log-vraisemblance conditionnelle et K_i le nombre de paramètres libres, ici la dimension du vecteur θ_i . L'application qui nous concerne est assimilable à l'estimation conjointe symbole canal *Simple Input Simple Output (SISO)* large bande en communication numérique. A notre connaissance, les travaux publiés les plus proches, (Liavas *et al.*, 1999 ; Gerstacker, Taylor, 2003 ; Via, Santamaria, 2006) sont basés sur des techniques de décomposition en sous-espaces, et donc adaptés à des canaux SIMO ou SISO avec suréchantillonnage à la réception. Ils ne sont donc pas adaptés à notre problématique.

3. Application à la problématique d'estimation du canal et de détection de symboles

L'objectif est d'estimer le canal et de détecter les symboles à partir d'une observation flouée et bruitée. Les données manquantes ne sont pas observées mais sont estimées par la stratégie du maximum de la marginale a posteriori (MPM) faisant suite à l'algorithme Forward-Backward (Bahl *et al.*, 1974).

3.1. Modèle de communication large bande

Soit $\mathbf{X} = (X_0, \dots, X_{N-1})$ une séquence de symboles transmis à travers un canal modélisé par un filtre de Réponse Impulsionnelle Finie (RIF) $\mathbf{h} = (h_0, h_1, \dots, h_{L-1})^T$ de longueur L et un bruit blanc Gaussien de puissance σ^2 noté W_n . Les échantillons observés et les interférences entre symboles forment un modèle de Markov caché défini par les équations :

$$\mathbf{S}_n = \mathbf{b}\mathbf{S}_{n-1} + \mathbf{V}_n \quad (6)$$

$$Y_n = \mathbf{h}^T \mathbf{S}_n + W_n \quad (7)$$

où $\mathbf{S}_n = (X_n, \dots, X_{n-L+1})^T$, L est la longueur du canal, $\mathbf{b} = \begin{pmatrix} 0 & 0 \\ I_{L-1} & 0 \end{pmatrix}$

et $\mathbf{V}_n = (x_n \ 0 \ \dots \ 0)^T$ est un processus composé de variables aléatoires indépendantes et indépendant de W_n . D'après l'Eq. (6), \mathbf{S} est markovien et défini dans l'ensemble des états $\{\xi_0, \dots, \xi_{M-1}\}$ de cardinal $M = 2^L$ dans le cas de symboles binaires. Étant donnée la markovianité de $\mathbf{S} = (\mathbf{S}_0, \dots, \mathbf{S}_{N-1})$, la loi du couple (\mathbf{S}, \mathbf{Y}) est donnée par :

$$P_{\theta}(\mathbf{S}, \mathbf{Y}) = P(\mathbf{s}_0) \prod_{n=1}^{N-1} P(\mathbf{s}_n | \mathbf{s}_{n-1}) \prod_{n=0}^{N-1} f_{Y_n | \mathbf{s}_n, \theta}(y_n) \quad (8)$$

avec $f_{Y_n | \mathbf{s}_n, \theta}(y_n) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{1}{2} \left(\frac{y_n - \mathbf{h}^T \mathbf{s}_n}{\sigma} \right)^2$, $P(\mathbf{s}_n | \mathbf{s}_{n-1})$ est la probabilité de transition et $\theta = (\mathbf{h}, \sigma^2)^T$ est le vecteur de paramètres du modèle.

3.2. Estimation du canal et détection de symboles

L'estimation du canal est basée sur le critère du maximum de vraisemblance :

$$\hat{\theta} = \arg \max_{\theta} f_{\mathbf{Y}, \theta}(\mathbf{y}) \quad (9)$$

avec $f_{\mathbf{Y}, \theta}$ est la vraisemblance de l'observation. Comme la vraisemblance n'est pas calculable analytiquement, nous proposons d'approcher le maximum par l'algorithme

Espérance-Maximisation (EM). Il consiste en une maximisation itérative de la fonction auxiliaire $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(i-1)})$ donnée par :

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(i-1)}) = E_{\mathcal{S}|\mathbf{y}, \boldsymbol{\theta}^{(i-1)}}[\log(P_{\boldsymbol{\theta}}(\mathcal{S}, \mathbf{Y}))] \quad (10)$$

Les estimateurs des coefficients du canal et de la puissance du bruit sont calculés dans (Kaleh, Vallet, 1994). Ils sont donnés par :

$$\left\{ \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} P_{\boldsymbol{\theta}^{(i-1)}}(\mathcal{S}_n = \boldsymbol{\xi}_m | \mathbf{y}) \boldsymbol{\xi}_m^T \boldsymbol{\xi}_m \right\} \mathbf{h}^{(i)} = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} P_{\boldsymbol{\theta}^{(i-1)}}(\mathcal{S}_n = \boldsymbol{\xi}_m | \mathbf{y}) y_n \boldsymbol{\xi}_m \quad (11)$$

$$\sigma^{2(i)} = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} P_{\boldsymbol{\theta}^{(i-1)}}(\mathcal{S}_n = \boldsymbol{\xi}_m | \mathbf{y}) \times |y_n - (\mathbf{h}^T)^{(i)} \boldsymbol{\xi}_m|^2 \quad (12)$$

