

# Sélection de modèles en restauration d’image

## Approche bayésienne dans le cas gaussien

Benjamin HARROUÉ<sup>1</sup>, Jean-François GIOVANNELLI<sup>1</sup>, Marcelo PEREYRA<sup>2</sup>

<sup>1</sup>IMS (Univ. Bordeaux, CNRS, B-INP)

<sup>2</sup>MACS (Heriot-Watt University, Édimbourg, Royaume-Uni)

benjamin.harroue@ims-bordeaux.fr, giova@ims-bordeaux.fr, m.pereyra@hw.ac.uk

**Résumé** – La sélection de modèles est un problème peu abordé dans l’inversion. Les modèles sont généralement choisis *a priori*, par un expert en supervision qui se base sur des données de références. Tester à la main tous les modèles *a priori* pour sélectionner le bon peut être très coûteux. Nous proposons d’aborder ce problème de manière plus automatique avec des méthodes bayésiennes, notamment pour calculer les probabilités *a posteriori*. Ces dernières reposent sur le calcul d’une quantité phare : l’évidence. Il s’agit ici d’une première mise en œuvre dans le cas gaussien circulant, avec paramètres inconnus qui plus est, qui demeure un cas essentiel en signal-image. La sélection de modèles concernera ici la forme des matrices de covariance pour la loi *a priori* de l’image et pour la distribution du bruit.

**Abstract** – This paper considers the objective comparison of statistical models to solve imaging inverse problems, with a focus on model selection. Imaging experts often address model comparison and selection in a supervised manner, which can be difficult and very time consuming. Here we adopt an unsupervised Bayesian approach and objectively compare the models based on their posterior probabilities. These probabilities directly depend on the marginal likelihood or “evidence” of the models. In this paper we focus on the family of Gaussian models with circulant covariance matrices and unknown hyper-parameters, which is predominant in signal and image processing. Here, we compare different types of covariance matrices for the image prior and measurement noise.

## 1 Introduction

La restauration d’image est un sujet d’intérêt dans de nombreux domaines : astronomie, imagerie médicale, radar, *etc.* et la littérature est vaste [1, 2]. Les difficultés récurrentes proviennent de la grande dimension des images traitées et du caractère mal-posé. Pour les résoudre, on s’appuie sur des modèles, notamment statistiques, comme des processus AR, des mélanges de distributions, des modèles gaussiens, *etc.* Il est nécessaire de connaître le paramétrage de ces modèles : l’ordre du processus AR, le nombre de composantes du mélange, le type de structure de covariance de la gaussienne, *etc.* L’intérêt de la sélection de modèles (SdM) apparaît de façon évidente. Le modèle est souvent fixé à la main : une des méthodes classiques consiste à disposer de données de référence pour les différents modèles candidats pour pouvoir ensuite les comparer. Cependant ces données de références ne reflètent pas toujours la réalité terrain, et cet écart peut engendrer nombre d’erreurs. De plus, cette approche requiert beaucoup de temps et un expert en supervision. Choisir des modèles *a priori* est souvent difficile. Nous proposons ici une méthode de comparaison automatique à partir des données, en considérant le modèle comme une quantité inconnue et d’utiliser l’approche bayésienne pour le sélectionner. Les modèles sont ici stationnaires et nous nous plaçons dans un cadre gaussien circulant. Cette hypothèse, qui permet de

faciliter les calculs numériques (notamment par FFT), demeure très présent en signal-image.

Il existe de nombreux critères [3–5] sur lesquels s’appuyer pour sélectionner un modèle comme les probabilités *a posteriori* ou les critères d’information (IC) - Information Criterion en anglais - : An IC (ou Akaike IC), Bayesian IC, Brigde Criterion, Deviance IC, Predictive BIC, Generalized IC, Widely Applicable BIC, *etc.* Nous sélectionnerons le modèle en utilisant une approche de type théorie de la décision bayésienne [6] : celui ayant la plus forte probabilité *a posteriori* parmi les modèles candidats. L’obtention de ces probabilités repose sur le calcul d’une intégrale appelée *evidence*. L’obstacle majeur provient du calcul difficile de cette quantité.

