



HAL
open science

BAYESIAN TEXTURE MODEL SELECTION BY HARMONIC MEAN

Cornelia Vacar, Jean-François Giovannelli, Ana-Maria Roman

► **To cite this version:**

Cornelia Vacar, Jean-François Giovannelli, Ana-Maria Roman. BAYESIAN TEXTURE MODEL SELECTION BY HARMONIC MEAN. International Conference on Image Processing, Oct 2012, Orlando, United States. pp.2003-2006. hal-00781487

HAL Id: hal-00781487

<https://hal.science/hal-00781487>

Submitted on 28 Jan 2013

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.

BAYESIAN TEXTURE MODEL SELECTION BY HARMONIC MEAN

Cornelia VACAR[†], Jean-François GIOVANNELLI[†], Ana-Maria ROMAN^{*}

[†]Univ. Bordeaux, IMS, UMR5218, F-33400 Talence, France

^{*}Universitatea Tehnică din Cluj-Napoca, Departamentul de Comunicații, Cluj-Napoca, România

ABSTRACT

The paper presents a model selection method for texture images, more specifically, it finds the most adequate model for the pixels' interaction. This approach relies on a Bayesian framework, that probabilizes all the quantities and determines the joint *a posteriori* law for the models and the parameters. In order to compute the *a posteriori* model probabilities, the model parameters are marginalized by means of sampling, performed independently for each model in a within-model sampling using a Metropolis-Hastings (M-H) algorithm. The resulting chains are used to compute the evidence of each model by an harmonic averaging of the likelihoods computed for the aforementioned sampled values. The work presented in the following represents a complex formalism based on state of the art methods for parameter estimation, model selection techniques and sampling algorithms, the novelty being the design of such an approach for a texture model selection problem. An image processing application of this kind raises serious difficulties regarding the large amount of data, the data correlation and the highly non-linear dependencies of the data with respect to the parameters. Despite these challenges, our method successfully solves the problem of texture model selection and parameter estimation.

Index Terms— Texture, model choice, harmonic mean, Bayes, Metropolis-Hastings

1. INTRODUCTION

Model selection can prove to be an important tool in data analysis, signal processing and, as shown in the following, even in image processing. The goal is to select the model that best describes the data, concurrently with computing the parameters of that model. In our case, the data is the observed image, denoted by \mathbf{X} , the models are $\mathcal{M} = k$, with $k = 1, \dots, K$, and each model is driven by the parameters θ_k . In such a setting, the Bayesian framework is conceptually well adapted and allows the development of a unified approach to solve the twofold problem.

The use of the Bayesian paradigm naturally leads to handling the joint law for the models, the parameters and the data, from which the *a posteriori* law for the models and the parameters can be deduced. The information contained by this law is extracted using numerical methods, sampling to be more exact and, regarding these methods, the literature is divided into two types of approaches: across-model simulation and **within-model** simulation.

The across-model approach consists in jointly sampling the model and the parameters corresponding to that model, thus allowing the algorithm to visit the models and jump from one to another concurrently with choosing adapted values for the parameters. This principle guides the state of the art algorithm of Reversible Jump Monte Carlo Markov Chain (RJCMC), first introduced in [1] and then extensively studied and expanded in numerous papers, as shown in [2, 3]. One of the immediate applications of this algorithm to a

Bayesian texture analysis problem, with an interesting perspective on texture modeling by Beta mixtures, is presented in [4].

The within-model simulation technique consists in computing the model probabilities by marginalizing out the parameters from the joint *a posteriori* law. This is done by sampling the parameters conditionally to each model, the values obtained being used to compute the “evidence”, *i.e.*, the probability of the data, given the model. The evidence for model $\mathcal{M} = k$ can be obtained using a harmonic mean of likelihood values computed for the parameter samples, drawn conditionally with respect to the data and under the assumption that $\mathcal{M} = k$ is the generating model. A detailed description of this principle can be found in [5, 6] and more recently in [7], where a survey of within-model simulation methods is presented.

