

---

# Reconnaissance supervisée et non supervisée de lois à partir d'échantillons finis

Olivier Alata<sup>1</sup>, Christian Olivier<sup>2</sup>, Zhan Jin<sup>2</sup>, Yannis Pousset<sup>2</sup>

1. Lab. H. Curien, UMR CNRS 5516, Université Jean Monnet Saint-Etienne  
Bât. F 18, rue du Prof. Benoît Luras - F-42000, Saint-Etienne  
olivier.alata@univ-st-etienne.fr

2. Institut XLIM, dépt. SIC, UMR CNRS 7252, Université de Poitiers  
Bât. SP2MI, BP 30179 - 86962 Futuroscope-Chasseneuil cedex  
{yannis.pousset,christian.olivier}@univ-poitiers.fr

---

*RÉSUMÉ.* Dans cet article, nous abordons le problème de la reconnaissance de lois de probabilité à partir d'échantillons variant de 100 à 10 000 ou plus. Le contexte applicatif porte sur la modélisation de canaux radio-mobile en situation de visibilité ou de non-visibilité directe entre émetteur et récepteur. Ce problème est crucial pour améliorer les communications numériques. Dans la communauté des transmissions numériques, il est courant d'utiliser la distance de Kolmogorov-Smirnov. Plus rarement, une méthode à noyau est considérée avant le test comparatif. Nous proposons d'utiliser les critères d'information (IC), d'une part pour approcher les lois de probabilité par un histogramme, et d'autre part pour sélectionner le meilleur modèle de loi. Nous étudions les cas supervisé et non supervisé et comparons les méthodes dans ces situations réalistes. Les résultats montrent l'intérêt d'utiliser les méthodes exploitant les IC.

*ABSTRACT.* In this paper, we address the problem of law recognition from samples with size varying from 100 to 10000 or more. The application context concerns the modelling of mobile radio channel for configurations where transmitter and receiver are either in "line of sight" or in "non line of sight" situations. This problem is crucial to improve the digital communication. In the digital transmission community, the Kolmogorov-Smirnov's distance is commonly used. More seldomly in this context, a kernel method is considered to approach the probability laws before the comparative test. We suggest here using the information criteria, on one hand to approach the probability laws by a histogram, and on the other hand to select the best law model. We shall study the supervised and unsupervised cases and shall compare the methods in these realistic situations. Results show the interest of using information criteria based methods.

*MOTS-CLÉS :* reconnaissance de lois, approches supervisée et non supervisée, critères d'information, estimateur de densité par noyau, communications numériques.

*KEYWORDS :* law recognition, supervised and unsupervised methods, information criteria, kernel density estimator, digital communications.

---

DOI:10.3166/TS.29.101-121 © 2012 Lavoisier

**Extended abstract**

The problem of the selection of a statistical distribution studied in this paper is linked to the wireless transmission context. In such a context, the electromagnetic wave undergoes multiple interactions like reflection, diffraction and their combinations on obstacles such as buildings, vehicles, etc. Thus, we speak about the “multi-path phenomenon” between transmitter and receiver antennae. It induces that the amplitude of the received signal, resulting from the sum and from the complexity of the paths, has undergone small scale fading in the spatio-temporal domain. This phenomenon is very penalizing for the quality of the digital communications (Goldsmith, 2005). The true knowledge of this channel behavior is thus crucial.

It is possible to describe it either in a rigorous way *via* deterministic propagation models, or *via* stochastic models thanks to adequate statistical distributions (Sarkar *et al.*, 2003 ; Iskander, Yun, 2002). The first way is accurate in estimation but very excessive in computation time. Thus, the second way proposes very fast models in computation compared to the previous ones. Nevertheless, it is essential for a good estimation, to be able to choose the best statistical model connected to the considered configuration and for a limited, even weak, number of measured samples. In other words, the challenge consists in finding an accurate identification of the adequate small scale fading distribution in considered configuration, to improve the quality of the digital communications.

In this application context of digital communications, many researches have focused on this problem of recognizing the adequate probability law to model the fast variations. We can mention the works of (Barbot *et al.*, 1992 ; Bultitude *et al.*, 1998 ; Babich, Lombardi, 2000 ; Hashemi, Tholl, 1994 ; Seitadji, Levy, 1994) as well as those of (Fryziej, 2001). However, these last ones consider only the simple Kolmogorov-Smirnov (KS) distance which is uniquely applied to the cumulative distribution functions (cdf) of the statistical distributions. Only (Santamaria *et al.*, 2002) and (Zayen, Hayar, 2011) propose works based on the information criterion AIC and on a kernel method respectively. These methods were only used as tools allowing to estimate either the adequate probability density functions (pdf) of signal to noise ratio (AIC's method) or the fast variations of the received signal (kernel method). However, these last ones have not been statistically compared with the existing selection methods such as those we evoked previously.

We distinguish Line Of Sight (LOS) and Non Line Of Sight (NLOS) configurations, considering transmitter antenna visible or not from receiver one. In literature, there are several statistical distributions modeling the propagation small scale fading of the channel. Nevertheless, the most pertinent laws in this propagation context that we will consider are the Rayleigh, Weibull and Nakagami distributions (Sarkar *et al.*, 2003). Choosing between these different laws is not an easy task because Weibull and Nakagami laws can be viewed as different generalizations of Rayleigh law: some particular values of the two parameters that define the pdf of Weibull and Nakagami laws allow to describe the different shapes of Rayleigh's pdf, parameterized with just one parameter. Then, samples issued from the Rayleigh law may be recognized as issued

from Weibull or Nakagami laws if the complexity of the model is not taken into account in the recognition method. This is the reason why the complexity of the chosen model is one of the main topics of our work.

Two recognition strategies may be envisaged:

- by considering the data as a significant representation of the laws. As previously said, the most used method in the field of digital communications is based on the computation of the empirical cdf of the data. Then the minimization of the KS distance allows selecting a law which belongs to a set of hypothesized laws;
- by firstly approximating the pdf. To our knowledge, the main non-parametric methods are the kernel based methods and the histograms. The Kulback-Leibler divergence (Basseville, 1989) can then be used for recognition tests for example.

In the paper, we compare the behaviour of the three methods in the context previously given. The kernel methods are often applied in a context of radio channel (Santamaria *et al.*, 2002). Due to its accuracy, we chose the Generalized Cross Entropy (GCE) method (Botev, 2007). Moreover, it has already been used in our application context. And we propose a IC based method to find a histogram that approximates pdf (Rissanen *et al.*, 1992 ; Coq, Alata, Pousset *et al.*, 2009).

We also study on one hand the supervised context, i.e. the parameters of the laws are known as they have been estimated from sets of data and we try to recognize a law from new data among these fixed laws. On the other hand we study the unsupervised context, i.e. one must recognize, in real configuration of digital communications, the most likely law from samples. Then, the law parameters are estimated for each new sample. The performances of the different methods are analyzed and compared, for the two different contexts (supervised and unsupervised), according to the number of samples of measured signal.

Information criteria are known criteria for their accuracy for the selection of parametric models (Olivier, Alata, 2009). Then they are also used for the unsupervised case as a tool allowing to recognize a law by taking into account the number of parameters of each hypothesized laws. We use the  $\Phi_\beta$  criterion (El Matouat, Hallin, 1996 ; Jouzel *et al.*, 1998),  $\beta \in [0, 1]$  for good asymptotic properties. This criterion can be seen as a generalization of the Rissanen's works (Rissanen, 1989) on the stochastic complexity. With particular values of  $\beta$ , the penalization of the criterion can be equal to the penalization of Akaike's criterion or the one of Bayesian Information Criterion (BIC) which is of interest and illustrated in the paper.

The experiments are conducted on simulated samples from Rayleigh, Weibull and Nakagami laws (450 samples for each laws) with different sizes ( $n = 100, 500, 1000, 5000, 10000$ ). The learning step provide us three sets of parameters (each one associated to the three studied laws) for each context (LOS and NLOS). the shapes of the obtained pdf for the NLOS context are closer than the ones obtained for the LOS context. Then, it gives us the possibility to study the resolution power of the different methods.