Pour calculer ces estimateurs, il faut d'abord calculer de la probabilité marginale a posteriori $P_{\boldsymbol{\theta}^{(i-1)}}(\mathcal{S}_n = \boldsymbol{\xi}_m | \mathbf{y})$ par l'algorithme Forward-Backward (Bahl *et al.*, 1974). Cette probabilité est également utilisée pour la détection de symboles qui se basent sur le Maximum de la Marginale a Posteriori (MPM). Pour chaque symbole x_n appartenant à l'ensemble $\{0, 1\}$, la décision est donnée par :

$$P_{\boldsymbol{\theta}}(x_n = 0 | \mathbf{y}) \stackrel{\hat{x}_n=1}{\underset{\hat{x}_n=0}{\lesseqgtr}} P_{\boldsymbol{\theta}}(x_n = 1 | \mathbf{y}) \quad (13)$$

avec,

$$P(x_n = i | \mathbf{y}, \boldsymbol{\theta}) = \sum_{\mathcal{S}_n} P(x_n = i, \mathcal{S}_n = \mathcal{S}_n | \mathbf{y}, \boldsymbol{\theta}) = \sum_{\mathcal{S}_n | x_n=i} P(\mathcal{S}_n = \mathcal{S}_n | \mathbf{y}) \quad (14)$$

4. Sélection de l'ordre du modèle

Les critères d'information trouvent naturellement un intérêt pour estimer l'ordre de modèles autoregressifs (Hannan, Quinn, 1979 ; Bhansali, Downham, 1977), ARMA (Stoica, Selen, 2004), pour la recherche du nombre des composantes sinusoïdales dans un signal périodique bruité, pour la taille d'un canal multitrajet en communication numérique (Liavas, Regalia, 2001 ; Gerstacker, Taylor, 2003) ou encore pour la détermination du nombre de classes en imagerie radar (Bouyahia *et al.*, 2008). Les travaux dédiés à la problématique d'estimation du canal et de détection de symboles supposent souvent connu l'ordre du filtre modélisant le canal (Anton-Haro, Fonollosa, 1997 ; Dridi *et al.*, 2010 ; Kaleh, Vallet, 1994). Dans plusieurs applications pratiques, cette information n'est pas disponible. Nous proposons dans cet article d'appliquer les trois critères définis par les équations (2, 4, 5) pour estimer l'ordre du canal. L'ordre du filtre ainsi que la puissance du bruit définissent le nombre de paramètres $K_i = L + 1$ alors que la log-vraisemblance calculée au maximum de vraisemblance s'écrit :

$$l_{M_i}(\tilde{\boldsymbol{\theta}}_i) = \ln f_{\mathbf{Y}|\tilde{\mathbf{s}}, \tilde{\boldsymbol{\theta}}_i, M_i}(\mathbf{y}) = -\frac{N}{2} \ln 2\pi\tilde{\sigma}^2 - \frac{1}{2\tilde{\sigma}^2} \sum_{n=1}^N (y_n - \tilde{\mathbf{h}}_i^T \tilde{\mathbf{s}}_n)^2 \quad (15)$$

où $\tilde{\boldsymbol{\theta}}_i$ est l'estimateur du maximum du vraisemblance. La méthode de sélection est décrite par l'algorithme 1.

Algorithme 1 Algorithme de sélection de l'ordre du modèle

-
- 1: Paramètres : K ensemble de valeurs possibles de L , ε seuil d'arrêt,
 - 2: **pour** $L_k \in K$ **faire**
 - 3: Calculer le nombre d'états et la matrice de transition de la CMC.
 - 4: Initialisation, $\mathbf{h}^{(0)}$, $(\sigma^2)^{(0)}$, pour L_k fixé.
 - 5: **tantque** $\left| \frac{Q(\boldsymbol{\theta}, \boldsymbol{\theta}^i) - Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(i-1)})}{Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(i-1)})} \right| > \varepsilon$ **faire**
 - 6: Calculer les probabilités Forward et Backward et la probabilité marginale a posteriori par l'algorithme Forward-Backward (Bahl *et al.*, 1974).
 - 7: Estimation de \mathbf{h} et σ^2 à partir des équations (11) et (12).
 - 8: **fin tantque**
 - 9: **fin pour**
 - 10: Détection de symboles comme décrit par les équations (13) et (14).
 - 11: Étant donné les paramètres $\tilde{\boldsymbol{\theta}} = (\tilde{\mathbf{h}}, \tilde{\sigma}^2)$ et les symboles estimés, calculer les coûts des critères AIC, AICc et BIC donnés par les équations (2, 4, 5).
 - 12: Sélectionner le vecteur de paramètres $\hat{\boldsymbol{\theta}}$ qui minimise le critère fixé.
-

5. Sélection de CMC

Lorsque plusieurs modèles paramétriques du canal sont en compétition, l'opération de sélection peut aussi être obtenue par une méthode AIC, BIC ou AICc. Dans ce cas, on ne cherche plus à définir le nombre de coefficients du canal mais sa forme parmi un ensemble de formes variées. Pour illustrer les performances de ces méthodes de sélection de modèles, nous nous plaçons dans le cadre de la lecture de codes graphiques pour lesquels les modèles de canaux à l'origine des différents flous sont connus.