An image processing application, more specifically texture analysis, will illustrate the model selection method. Our previous work [8], devoted to efficient sampling algorithms for the same types of textures, is thus continued and enriched by a method that uses the parameter samples to select the generating model.

2. DATA MODEL

Most of the applications of the above methods attempt to determine the best model for the data probability distribution. In a significant part of the literature, the observations consist of i.i.d. data, so these methods do not deal with data correlations and assume the observations have a common distribution. Our goal was to assemble a method to solve a more complex task: selecting the model for the pixel correlation structure in a texture image.

The work focuses on textures modeled as zero-mean stationary Gaussian Random Fields (GRFs), as they are rather versatile in generating diverse patterns (see Fig.1) and at the same time easy to manipulate and embed in a probabilistic framework. The model is completely described by its second order statistics and a distribution law:

$$f(\mathbf{x}|\mathbf{R}) = (2\pi)^{-N/2} \det(\mathbf{R})^{-1/2} \exp\left[-\frac{1}{2}\mathbf{x}^t \mathbf{R}^{-1} \mathbf{x}\right] \quad (1)$$

where \mathbf{x} collects the N pixels and \mathbf{R} is the correlation matrix. In the frequency domain, by Whittle approximation, the previous expression becomes separable in terms of the Fourier coefficients:

$$f(\hat{\mathbf{x}}|\boldsymbol{\lambda}) = (2\pi)^{-N/2} \prod_{n=1}^N \lambda_n^{1/2} \exp\left[-\frac{1}{2} \sum_{i=1}^N |\hat{x}_n|^2 \lambda_n\right] \quad (2)$$

In fact, every element λ_n is the variance of the n th coefficient of the image's Fourier transform. Thus, as a physical interpretation, $\boldsymbol{\lambda}$ describes the image's Power Spectral Density (PSD).

An important tool in the subsequent developments is the negative log-likelihood (NLL) of the data knowing the underlying model:

$$\text{NLL} = \frac{N}{2} \log 2\pi - \frac{1}{2} \sum_{n=1}^N \left(\log \lambda_n - |\hat{x}_n|^2 \lambda_n\right) \quad (3)$$

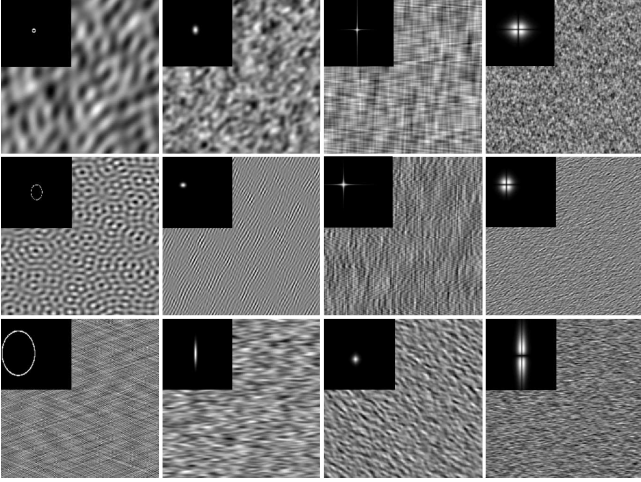


Fig. 1: Texture realizations using GRFs with structured PSDs [8].

This is where the notion of model intervenes, as the texture can be obtained by using any type of pixel interaction model, *i.e.*, any form for the PSD. The dictionary of models chosen in the current work consists of parametric forms: Generalized Gaussian ($\mathcal{M} = 1$), Student ($\mathcal{M} = 2$), Chi-Square ($\mathcal{M} = 3$), Ring ($\mathcal{M} = 4$). Each model $\mathcal{M} = k$ is driven by the set of parameters θ_k , Table 1 showing the dependencies $\lambda_k(\theta_k)$. Some parameters are common to all the models, such as the scale factor γ , the central frequencies ν_{x0}, ν_{y0} , that give the positioning in the frequency space, and the widths σ_x, σ_y that determine the spread of the function. There are also some model specific parameters: the interior and exterior radii r_1, r_2 and the power p , which gives the models the possibility to have a wide range of shapes according to its values.

