Model selection and comparison for independents sinusoids

Nielsen, Jesper Kjær; Christensen, Mads Græsbøll; Jensen, Søren Holdt

Published in:
IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014

DOI (link to publication from Publisher):
10.1109/ICASSP.2014.6853927

Publication date:
2014

Document Version
Early version, also known as pre-print

MODEL SELECTION AND COMPARISON FOR INDEPENDENTS SINUSOIDS

Jesper Kjær Nielsen[†], *Mads Græsbøll Christensen*[‡], *Søren Holdt Jensen*[†]

[†]Aalborg University
Signal and Inf. Process., Dept. of Electronic Systems
{jkn,shj}@es.aau.dk

[‡]Aalborg University
Audio Analysis Lab, Dept. of Arch., Design & Media Tech.
mgc@create.aau.dk

ABSTRACT

In the signal processing literature, many methods have been proposed for estimating the number of sinusoidal basis functions from a noisy data set. The most popular method is the asymptotic MAP criterion, which is sometimes also referred to as the BIC. In this paper, we extend and improve this method by considering the problem in a full Bayesian framework instead of the approximate formulation, on which the asymptotic MAP criterion is based. This leads to a new model selection and comparison method, the lp-BIC, whose computational complexity is of the same order as the asymptotic MAP criterion. Through simulations, we demonstrate that the lp-BIC outperforms the asymptotic MAP criterion and other state of the art methods in terms of model selection, de-noising and prediction performance. The simulation code is available online.

***Index Terms*—** Model comparison and selection, Bayesian information criterion, sinusoidal models, spectral estimation.

1. INTRODUCTION

Many physical and synthetic signals can accurately be modelled as a weighted sum of sinusoidal basis functions. Such signals are encountered in a large number of areas such as speech and music processing, electrocardiography, seismology, radar and sonar processing, astronomy, meteorology, and economics. Often, the number of sinusoidal basis functions is assumed known a priori, but in most practical cases this is not true since this number is typically unknown or time-varying. As the statistical performance of most estimators of the sinusoidal parameters depends on the assumed number of basis functions, detecting this number from the available data is an important, yet often overlooked, problem in the scientific literature [1].

We consider a noisy data set $\boldsymbol{x} = \begin{bmatrix} x(0) & \cdots & x(N-1) \end{bmatrix}^T$ consisting of N real- or complex-valued data points, and we assume that these data points originate from an unknown sinusoidal process. Since we are unsure about the true model, we select a set of K candidate parametric models $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_K$ which we wish to compare in the light of the data $\boldsymbol{x}$. We assume that the candidate models are weighted sums of l_k independent sinusoidal basis functions in additive noise. Therefore, the k'th model can be written as

$$\mathcal{M}_k: \quad x(n) = s_k(n) + e(n)\,, \quad n = 0, 1, \ldots, N-1 \tag{1}$$

where $s_k(n)$ and $e(n)$ form a Wold decomposition of the real- or complex-valued data $x(n)$ into a predictable part and a non-predictable part, respectively. At the sampling index n, the predictable part is given by

$$s_k(n) = \begin{cases} \sum_{i=1}^{l_k} \alpha_i \exp(j\omega_i n), & x(n) \in \mathbb{C} \\ \sum_{i=1}^{l_k} a_i \cos(\omega_i n) - b_i \sin(\omega_i n), & x(n) \in \mathbb{R} \end{cases} \tag{2}$$

where $\mathbb{C}$ and $\mathbb{R}$ denote the set of complex- and real-valued numbers, respectively, and $j = \sqrt{-1}$ is the imaginary unit[1]. The complex amplitude α_i, the linear weights a_i and b_i, the amplitude A_i, and the phase ϕ_i are related by $\alpha_i = a_i + jb_i = A_i \exp(j\phi_i)$. The number l_k of sinusoidal basis functions is simply called the model order, and many methods for detecting the sinusoidal model order has been proposed and used in the signal processing literature. These methods can roughly be dichotomised into information criteria and subspace methods. In the information criteria, a penalty term is added to a log-likelihood term to penalise model complexity and thereby prevent fitting the noise in the data. The penalty term can be derived in a number of ways, leading to a large collection of criteria such as the Akaike's information criterion (AIC) [3], the generalised AIC (GAIC) [4], the minimum description length (MDL) [5, 6], the Bayesian information criterion (BIC) [7], the asymptotic MAP criterion [8, 9], and recently the EVT in [10]. The subspace methods are based on an eigenvalue decomposition of the data covariance matrix [11], and the model order can be detected by analysing the eigenvalues [12], the eigenvectors [13], and recently the angles between subspaces [1]. From a computational viewpoint, the subspace methods require typically less resources than the information criteria since the computation of the maximum likelihood estimate of the sinusoidal frequency is a costly operation. On the other hand, the information criteria typically outperform the subspace methods in terms of detection, de-noising and prediction performance.