5.1. Modèles de CMC

Deux phénomènes sont à l'origine du flou dans les images capturées par des systèmes de lecture optique ; le flou optique qui provient d'une mauvaise focalisation du système et le flou de mouvement induit par la mobilité du lecteur relativement au code. Ces deux flous, qui peuvent se combiner, présentent des dégradations spécifiques et sont modélisés par des filtres dont la forme de la réponse impulsionnelle dépend de leur nature (Houni *et al.*, 2008). Ainsi un flou optique est en général modélisé par une réponse impulsionnelle Gaussienne dont la variance conditionne l'importance du flou, (Horn, 1986). La réponse impulsionnelle du flou de mouvement est une fonction porte dont la longueur dépend de la vitesse et de la durée de pause de l'équipement (Grennery, 1973 ; Cannon, 1976 ; Kundur, Hatzinakos, 1996). Le flou combinant une mauvaise focalisation et un mouvement de lecteur a une réponse impulsionnelle définie par la convolution des deux réponses impulsionnelles précédemment citées. Les performances de détection de symboles dépendent de la précision d'estimation des paramètres qui s'améliore nettement en sélectionnant le modèle adéquat du flou. Nous

proposons d'appliquer les méthodes AIC, BIC et AICc pour la sélection. Les résultats sont donnés pour le cas 1D par soucis de simplicité et peuvent se généraliser au cas multidimensionnel sans difficulté.

Un flou de mouvement de paramètre D est modélisé par la fonction rectangulaire :

$$h_m(t) = \begin{cases} \frac{1}{D} & \text{if } |t| < D/2 \\ 0 & \text{sinon} \end{cases} \quad (16)$$

La fonction Gaussienne modélise un flou résultant d'une mauvaise focalisation, avec une variance σ_h^2 et une moyenne m :

$$h_o(t) = \frac{1}{\sqrt{2\pi\sigma_h^2}} \exp - \frac{1}{2} \left(\frac{t-m}{\sigma_h} \right)^2 \quad (17)$$

Un flou combiné est la convolution de deux composantes h_m et h_o :

$$\begin{aligned} h_c(t) &= (h_o * h_m)(t) \\ &= \frac{1}{2D} \left\{ \operatorname{erf} \left(\frac{t-m+\frac{D}{2}}{\sqrt{2\sigma_h^2}} \right) - \operatorname{erf} \left(\frac{t-m-\frac{D}{2}}{\sqrt{2\sigma_h^2}} \right) \right\} \end{aligned} \quad (18)$$

5.2. Estimation des modèles par maximum de vraisemblance

L'estimateur du maximum de vraisemblance des paramètres est approché par l'algorithme EM, la pseudo-vraisemblance dans le cas d'une CMC s'écrivant :

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(i-1)}) = \text{cte} - \frac{1}{2} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} P(\mathbf{S}_n = \boldsymbol{\xi}_m | \mathbf{y}, \boldsymbol{\theta}^{(i-1)}) \left\{ \log(2\pi\sigma^2) + \frac{|y_n - \mathbf{h}_j^T \boldsymbol{\xi}_m|^2}{\sigma^2} \right\} \quad (19)$$

avec $j \in \{o, m, c\}$ et i l'itération de l'algorithme EM.

Dans le cas d'un flou de mouvement Eq. (16), la maximisation de la fonction Q à l'itération i donne :

$$\hat{D}^{(i)} = \frac{\sum_{n=0}^{N-1} \sum_{m=0}^{M-1} P(\mathbf{S}_n = \boldsymbol{\xi}_m | \mathbf{y}, \boldsymbol{\theta}^{(i-1)}) \boldsymbol{\xi}_m^T \boldsymbol{\xi}_m}{\sum_{n=0}^{N-1} \sum_{m=0}^{M-1} P(\mathbf{S}_n = \boldsymbol{\xi}_m | \mathbf{y}, \boldsymbol{\theta}^{(i-1)}) y_n \sum_{i=1}^L \boldsymbol{\xi}_{m,i}} \quad (20)$$

avec $\boldsymbol{\xi}_m$ le vecteur d'état de dimension L , $\boldsymbol{\xi}_{m,i}$ est la $i^{\text{ième}}$ composante de ce vecteur. Pour un flou combiné ou optique Eq. (17), les paramètres sont la moyenne m , la variance σ_h^2 et la puissance du bruit σ^2 . Pour m et σ_h^2 il n'est pas possible de

calculer l'expression analytique des estimateurs, le maximum est alors approché par la méthode de descente du gradient. La $k^{ième}$ itération de l'algorithme de la descente du gradient appliquée à l' $i^{ième}$ itération de l'algorithme EM :

$$\boldsymbol{\theta}^{(k)(i)} = \boldsymbol{\theta}^{(k-1)(i)} - \nu \nabla (Q(\boldsymbol{\theta}^{(k-1)(i-1)}, \boldsymbol{\theta}^{(i-1)})) \quad (21)$$

avec $\boldsymbol{\theta}^{(k)(i)} = ((\sigma_{\mathbf{h}}^2)^{(k,i)}, (\sigma^2)^{(i)})$, $\boldsymbol{\theta}^{(i-1)} = ((\sigma_{\mathbf{h}}^2)^{(i-1)}, (\sigma^2)^{(i-1)})$ et ν désigne le pas de l'algorithme. Étant donné les estimateurs de $\sigma_{\mathbf{h}}^2$, m et D on déduit une estimation du canal Eq. (17) et (??). Les algorithmes d'estimation du flou de mouvement, du flou optique ou du flou combiné sont décrits dans l'algorithme 2.