3. METHODOLOGY

3.1. Bayesian Setting

The model choice is a continuously active branch of statistics, that compares the models based on their structural properties and selects from a family of models the one that is best adapted to the available data, *i.e.*, the observations and the *a priori* information.

The Bayesian approach is fundamentally probabilistic, *i.e.*, a joint probability distribution is used to describe the relations between the unknowns and the data, and from this the *a posteriori* distribution is deduced, which extracts the useful information from the data.

In the current problem, we have a set of $K = 4$ models that will be tested for the data \mathbf{X} . For the k th model, \mathbf{X} has a law $f(\mathbf{X}|\theta_k, \mathcal{M} = k)$, as seen in (3). The Bayesian approach pro-

Table 1: Pixel correlation models for the textures

Model $\mathcal{M} = k$	Expression of $\lambda_k(\theta_k)$
Generalized Gauss	$\frac{1}{\gamma} \exp \frac{1}{2} \left[\frac{ \nu_x - \nu_{x0} ^p}{\sigma_x^2} + \frac{ \nu_y - \nu_{y0} ^p}{\sigma_y^2} \right]$
Student	$\frac{1}{\gamma} \left[1 + \frac{(\nu_x - \nu_{x0})^2}{\sigma_x^2} + \frac{(\nu_y - \nu_{y0})^2}{\sigma_y^2} \right]^p$
Chi-Square	$\frac{1}{\gamma} \nu_x - \nu_{x0} ^{1/2-p} \nu_y - \nu_{y0} ^{1/2-p} \exp \frac{1}{2} \left[\frac{ \nu_x - \nu_{x0} }{\sigma_x} + \frac{ \nu_y - \nu_{y0} }{\sigma_y} \right]$
Ring	$\begin{cases} \frac{1}{\gamma}, & \text{if } r_1^2 \leq \frac{(\nu_x - \nu_{x0})^2}{\sigma_x^2} + \frac{(\nu_y - \nu_{y0})^2}{\sigma_y^2} \leq r_2^2 \\ 10^5, & \text{otherwise} \end{cases}$

ceeds by attributing an *a priori* distribution $\pi(\theta_k|\mathcal{M} = k)$ to the parameters of model k , consisting of uniform laws on the interval $[-0.5, 0.5]$ for the central frequencies, as they represent the positions in the reduced frequency domain. For the widths we chose uniform *a priori* laws $[0.01, 1]$, as beyond these values the laws are either too narrow and will correspond to uncorrelated fields, either too wide and then all pixels will be correlated to all the rest. For p, r_1, r_2 and γ uniform *a priori* laws were also used, in accordance with their physical significance. The *a priori* probabilities for each model were set to $\Pr(\mathcal{M} = k) = 1/K$, so that no model is favored with respect to another and the data decides on the model adequacy.

Intuitively, this complete specification can be regarded as a three stage model of hierarchic structure to generate the data \mathbf{X} : firstly, the model k is chosen using $\Pr(\mathcal{M} = k)$, secondly, the parameter vector θ_k is obtained by using $\pi(\theta_k|\mathcal{M} = k)$ and, thirdly, the data is generated according to $f(\mathbf{X}|\theta_k, \mathcal{M} = k)$ [9].

3.2. Posterior Model Probabilities

The problem amounts to finding the most probable model, given the data, and for this the *a posteriori* probabilities of the models will be used. By applying Bayes' theorem, the probability of model k is:

$$\Pr(\mathcal{M} = k|\mathbf{X}) = \frac{\Pr(\mathcal{M} = k)f(\mathbf{X}|\mathcal{M} = k)}{f(\mathbf{X})} \quad (4)$$

In (4) there are two quantities that have to be computed in order to select the most probable model:

1. the probability distribution of the data, $f(\mathbf{X})$,
2. the evidence for each model, $f(\mathbf{X}|\mathcal{M} = k)$.