## 2 Notations et problématique

On considère un modèle convolutif avec bruit additif pour le système d’observations :  $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{e}$ . Les vecteurs  $\mathbf{y}$ ,  $\mathbf{x}$  et  $\mathbf{e} \in \mathbb{R}^P$  sont respectivement les données, l’objet original et le bruit.  $\mathbf{H} \in \mathbb{R}^{P \times P}$  est la matrice de convolution ( $P = N^2$  est le nombre de pixels).

### 2.1 Cadre gaussien

Dans nos travaux, les quantités  $\mathbf{x}$  et  $\mathbf{e}$  sont modélisées par des variables normales centrées de covariances  $\mathbf{R}_x$  et  $\mathbf{R}_e$ .

Elle se décompose en un produit d'une matrice et d'un scalaire :  $\mathbf{R}_x = \gamma_x^{-1} \mathbf{\Pi}_x$  et  $\mathbf{R}_e = \gamma_e^{-1} \mathbf{\Pi}_e$ . Les matrices  $\mathbf{\Pi}_x$  et  $\mathbf{\Pi}_e$  fixent la structure de la covariance, tandis que  $\gamma_x$  et  $\gamma_e$  quantifient les niveaux de signal et de bruit. Ces deux hyperparamètres, ici inconnus, sont stockés dans  $\gamma = [\gamma_x, \gamma_e]$ .

## 2.2 Matrices circulantes

On se place dans le cas où  $\mathbf{R}_x$ ,  $\mathbf{R}_e$  et  $\mathbf{H}$  sont circulantes : elles sont diagonalisables dans l'espace de Fourier. On note la matrice de transformée de Fourier discrète (TFD)  $\mathbf{F}$ , on a alors :

- $\mathbf{R}_x = \gamma_x^{-1} \mathbf{F}^\dagger \mathring{\mathbf{S}}_x \mathbf{F}$  où  $\mathring{\mathbf{S}}_x = \text{diag} \{ \mathring{s}_x(p), p = 1, \dots, P \}$
- $\mathbf{R}_e = \gamma_e^{-1} \mathbf{F}^\dagger \mathring{\mathbf{S}}_e \mathbf{F}$  où  $\mathring{\mathbf{S}}_e = \text{diag} \{ \mathring{s}_e(p), p = 1, \dots, P \}$
- $\mathbf{H} = \mathbf{F}^\dagger \mathring{\mathbf{H}} \mathbf{F}$  où  $\mathring{\mathbf{H}} = \text{diag} \{ \mathring{h}(p), p = 1, \dots, P \}$

où  $\dagger$  représente le transposé conjugué. Les matrices  $\mathring{\mathbf{S}}_x$  et  $\mathring{\mathbf{S}}_e$  correspondent aux densités spectrales de puissance (DSP) objet et bruit, aux facteurs  $\gamma_x$  et  $\gamma_e$  près.

## 2.3 Modèles de DSP

Nous nous intéressons à la sélection de modèle de DSP objet ( $I$  modèles) et bruit ( $J$  modèles), désignés respectivement par les indices  $i = 1, \dots, I$  et  $j = 1, \dots, J$ . Il y a donc  $K = IJ$  combinaisons modèle objet/modèle bruit possibles. Elles sont désignées la variable discrète  $\mathcal{M}$ , prenant la valeur  $k$ , où  $k = 1, \dots, K$ . Chaque valeur de  $k$  renvoie à une combinaison unique  $(i, j)$ . On ajoutera par la suite les indices  $i$  et  $j$  aux matrices associées :  $\mathring{\mathbf{S}}_x^i$  et  $\mathring{\mathbf{S}}_e^j$ .