In the supervised context, the best results have been obtained with the histogram based method but the two other methods provide almost similar results: for the laws associated to the LOS context, the recognition rate is 99.5% for  $n = 5000$ ; for the laws associated to the NLOS context, the recognition rate is 96.7% for  $n = 10000$ . In the unsupervised case, the IC method is the only one among the tested methods that allows a correct recognition of the Rayleigh law: for the laws associated to the LOS context, the recognition rate is 99.8% for  $n = 5000$  (less than 78.7% for the three other methods); for the laws associated to the NLOS context, the recognition rate is 83.7% for  $n = 10000$  (less than 64.4% for the three other methods).

These results globally show the interest of the use of IC for the law recognition in the supervised case, by approximating the pdf by an optimal histogram, and in the unsupervised case, by minimizing an IC. These results may interest persons working in numerous application contexts in signal and in image domains as soon as it is necessary to use an approximation of the distribution of the samples.

## 1. Introduction

Le problème étudié dans cet article porte sur la reconnaissance de distributions statistiques dans un contexte de transmission sans fil. Cette reconnaissance est cruciale pour une bonne communication, sachant que celle-ci subit notamment des évanouissements rapides (*fast fading*) dus aux différents échos de l'onde électromagnétique dans l'environnement, qui perturbent la transmission (Goldsmith, 2005). Il est possible de prédire ce phénomène électromagnétique, soit de façon rigoureuse *via* des modèles de propagation déterministes, soit de manière stochastique par le biais de distributions statistiques adéquates (Sarkar *et al.*, 2003 ; Iskander, Yun, 2002). La première approche est précise dans l'estimation mais très coûteuse en temps de calcul. Ainsi, la seconde propose des modèles très rapides en temps de calcul comparativement aux précédents. Le challenge consiste donc à identifier avec précision la distribution adéquate propre à un "fast fading" à partir de mesures de terrain, dans le but par la suite, d'améliorer la réception de la communication. En effet, d'une bonne estimation de ces évanouissements du canal de propagation, pour une configuration d'étude donnée, résulte une stratégie de transmission (codage, modulation ...) adaptée (Proakis, 1995 ; Goldsmith, 2005 ; Rappaport, 1996). Dans ce contexte applicatif des communications numériques, de nombreux travaux de recherche ont porté sur cette problématique qu'est de reconnaître la loi statistique adéquate pour modéliser les variations rapides. Nous pouvons citer les travaux de (Barbot *et al.*, 1992 ; Bultitude *et al.*, 1998 ; Babich, Lombardi, 2000 ; Hashemi, Tholl, 1994 ; Seitadji, Levy, 1994) ainsi que ceux de (Fryziej, 2001). Toutefois, ces derniers ne considèrent que le test de Kolmogorov-Smirnov qui ne s'applique qu'aux fonctions de répartitions des distributions statistiques. Seuls (Zayen, Hayar, 2011) et (Santamaria *et al.*, 2002) proposent respectivement des travaux reposant sur le critère d'information AIC (*Akaike's Information Criterion*) et sur la méthode à noyau. Ces méthodes n'ont été utilisées que comme de simples outils permettant d'estimer les densités de probabilité adéquates du rapport signal sur bruit (méthode AIC) ou tout simplement des variations rapides

du signal reçu (méthode à noyau). Toutefois, ces dernières n'ont pas été comparées statistiquement aux méthodes de sélection existantes telles que celles que nous évoquerons par la suite. Ainsi, aucune étude statistique telle que celle que nous proposons n'a été, jusqu'ici et à notre connaissance, réalisée dans ce contexte de communication.

Nous distinguerons d'une part les configurations LOS (*Light Of Sight*) et les configurations NLOS (Non LOS), selon que les émetteurs et récepteurs sont en visibilité ou non-visibilité les uns des autres. Les modèles de distribution statistique les plus couramment acceptés dans notre contexte d'étude sont les lois de Rayleigh, de Weibull et de Nakagami (Sarkar *et al.*, 2003). Il existe, bien entendu, d'autres hypothèses de lois associées à des contextes applicatifs très marginaux que nous n'évoquerons pas ici.

Deux stratégies de reconnaissance sont possibles :

- considérer les données de terrain comme représentation significative des lois. Comme on l'a dit précédemment, la méthode la plus utilisée en communication numérique adopte cette stratégie en considérant les fonctions de répartition directement obtenues par les mesures de terrain. Une simple distance de Kolmogorov-Smirnov permet alors la reconnaissance des distributions (Bultitude *et al.*, 1998 ; Babich, Lombardi, 2000) ;

- une autre stratégie consiste à estimer tout d'abord les fonctions de densité de probabilité de façon optimale. Citons pour principales méthodes non-paramétriques : la méthode à noyau et la méthode par histogramme. Une fois une représentation des lois choisie, il s'agit là encore de considérer des tests de reconnaissance. Nous choisissons la divergence de Kullback-Leibler comme la plus usuelle et la plus classique des  $\Phi$ -divergences (Basseville, 1989).

L'objectif de cet article est de montrer le comportement des trois méthodes dans le contexte présenté ci-avant. La méthode à noyau est souvent appliquée pour l'estimation de canal radio (Santamaria *et al.*, 2002). Parmi le panel de méthodes déclinées sur le principe de la méthode à noyau, nous choisissons la méthode GCE (*Generalized Cross Entropy*) comme très performante (Botev, 2007), et déjà considérée dans ce contexte applicatif. La méthode d'approximation par histogrammes proposée exploite les critères d'information (IC pour *Information Criterion*) (Rissanen *et al.*, 1992 ; Coq, Alata, Pousset *et al.*, 2009) sur lesquels nous allons revenir dans cette introduction. Nous nous plaçons d'une part dans le contexte supervisé, où les lois sont connues et où nous cherchons à les reconnaître, et d'autre part dans le contexte non supervisé, où il s'agit de reconnaître en situation réelle de communication numérique la loi la plus probable au regard des enregistrements du comportement du canal. Les cas LOS et NLOS seront à chaque fois considérés, et les performances de ces méthodes analysées suivant le nombre d'échantillons de mesure du signal. Enfin, les IC sont des outils efficaces en sélection de modèles paramétriques, palliant ainsi les insuffisances du maximum de vraisemblance (MV) (Olivier, Alata, 2009). Nous les utiliserons dans le cas non supervisé, comme outils de sélection de différentes lois, suivant le nombre de paramètres les définissant.

L'emploi des IC s'appuie sur la considération suivante : le maximum de vraisemblance utilisé pour estimer la dimension d'un modèle paramétrique surestime le nombre de paramètres libres. C'est dans ce contexte, précisément sur les modèles autorégressifs (AR), que les IC furent proposés pour la première fois sous la forme du critère AIC déjà évoqué (Akaike, 1974). Ce critère consiste à minimiser la somme du terme de vraisemblance ( $-\log MV$ ) et d'un terme dit de pénalité. Notons que AIC est lui-même une amélioration du critère FPE (pour *Final Prediction Error*) du même auteur, et est basé sur une minimisation de l'information de Kullback-Leibler. Malheureusement, ce critère AIC conduit encore à une sur-paramétrisation de l'ordre du modèle AR (Shibata, 1976 ; Alata, Olivier, 2003). Pour corriger cette insuffisance de AIC, divers auteurs proposèrent d'autres critères comme le critère  $\Phi$  faiblement consistant (convergence seulement en probabilité) (Hannan, Quinn, 1979), ou le célèbre critère BIC (*Bayesian Information Criterion*) (Schwarz, 1978). J. Rissanen introduit dans le contexte de codage de l'information le critère MDL (pour *Minimum description Length*) (Rissanen, 1978 ; Grünwald, 2005). MDL est assez comparable au critère BIC en sélection de modèle, dans la mesure où différents termes additifs le distinguent de BIC (comme l'information de Fischer), mais que ces termes sont souvent négligeables. BIC et MDL sont fortement consistants (convergence presque sûre) en pénalisant le terme de vraisemblance plus que AIC et  $\Phi$ . Un autre critère que nous utiliserons est le critère  $\Phi_\beta$  (El Matouat, Hallin, 1996 ; Jouzel *et al.*, 1998), où  $\beta$  est un paramètre variant de 0 à 1 pour avoir un bon comportement asymptotique. Cet intervalle de variation a d'ailleurs été affiné dans (Olivier *et al.*, 1999). Ce critère peut être vu comme une généralisation des travaux de J. Rissanen sur la complexité stochastique (Rissanen, 1989). Ce critère est encore fortement consistant et offre un bon compromis entre BIC/MDL et  $\Phi$ . A partir de  $\Phi_\beta$ , il est d'ailleurs possible de retrouver les différents termes de pénalité des autres critères comme nous le verrons par la suite. Enfin, d'autres critères existent mais sont utilisés dans des configurations éloignées des nôtres ou très spécifiques, comme par exemple lorsqu'on dispose de peu de données d'observation pour un nombre élevé de paramètres d'un modèle autorégressif (Hurvich, Tsai, 1989 ; Broersen, 2000). Le plus souvent, les critères utilisés ne sont alors pas consistants.