Algorithme 2 Algorithme d'estimation d'un flou de mouvement, d'un flou optique ou d'un flou combiné

- 1: Calculer le nombre d'états et la matrice de transition de la CMC.
 - 2: Initialisation flou de mouvement $D^{(0)}$, $(\sigma^2)^{(0)}$, flou optique $m^{(0)}$, $(\sigma_{\mathbf{h}}^2)^{(0)}$, $(\sigma^2)^{(0)}$, flou combiné $D^{(0)}$, $(\sigma^2)^{(0)}$, $m^{(0)}$, $(\sigma_{\mathbf{h}}^2)^{(0)}$, $(\sigma^2)^{(0)}$.
 - 3: **tantque** $\left| \frac{Q(\boldsymbol{\theta}, \boldsymbol{\theta}^i) - Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(i-1)})}{Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(i-1)})} \right| > \varepsilon$ **faire**
 - 4: Calculer les probabilités Forward et Backward et déduire la probabilité marginale a posteriori par l'algorithme proposé dans (Bahl *et al.*, 1974).
 - 5: Si flou de mouvement, calculer \hat{D} à partir de (20). Si flou optique, estimer m et $\sigma_{\mathbf{h}}^2$ par l'algorithme du gradient (21). Si flou combiné, estimation de \hat{D} , m et $\sigma_{\mathbf{h}}^2$ par l'algorithme du gradient (21).
 - 6: Si flou de mouvement, \mathbf{h}_m obtenu par (16). Si flou optique, calcul de \mathbf{h}_o par (17). Si flou combiné, \mathbf{h}_e obtenu par (??).
 - 7: Estimation de la puissance du bruit σ^2 selon (12).
 - 8: **fin tantque**
-

6. Résultats

Dans cette partie, nous étudions les performances de différents critères de sélection (BIC, AIC, AICc). Dans une première partie ils sont utilisés pour estimer l'ordre du modèle, et dans une deuxième partie pour choisir un modèle paramétrique pour le canal parmi plusieurs.

6.1. Sélection de l'ordre

Nous évaluons en particulier l'impact de la forme du canal (symétrique, exponentielle,...) ainsi que l'influence de la taille de la séquence observée. Nous considérons $N_s = 100$ simulations, chacune consiste à générer un nouveau signal avec $N = 100$ symboles et un nouveau canal d'ordre L . Nous présentons d'abord les résultats pour un canal dont la RIF est symétrique d'ordre $L = 4$ et le rapport $E_b/N_0 = 5 \text{ dB}, 10 \text{ dB}, 15 \text{ dB}$. Le but est d'estimer la longueur du canal à partir d'une liste de valeurs candidates $\{1, \dots, 6\}$ à l'aide de l'algorithme 1. Nous

présentons également les résultats dans le cas où les symboles X sont connus, notés respectivement (BIC-XC, AIC-XC, AICc-XC). Ces résultats servent de référence pour apprécier l'impact de l'estimation de symboles sur le critère de sélection.

6.1.1. Cas d'une réponse impulsionnelle symétrique

Pour chaque critère, le nombre d'erreurs de sélection est donné dans le tableau 1, pour différents rapports signal sur bruit.

Nous constatons que le nombre d'erreurs diminue en augmentant E_b/N_0 , ceci est dû à l'amélioration de la qualité de l'estimation des paramètres du modèle et des symboles qui impactent directement le coût du critère. Par ailleurs, il est clair pour les algorithmes supervisés (BIC-XC, AIC-XC, AICc-XC) que le nombre de fois où l'algorithme se trompe est plus faible. Remarquons que le critère AICc n'améliore pas vraiment les performances, en effet les gains en performance comparés aux méthodes de sélection de modèles AIC ou BIC sont généralement plus significatifs quand le nombre de paramètres croît comparativement à la taille de l'échantillon ce qui n'est pas le cas pour notre application. L'impact de la sélection du modèle sur le taux d'erreur binaire (TEB) est illustré sur la figure 1. La courbe en trait continu présente le vrai modèle pour lequel l'ordre du canal est connu. L'écart entre cette courbe et les autres est dû aux erreurs commises par les algorithmes de sélection de l'ordre du modèle. Remarquons que cet écart décroît en augmentant E_b/N_0 . Enfin, nous notons la supériorité de l'algorithme BIC pour $E_b/N_0 \geq 10$.

Tableau 1. Taux d'erreur (%) en fonction de E_b/N_0 , pour un canal symétrique d'ordre $L = 4$, $N = 100$, Nombre de simulations $N_s = 100$

Algorithme	BIC	AIC	AICc	BIC-XC	AIC-XC	AICc-XC
$E_b/N_0 = 5 \text{ dB}$	97	91	91	18	19	19
$E_b/N_0 = 10 \text{ dB}$	40	44	39	13	13	13
$E_b/N_0 = 15 \text{ dB}$	6	12	11	1	1	1

L'étude des performances en fonction de la taille de la séquence observée N est donnée tableau 2. Nous constatons une nette dégradation des performances pour tous les critères pour $N = 10$. En effet, les critères sont asymptotiques ce qui nécessite des séquences observées suffisamment grandes. En outre, les performances sont améliorées en augmentant la taille de l'échantillon ainsi un taux d'erreur $< 15\%$ est atteint pour $N \geq 100$.