Dans notre cas, les DSP peuvent suivre quatre modèles différents (cf. figure 1) :

- Lorentz :  $1/[(\pi L^2)(1 + [\nu_x/L]^2)(1 + [\nu_y/L]^2)]$
- Gauss :  $(2\pi L^2)^{-1} \exp[-(\nu_x + \nu_y)^2/(2L^2)]$
- Laplace :  $(4L^2)^{-1} \exp[-(|\nu_x| + |\nu_y|)/L]$
- Blanc :  $\mathbb{1}(\nu_x, \nu_y)$

où  $L$  pilote la largeur (fixée arbitrairement) et  $(\nu_x, \nu_y)$  le couple de variables de fréquence réduites.

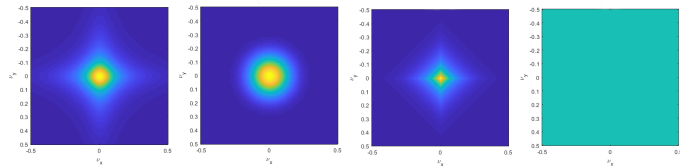


FIGURE 1 – Différentes structures de DSP. De gauche à droite : Lorentz - Gauss - Laplace - Blanc.

## 2.4 Sélection de modèles : approche bayésienne

Pour opérer la sélection de modèles, nous nous intéressons aux probabilités *a posteriori*  $p(\mathcal{M} = k | \mathbf{y})$ .

$$p(\mathcal{M} = k | \mathbf{y}) = \frac{f(\mathbf{y} | \mathcal{M} = k) p(\mathcal{M} = k)}{f(\mathbf{y})} \quad (1)$$

$$= \frac{f(\mathbf{y} | \mathcal{M} = k) p(\mathcal{M} = k)}{\sum_{l=1}^K f(\mathbf{y} | \mathcal{M} = l) p(\mathcal{M} = l)} \quad (2)$$

Ces probabilités reposent sur une quantité primordiale :  $f(\mathbf{y} | \mathcal{M} = k)$ , appelée l'**évidence** du modèle  $k$  :

$$f(\mathbf{y} | \mathcal{M} = k) = \int_{\gamma} \int_{\mathbf{x}} f(\mathbf{y}, \mathbf{x}, \gamma | \mathcal{M} = k) d\gamma d\mathbf{x} \quad (3)$$

$$= \int_{\gamma} f(\mathbf{y}, \gamma | \mathcal{M} = k) d\gamma \quad (4)$$

L'évidence est une vraisemblance marginale, où tous les paramètres inconnus ont été intégrés. L'équation (4) diffère de (3) par sa marginalisation en  $\mathbf{x}$ . Travailler sur (4) nécessite d'intégrer par rapport à l'objet  $\mathbf{x}$ . Quand c'est possible, on peut alors éviter le recours à l'échantillonnage, ce qui est un avantage considérable.

## 3 Calcul d'évidence

### 3.1 Distribution des données

Nous allons travailler avec la forme (4). Le système d'observations étant linéaire et les variables  $\mathbf{x}$  et  $\mathbf{e}$  étant normales, la distribution des données est également normale de moyenne et de covariance :

$$\mathbb{E}[\mathbf{y}] = 0$$

$$\mathbf{R}_y^k = \mathbb{E}[\mathbf{y}\mathbf{y}^\dagger] = \mathbf{H}\mathbf{R}_x^i\mathbf{H}^\dagger + \mathbf{R}_e^j$$

Les matrices  $\mathbf{R}_x^i$ ,  $\mathbf{R}_e^j$  et  $\mathbf{H}$  étant circulantes, la matrice  $\mathbf{R}_y^k$  l'est également :

$$\begin{aligned} \mathbf{R}_y^k &= \mathbf{H}\mathbf{R}_x^i\mathbf{H}^\dagger + \mathbf{R}_e^j \\ &= \mathbf{F}^\dagger (\gamma_x^{-1} \mathring{\mathbf{H}} \mathring{\mathbf{S}}_x^i \mathring{\mathbf{H}}^\dagger + \gamma_e^{-1} \mathring{\mathbf{S}}_e^j) \mathbf{F} = \mathbf{F} \mathring{\mathbf{S}}_y^k \mathbf{F}^\dagger \end{aligned}$$

où  $\mathring{\mathbf{S}}_y^k = \text{diag} \left\{ \mathring{s}_y^k(p), p = 1, \dots, P \right\}$  représente la DSP des données et

$$\mathring{s}_y^k(p) = \gamma_x^{-1} |\mathring{h}(p)|^2 \mathring{s}_x^i(p) + \gamma_e^{-1} \mathring{s}_e^j(p) \quad (5)$$