Cet article contient deux sections suivies d'une conclusion. La section 2 expose la méthodologie proposée et rappelle les méthodes employées comparativement :

- fonctions cumulatives et distance de Kolmogorov-Smirnov,
- approximations de lois par histogrammes construits par IC et divergence de Kullback-Leibler,
- approximations de lois par méthode à noyau et divergence de Kullback-Leibler,
- sélection de modèles par IC.

La section 3 présente des résultats comparatifs suivant les méthodes employées. La section 4 conclut cet article.

## 2. Méthodologies

L'observation  $x^n = (x_i), i = 1, \dots, n$ , est supposée être la réalisation d'un processus  $X$  indépendant et identiquement distribué (i.i.d.) pour lequel chaque variable aléatoire admet une densité de probabilité inconnue  $f(x)$ , absolument continue par rapport à la mesure de Lebesgue.

### 2.1. Approximation de densités de probabilité

Dans cette section, nous présentons deux méthodes d'approximation de densités de probabilité à partir des échantillons.

#### 2.1.1. Méthode à noyau GCE

Il est toujours possible de définir une valeur minimum et maximum pour les échantillons étudiés, et nous supposons ici ces valeurs normalisées entre 0 et 1. Il n'y a en effet pas perte de généralité que de considérer cette hypothèse car une simple transformation inversible permet de passer des valeurs nécessairement bornées du processus  $X$  (nous sommes dans un contexte applicatif) à l'intervalle  $[0,1]$ . Ainsi, toutes les variables aléatoires du processus i.i.d. peuvent être considérées comme admettant l'espace de réalisation  $\Omega = [0, 1]$ .

A l'aide de la méthode à noyau, la manière la plus directe d'approcher  $f(x)$  à partir de l'observation consiste à considérer la distribution empirique définie par  $\Delta(x) = \frac{1}{n} \sum_{i=1}^n \delta(x - x_i)$  ce qui est bien entendu sous-optimal. Dans (Botev, 2007), un estimateur par méthode à noyau a été proposé :

$$\hat{f}(t, x) = p(x) \sum_{k=0}^{+\infty} e^{\lambda_k t/2} \hat{\varphi}_k \phi_k(x) \quad (1)$$

avec :

- $p(x)$  est la densité *a priori* des variables aléatoires contenues dans  $X$  ;
- $\{1, \phi_1, \phi_2 \dots\}$  sont les fonctions propres normalisées du problème de Sturm-Liouville sur  $[0, 1]$ . Elles forment une base orthonormale de  $L^2([0, 1], p(x)dx)$  ;
- $\lambda_k$  est la valeur propre associée à  $\phi_k$  et  $\lambda_0 = 0 > \lambda_1 > \lambda_2 > \dots$  ;
- $\hat{\varphi}_k = \mathbb{E}_\Delta[\phi_k(x)] = \frac{1}{n} \sum_{i=1}^n \phi_k(x_i)$  ;
- $t$  est un paramètre inconnu (la variance du noyau) et  $x_i$  la moyenne de chaque noyau.

Suivant des considérations asymptotiques et en faisant l'hypothèse d'un noyau gaussien, Botev a montré que l'estimateur  $\hat{f}(t, x)$  est consistant pour le critère de la moyenne de l'erreur quadratique intégrée (*Mean Integrated Squared Error*: MISE)

(Botev, 2007). Afin d'obtenir  $\hat{f}(t, x)$ , il suffit de déterminer  $t$  en minimisant le critère MISE ce qui donne la formule suivante :

$$\hat{t} = (2n\sqrt{\pi} \|(f/p)''\|^2)^{-2/5} \quad (2)$$

Comme (2) dépend de la densité inconnue  $f$ , Botev suggère d'estimer  $t$  en recherchant l'unique solution positive  $t^*$  de l'équation non linéaire suivante (le premier terme de l'égalité est en effet une fonction monotone de  $t$ ):

$$\sum_{k=1}^{+\infty} \frac{\mathbb{E}_{\Delta} [\hat{\varphi}_k^2]}{-\lambda_k} (e^{\lambda_k t/2} - 1)^2 + \frac{1}{n} \sqrt{\frac{t}{\pi}} = \frac{1}{n} \int_0^1 F_{\Delta}(x)[1 - F_{\Delta}(x)] dx \quad (3)$$

où  $F_{\Delta}(x)$  est la fonction de répartition associée à  $\Delta(x)$ .

On obtient l'algorithme suivant en initialisant  $p(x)$  à la densité uniforme sur  $[0, 1]$  :

1. avec la solution  $t^*$  issue de (3), calculer  $\hat{f}_{GCE}(t^*, x)$  à l'aide de (1);
2. utiliser  $\hat{f}_{GCE}(t^*, x)$  dans (2) afin d'obtenir une nouvelle estimation de  $t$  appelée  $\hat{t}$ ;
3. l'approximation de la ddp est obtenue en utilisant maintenant  $\hat{t}$  et  $\hat{f}_{GCE}(t^*, x)$  à la place de  $p(x)$  dans (1).

### 2.1.2. Approximation par histogramme obtenu par critère d'information

Comme nous l'avons signalé en introduction, la plupart des critères d'information (AIC, BIC) ont été établis pour la sélection de modèles paramétriques comme les modèles AR. L'utilisation des critères fut ensuite étendue à la recherche du nombre de classes d'un histogramme, approximation d'une densité de probabilité théorique  $f(x)$  d'une variable aléatoire  $X$ . Supposons que l'espace de réalisation de  $X$  est un intervalle  $I$  et que la seule autre information connue est l'observation  $x^n$ . L'objectif est de trouver le meilleur histogramme qui résume  $f(x)$  au sens du critère : nombre et tailles optimales des intervalles ou classes de cet histogramme. Soit  $(P_k)$  une partition en  $k$  intervalles  $I_j$  de  $I$ . Les critères IC s'écrivent alors :

$$IC(P_k) = - \sum_{j=1}^k n_j \log \frac{n_j}{nL_j} + kC(n) \quad (4)$$

où  $L_j, j = 1, \dots, k$  est la taille de  $I_j$  et  $n_j, j = 1, \dots, k$  est le nombre de  $x_i$  dans  $I_j$ . Rappelons que le premier terme de (4) est le terme de log-vraisemblance et que  $C(n)$  est la pénalité. La partition à  $k$  intervalles retenue est celle qui minimise IC. Elle est obtenue par programmation dynamique comme dans (Rissanen *et al.*, 1992), ce qui permet une complexité raisonnable de l'algorithme et peu gourmande en temps de calcul (voir l'Annexe).



La pénalité  $C(n)$  diffère suivant les critères (tableau 1). Nous choisissons le critère  $\Phi_\beta$  dont la pénalité est égale à  $C(n) = n^\beta \log \log n$ , avec  $0 < \beta < 1$  (El Matouat, Hallin, 1996). Nous avons donné dans (Olivier *et al.*, 1999) la condition supplémentaire  $\beta_{\min} = \frac{\log \log n}{\log n} \leq \beta < 1$  qui permet de réajuster  $\beta$  suivant le nombre des observations  $x^n$ . De plus, à partir de la pénalité de  $\Phi_\beta$ , il est possible de calculer celle du critère BIC ou AIC en affectant les valeurs suivantes à  $\beta$ :  $\beta_{\text{BIC}} = \frac{\log \log N - \log \log \log N}{\log N}$  et  $\beta_{\text{AIC}} = \frac{\log 2 - \log \log \log N}{\log N}$ . Ces résultats sont vrais quel que soit le domaine applicatif des IC (histogrammes ou sélection de modèles).