6.1.2. Cas d'une réponse impulsionnelle exponentielle

Nous considérons un canal dont la réponse impulsionnelle est exponentielle de la forme $h_i = \lambda^i$ pour $i \in \{1, \dots, 4\}$ avec $\lambda = 0,8$ et d'ordre $L = 4$. En comparant les résultats donnés par le tableau 3 avec ceux qui correspondent à une réponse impulsionnelle symétrique tableau 1, nous constatons que le nombre d'erreur d'un même critère augmente. En effet pour une réponse impulsionnelle exponentielle, le coût moyen du critère décroît avec l'ordre du modèle et stagne à partir de la vraie

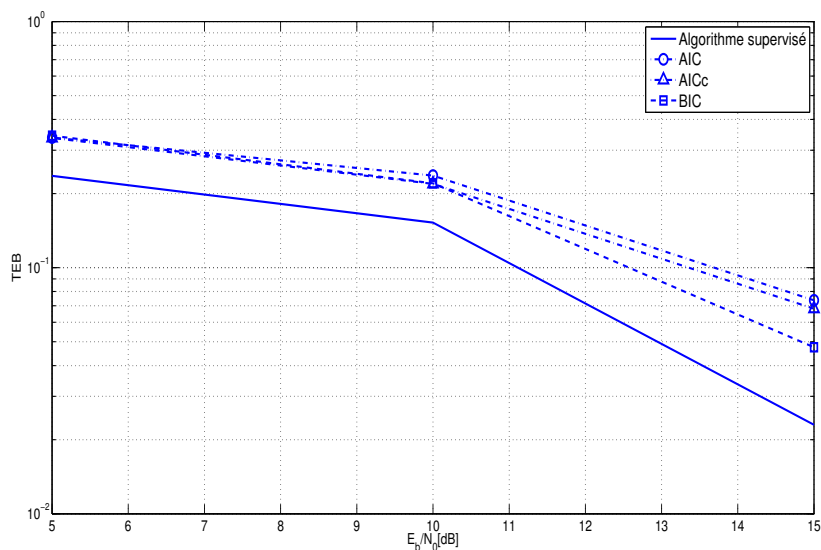


Figure 1. TEB en fonction de E_b/N_0 , pour une réponse impulsionnelle symétrique d'ordre $L = 4$ avec $N = 100$

Tableau 2. Taux d'erreur (%) de sélection en fonction de N , pour une réponse impulsionnelle symétrique d'ordre $L = 4$. Nombre de simulations $N_s = 100$

Algorithme	$E_b/N_0 = 10$			$E_b/N_0 = 15$		
	BIC	AIC	AICc	BIC	AIC	AICc
$N = 300$	11	33	31	0	6	6
$N = 100$	40	44	39	6	12	11
$N = 10$	85	84	100	63	64	91

valeur de $L = 4$. Les écarts des coûts sont minimales pour des ordres différents, ce qui rend la sélection incertaine figure 3. Toutefois, ce résultat ne dégrade pas vraiment les performances de détection de symboles comme le montre la figure 2. Le TEB est comparable à celui obtenu avec l'algorithme supervisé.

Tableau 3. Taux d'erreur (%) de sélection pour les différents critères en fonction de E_b/N_0 , pour une réponse impulsionnelle exponentielle d'ordre $L = 4$, avec $N = 100$, Nombre de simulations $N_s = 100$

Algorithme	BIC	AIC	AIC-c	BIC-XC	AIC-XC	AICc-XC
$E_b/N_0 = 5 \text{ dB}$	79	84	83	34	37	35
$E_b/N_0 = 10 \text{ dB}$	24	51	46	12	18	17
$E_b/N_0 = 15 \text{ dB}$	8	27	23	6	19	16

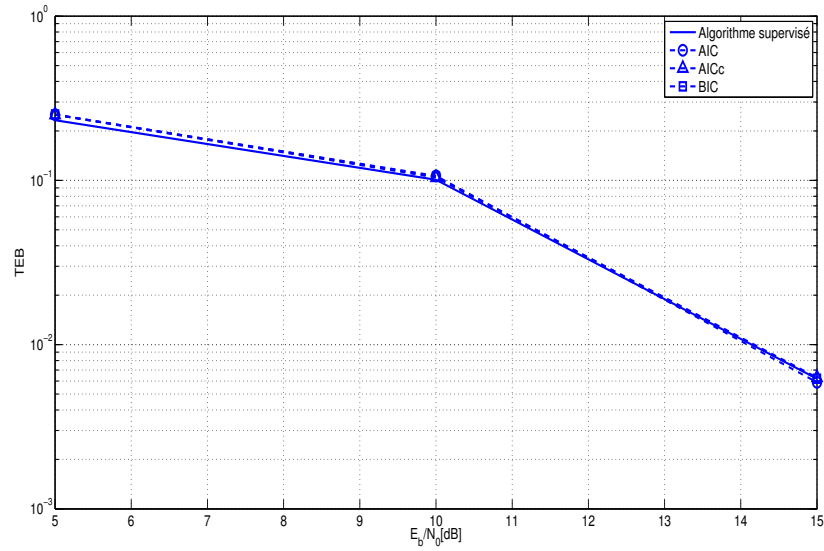


Figure 2. TEB en fonction de E_b/N_0 , pour une réponse impulsionnelle exponentielle d'ordre $L = 4$ avec $N = 100$