La vraisemblance  $f(\mathbf{y} | \gamma_x, \gamma_e, \mathcal{M} = k)$ , indispensable dans le calcul de l'évidence (voir l'équation (8)), s'écrit donc dans le domaine fréquentiel. Nous allons simplifier le terme du déterminant et celui de la norme de  $\mathbf{y}$  :

$$\det(\mathbf{R}_y^k) = \prod_{p=1}^P \mathring{s}_y^k(p) \quad \text{et} \quad \mathbf{y}^\dagger (\mathbf{R}_y^k)^{-1} \mathbf{y} = \sum_{p=1}^P \frac{|\mathring{y}(p)|^2}{\mathring{s}_y^k(p)} \quad (6)$$

où  $\hat{\mathbf{y}} = \mathbf{F}\mathbf{y}$  est la TFD des données  $\mathbf{y}$ .

Par la suite, nous utiliserons la log vraisemblance, dans laquelle on injecte (6) :

$$\begin{aligned} \log f(\mathbf{y}|\gamma, \mathcal{M} = k) &= -\frac{1}{2}(P \log(2\pi) + \log \prod_{p=1}^P s_y^{\circ k}(p) + \sum_{p=1}^P \frac{|\hat{y}(p)|^2}{s_y^{\circ k}(p)}) \\ &= -\frac{1}{2}\left(P \log(2\pi) + \sum_{p=1}^P \left(\log(s_y^{\circ k}(p)) + \frac{|\hat{y}(p)|^2}{s_y^{\circ k}(p)}\right)\right) \quad (7) \end{aligned}$$

### 3.2 Évidence : expression finale

D'après (4), on peut écrire l'évidence :  $f(\mathbf{y}|\mathcal{M} = k)$

$$\begin{aligned} &= \int_{\gamma_e, \gamma_x} f(\mathbf{y}, \gamma_x, \gamma_e | \mathcal{M} = k) d\gamma_x d\gamma_e \\ &= \int_{\gamma_e, \gamma_x} f(\mathbf{y}|\gamma_x, \gamma_e, \mathcal{M} = k) \underbrace{\pi(\gamma_x, \gamma_e | \mathcal{M} = k)}_{\gamma_x, \gamma_e \text{ et } \mathcal{M} \text{ indépendants}} d\gamma_x d\gamma_e \quad (8) \\ &= \int_{\gamma_e, \gamma_x} \underbrace{f(\mathbf{y}|\gamma_x, \gamma_e, \mathcal{M} = k) \pi(\gamma_x) \pi(\gamma_e)}_{\exp[\log J(k)]} d\gamma_x d\gamma_e \end{aligned}$$

où le logarithme de la loi jointe est :

$$\begin{aligned} \log J(k) &= \log f(\mathbf{y}|\gamma_x, \gamma_e, \mathcal{M} = k) + \log \pi(\gamma_x) + \log \pi(\gamma_e) \quad (9) \end{aligned}$$

Dans l'étude analytique, les variables  $\gamma_x$  et  $\gamma_e$  sont continues. Numériquement, ces intégrales sont approchées par des sommes discrètes sur des grilles de valeurs de  $\gamma_x$  et  $\gamma_e$ . La forme (9) est utilisée pour les calculs numériques, permettant de manipuler des valeurs moins grandes.