Tableau 1. Les critères utilisés et leur pénalité

Critère	Pénalité $C(n)$
AIC	$C(n) = 2$
BIC	$C(n) = \log n$
$\Phi_\beta$	$C(n) = n^\beta \log \log n$

## 2.2. Procédures de reconnaissance de lois

### 2.2.1. Présentation des lois utilisées

Donnons-nous une famille de densités de probabilité paramétrées. Comme indiqué en introduction, nous considérons la famille constituée des lois de Rayleigh, Weibull et Nakagami, de densités de probabilité respectivement données par :

$$f_R(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right),$$

$$f_W(x) = \left(\frac{k_w}{\lambda}\right) \left(\frac{x}{\lambda}\right)^{(k_w-1)} \exp\left(-\left(\frac{x}{\lambda}\right)^{k_w}\right),$$

$$f_N(x) = \frac{2\mu^\mu x^{2\mu-1}}{\Gamma(\mu)\Omega^\mu} \exp\left(-\frac{\mu}{\Omega}x^2\right),$$

où  $\sigma$ ,  $k_w$ ,  $\lambda$ ,  $\mu$  et  $\Omega$  sont les paramètres de forme de ces lois. On remarque que, pour  $k = 2$  ou  $\mu = 1$ , les lois de Weibull ou de Nakagami correspondent à celle de Rayleigh.

Notons que la loi de Rayleigh est utilisée lorsque les multitrajets constituant le canal de propagation sont environ de même amplitude. Aussi, les lois de Weibull et de Nakagami sont plus générales que la précédente et incluent, en plus du cas précédent, celui ou un des trajets peut être prédominant en amplitude par rapport aux autres. Nous n'avons pas choisi d'utiliser la distribution de Rice, généralement considérée dans le cas d'un trajet prédominant, de part la remarque de (Dholer, 2003) qui souligne que ce modèle n'est pas approprié dans notre contexte d'application.

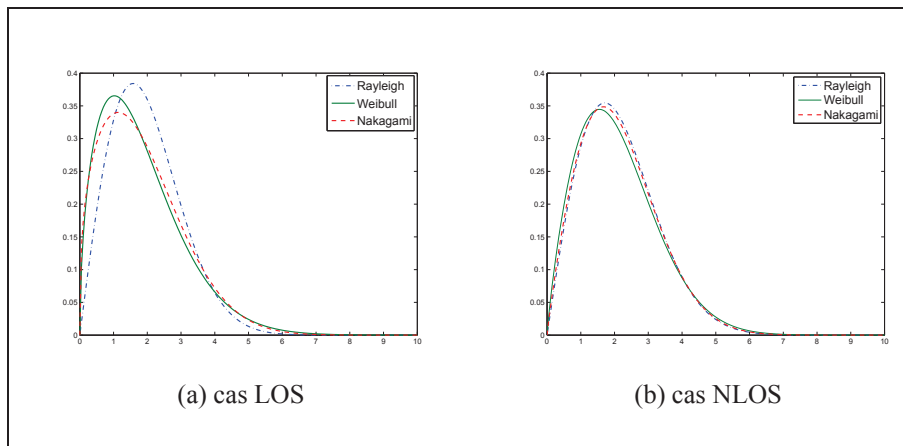


Figure 1. Distribution des lois de Rayleigh, Weibull et Nakagami, suivant les cas LOS (a) et NLOS (b)

Tableau 2. Paramètres des différentes lois appris à partir d'échantillons simulés

Loi et paramètre(s)	cas LOS	cas NLOS
Rayleigh ( $\sigma$ )	1.579	1.711
Weibull ( $k, \lambda$ )	1.529, 2.050	1.845, 2.367
Nakagami ( $\mu, \Omega$ )	0.675, 4.985	0.956, 5.854

Nous nous intéresserons aux cas où les émetteurs et récepteurs sont en position de visibilité LOS ou de non-visibilité NLOS. Le cas NLOS est le plus délicat. En effet, une simulation de la propagation des ondes radio en configuration NLOS montre que les densités de probabilité candidates (voir la figure 1) sont plus proches visuellement que dans la configuration LOS. Notons que ces deux configurations (LOS et NLOS) résument l'ensemble des cas rencontrés où les distributions présentent des densités de probabilité similaires ou non.

Les paramètres des trois lois de Rayleigh, Nakagami et Weibull, donnés au tableau 2, ont donc été appris à partir des simulations. Nous nous proposons, dans un premier temps, à partir de l'observation  $x^n$  d'une des 3 lois définies par les paramètres du tableau 2, de valider statistiquement les méthodes de sélection proposées, suivant les deux contextes supervisé et non supervisé.

### 2.2.2. Cas supervisé

Dans ce contexte, on cherche à savoir quelle est la densité la plus appropriée parmi les trois densités connues (voir tableau 2 pour les cas LOS et NLOS). Il n'y a donc pas d'estimation de paramètres à partir de  $x^n$ . De ce fait, une approche exploitant la notion de maximum de vraisemblance ne peut pas être utilisée car les paramètres appris ne sont *a priori* pas ceux qui vont maximiser la vraisemblance de la nouvelle observation.

Pour comparer deux processus  $X$  et  $Y$ , il est courant d'utiliser la distance de Kolmogorov-Smirnov (KS) entre fonctions de répartition  $F_X$  et  $F_Y$  :

$$\text{KS}(X, Y) = \sup_x \{|F_X(x) - F_Y(x)|\} \quad (5)$$

Bien que facile à calculer et très utilisée dans le contexte applicatif considéré, son faible pouvoir discriminant est connu.

Les  $\Phi$ -divergences sont des outils classiques pour mesurer la distance entre densités de probabilité. Nous choisissons la divergence de Kullback-Leibler (KL), qui est à l'origine, par exemple, de l'obtention du critère AIC :

$$\text{KL}(X, Y) = \frac{1}{2} \int (f_X - f_Y) \log(f_X/f_Y) dx \quad (6)$$

Cependant, alors qu'il est presque immédiat de calculer KS à partir de l'observation  $x^n$  du processus  $X$ , nous devons estimer une densité de probabilité pour utiliser KL. Pour cela, nous prendrons soit la méthode à noyau présentée en (2.1.1), ou la méthode par histogramme optimisé par IC présentée en (2.1.2).

Dans le contexte supervisé, l'algorithme de reconnaissance fonctionne donc de la manière suivante, dans le cas de l'utilisation de la distance de Kolmogorov-Smirnov (5) :

1. à partir d'une observation de taille  $n$ ,  $x^n$ , calculer les trois distances de Kolmogorov-Smirnov entre la fonction de répartition empirique et les trois fonctions de répartition des lois dont les paramètres sont fournis tableau 2 ;
2. attribuer à l'observation la loi qui renvoie la distance la plus faible.

Si une approximation de la densité de probabilité est utilisée (voir section 2.1), la distance de Kolmogorov-Smirnov est remplacée par la divergence de Kullback-Leibler (6), qui mesure alors l'écart entre la densité de probabilité estimée par la méthode à noyau, ou à l'aide d'un histogramme optimal, et la densité d'une des trois lois dont les paramètres sont fournis tableau 2.

### 2.2.3. Cas non supervisé

Dans ce contexte, nous nous donnons juste comme *a priori* l'ensemble des densités de probabilité avec lesquelles nous voulons approcher la densité. Pour l'observation  $x^n$ , les paramètres associés aux trois densités sont estimés au sens du maximum de vraisemblance.

Comme précédemment, nous utilisons la distance KS entre fonctions de répartition ainsi que la divergence KL entre densités de probabilité estimées par la méthode GCE ou par histogramme optimisé par IC.

L'algorithme de reconnaissance fonctionne alors de la manière suivante, dans le cas de l'utilisation de la distance de Kolmogorov-Smirnov (5) :

1. estimer au sens du maximum de vraisemblance les paramètres des trois lois à partir d'une observation de taille  $n$ ,  $x^n$  ;

2. calculer les trois distances de Kolmogorov-Smirnov entre la fonction de répartition empirique et les trois fonctions de répartition des lois dont les paramètres ont été estimés à l'étape précédente ;

3. attribuer à l'observation la loi qui renvoie la distance la plus faible.

Si une approximation de la densité de probabilité est utilisée (voir section 2.1), la distance de Kolmogorov-Smirnov est remplacée par la divergence de Kullback-Leibler (6), qui mesure alors l'écart entre la densité de probabilité estimée par méthode à noyau, ou à l'aide d'un histogramme optimal, et la densité d'une des trois lois dont les paramètres ont été estimés à l'étape 1 de l'algorithme.