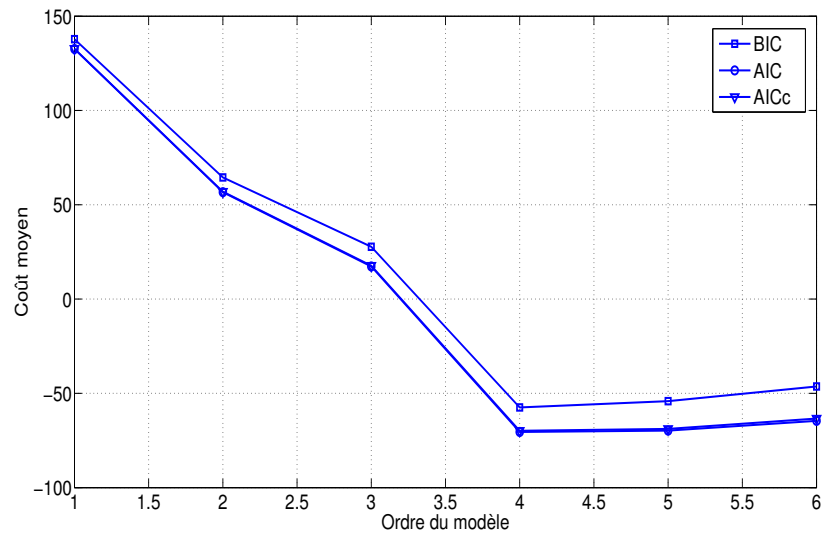


Figure 3. Moyenne des coûts des différents critères pour une réponse impulsionnelle exponentielle, la moyenne est calculée sur les $N_s = 100$ simulations. Nombre de symboles émis $N = 100$, $L = 4$ et $E_b/N_0 = 15$

6.1.3. Cas d'une réponse impulsionnelle uniforme

Nous considérons une réponse impulsionnelle uniforme d'ordre $L = 4$, pour chaque expérience un nouveau canal aléatoire est généré dont les coefficients sont égaux ; $h = [c \ c \ c \ c]$ avec c un nombre aléatoire. Les résultats sont présentés dans le tableau 4.

En comparant avec les précédents résultats, nous remarquons que les performances sont meilleures pour un canal dont la réponse impulsionnelle est uniforme par rapport à une réponse impulsionnelle symétrique ou exponentielle. Tous les coefficients du canal ayant le même poids pour le critère (puisque'ils sont égaux), pour différents ordres les écarts entre les coûts sont alors significatifs et la sélection du vrai ordre devient plus facile, figure 4.

Tableau 4. Nombre d'erreurs en fonction de E_b/N_0 , pour une réponse impulsionnelle uniforme avec, les paramètres $N = 100$, $L = 4$. Nombre de simulations $N_s = 100$

Algorithme	BIC	AIC	AICc	BIC-XC	AIC-XC	AICc-XC
$E_b/N_0 = 5 \text{ dB}$	92	83	83	3	3	3
$E_b/N_0 = 10 \text{ dB}$	8	8	8	0	0	0
$E_b/N_0 = 15 \text{ dB}$	0	0	0	0	0	0

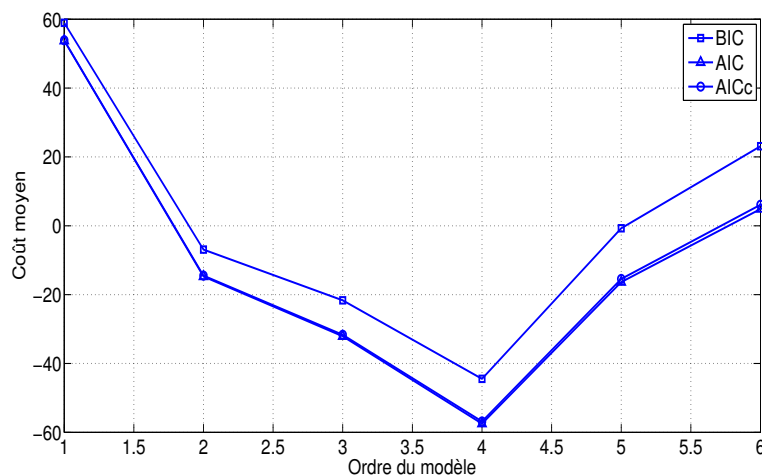


Figure 4. Moyenne des coûts des différents critères pour une réponse impulsionnelle uniforme, la moyenne est calculée sur les $N_s = 100$ simulations. Nombre de symboles émis $N = 100$, $L = 4$ et $E_b/N_0 = 10$

6.2. Sélection du modèle de flou

Nous proposons dans cette section d'appliquer les critères de sélection du modèle pour identifier le modèle le plus en adéquation avec les données observées. Les algorithmes suivants sont évalués en termes de REQM (racine carrée de l'erreur quadratique moyenne de la réponse impulsionnelle) et de TEB :

- Un algorithme supposant connu le type de dégradation et qui estime les paramètres du modèle utilisé pour générer les observations par l'algorithme EM. Appelé «vrai modèle», cet algorithme sert de référence aux autres algorithmes.
- Un algorithme qui n'inclut pas la sélection du modèle, il considère toujours un modèle du flou combiné, appelé «Flou combiné».
- Des algorithmes qui procèdent à une sélection du modèle utilisant les différents critères proposés, et décrits par l'algorithme 3.

