



**HAL**  
open science

## Méthodes comparatives d'utilisation des critères d'information pour la sélection de modèles.

Guilhem Coq, Olivier Alata, Christian Olivier, Marc Arnaudon

► **To cite this version:**

Guilhem Coq, Olivier Alata, Christian Olivier, Marc Arnaudon. Méthodes comparatives d'utilisation des critères d'information pour la sélection de modèles.. TAIMA, May 2009, Hammamet, Tunisie. pp.241-246. hal-00367587

**HAL Id: hal-00367587**

**<https://hal.science/hal-00367587>**

Submitted on 11 Mar 2009

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

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

# Méthodes comparatives d'utilisation des critères d'information pour la sélection de modèles.

Guilhem Coq<sup>1</sup>, Olivier Alata<sup>2</sup>, Christian Olivier<sup>2</sup>, et Marc Arnaudon<sup>1</sup>

<sup>1</sup> Laboratoire de Mathématiques et Applications, UMR CNRS 6086  
BP 30179 - 86962 Futuroscope Chasseneuil Cedex France  
Tél : 05 49 49 68 97 Fax : 05 49 49 69 01  
coq,arnaudon@math.univ-poitiers.fr

<sup>2</sup> Laboratoire Signal Image et Communication, UMR CNRS 6172  
BP 30179 - 86962 Futuroscope Chasseneuil Cedex France  
Tél. : 05 49 49 65 67 Fax : 05 49 49 65 70  
alata,olivier@sic.sp2mi.univ-poitiers.fr

**Résumé** Les problèmes de sélection de modèle sont couramment traités à l'aide de critères d'information. Il est cependant moins commun qu'il existe différentes façons d'utiliser ces critères. Nous présentons ici deux méthodes comparatives d'utilisation des critères d'information et les comparons avec la méthode globale, utilisée habituellement. Cette comparaison est effectuée de manière théorique sur des autorégressions unidimensionnelles ainsi que de manière appliquée à la modélisation de textures par des autorégressions bidimensionnelles.

**Mots clés** Critères d'informations, complexité d'utilisation, support d'autorégression

## 1 Introduction

Etant donnée une observation  $x_1, \dots, x_n$  d'un processus stochastique  $(X_i)_{1 \leq i \leq n}$  et une famille paramétrée  $\{f(\cdot|\theta), \theta \in \Theta\}$  de densités de probabilité, la méthode du Maximum de Vraisemblance (MV) est un outil efficace et largement utilisé pour estimer un paramètre  $\hat{\theta}$  décrivant au mieux ces observations. Cependant, le problème de la sélection du modèle est d'un plus grand intérêt. Citons comme exemple la détermination de l'ordre d'une autorégression [3], d'un histogramme estimant une densité de probabilité [6] ou des fonctions nécessaires à réaliser une régression gaussienne [4]. Dans toutes ces situations, la méthode MV tombe en défaut et tend à surparamétrer le modèle recherché.

Une alternative est donnée par les critères d'information (IC pour Information Criteria). Ils sont de la forme  $IC = -\log(MV) + p(n)$ , où  $p(n)$  est une fonction pénalisante qui croît à mesure que l'on autorise le paramètre à devenir complexe. Le terme  $-\log(MV)$  ayant un comportement inverse, la minimisation du critère permet de réaliser un compromis entre l'adéquation aux données et la complexité du paramètre choisi.

En vue de résoudre un problème de sélection de modèles à l'aide d'un IC, un utilisateur doit choisir une pénalité  $p(n)$  ainsi qu'une méthode d'utilisation de l'IC.

On trouve différents types de pénalités. Se reposant sur la minimisation d'un risque de Kullback, Akaike [1] a introduit le critère AIC ; Schwarz [12] a proposé le critère BIC fondé

sur l'estimation Bayésienne. Rissanen s'appuie sur la notion de complexité stochastique [11] pour justifier un critère qui prend la même forme que BIC. Poursuivant cette approche, El-Matouat et Hallin [7] ont introduit la famille de critères  $\varphi_\beta$  utilisée ici.

Pour le deuxième choix, nous introduirons trois méthodes d'utilisation des IC : la méthode globale et deux méthodes comparatives. Ces dernières constituent l'originalité de ce travail. Elles ont l'avantage de donner une description du modèle aussi fine que la méthode globale tout en présentant une complexité de calcul beaucoup moins élevée.