De plus, afin d'améliorer la discrimination de densités possédant un nombre différent de paramètres (ici, la loi de Rayleigh possède un paramètre alors que les deux autres en possèdent deux), nous proposons d'utiliser aussi les IC mais dans le contexte de sélection de modèles. Ainsi, si une densité  $f_m$  à  $m$  paramètres libres est un de ces modèles et si  $\hat{f}_m$  dénote cette densité de probabilité configurée avec les paramètres maximisant la vraisemblance, les critères IC s'écrivent :

$$IC(f_m) = -2 \log \left( \hat{f}_m(x^n) \right) + mC(n) \quad (7)$$

avec  $C(n)$  défini comme dans la section 2.1.2.

L'algorithme de reconnaissance fonctionne alors de la manière suivante :

1. estimer au sens du maximum de vraisemblance les paramètres des trois lois à partir d'une observation de taille  $n$ ,  $x^n$  ;

2. calculer les trois critères d'information associés aux trois lois dont les paramètres ont été estimés à l'étape précédente ;

3. attribuer à l'observation la loi qui renvoie le critère d'information le plus faible.

### 3. Résultats expérimentaux

Au cours de nos expériences, nous avons étudié l'influence du nombre  $n$  d'échantillons sur les résultats de reconnaissance en faisant varier  $n$  de 100 à 10 000. Nous montrons les résultats pour les valeurs 100, 500, 1 000, 5 000 et 10 000 échantillons. En plus des taux de bonne reconnaissance de chaque loi, nous donnons un paramètre statistique de confiance, la VPP, pour valeur prédictive positive, égale à la probabilité d'avoir été généré par une loi lorsqu'on est reconnu comme cette loi. Le résultat en taux de reconnaissance sera d'autant plus crédible que la VPP sera proche de 1. D'autres statistiques ont été utilisées dans (Coq, Alata, Olivier *et al.*, 2009) pour juger de la qualité des méthodes employées comme les facteurs de sensibilité, de spécificité, de précision ainsi que la VP négative. De plus, nous avons montré que, dans certains cas, la méthode par histogramme optimal était significativement la plus efficace en termes de disjonction des intervalles de confiance que la méthode exploitant la distance de Kolmogorov-Smirnov en particulier pour les lois de Weibull et Nakagami. Nous renvoyons le lecteur à l'article cité précédemment. Nous n'avons pas fourni ici

d'étude sur les intervalles de confiance mais des résultats moyens à partir de 450 expériences pratiquées pour chaque loi à identifier.

Ainsi si  $a$  est le nombre de fois où une loi considérée  $L$  a été bien reconnue (taux de bonne reconnaissance  $a/450$ ) et si  $b$  est le nombre de fois où  $L$  est reconnue alors que les échantillons sont issus d'une autre loi, alors  $VPP = a/(a + b)$ .

Nous notons, dans les tableaux 3, 4 et 5, "Histo" et "Kernel" pour les méthodes utilisant l'approximation des densités de probabilité par histogramme ou par méthode à noyau et la distance KL, "KS" pour celle utilisant la distance KS sur les fonctions de répartition. Dans les tableaux 4 et 5, nous rajoutons "IC" pour la méthode de sélection par IC (7). On notera par R, W, N, les trois lois de Rayleigh, de Weibull et de Nakagami, respectivement.

### 3.1. Cas supervisé

Dans cette section, nous donnons des résultats moyens de taux de bonne reconnaissance sur des échantillons pour lesquels la loi dont ils sont issus est connue (cas supervisé) afin de comparer les performances des différentes méthodes en LOS et NLOS (tableau 3). Nous donnons les résultats pour la méthode "Histo" avec le critère  $\Phi_\beta$  donnant les meilleures performances et correspondant en fait au critère BIC en LOS et NLOS. La figure 2 justifie ce choix de  $\Phi_\beta$  : elle montre le critère  $\Phi_\beta$  pour  $\beta_{\min}$  (jaune), BIC (magenta) et AIC (vert). Nous pouvons observer que  $\Phi_{\beta_{\text{BIC}}}$  donne pratiquement le taux optimal dans les deux cas présentés.

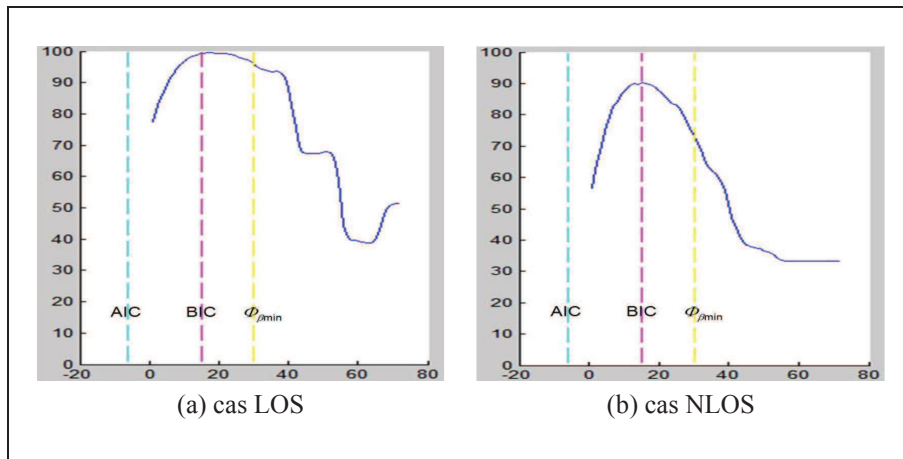


Figure 2. Evolution du taux de reconnaissance (en ordonnées) pour les cas LOS (a) et NLOS (b) suivant  $\beta$  (en abscisse) qui varie de 0 à 0.7 en 75 valeurs. Les résultats avec AIC, BIC et min sont indiqués en vert, magenta et jaune respectivement

Pour  $n > 500$  et le cas LOS, la méthode "Histo" donne des résultats supérieurs aux deux autres méthodes, avec une VPP presque toujours proche de un. Pour de petits

échantillons ( $n = 100$ ), aucune méthode n'est pertinente, mais dans ce cas limite, il peut être préférable de prendre la méthode classique "KS", la méthode "Histo" étant basée sur des critères asymptotiques d'où son très mauvais comportement. Dans le cas  $n = 500$ , les méthodes sont équivalentes en LOS, avec la loi de Weibull difficilement reconnaissable.

En NLOS, pour  $n < 1000$ , les taux sont globalement insuffisants et donc inexploitable pour  $n$  jusqu'à 1000 échantillons. Les lois sont en effet trop proches (figure 1) pour être identifiées quand le nombre d'observations est insuffisant. Rappelons que les méthodes "Kernel" et "Histo" sont des méthodes asymptotiques. Néanmoins, dans le cas de petits échantillons, il existe d'autres critères d'information comme (Hurvich, Tsai, 1989; Broersen, 2000), mais à notre connaissance non justifiables dans le cas des histogrammes.

Pour résumer, à la vue des résultats moyens, il semble que la méthode par histogramme donne globalement les meilleurs résultats, les meilleurs écarts étant obtenus pour le cas NLOS avec  $n \geq 1000$ . Les résultats présentés ne permettent néanmoins pas de prouver statistiquement parlant une supériorité de cette méthode sur les autres.