Fortunately, $f(\mathbf{X})$ is model independent and thus it can be numerically computed by factorizing it and imposing that the model probabilities add to 1:

$$f(\mathbf{X}) = \sum_{k=1}^K f(\mathbf{X}|\mathcal{M} = k)\Pr(\mathcal{M} = k) \quad (5)$$

The second quantity of interest, the evidence, is determined from the joint law, by marginalizing the parameters and amounts to calculating the integral:

$$f(\mathbf{X}|\mathcal{M} = k) = \int f(\mathbf{X}|\theta_k, \mathcal{M} = k)\pi(\theta_k|\mathcal{M} = k)d\theta_k \quad (6)$$

which is mathematically intractable due to the very complex dependencies with respect to the parameters [8]. Thus, the integral must be computed numerically, as detailed in Section 3.3.

Remark: In the final stage, the pairwise model comparison, say $\mathcal{M} = k$ and $\mathcal{M} = k'$ could be summarized by:

$$\underbrace{\frac{\Pr(\mathcal{M} = k|\mathbf{X})}{\Pr(\mathcal{M} = k'|\mathbf{X})}}_{A \text{ posteriori chances}} = \underbrace{\frac{f(\mathbf{X}|\mathcal{M} = k)}{f(\mathbf{X}|\mathcal{M} = k')}}_{\text{Bayes Factors}} \times \underbrace{\frac{\Pr(\mathcal{M} = k)}{\Pr(\mathcal{M} = k')}}_{A \text{ priori chances}} \quad (7)$$

3.3. Harmonic Mean

As previously seen, the within-model analysis requires the computation of the evidences, $f(\mathbf{X}|\mathcal{M} = k)$. If these could be calculated explicitly, the *a posteriori* model probability would result immediately. However, this is not possible and a numerical computation must be used, meaning that the usual difficulty of the Bayesian approach, *i.e.*, computing integrals of the type (6), will be eliminated by using MCMC methods.

Relation (6) can be seen as an expectation of the likelihood under the *a priori* law. In numerical terms, this can be computed by drawing samples from the *a priori* and averaging the likelihood computed

for these samples. However, the laws $f(\mathbf{X}|\boldsymbol{\theta}_k, \mathcal{M} = k)$ for our texture images are very peaked, and the *a priori* $\pi(\boldsymbol{\theta}_k|\mathcal{M} = k)$ for the parameters are uniform, meaning that if we were to sample under the *a priori* there would be a high risk of drawing a considerable amount of samples from regions of very low probability, leading to a high rejection rate and a slow sampling process.

In order to avoid this, in a similar manner as with importance sampling, the integral is computed by introducing a more adequate sampling distribution and then compensating for this change. Thus, integral (6) can be rewritten as:

$$f(\mathbf{X}|\mathcal{M} = k) = \mathbb{E}_g \left[f(\mathbf{X}|\boldsymbol{\theta}_k, \mathcal{M} = k) \frac{\pi(\boldsymbol{\theta}_k|\mathcal{M} = k)}{g(\boldsymbol{\theta}_k)} \right] \quad (8)$$

where g is an auxiliary function, not necessarily the sampling distribution. Its role is to make possible the sampling under a more adapted law that will allow a good exploration of the parameter space, but will also take into account the form of the likelihood. Such a law is the *a posteriori* and then (8) becomes:

$$f(\mathbf{X}|\mathcal{M} = k) = \left\{ \mathbb{E}_{\boldsymbol{\theta}_k|\mathbf{X}, \mathcal{M}=k} [f(\mathbf{X}|\boldsymbol{\theta}_k, \mathcal{M} = k)^{-1}] \right\}^{-1} \quad (9)$$