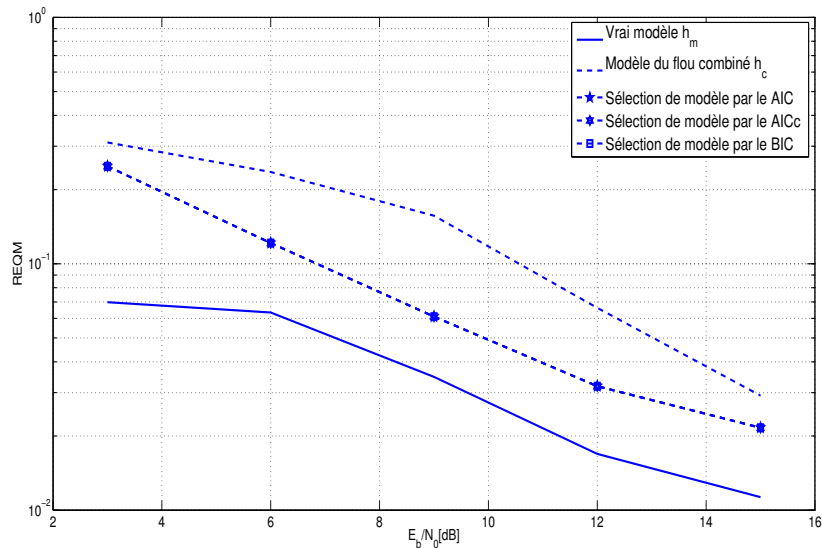


Figure 5. REQM en fonction de E_b/N_0 pour un flou de mouvement de taille $L = 3$, avec $N = 100$

Les performances des algorithmes de sélection sont données pour un flou de mouvement de paramètre $D = 3$. Pour chaque E_b/N_0 , $N_s = 100$ séquences de taille $N = 100$ sont simulées. Les figures 5 et 6 montrent que les algorithmes utilisant les critères de sélection de modèle offrent des meilleures performances en terme de la REQM et du TEB par rapport à un algorithme qui estime directement un modèle du flou combiné. Cet écart est expliqué par le fait qu'un modèle combiné inclut une composante du flou optique qui n'est pas présent dans le vrai modèle. Par ailleurs,

nous remarquons que les performances des algorithmes de sélection s'approchent de celles de l'algorithme estimant le vrai modèle.

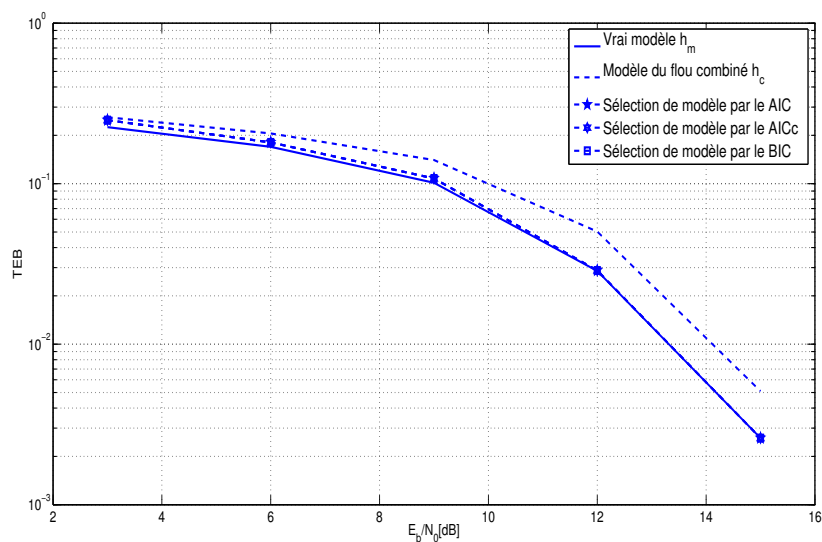


Figure 6. TEB en fonction de E_b/N_0 pour un flou de mouvement de taille $L = 3$, avec $N = 100$

Algorithme 3 Algorithme de sélection basé sur les données complètes

- 1: Paramètres $(\mathbf{h}_j)_{j \in \{m, o, c\}}$ les trois modèles candidats du flou, ε seuil d'arrêt,
 - 2: **pour** $j \in \{m, o, c\}$ **faire**
 - 3: Estimer la réponse impulsionnelle du canal \mathbf{h}_j et la puissance du bruit σ^2 en utilisant l'algorithme 2.
 - 4: Étant donné les paramètres $\tilde{\boldsymbol{\theta}} = (\tilde{h}, \tilde{\sigma}^2)$ et symboles estimés, calculer les coûts de critères AIC, AIC-c et BIC donnés par les équations (4, 5, 2).
 - 5: Détection de symboles comme décrit dans la section 3.2.
 - 6: **fin pour**
 - 7: Sélectionner le vecteur des paramètres $\tilde{\boldsymbol{\theta}}$ qui minimisent les trois coûts.
-

7. Conclusion

En se basant sur les données observées et sur l'estimation des données cachées, les critères de sélection de modèles AIC et BIC ont été adaptés aux CMC. Ces critères ont été ensuite appliqués à l'estimation aveugle en communication numérique à large bande en considérant le canal comme un ensemble indéterminé de coefficients inconnus puis en considérant plusieurs familles de modèles de canaux (réponse impulsionnelle gaussienne, uniforme, ou combinée entre les deux). Les performances des méthodes de sélection d'ordre et celles de sélection de modèles ont été étudiées dans plusieurs conditions d'expérimentation. Les résultats montrent que la réponse impulsionnelle du canal impacte de manière significative les performances des algorithmes en fonction de sa forme (symétrique, exponentielle, uniforme) et en particulier que le critère BIC offre les meilleures performances pour les canaux symétriques. L'importance de la taille de l'échantillon pour respecter les contraintes asymptotiques des critères a été analysée. Le taux d'erreur de 5 % est obtenu pour un échantillon de taille $N \geq 100$. Enfin, la comparaison de l'estimation aveugle basée sur le modèle issu d'une sélection à l'estimation aveugle du modèle combinant l'ensemble des modèles (convolution des deux réponses impulsionnelles) donne nettement l'avantage à l'algorithme associant une méthode de sélection.