Tableau 3. Comparaison des méthodes en supervisé. Le premier nombre donne le taux de bonne reconnaissance (en %), le second entre parenthèses donne la VPP

cas supervisé		LOS			NLOS		
		KS	Kernel	Histo	KS	Kernel	Histo
100	R	81.5(0.80)	35.3(0.98)	51.3(0.65)	46.6(0.43)	11.1(0.62)	40.4(0.43)
	W	59.3(0.54)	44.7(0.67)	52.9(0.46)	64.6(0.43)	95.3(0.40)	67.1(0.39)
	N	45.6(0.51)	82.0(0.41)	45.3(0.43)	16.6(0.38)	15.3(0.40)	8.9(0.37)
	<b>Moy</b>	<b>62.2</b>	54.0	49.8	<b>42.7</b>	40.6	38.1
500	R	99.3(0.99)	96.7(1)	98.2(0.99)	54.7(0.55)	20.2(0.85)	40.9(0.66)
	W	67.8(0.67)	52.2(0.9)	55.8(0.84)	71.1(0.57)	96.7(0.52)	83.8(0.56)
	N	65.3(0.67)	94.4(0.65)	89.5(0.67)	27.8(0.37)	36.2(0.4)	36.2(0.41)
	<b>Moy</b>	77.4	81.1	<b>81.2</b>	51.1	51	<b>53.6</b>
1000	R	100(1)	100(1)	100(1)	59.6(0.61)	22.4(0.89)	56.4(0.66)
	W	77.6(0.75)	59.6(0.99)	76.2(0.92)	79.8(0.7)	97.3(0.63)	83.8(0.73)
	N	74.4(0.77)	99.1(0.71)	93.3(0.8)	38.4(0.44)	54.4(0.45)	50.2(0.5)
	<b>Moy</b>	84.0	86.2	<b>89.8</b>	59.2	58.1	<b>63.5</b>
5000	R	100(1)	100(1)	100(1)	78.7(0.77)	45.1(0.97)	81.3(0.91)
	W	96.0(0.95)	91.3(1)	98.7(1)	95.8(0.98)	100(0.95)	98.9(0.98)
	N	95.3(0.96)	100(0.92)	100(0.99)	74.4(0.74)	93.1(0.63)	90.2(0.82)
	<b>Moy</b>	97.1	97.1	<b>99.5</b>	83.0	79.4	<b>90.1</b>
10000	R	100(1)	100(1)	100(1)	85.1(0.85)	62.7(1)	95.5(0.95)
	W	99.1(0.99)	98.5(1)	100(1)	99.1(1)	100(1)	99.5(1)
	N	98.7(0.99)	100(0.98)	100(1)	84.4(0.84)	99.5(0.73)	94.9(0.95)
	<b>Moy</b>	99.2	99.5	<b>100</b>	89.5	87.4	<b>96.7</b>

### 3.2. Cas non supervisé

Dans ce cas, les échantillons sont issus des trois lois, mais, comme les paramètres ne sont pas connus, ils sont estimés au sens du maximum de vraisemblance à partir de l'observation  $x^n$ . Nous pouvons introduire ici la méthode de sélection de modèle dite "IC" (7) car utilisable dans ce contexte, ce qui n'était pas cohérent dans le contexte supervisé. Les deux méthodes basées IC, soit "Histo" et "IC", sont encore données avec le critère  $\Phi_\beta$  produisant les taux optimaux, soit avec  $\Phi_\beta$  proche du critère BIC.

Tableau 4. Comparaison des méthodes en non supervisé pour le cas LOS. Le premier nombre donne le taux de bonne reconnaissance (en %), le second entre parenthèses donne la VPP

cas non-supervisé		LOS			
		KS	Kernel	Histo	IC
100	R	21.5(0.88)	43.8(0.99)	44.0(0.99)	92.7(0.88)
	W	77.1(0.43)	39.8(0.54)	40.4(0.53)	62.9(0.64)
	N	47.3(0.49)	86.2(0.47)	81.8(0.46)	61.3(0.63)
	<b>Moy</b>	48.7	40.6	55.4	<b>72.3</b>
500	R	26.0(1)	44.7(1)	34.7(0.99)	100(1)
	W	86.2(0.5)	43.1(0.64)	54.4(0.58)	82.9(0.84)
	N	64.2(0.64)	97.3(0.52)	90.0(0.52)	84.0(0.83)
	<b>Moy</b>	58.8	61.7	59.7	<b>88.9</b>
1000	R	23.6(1)	47.8(1)	38.4(1)	100(1)
	W	92.0(0.55)	53.3(0.72)	62.4(0.68)	90.2(0.92)
	N	78.7(0.71)	99.3(0.56)	97.6(0.58)	91.8(0.9)
	<b>Moy</b>	64.74	66.81	66.14	<b>94</b>
5000	R	22.0(1)	48.7(1)	41.3(1)	100(1)
	W	99.1(0.65)	87.1(0.83)	94.9(0.79)	100(0.99)
	N	96.2(0.77)	100(0.68)	100(0.72)	99.6(1)
	<b>Moy</b>	72.4	78.6	78.7	<b>99.8</b>
10000	R	21.6(1)	49.6(1)	48.9(1)	100(1)
	W	100(0.66)	97.3(0.85)	98.9(0.83)	100(1)
	N	99.1(0.78)	100(0.73)	100(0.76)	100(1)
	<b>Moy</b>	73.6	82.3	82.6	<b>100</b>

Tableau 5. Comparaison des méthodes en non supervisé pour le cas NLOS. Le premier nombre donne le taux de bonne reconnaissance (en %), le second entre parenthèses donne la VPP

cas non-supervisé		NLOS			
		KS	Kernel	Histo	IC
100	R	24.4(0.40)	50.0(0.49)	49.5(0.49)	74.6(0.36)
	W	56.4(0.36)	35.6(0.44)	34.0(0.46)	24.0(0.50)
	N	24.7(0.30)	36.9(0.31)	41.3(0.32)	13.3(0.30)
	<b>Moy</b>	35.2	40.8	41.6	37.3
500	R	27.3(0.56)	50.7(0.75)	45.3(0.74)	83.6(0.49)
	W	64.9(0.4)	34.4(0.45)	40.4(0.44)	48.7(0.68)
	N	32.0(0.35)	57.8(0.37)	54.7(0.37)	16.7(0.29)
	<b>Moy</b>	41.4	47.6	46.8	<b>49.6</b>
1000	R	22.9(0.57)	43.1(0.75)	45.6(0.75)	84.7(0.56)
	W	67.6(0.4)	33.1(0.43)	36.2(0.44)	61.3(0.71)
	N	33.3(0.38)	62.7(0.38)	59.8(0.38)	20.4(0.33)
	<b>Moy</b>	41.2	46.3	47.2	<b>55.5</b>
5000	R	25.6(0.86)	48.9(0.97)	39.3(0.97)	90.9(0.81)
	W	79.1(0.45)	34.0(0.54)	56.9(0.54)	82.2(0.72)
	N	47.6(0.51)	87.3(0.54)	78.2(0.54)	49.1(0.68)
	<b>Moy</b>	50.7	56.7	58.1	<b>74.1</b>
10000	R	24.2(0.95)	45.8(1)	45.1(0.99)	95.8(0.93)
	W	87.8(0.49)	38.7(0.59)	60.4(0.62)	91.1(0.74)
	N	56.7(0.59)	94.4(0.50)	87.5(0.56)	64.4(0.86)
	<b>Moy</b>	56.2	59.6	64.4	<b>83.7</b>

Nous donnons dans les tableaux 4 et 5, les taux de bonne reconnaissance et la VPP, suivant le nombre  $n$  d'échantillons.

On voit que la méthode de sélection de modèle "IC" donne les meilleurs résultats dans tous les cas, suivie de loin par les trois autres méthodes dans le cas le plus simple LOS (tableau 4). Comme dans le cas supervisé, et pour  $n = 100, 500, 1\ 000$ , ces trois méthodes donnent des résultats comparables mais insuffisants, tant en termes de taux de reconnaissance qu'en termes de crédibilité (la VPP) avec un avantage pour la méthode asymptotique "Histo" pour  $n > 1\ 000$ .

Pour le cas NLOS (tableau 5), si nous excluons le cas limite de très peu d'échantillons ( $n = 100$ ) où les résultats manquent de crédibilité avec les quatre méthodes, la méthode "IC" est aussi meilleure en termes de taux de reconnaissance, mais les VPP sont encore trop basses pour permettre de conclure sur la nature de la loi pour  $n$  jusqu'à 1 000 échantillons. Pour  $n = 5\ 000$  échantillons, nous pouvons juger que seuls les résultats par la méthode "IC" sont crédibles en termes de VPP, ce qui se confirme également pour 10 000 échantillons.

#### 4. Conclusion

Nous avons montré dans cet article l'intérêt, dans un contexte supervisé comme dans un contexte non supervisé, de l'usage des critères d'information (IC) comme outils pouvant intervenir dans la reconnaissance de lois de probabilité, d'une part pour la construction d'histogrammes optimaux, et d'autre part, pour la sélection de modèles. Pour cela, nous nous sommes appuyés sur un contexte applicatif particulier et difficile de reconnaissance de lois en communication numérique, avec ou sans visibilité entre émetteur et récepteur (cas LOS et NLOS), où la connaissance de la loi de propagation est cruciale si on veut assurer une qualité de service garantie à la réception. Nous avons pu constater la bonne tenue des deux méthodes basées IC comparativement à une méthode à noyau performante, appelée GCE, et à la méthode usuelle employée en propagation numérique. Les résultats présentés ne se limitent pas aux communications numériques et sont susceptibles d'intéresser des personnes travaillant dans de nombreux contextes applicatifs en signal et en image dès qu'il faut faire appel à la description de la distribution des valeurs des échantillons.