Practically, knowing that $\boldsymbol{\theta}_k^{(t)}$ with $t = 1, \dots, T$ are samples from the *a posteriori* law, the evidence can be numerically computed as:

$$\tilde{f}(\mathbf{X}|\mathcal{M} = k) = \left\{ \frac{1}{T} \sum_{t=1}^T [f(\mathbf{X}|\boldsymbol{\theta}_k^{(t)}, \mathcal{M} = k)]^{-1} \right\}^{-1} \quad (10)$$

This is in fact the harmonic mean of the likelihood values computed for the samples $\boldsymbol{\theta}_k^{(t)}$. For this estimator, the g function is equivalent to the inverse of an importance function and thus the estimator performs better when g has lighter tails than $\pi(\boldsymbol{\theta}_k|\mathbf{X}, \mathcal{M} = k)$ [7].

Even though $\tilde{f}(\mathbf{X}|\mathcal{M} = k)$ converges to the true value $f(\mathbf{X}|\mathcal{M} = k)$ when $T \rightarrow \infty$, it does not generally satisfy the Gaussian central limit theorem. To be more exact, occasionally, a value of $\boldsymbol{\theta}_k^{(t)}$ may occur that has low probability and thus has important effects on the final result [6]. These reservations were reinforced by [10], making the harmonic mean a rather avoided algorithm, since in some cases it may exhibit infinite variance. However, its computational efficiency and ease of implementation have led to its use in a series of topic modeling papers such as [11, 12].

All these considerations are true for the general case, however, for the current application, the probability of drawing a sample of very low probability region is extremely small, as the sampling law (the *a posteriori*) is very peaked. Moreover, even if such a sample from a low probability region were to occur, in our case it would only do so during the burn-in period, as our sampler allows a thorough exploration of the law, but especially its high probability regions. This actually means that the general doubts regarding the infinite variance do not apply in our case due to the form of the posterior law, as long as we make sure the sampler has reached convergence before using the values in the computations.

3.4. Parameter Estimates

The samples drawn from the posterior distribution are not only used for computing the harmonic mean, and thus the evidence, but also to compute a parameter estimate. In this case, the *A Posteriori* Expectation estimator was chosen and it consists in calculating the expectation:

$$\hat{\boldsymbol{\theta}} = \mathbb{E}_{\boldsymbol{\theta}|\mathbf{X}, \mathcal{M}=k}[\boldsymbol{\theta}] \quad (11)$$

numerically computed by averaging the samples drawn from the *a posteriori* distribution.

3.5. Metropolis-Hastings Sampling

The evidences are the key ingredients for computing the model probabilities and they are obtained by a harmonic mean of likelihoods computed for a series of samples $\boldsymbol{\theta}^{(t)}$, drawn from the *a posteriori*.

At this point, the difficulty of sampling these laws must be pointed out, due to the complicated dependencies with respect to $\boldsymbol{\theta}$, as shown in [8]. For this reason, direct sampling is infeasible and thus the M-H sampler will be used. Among the class of M-H samplers we chose the Fisher M-H, introduced in [8] due to its very good speed performances when applied to texture parameter sampling. Although it adds an extra layer of complexity to an already multi-stage multi-algorithm combining approach, it exploits particularities of the problem and enhances the sampling speed.

The algorithm starts by initializing the parameter and at each iteration it makes a proposal, by a directional perturbation of the current parameter value. This proposal is either accepted, or rejected (the current value is repeated) according to an acceptance probability dictated by the proposal law and the *a posteriori* law.

4. ALGORITHM

The methodological and algorithmic ingredients previously presented are mainly state of the art, being complex tools in probability theory, random fields analysis and sampling. They are used to assemble the model selection method for texture images, a composite algorithm that inherits their capabilities, but also their complexity.