## 2 Sélection de modèles par critères d'information

Soit  $\{\Omega^n; \mathcal{A}^n; f(\cdot|\theta), \theta \in \Theta\}$  une structure statistique, où  $\Theta$  est une partie de  $\mathbb{R}^m$ , et  $x = x_1, \dots, x_n$  une réalisation d'une densité inconnue  $f(\cdot|\theta)$ . Notons  $S^*$  le support de  $\theta$  :

$$S^* = \{j \in \llbracket 1, m \rrbracket \mid \theta_j \neq 0\}.$$

Dans ce cadre, sélectionner le modèle c'est donner, à partir de  $x$ , une estimation  $\widehat{S}$  de  $S^*$ .

### 2.1 Critères d'information

Les IC sont un outil approprié à l'estimation du support  $S^*$ . Pour  $S \subset \llbracket 1, m \rrbracket$  un support quelconque, ils se présentent sous la forme générale suivante :

$$\text{IC}(S) = -2 \log f(x^n | \widehat{\theta}_S) + |S| \alpha(n) \quad (1)$$

où  $|S|$  est le cardinal de  $S$  et  $\widehat{\theta}_S$  est l'estimé de  $\theta$  au sens du maximum de vraisemblance au sein de la classe des paramètres de support  $S$ .

Ici, nous utiliserons le critère  $\varphi_\beta$  pour lequel  $\alpha(n)$  est donnée, pour  $\beta \in ]0, 1[$  fixé, par :

$$\alpha(n) = n^\beta \log \log n. \quad (2)$$

A  $n$  fixé, l'ajustement de la valeur de  $\beta$  dans (2) permet de retrouver les critères usuels AIC et BIC pour lesquels  $\alpha(n)$  doit valoir 2 et  $\log(n)$  respectivement ; dans [9], les auteurs proposent également les valeurs particulières  $\beta_{\min}$  et  $\beta_{\max}$  pour  $\beta$  :

$$\begin{aligned} \beta_{\text{AIC}} &= (\log 2 - \log \log \log n) / \log n \\ \beta_{\text{BIC}} &= (\log \log n - \log \log \log n) / \log n. \\ \beta_{\min} &= \log \log n / \log n = 1 - \beta_{\max}. \end{aligned} \quad (3)$$

### 2.2 Les méthodes étudiées

**Méthode globale.** Le support inconnu  $S^*$  est estimé par minimisation du critère sur tous les supports envisagés. C'est le cas usuel d'utilisation :

$$\widehat{S} = \text{Argmin} (\text{IC}(S), S \subset \llbracket 1, m \rrbracket). \quad (4)$$

**Méthode comparative.** Proposée par Nishii [8], elle permet la sélection de n'importe quel support de  $\llbracket 1, m \rrbracket$  tout en nécessitant nettement moins de calculs que (4). Fixons la valeur du critère lorsque tous les paramètres sont libres comme référence et estimons le support  $S^*$  par  $\widehat{S}$  défini comme suit :

$$\begin{cases} \text{IC}_{\text{ref}} = \text{IC}(\llbracket 1, m \rrbracket) \\ \widehat{S} = \{j \in \llbracket 1, m \rrbracket, \text{IC}_{\text{ref}} \leq \text{IC}(\llbracket 1, m \rrbracket \setminus \{j\})\} . \end{cases} \quad (5)$$

**Méthode comparative descendante.** Elle élimine les composantes jugées non utiles par étape. L'étape 0 consiste à fixer  $S^0 = \llbracket 1, m \rrbracket$ ,  $\text{IC}_{\text{ref}}^0 = \text{IC}(S^0)$ . L'étape  $k \geq 0$  ayant été effectuée, on procède à l'étape  $k + 1$  en posant :

$$\begin{aligned} C^{k+1} &= \left\{ j \in S^k, \text{IC}(S^k \setminus \{j\}) \leq \text{IC}_{\text{ref}}^k \right\}, J^{k+1} = \text{Argmin}(\text{IC}(S^k \setminus \{j\}), j \in C^{k+1}), \\ S^{k+1} &= S^k \setminus \{J^{k+1}\}, \text{IC}_{\text{ref}}^{k+1} = \text{IC}(S^{k+1}). \end{aligned}$$