#### Bibliographie

- Akaike H. (1974). A new look at the statistical model identification. *IEEE Trans. on Automatic Control*, vol. 19, n° 6, p. 716-723.
- Alata O., Olivier C. (2003). Choice of a 2d causal ar texture model using information criteria. *Pattern Recognition Letters*, vol. 24, n° 9-10, p. 1191-1201.
- Babich F., Lombardi G. (2000). Statistical analysis and characterization of the indoor propagation channel. *IEEE Trans. on Communications*, vol. 48, n° 3, p. 455-464.
- Barbot J. P., Levy A. J., Bic J. C. (1992, June). Estimation of fast fading distribution functions. In *proc. of the ursi commission f symposium*. Ravenscar, (U.K.).



- Basseville M. (1989). Distance measures for signal processing and pattern recognition. *Signal Processing*, vol. 18, n° 4, p. 349-369.
- Botev Z. I. (2007). *A novel nonparametric density estimator*. Technical Report. University of Queensland, Brisbane (Australia).
- Broersen P. M. T. (2000, December). Finite Sample Criteria for Autoregressive Order Selection. *IEEE Trans. on Signal Processing*, vol. 48, n° 12, p. 3550-3558.
- Bultitude R. J. C., Hahn R. F., Davies R. J. (1998). Propagation considerations for the design of an indoor broad-band communications system at ehf. *IEEE Trans. on Vehicular Technology*, vol. 47, n° 1, p. 235-245.
- Coq G. (2008). *Utilisation d'approches probabilistes basées sur les critères entropiques pour la recherche d'information sur support multimédia*. Thèse de doctorat, Université de Poitiers (France).
- Coq G., Alata O., Olivier C., Pousset Y., Li X. (2009, September). Reconnaissance de loi via une estimation de densités par histogramme pour la modélisation de canaux de transmission. In *Actes de la conférence gretsi*. Dijon (France).
- Coq G., Alata O., Pousset Y., Li X., Olivier C. (2009, April). Law recognition via histogram-based estimation. In *in proc. ieee iccasp*, p. 3425-3428. Taïpei (Taïwan).
- Coq G., Olivier C., Alata O., Arnaudon M. (2007, Sept.). Information criteria and arithmetic codings: an illustration on raw images. In *in proc. 15th eusipco*, p. 634-638. Poznan (Poland).
- Dholer M. (2003). *Virtual antenna arrays*. PhD, King's College London, London (UK).
- El Matouat A., Hallin M. (1996). Order selection, stochastic complexity and Kullback-Leibler information. In *Athens conference on applied probability and time series analysis, vol. ii (1995)*, vol. 115, p. 291-299. New York, Springer.
- Fryziej M. (2001). *Caractérisation large bande du canal radio intra-bâtiment à 60 ghz*. Thèse de doctorat, CNAM (Paris, France).
- Goldsmith A. (2005). *Wireless communications*. Cambridge University Press.
- Grünwald P. (2005). A tutorial introduction to the minimum description length principle. In *Advances in minimum description length: Theory and applications*. MIT Press.
- Hannan E., Quinn B. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistic Society*, vol. 41, n° 2, p. 190-195.
- Hashemi H., Tholl D. (1994). Statistical modeling and simulation of the rms delay spread of indoor radio propagation channels. *IEEE Transactions on Vehicular Technology*, vol. 43, n° 1, p. 110-120.
- Hurvich C. M., Tsai C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, vol. 76, n° 2, p. 297-307.
- Iskander M., Yun Z. (2002). Propagation prediction models for wireless communication systems. *IEEE Trans. on Microwave Theory and Techniques*, vol. 50, n° 3, p. 662-673.
- Jouzel F., Olivier C., El Matouat A. (1998, September). Information criteria based edge detection. In *EUSIPCO-signal processing IX, Rhodes (Greece)*, vol. 2, p. 997-1000. Elsevier.

- Olivier C., Alata O. (2009, July). The information criteria: examples of applications in image and signal processing. In *Optimisation in image and signal processing, desip series*, p. 105-136. Wiley & ISTE Ltd ed.
- Olivier C., Jouzel F., Matouat A. E. (1999, May). Choice of the Number of Component Clusters in Mixture Models by Information Criteria. In *Proc. iapr vision interface*, p. 74-81.
- Proakis J. (1995). *Digital communications, 3th edition*. MacGraw-Hill, New-York.
- Rappaport T. (1996). *Wireless communications : Principle and practice*. Englewood Cliffs, NJ: Prentice-Hall.
- Rissanen J. (1976). Generalized kraft inequality and arithmetic coding. *IBM Journal of Research and Development*, vol. 20, n° 3.
- Rissanen J. (1978). Modeling by the shortest data description. *Automatica*, vol. 14, p. 465-471.
- Rissanen J. (1989). *Stochastic complexity in statistical inquiry* (vol. 15). Teaneck, NJ, World Scientific Publishing Co. Inc.
- Rissanen J., Speed T. P., Yu B. (1992). Density estimation by stochastic complexity. *IEEE Trans. on Information Theory*, vol. 38, n° 2, p. 315-323.
- Santamaria I., Erdogmus D., Principe J. (2002). Entropy minimization for supervised digital communications channel equalization. *IEEE Trans. on Signal Processing*, vol. 50, n° 5, p. 1184-1192.
- Sarkar T. K., Zhong J., Kyungjung K., Medouri A., Salazar-Palma M. (2003). A survey of various propagation models for mobile communication. *IEEE Trans. on Antennas and Propagation Magazine*, vol. 45, n° 3, p. 51-82.
- Schwarz G. (1978). Estimating the dimension of a model. *Ann. Statist.*, vol. 6, n° 2, p. 461-464.
- Seitadji B., Levy A. J. (1994). A statistical model for the simulation of time-varying multipath mobile radio propagation channel. In *In proc. ieee int. conf. on acoustics, speech and signal processing (ICASSP)*, vol. 6, p. 149-152.
- Shibata R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika*, vol. 63, n° 1, p. 117-126.
- Zayen B., Hayar A. (2011). A performance study of kullback-leibler distance-based spectrum sensing algorithm. In *In proc. 3rd ieee international conference on ultra modern telecommunications (icumt)*, p. 1-5.

## Annexe A.

Rappelons que nous travaillons avec une séquence de  $n$  observations,  $x^n = (x_i)$ ,  $i = 1, \dots, n$ ,  $x_i \in I = [a, b]$ , issues d'un processus i.i.d.  $X$  dont la densité de probabilité est notée  $f$ . A partir d'une partition de  $I$  en  $k$  intervalles,  $P = \{I_j\}_{j=1, \dots, k}$ , il est possible de définir un histogramme approchant  $f$  :

$$\hat{f}_P = \sum_{j=1}^k \frac{n_j}{nL_j} 1_{I_j}$$

avec  $1_Y$ , la fonction indicatrice de l'ensemble  $Y$ ,  $L_j$ ,  $j = 1, \dots, k$ , la longueur de  $I_j$ , et  $n_j$ ,  $j = 1, \dots, k$ , le nombre de réalisations tombant dans  $I_j$ . L'objectif principal d'une méthode de sélection d'histogramme est de déterminer la partition  $P$  qui va permettre une estimation optimale de  $f$  par  $\hat{f}_P$ . A cet effet, nous définissons une partition  $P_{\max}$  contenant  $R$  intervalles de même longueur  $r$  :

$$a < a + r < a + 2r < \dots < a + (R - 1)r < a + Rr = b.$$

A partir de cette partition, il est possible de créer  $2R - 1$  sous-partitions dont les intervalles sont les unions d'intervalles adjacents de  $P_{\max}$ .