## 4 Calculs numériques et résultats

### 4.1 Simulation de $x$ , $e$ et $y$

Pour les simuler les données  $\mathbf{y}$ , on procède de la façon suivante : on génère  $\hat{\mathbf{y}} = \hat{\mathbf{H}}\hat{\mathbf{x}} + \hat{\mathbf{e}}$ , puis on obtient  $\mathbf{y}$  par TFD inverse. Pour obtenir  $\hat{\mathbf{y}}$ , il faut au préalable simuler  $\hat{\mathbf{x}}$  et  $\hat{\mathbf{e}}$ .

Comme la variable  $x$  suit une loi normale, la variable  $\hat{x}$  en suit également une, dont on peut déterminer la moyenne et la covariance :

$$\begin{aligned} \mathbb{E}[\hat{x}] &= \mathbb{E}[\mathbf{F}\mathbf{x}] = \mathbf{F}\mathbb{E}[\mathbf{x}] = 0 \\ \mathbf{R}_{\hat{x}} &= \mathbb{E}[\hat{x}\hat{x}^\dagger] = \mathbf{F}\mathbb{E}[\mathbf{x}\mathbf{x}^\dagger]\mathbf{F}^\dagger = \mathbf{F}\mathbf{R}_x\mathbf{F}^\dagger = \gamma_x^{-1}\mathbf{S}_x^i \end{aligned}$$

$\mathbf{S}_x^i$  étant diagonale, les variables  $\hat{x}(p)$  sont indépendantes. Elles sont également centrées, de variance respective  $\gamma_x^{-1}s_x^i(p)$ .

On simule donc  $\hat{x}(p) \sim \mathcal{N}(0, \gamma_x^{-1}s_x^i(p))$  pour obtenir  $\hat{\mathbf{x}}$ . On effectue ensuite la convolution dans le domaine fréquentiel, en multipliant  $\hat{\mathbf{x}}$  par le gain  $\hat{\mathbf{H}}$ . On tire le bruit  $\hat{\mathbf{e}}$  de la même manière que  $\hat{\mathbf{x}}$ , pour enfin construire  $\hat{\mathbf{y}} = \hat{\mathbf{H}}\hat{\mathbf{x}} + \hat{\mathbf{e}}$ .

### 4.2 Calcul numérique

Pour la mise en œuvre numérique, nous avons généré 100 jeux de données comportant une réalisation, pour chacun des  $K = 16$  modèles. On fixe  $N = 128$ , le gain à un sinus cardinal pour la convolution et le prior  $\pi(\gamma_x, \gamma_e)$  à une loi uniforme. Les vrais niveaux objet et bruit sont fixés à  $\gamma_x^* = 6$  et  $\gamma_e^* = 4$ .

Sur la Figure 2 sont affichées les lois *a posteriori* jointes (à une constante près) de  $(\gamma_x, \gamma_e)$  pour l'une des réalisations du modèle suivant : DSP objet Lorentzienne et DSP bruit blanc. Le modèle bruit est fixé à Blanc, et on calcule les distributions pour les différents modèles objets candidats.

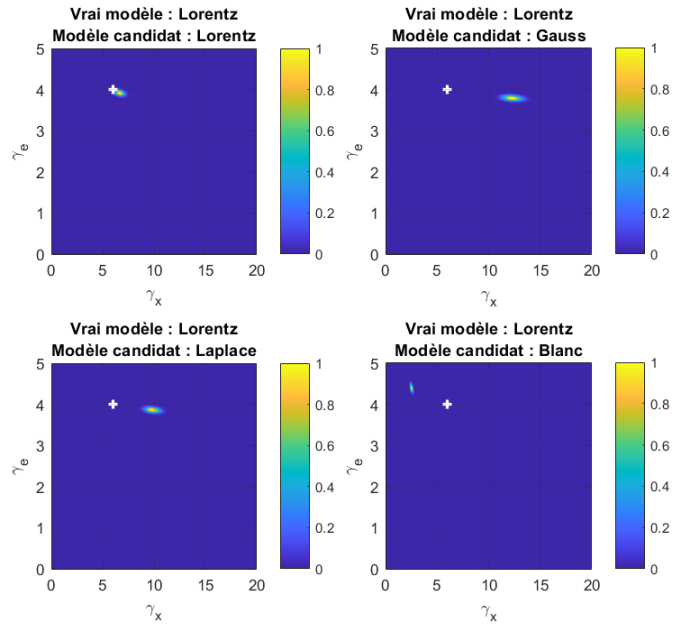


FIGURE 2 – Lois *a posteriori* jointes pour différents modèles objet candidats. Les vraies valeurs  $(\gamma_x^*, \gamma_e^*)$  sont repérées par une croix blanche.