Bibliographie

- Akaike H. (1974, Dec). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, vol. 19, n° 6, p. 716 - 723.
- Allen D. (1971). The prediction sum of square as a criterion of selecting predicting variables. *Technical report. Univ of Kentucky, Dep of Statistics.*, n° 23.
- Anton-Haro C., Fonollosa J. (1997, Jan). Blind channel estimation and data detection using hidden Markov models. *IEEE Transactions on Signal Processing*, vol. 45, n° 1, p. 241-247.
- Bahl L., Cocke J., Jelinek F., Raviv J. (1974, Mar). Optimal decoding of linear codes for minimizing symbol error rate (corresp). *IEEE Transactions. Information. Theory*, vol. 20, n° 2, p. 284 - 287.
- Bhansali R. J., Downham D. Y. (1977). Some properties of the order of an autoregressive model selected by a generalization of Akaike's epf criterion. *Biometrika*, vol. 64, n° 3, p. 547-551.
- Bouyahia Z., Benyoussef L., Derrode S. (2008). Change detection in synthetic aperture radar images with a sliding hidden Markov chain model. *J. of Applied Remote Sensing*, vol. 2, n° 1, p. 392 - 403.
- Cannon M. (1976, Feb). Blind deconvolution of spatially invariant image blurs with phase. *IEEE Transactions. Acoustics, Speech, Signal Processing*, vol. 24, n° 1, p. 58-63.
- Coq G., Olivier C., Alata O., Arnaudon M. (2007). Information criteria and arithmetic codings: an illustration on raw images. *European Signal Processing Conference*, p. 634-638.
- Dridi N., Delignon Y., Sawaya W., Septier F. (2010, Dec). Blind detection of severely blurred 1d barcode. *IEEE Global Communications Conference, Miami, FL, USA*, p. 1-5.

- Fjortoft R., Delignon Y., Pieczynski W., Sigelle M., Tupin F. (2003). Unsupervised segmentation of radar images using hidden Markov chains and hidden Markov random fields. *IEEE Transactions. on Geoscience and Remote Sensing*, vol. 3, n° 41, p. 675-686.
- Gerstacker W., Taylor D. (2003, Feb). Blind channel order estimation based on second-order statistics. *IEEE Signal Processing Letters*, vol. 10, n° 2, p. 39 -42.
- Grennery D. B. (1973, Dec). Determination of optical transfer function by inspection of frequency-domain plot. *Journal of The optical society of America*, vol. 63, n° 12, p. 1571-1577.
- Hannan E., Quinn B. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 41, n° 2, p. 190-195.
- Horn B. (1986). *Robot vision*. MIT Press, McGraw-Hill, Book Company Cambridge, Mass.
- Houni K., Sawaya W., Delignon Y. (2008, Mar). One-dimensional barcode reading : an information theoretic approach. *Journal of Applied Optics*, vol. 47, p. 1025-1036.
- Hurvich C., Chih Ling T. (1989). Regression and time series model selection in small samples. *Biometrika*, vol. 76, n° 2, p. 297-307.
- Kaleh G. K., Vallet R. (1994, Jul). Joint parameter estimation and symbol detection for linear or nonlinear unknown channels. *IEEE Transactions. Communications*, vol. 42, n° 7, p. 2406-2413.
- Kundur D., Hatzinakos D. (1996, May). Blind image deconvolution. *IEEE Signal Processing Magazine*, vol. 13, n° 3, p. 43 -64.
- Liavas A., Regalia P. (2001, Aug). On the behavior of information theoretic criteria for model order selection. *IEEE Transactions on Signal Processing*, vol. 49, n° 8, p. 1689 -1695.
- Liavas A., Regalia P., Delmas J. (1999, Dec). Blind channel approximation: effective channel order determination. *IEEE Transactions on Signal Processing*, vol. 47, n° 12, p. 3336 -3344.
- Olivier C., Paquet T., Avila M., Lecourtier Y. (1997). Optimal order of markov models applied to bankchecks. *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 11, n° 05, p. 789-800.
- Raftery A. (1995). Bayesian model selection in social research. *Sociological Methodology*, vol. 25, p. 111-196.
- Schwarz G. (1978, Mar). Estimating the dimension of a model. *Annals of statistics*, vol. 6, n° 2, p. 461-464.
- Stoica P., Selen Y. (2004, July). Model-order selection: a review of information criterion rules. *IEEE Signal Processing Magazine*, vol. 21, n° 4, p. 36 - 47.
- Sugiura N. (1978). Further analysts of the data by Akaike's information criterion and the finite corrections. *Journal of Communications in Statistics - Theory and Methods*, vol. 7, n° 1, p. 13-26.
- Tison C., Nicolas J., Tupin F., Maitre H. (2004, Oct). A new statistical model for markovian classification of urban areas in high-resolution sar images. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, n° 10, p. 2046 - 2057.
- Tong H. (1975, Sep). Determination of the order of a markov chain by akaike's information criterion. *Journal of Applied Probability*, vol. 12, n° 3, p. 488 - 497.

400 TS. Volume 31 – n° 3-4/2014

Via J., Santamaria J., I. and Perez. (2006, Sept). Effective channel order estimation based on combined identification/equalization. *IEEE Transactions on Signal Processing*, vol. 54, n° 9, p. 3518 -3526.