Parmi ces sous-partitions, nous recherchons celle qui va permettre d'encoder les échantillons  $x^n$  de manière à obtenir une chaîne de longueur minimale *via* le principe du *Minimum Description Length* (Grünwald, 2005). Nous présentons maintenant comment nous obtenons un critère à partir de considérations autour du codage de  $x^n$  permettant de trouver une partition "optimale" par minimisation. Le codage de  $x^n$  est réalisé en deux étapes.

#### A.1. 1<sup>re</sup> étape : codage arithmétique

Tout d'abord, nous commençons par transformer la séquence  $x^n$  en une séquence  $y^n$  :

$$y_i = \sum_{j=1}^k j \cdot 1_{\{x_i \in I_j\}}, i = 1, \dots, n.$$

En d'autres termes,  $y_i$  désigne le numéro de l'intervalle de  $P$  dans lequel tombe  $x_i$ . Afin d'encoder  $y^n$ , nous utilisons une version de la technique de codage arithmétique (Rissanen, 1976) nommée *Predictive Adaptive Arithmetic Coding* d'ordre 0 (0-PAAC) (Coq *et al.*, 2007). Par la suite,  $L(y^n | P)$  est la longueur de la chaîne binaire issue du codage de  $y^n$ . La technique PAAC est liée aux travaux de Rissanen car  $L(y^n | P)$  peut être estimée asymptotiquement par la complexité stochastique de  $y^n$  :

$$L(y^n | P) \approx - \sum_{j=1}^k n_j \log \frac{n_j}{n} + \frac{(k-1)}{2} C(n)$$

$C(n)$  étant ici la pénalité du critère  $\Phi_\beta$  (voir section 2.1.2). Pour finir, remarquons que, quand  $k$  augmente,  $L(y^n | P)$  a tendance à augmenter car il devient de plus en plus coûteux d'encoder les symboles  $y_i$  qui peuvent prendre  $k$  différentes valeurs.

#### A.2. 2<sup>e</sup> étape : codage uniforme

L'information  $y_i = j$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, k$ , ne permet pas à elle seule de pouvoir retrouver  $x_i$ . Afin de retrouver précisément la localisation de  $x_i$  à l'intérieur de l'intervalle  $I_j$ , à une précision  $r$  donnée, nous utilisons un codage uniforme de

l'intervalle. L'intervalle  $I_j$  contient alors  $L_j/r$  nombres réels et la longueur du codage des  $x_i$ , à l'intérieur des intervalles associés est :

$$L(x^n | y^n) = \sum_{j=1}^k n_j \log \left( \frac{L_j}{r} \right)$$

Contrairement à  $L(y^n | P)$ ,  $L(x^n | y^n)$  va décroître quand  $k$  augmente. Plus  $k$  est important, plus les intervalles de  $P$  sont petits et il devient alors plus facile de retrouver  $x_i$  dans chacun d'entre eux.

### A.3. Optimisation du critère d'information

En additionnant les deux longueurs de code,  $L(y^n | P)$  et  $L(x^n | y^n)$ , l'IC de l'équation (4) peut être obtenu en supprimant les termes constants dépendant de  $r$ . Par contre, pour pouvoir développer la procédure de programmation dynamique permettant de trouver la partition qui minimise le critère, il faut partir de sa forme complète :

$$IC(x^n, P) = - \sum_{j=1}^k n_j \log \frac{n_j}{n L_j} + \left( \frac{k-1}{2} \right) C(n) - n \log r$$

avec  $k \leq R$ . Pour  $\tau \in [1, R]$ ,  $\tau \in \mathbb{N}$ , nous définissons les ensembles suivants :  $I(\tau) = [a, a + \tau r]$  et  $x(\tau) = \{x_i \text{ tels que } x_i \in I(\tau)\}$ . Notons  $n(\tau) = |x(\tau)|$ , avec  $|\cdot|$  le cardinal d'un ensemble, et  $P_{\max}(\tau)$ , la plus grande partition de l'intervalle tronqué  $I(\tau)$  :

$$a < a + r < a + 2r < \dots < a + (R-1)r < a + \tau r.$$

Il est alors possible d'écrire un critère associé à l'intervalle tronqué, avec  $k \leq \tau$  :

$$IC(x(\tau), P(\tau)) = - \sum_{j=1}^k n_j \log \frac{n_j}{n(\tau) L_j} + \left( \frac{k-1}{2} \right) C(n(\tau)) - n(\tau) \log r$$

où  $P(\tau)$  désigne une sous-partition de  $P_{\max}(\tau)$  avec  $k \leq \tau$  intervalles. C'est la forme complète du critère lorsqu'on oublie les échantillons qui sont en dehors de l'intervalle  $I(\tau)$ , avec  $\tau$  restreint à l'intervalle  $[1, R-1]$ . Soit  $\sigma \in [\tau, R]$ ,  $\sigma \in \mathbb{N}$ , et  $P(\tau, \sigma)$  une partition à  $k+1$  intervalles de  $I(\sigma)$ ,  $P(\tau, \sigma) = P(\tau) \cup I_{k+1}$  avec  $I_{k+1} = [a + \tau r, a + \sigma r]$ . Alors,

$$IC(x(\sigma), P(\tau, \sigma)) = - \sum_{j=1}^{k+1} n_j \log \frac{n_j}{n(\sigma) L_j} + \frac{k}{2} C(n(\sigma)) - n(\sigma) \log r$$

et la différence entre les deux critères donne :

$$\begin{aligned} \Delta IC(\tau, \sigma) &= IC(x(\sigma), P(\tau, \sigma)) - IC(x(\tau), P(\tau)) \\ &= (n(\sigma) - n(\tau)) \log \left( \frac{n(\sigma)(\sigma - \tau)}{n(\sigma) - n(\tau)} \right) + n(\tau) \log \left( \frac{n(\sigma)}{n(\tau)} \right) \\ &\quad + \frac{k}{2} C(n(\sigma)) - \left( \frac{k-1}{2} \right) C(n(\tau)) \end{aligned}$$

Cette différence ne dépend pas de la sous-partition  $I(\tau)$ . Cette propriété permet de construire un algorithme de programmation dynamique contenant trois boucles “For” imbriquées : sur  $k$ , le nombre d’intervalles variant de 2 à  $R$ , sur  $\sigma$  qui varie de  $k$  à  $R$ , et sur  $\tau$  qui varie de  $k_1$  à  $\sigma - 1$  (Coq, 2008).

Olivier Alata a obtenu en 1992 le diplôme d’ingénieur en électronique de l’ENSERB (Ecole Nationale Supérieure d’Electronique et Radio Electricité de Bordeaux) puis le Doctorat de l’Université de Bordeaux en Janvier 1998. De 1998 à 2011, il a été Maître de Conférences à l’Université de Poitiers. Depuis 2011, il est Professeur à Telecom Saint-Etienne, Université Jean Monnet, où il effectue ses recherches au Laboratoire Hubert Curien. Ses travaux concernent principalement la modélisation stochastique et le développement de méthodes pour l’analyse et le traitement des signaux et des images.

Christian Olivier a obtenu un doctorat en Probabilité et Statistique en 1979. Il a été Maître de Conférence à l’Université de Rouen de 1987 à 1997, où ses activités de recherche portèrent sur le développement d’outils mathématiques pour la segmentation d’images, l’analyse et la reconnaissance de formes. Il rejoint ensuite l’Université de Poitiers comme Professeur à l’Institut de recherche XLIM, département SIC (Signal, Image et Communication). Son activité de recherche porte toujours sur le développement de modèles, dans l’objectif d’une représentation optimale de l’information, cette représentation s’intégrant dans une stratégie conjointe de codage source/canal pour les transmissions sans fil d’images fixes et de vidéos.

Zhan Jin a fait quatre années d’études (2002-2006) en électronique à l’Université de Nanchang, Chine. Arrivé en France en 2007, il a obtenu son diplôme de Master en Télécommunication, Multimédia et Réseaux de l’Université de Poitiers en 2010. Il est actuellement ingénieur dans la société Fullhan à Shanghai, République Populaire de Chine.

Yannis Pousset a obtenu son diplôme de doctorat sur les radiocommunications mobiles à l’Université de Poitiers en 1998. Il est, depuis 2000, Maître de Conférences à l’Université de Poitiers et réalise ses activités de recherche au sein du laboratoire XLIM-SIC (Signal Image et Communication). Sa thématique de recherche est principalement axée sur l’étude de la propagation des ondes électromagnétiques pour les systèmes de communication sans fil SISO et MIMO.