Ainsi, à chaque étape, nous écartons définitivement la composante  $J^{k+1}$  jugée la moins utile par le critère. A une certaine étape  $k_f + 1$ , on obtiendra  $C^{k_f+1} = \emptyset$ . Dans ce cas, il n'y a plus de composante inutile dans  $S^{k_f}$  qui est donc choisi pour estimation de  $S^*$  :

$$\widehat{S} = S^{k_f} \quad (6)$$

Notons que toutes les quantités  $C, J, S, \text{IC}_{\text{ref}}, k_f$  produites par cette méthode sont aléatoires. Le nombre de critères qu'il est nécessaire de calculer pour achever cette méthode est donc lui aussi aléatoire, mais borné par  $m(m+1)/2$ .

**Complexités des méthodes.** Le tableau 1 présente le nombre de critères à calculer pour la sélection de  $\widehat{S}$ . La méthode globale, présentant une complexité exponentielle, est beaucoup plus coûteuse en terme de calcul que les méthodes comparatives.

**TAB. 1.** Les trois méthodes et leur complexité.

Globale (4)	Comparative (5)	Descendante (6)
$2^m$	$m + 1$	$\leq m(m+1)/2$

### 3 Simulations sur des processus autorégressifs AR

#### 3.1 Processus AR en dimension 1

Un processus autorégressif d'ordre  $P \in \mathbb{N}$  est de la forme  $X_t = -\sum_{i=1}^P a_i X_{t-i} + E_t$  où  $E$  est un bruit blanc gaussien décorréolé de variance  $\sigma^2$ . Les  $P + 1$  paramètres de ce modèle

sont le vecteur de régression  $a = (a_1, \dots, a_P)$  et la variance  $\sigma^2$ . Déterminer le support d'un tel processus, c'est trouver les indices  $j$  pour lesquels  $a_j \neq 0$ . Se donnant un ordre de régression maximal  $m$ , il est classique que le critère (1) s'écrit, pour  $S \subset \llbracket 1, m \rrbracket$  :

$$\text{IC}(S) = n \log \hat{\sigma}_S^2 + |S|\alpha(n)$$

où  $\hat{\sigma}_S^2$  est la variance estimée en supposant que le support est  $S$ .

Nous synthétisons un processus autorégressif, stable au sens entrée-bornée/sortie-bornée, d'ordre  $P = 15$ , dont les paramètres sont  $a = (0.5, 0.4, 0, \dots, 0, 0.45)$ ,  $\sigma^2 = 1$ . Nous appliquons ensuite les méthodes (4) (5) et (6) d'utilisation du critère  $\varphi_\beta$  (2) avec  $m = 20$ . On compte un succès lorsque le support  $\hat{S}$  choisi est  $\{1, 2, 15\}$ . L'expérience est répétée 100 fois et les pourcentages de succès des méthodes sont représentés, en fonction de la valeur de  $\beta$ , sur la figure 3.1 pour  $n = 100$  et 10.000. Les barres verticales correspondent, de gauche à droite, aux valeurs de  $\beta_{\text{AIC}}$ ,  $\beta_{\text{BIC}}$ ,  $\beta_{\text{min}}$  et  $\beta_{\text{max}}$  données en (3). Pour  $n = 10.000$ , on obtient  $\beta_{\text{AIC}} < 0$ , cette valeur n'apparaît donc pas sur la figure.

Les résultats des méthodes sont satisfaisants et comparables à faible comme à grand nombre d'échantillons. Cependant, pour obtenir ces résultats, les méthodes comparative et descendante ont été environ 4000 fois plus rapide que la méthode globale. Le critère  $\varphi_{\beta_{\text{min}}}$  donne toujours des pourcentages de succès satisfaisants.

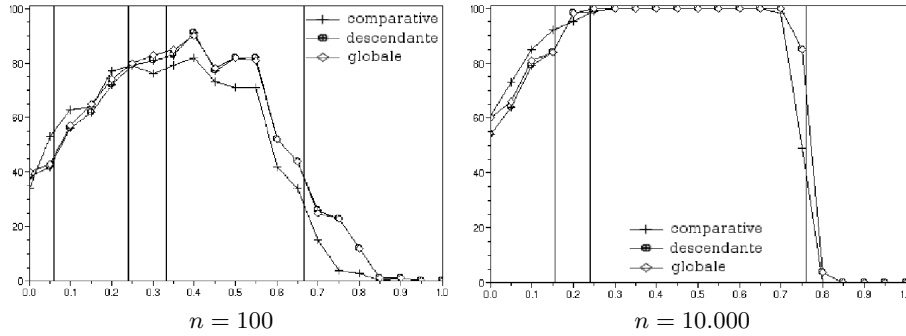


FIG. 1. Pourcentage de succès des méthodes comparatives et globale en fonction de  $\beta$ .