The implementation of the algorithm posed a series of numerical problems, mainly due to the exponential form of the likelihood law, amounting to quantities that overpass Matlab's representation capabilities and are set to zero or infinity. To be more exact, we would need to compute exponentials with arguments superior to 10^3 or inferior to -10^3 . For this reason NLLs are computed and handled in the sampling phase instead of the likelihoods themselves.

This solves only partially the problem, as when computing the harmonic means we need to use the exponential quantities. To overcome this standstill, a normalization was in order: we determined the minimum NLL for each chain and we normalized the chain by this value. However, the value by which we normalized had to be reintroduced in order to obtain correct results and this was done in the final phase of the *a posteriori* model probability computation.

Remark: The values of the NLL for each parameter sample were already computed during the sampling stage.

Algorithm 1 Model Choice Algorithm

```

% Initialize the a priori model probabilities
Pr_M(k) = 1/K, k = 1...K;
% For each model k generate T samples from pi(theta_k|X, M = k)
for k = 1 -> K do
    [Theta_k, NLL_k] = MetropolisHastings(T);
    mNLL(k) = min(NLL_k);
    normNLL_k = NLL_k - mNLL(k);
    Evid(k) = HarmonicMean(normNLL_k);
end for
for k = 1 -> K do
    % Compute the a posteriori model probabilities

```

$$Pr_{\mathcal{M}|\mathbf{X}}(k) = \frac{Pr_{\mathcal{M}}(k) \cdot Evid(k)}{\sum_{l=1}^K Pr_{\mathcal{M}}(l) \cdot Evid(l) \cdot e^{mNLL(k) - mNLL(l)}} \quad (12)$$

```
end for
```

Table 2: The probability (in %) of the method selecting a certain model, on the vertical, the true model used to generate the data and on the horizontal the model chosen by the method

True model	Estimated model			
	GG	St	CS	Ro
GG	96.3	1.8	1.6	0.3
St	3.3	92.5	0.7	3.5
CS	5.8	1.9	90.1	2.2
Ro	1.8	6.1	0.8	91.3

Consequently, the exponentials intervene only in the final stage, and they are factored, as shown by expression (12). This is done in order to replace computing the exponentials of the NLL chains minima with the exponentials of these differences, since the exponentials of the minima will be seen by Matlab as zero, yielding indeterminateness, while the exponentials of the differences will result in finite non trivial values. Moreover, even if the values of the differences are not finite, non trivial values, we will still have eliminated the indeterminateness: for instance if for a model l we have $\exp[mNLL(k) - mNLL(l)] = \infty$, then $Pr_{\mathcal{M}|\mathbf{X}}(l) = 1$. Conversely, if for model k , $\exp[mNLL(k) - mNLL(l)] = 0$, all the models $l \neq k$ will have $Pr_{\mathcal{M}|\mathbf{X}}(l) = 0$ and $Pr_{\mathcal{M}|\mathbf{X}}(k) = 1$. This solution eliminates the numerical problems without introducing any approximation, only by conveniently handling the quantities.

5. EXPERIMENTAL RESULTS

This section is devoted to the performances in terms of speed and accuracy of our model selection method. The true model and the values of its parameters used to generate the data are known since we worked on synthetic images. Consequently, we can directly evaluate the accuracy of the model probabilities and parameter estimates.

The concerns regarding the behavior of the harmonic mean estimator, more specifically the fact that infinite variance situations may be encountered, have lead us to monitor the empirical variance all throughout the simulations and we report that with no exception, the variance is not only finite, but also tends to zero.

The implementation was done in Matlab, with $T = 1000$, resulting in a total run time of about 4 minutes on a 3GHz processor.

The algorithm performs very well, managing to correctly yield a strong probability for the model that generated the texture and very weak probabilities for the other models. An interesting fact, highlighting the strength of the method, is that it successfully distinguishes the models even if the observed textures are visually similar.