**Remarque** : lorsque l'on considère le vrai modèle, le pic de la distribution se trouve aux coordonnées  $(\gamma_x^*, \gamma_e^*)$ .

On marginalise ensuite ces distributions par rapport à  $\gamma_x$  et  $\gamma_e$ . Pour éviter les problèmes de précision numérique, on calcule pour chaque modèle  $k$  :

$$\tilde{e}_k = \int_{\gamma_e, \gamma_x} \exp[\log J(k) - \Delta(k)] d\gamma_x d\gamma_e$$

où  $\Delta(k) = \max_{\gamma_x, \gamma_e} \log J(k)$ . Le maximum de la loi jointe est alors normalisé à 1 (voir Figure 2).

La quantité  $\tilde{e}_k$  est telle que  $f(\mathbf{y}|\mathcal{M} = k) = \tilde{e}_k \cdot \exp[\Delta(k)]$ . Le prior sur les modèles  $p(\mathcal{M} = k)$  étant supposé uniforme, on calcule ensuite les probabilités *a posteriori* ainsi (d'après (2)) :

$$p(\mathcal{M} = k | \mathbf{y}) = \tilde{e}_k \cdot \left[ \sum_{l=1}^K \tilde{e}_l \cdot \exp[\Delta(l) - \Delta(k)] \right]^{-1}$$

Les distributions sont calculées sur une grille de valeurs  $(\gamma_x, \gamma_e)$ , de taille  $55 \times 50$  points parcourant  $[0, 15] \times [0, 14]$ . Pour sélectionner le modèle  $\mathcal{M}$  à partir d'une réalisation, il faut environ 9,21 secondes.

Les performances de sélections sont affichées en Figure 3. On peut voir que pour les seize configurations différentes, on obtient de très bons résultats. Sur 100 réalisations, certains cas atteignent 100% de bonnes sélections, lorsque les modèles objets sont gaussiens et pour les combinaisons Blanc/Laplace et Blanc/Blanc. Les cas Lorentz/Gauss et Lorentz/Blanc affichent quant à eux les plus basses performances, qui n'en restent pas moins élevées avec 91% de sélections correctes. Dans les configurations restantes, les proportions de bonnes sélections s'élèvent à plus de 95%. Il est à noter qu'il y a peu de cas, où les modèles objet et bruit sont tous deux mal sélectionnés : moins de 2%.

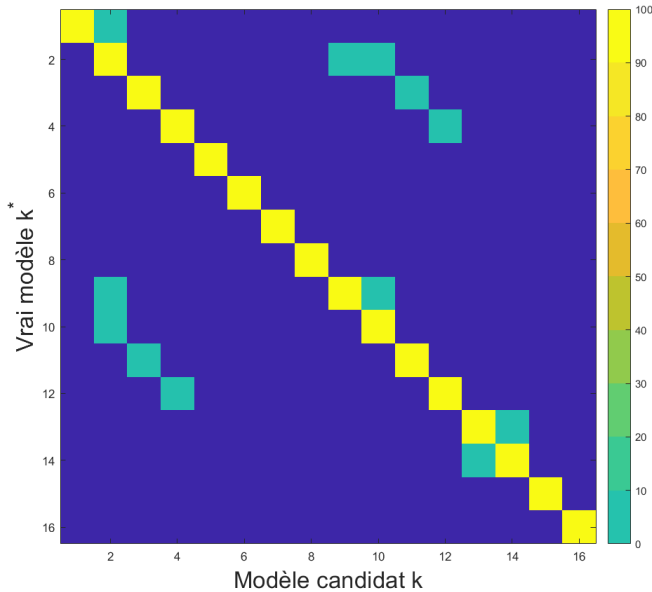


FIGURE 3 – Nombre de sélections pour chaque modèle candidat (abscisse) sachant le vrai modèle (ordonnée). La couleur bleue foncée correspond à la valeur 0.