### 3.2 Processus AR en dimension 2

**Présentation.** Le support d'un modèle AR 2D est défini par un couple d'entier  $(P_1, P_2) \in \mathbb{N}^2$  et différents types de géométries [10] : Quart de Plan causal (QP), Demi-Plan Non Symétrique causal (DPNS), semi-causal et non-causal. Nous utilisons ici le support DPNS car il facilite la présentation des résultats expérimentaux. Il est classiquement défini de la manière suivante :

$$S_{P_1, P_2} = \left\{ (i_1, i_2) \left| \begin{array}{l} 1 \leq i_2 \leq P_2 \text{ et } i_1 = 0, \\ -P_2 \leq i_2 \leq P_2 \text{ et } 1 \leq i_1 \leq P_1 \end{array} \right. \right\}$$

Un processus AR 2D DPNS d'ordre  $(P_1, P_2)$  s'écrit :

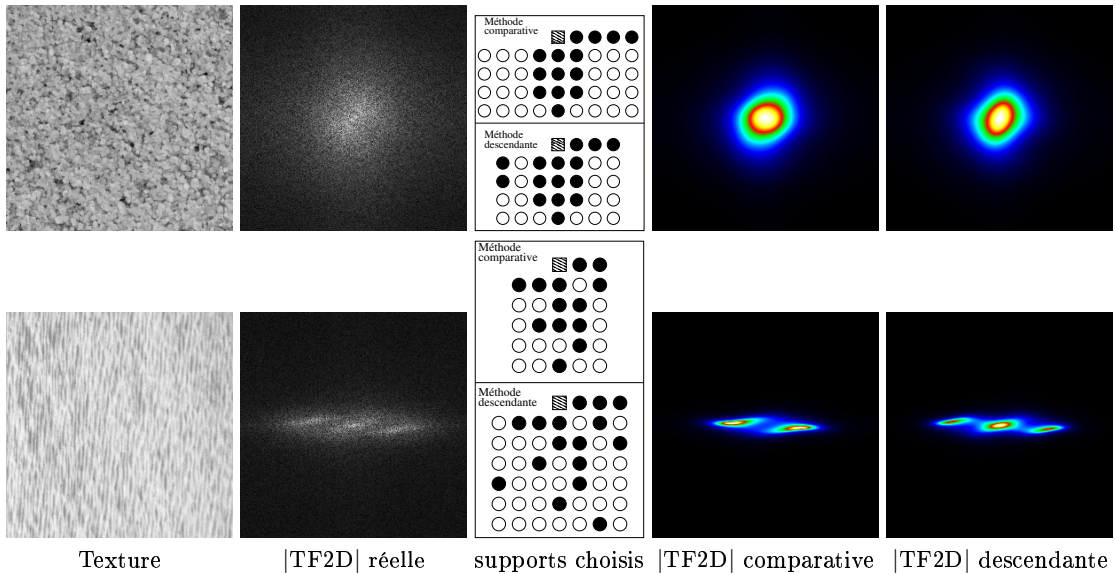
$$X_{t_1, t_2} = - \sum_{(i_1, i_2) \in S_{P_1, P_2}} a_{i_1, i_2} X_{t_1 - i_1, t_2 - i_2} + E_{t_1, t_2}. \quad (7)$$

Sélectionner le support d'un tel processus, c'est déterminer les indices  $(i_1, i_2) \in S_{P_1, P_2}$  pour lesquels  $a_{i_1, i_2} \neq 0$ .

Nous allons montrer dans la partie suivante que les méthodes présentées dans l'article permettent un choix des paramètres pertinents à l'intérieur d'un support DPNS de taille maximum tout en rejetant les paramètres non pertinents.