The experimental setting is the following: we chose 20 sets of parameter values, θ , and using these sets we generated the 80 corresponding textures, by using every set θ for each model. We ran the algorithm 10 times for each texture and we counted the number of times the algorithm associates the highest probability to the true model. The results are synthesized in Table 2, where we can see that the method correctly identifies the true model in the large majority of cases, as shown by the strong probabilities on the diagonal. There are a series of misclassification situations, quantified by the other percentages, which occur mainly in situations where the shapes of the PSDs for those specific values of the parameters are very similar.

The parameter estimation for each model is very accurate, yielding values situated within the $\pm 3\%$ relative error interval with respect to the true parameter values, as also discussed in [8].

In Fig.2 we have shown an example of an original texture and its reconstruction using our model selection and parameter estimation method. We can observe their strong similarity as proof of the

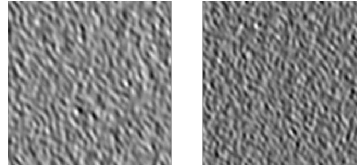


Fig. 2: Observed image and reconstructed texture

successful model identification and accurate parameters estimation.

6. CONCLUSION AND PERSPECTIVES

The paper presents the mathematical formulation and the performances of a model selection with joint parameter estimation method for texture images modeled as GRFs. This approach assembles state of the art algorithms, such as M-H, within-model simulation, harmonic means, in a fully Bayesian framework. The joint probability law synthesizes the information contained by the data and all prior information available on the models and the parameters.

This method uses a within-model sampling technique, the very nature of the tackled problem having lead us towards choosing this type of sampling rather than RJMCMC. The relatively small number of models and the fact that, based on our work in [8], we were able to efficiently sample the model parameters reinforced our choice.

We foresee a number of future developments for this algorithm, among which we consider its integration in a more complex application for textured images segmentation with model selection and parameter estimation for each region. We also intend to widen the set of texture models beyond the class of GRFs, perspectives in this direction being non-Gaussian models such as: fractal models, Wold models, Markov Random Fields.

7. REFERENCES

- [1] P.J. Green, "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, vol. 82, pp. 711–732, 1995.
- [2] P.J. Green, "Trans-dimensional Markov chain Monte Carlo," 2003.
- [3] P.J. Green and D.I. Hastie, "Reversible Jump MCMC," 2009.
- [4] T. Elguebaly and N. Bouguila, "A Bayesian Approach for Texture Images Classification and Retrieval," in *Proceedings of the ICMCS*, 2011, pp. 1–6.
- [5] A.E. Gelfand and D.K. Dey, "Bayesian Model Choice: Asymptotics and Exact Calculations," *J. Royal Stat. Society B*, vol. 56, no. 3, pp. 501–514, 1994.
- [6] M.A. Newton and A.E. Raftery, "Approximate Bayesian Inference with the Weighted Likelihood Bootstrap," *J. Royal Stat. Society B*, vol. 56, no. 1, pp. 3–48, 1994.
- [7] C.P. Robert and D. Wraith, "Computational methods for Bayesian model choice," 2009.
- [8] C. Vacar, J.-F. Giovannelli, and Y. Berthoumieu, "Langevin and Hessian with Fisher approximation stochastic sampling for parameter estimation of structured covariance," in *Proceedings of the IEEE ICASSP, Prague*, 2011, pp. 3964–3967.
- [9] H. Chipman, E.I. George, and R.E. McCulloch, "The practical implementation of Bayesian model selection," in *Institute of Mathematical Statistics*, 2001, pp. 65–134.
- [10] R. Neal, "J. Royal Stat. Society B", vol. 56, pp. 41–42, 1994.
- [11] T.L. Griffiths and M. Steyvers, "Finding scientific topics," in *Proceedings of the Nat. Ac. of Sc. USA*, 2004, pp. 5228–5235.
- [12] H.M. Wallach, "Topic modeling: beyond bag-of-words," in *Proceedings of the ICML*, 2006, pp. 977–984.