A titre comparatif, nous avons traité des images  $64 \times 64$  et les performances chutent jusqu'à 61% seulement de bonnes sélections pour certains modèles.

## 5 Synthèse et perspectives

Nous avons présenté ici une première contribution de la sélection automatique de modèles basée sur une approche

bayésienne. Notre stratégie de sélection de modèle, basée sur le MAP, est optimale au sens de l'approche bayésien. Dans notre approche, nous cherchons à estimer le modèle à partir des données observées. On souhaite évaluer la probabilité  $p(\mathcal{M} = k | \mathbf{y})$  pour tous les modèles candidats, où l'*a priori* sur les modèles  $p(\mathcal{M} = k)$  est considéré uniforme. On a alors besoin de calculer l'évidence à partir de sa définition donnée en (8) : on explicite les équations (7) et (9), puis on marginalise  $\exp[\log J(k)]$  par rapport aux hyperparamètres  $\gamma_x$  et  $\gamma_e$ . Bien qu'il s'agisse du cas gaussien circulant, il a fallu pallier quelques difficultés, notamment numériques : l'intégration des hyperparamètres sur une grille de valeurs, la grande dimension des images, etc. Ces premiers travaux donnent de très bonnes performances en terme de sélection, ce qui nous encourage à étendre notre cadre d'étude. Dans une version étendue de cet article, nous établirons une comparaison complète des différentes méthodes existantes : Chib, moyenne harmonique, RJMCMC, WBIC, etc qui permettent de calculer les mêmes probabilités *a posteriori* des modèles. Il sera également intéressant de les mettre en regard avec des critères d'information comme le AIC ou le BIC.

Il est très souvent possible, pour des distributions non-usuelles, de les approcher avec des lois normales ou des mélanges de lois normales, et donc de se ramener en partie à notre étude. Parmi les perspectives envisageables, il s'agira d'étendre le problème au cas gaussien non circulant, puis au cas non-gaussien. Il sera possible de coupler la SdM à des méthodes d'échantillonnage, pour des distributions difficiles ou de grande dimension [7]. Il sera également possible d'étendre les travaux sur la manière de marginaliser les hyperparamètres  $\gamma_x$  et  $\gamma_e$ . Nous pourrions également intégrer au problème de nouveaux hyperparamètres inconnus, comme la largeur des DSP objet et bruit  $L$ . Nous pourrions alors passer à la SdM sur données réelles, dont les paramètres comme  $\gamma_x$  et  $L$  sont souvent inconnus.

## Références

- [1] J.-F. Giovannelli et J. Idier. *Méthodes d'inversion appliquées au traitement du signal et de l'image*, Traité IC2, Hermès, Nov 2013.
- [2] J. Kaipio *Statistical and Computational Inverse Problems (Applied Mathematical Sciences)*, Springer, 2005.
- [3] T. Ando. *Bayesian Model Selection and Statistical Modeling*, Chapman and Hall/CRC, 2010.
- [4] J. Ding, V. Tarokh et Y. Yang. *Model Selection Techniques - An overview*, IEEE Signal Processing Magazine, Nov 2018.
- [5] S. Watanabe. *A Widely Applicable Bayesian Information Criterion*, Journal of Machine Learning Research 14 (2013) 867-897, 2013.
- [6] C. P. Robert. *Le choix bayésien - Principes et pratique*, Springer Editions, 2010.
- [7] A. Durmus, E. Moulines et M. Pereyra. *Efficient Bayesian computation by proximal Markov chain Monte Carlo : when Langevin meets Moreau*, SIAM Journal on Imaging Sciences, vol. 11, no. 1, 473-506. Mar. 2018