**Resultats et discussion.** Pour réaliser nos simulations, nous avons utilisé deux textures de taille  $256 \times 256$  issues de l'album de Brodatz [5]. A l'aide des méthodes comparative (5) et descendante (6), des supports de forme non contrainte à l'intérieur d'un support DPNS d'ordre maximal  $(7, 7)$  ont été obtenus en utilisant le critère  $\varphi_{\beta_{\min}}$ .

La figure 3.2 présente les textures utilisées, les supports obtenus avec les méthodes proposées ainsi que les analyses spectrales qu'il est possible d'obtenir par le biais des modèles estimés. Ces analyses sont présentées sous la forme des modules des Transformées de Fourier 2D, notés  $|TF2D|$ , qui fournissent une estimation de la Densité Spectrale de Puissance (DSP) [2].



**FIG. 2.** Analyse spectrale de deux textures issues de l'album de Brodatz : d29 pour la 1ère ligne et d38 pour la 2ème ligne. Pour la troisième colonne, le site carré est le site courant et les sites décrits par des ronds noirs sont les sites choisis. Les 4ème et 5ème colonnes présentent le module de la TF2D associée aux modèles synthétiques choisis par les méthodes comparatives et descendantes respectivement.

Classiquement, on s'intéresse non pas à l'estimation du support mais à celle de l'ordre  $(P_1, P_2)$  d'un processus AR2D (7) ; l'ordre étant un support contraint à avoir une forme de type DPNS. Nous avons effectué cette estimation de l'ordre à l'aide du critère  $\varphi_{\beta_{\min}}$ . Pour la texture d29, un ordre (4, 2), contenant 22 paramètres, a été obtenu. Pour la texture d38, un ordre (5, 3) contenant 38 paramètres. Certains de ces paramètres estimés sont inférieurs à  $10^{-2}$  et ont été rejetés par les méthodes comparatives. A l'inverse des paramètres sélectionnés par les méthodes comparative et descendante ont été exclus par la sélection de l'ordre.

Pour finir, il est possible de noter que les supports obtenus avec les deux méthodes diffèrent et que la méthode descendante a tendance à fournir plus de paramètres que la méthode comparative. Si on se réfère aux images d'analyse spectrale, le spectre obtenu avec la méthode descendante sur la texture d38 semble améliorer celui obtenu avec la méthode comparative. Dans la suite de cette étude, nous comptons déterminer quantitativement la méthode qui permet de réaliser les meilleures analyses spectrales en s'appuyant sur des expérimentations du type de celles produites dans [3].

## Références

1. H. Akaike. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19 :716–723, 1974.
2. O. Alata, P. Baylou, and M. Najim. A New 2-D Spectrum Estimate using Multichannel AR Approach of 2-D Fast RLS Algorithms. In *Proc. IEEE ICIP*, pages 442–445, October 1997.
3. O. Alata and C. Olivier. Choice of a 2-d causal autoregressive texture model using information criteria. *Pattern Recognition Letters*, 24(9-10) :1191–1201, 2003.
4. Y. Baraud. Model selection for regression on a random design. *ESAIM Probab. Statist.*, 6 :127–146 (electronic), 2002.
5. P. Brodatz. *Texture : a Photographic Album for Artists and Designers*. New York, Dover, 1966.
6. G. Coq, C. Olivier, O. Alata, and M. Arnaudon. Information criteria and arithmetic codings : an illustration on raw image. In *15th European Signal Processing Conference proceedings*, pages 634–638, Poznan, Poland, 2007.
7. A. El-Matouat and M. Hallin. Order selection, stochastic complexity and Kullback-Leibler information. *Time series analysis*, 2 :291–299, 1996.
8. R. Nishii. Maximum likelihood principle and model selection when the true model is unspecified. *J. Multivariate Anal.*, 27(2) :392–403, 1988.
9. C. Olivier and O. Alata. *Optimization in Signal and Image Processing*, chapter 4, Information Criteria : examples of applications for the processing of signals and images. Wiley et ISTE LTD editions, To appear March 2009.
10. S. Ranganath and A. K. Jain. Two-Dimensional Linear Prediction Models - part I : Spectral Factorization and Realization. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-33(1) :280–299, February 1985.
11. J. Rissanen. Stochastic complexity and modeling. *Ann. Statist.*, 14(3) :1080–1100, 1986.
12. G. Schwarz. Estimating the dimension of a model. *Ann. Statist.*, 6(2) :461–464, 1978.