The asymptotic MAP criterion [8, 9], which is sometimes also referred to as the Bayesian information criterion [11, 14], is arguably the most popular model order selection method, and it can be derived in many ways based on MDL techniques [15], Kullback-Leibler information [11], and the Bayesian approach [8, 9]. As in the latter approach, we here also consider model order selection for the sinusoidal model in (1) in a Bayesian framework. In contrast to the asymptotic MAP criterion proposed in [8, 9], however, we formulate the problem in the full Bayesian framework recently proposed in [16]. By using a few simple assumptions, we derive a new model order selection and comparison method, the lp-BIC, which has a computational complexity similar to the information criteria and contains only one user-defined parameter. Moreover, the asymptotic MAP criterion can also be derived in our framework as a special case. Through simulations, we show that the lp-BIC in general outperforms all other methods for both white and mildly coloured Gaussian noise in terms of detection, de-noising, and prediction performance.

[1]Note that the real-valued signal model can be cast into the form of the complex-valued signal model by computing its discrete-time "analytic" signal, which can also be down-sampled by a factor of two [2].

2. SINUSOIDAL MODEL COMPARISON

In this section, we derive the the lp-BIC which can be used for both model order selection and comparison. In contrast to model selection, where only the most likely model is found, a probability for each candidate model is computed in model comparison. Consequently, all of the models can be used for model averaged parameter estimation, de-noising, and prediction. For both model selection and comparison, the elicitation of proper prior distributions on the model parameters is very important as improper prior distributions such as a flat prior on the linear parameters cause the simplest model to be preferred, regardless of the information in the data [16,17]. We therefore give a few arguments for the prior model in Sec. 2.1, but a much more thorough derivation can be found in [16].

2.1. Bayesian Model

2.1.1. The Observation Model

We assume a (complex-valued) normal distribution with probability density function (pdf)

$$p(\boldsymbol{e}|\sigma^2) = \frac{\exp\left(-\frac{\boldsymbol{e}^H\boldsymbol{e}}{r\sigma^2}\right)}{[r\pi\sigma^2]^{N/r}} = \begin{cases} \mathcal{CN}(\boldsymbol{e};\boldsymbol{0},\sigma^2\boldsymbol{I}_N)\,, & r=1 \\ \mathcal{N}(\boldsymbol{e};\boldsymbol{0},\sigma^2\boldsymbol{I}_N)\,, & r=2 \end{cases} \tag{3}$$

for the non-predictable part $\boldsymbol{e}$ which is defined analogously to $\boldsymbol{x}$.The notation $(\cdot)^H$ denotes conjugate transposition, and $\boldsymbol{I}_N$ is the $N\times N$ identity matrix. To simplify the notation, we use the non-standard notation $\mathcal{N}_r(\cdot)$ to refer to either the complex-valued normal distribution with pdf $\mathcal{CN}(\cdot)$ for $r=1$ or the real-valued normal distribution with pdf $\mathcal{N}(\cdot)$ for $r=2$. Besides being mathematically tractable, arguments such as maximisation of the entropy [18,19] and the Cramér-Rao bound [20] favour the white Gaussian noise (WGN) assumption on $\boldsymbol{e}$ [21]. If the noise is known to be coloured, the methods in this paper are still useful if combined with a linear pre-filter. The WGN assumption implies that the observation model is

$$p(\boldsymbol{x}|\boldsymbol{\alpha}_k,\boldsymbol{\omega}_k,\sigma^2,\mathcal{M}_k) = \mathcal{N}_r(\boldsymbol{x};\boldsymbol{Z}_k\boldsymbol{\alpha}_k,\sigma^2\boldsymbol{I}_N) \tag{4}$$

where we have defined

$$\boldsymbol{\omega}_k \triangleq \begin{bmatrix}\omega_1 & \cdots & \omega_{l_k}\end{bmatrix}^T \tag{5}$$

$$\boldsymbol{\alpha}_k \triangleq \begin{cases} \begin{bmatrix}\alpha_1 & \cdots & \alpha_{l_k}\end{bmatrix}^T, & \boldsymbol{x}\in\mathbb{C}^N \\ \begin{bmatrix}a_1 & \cdots & a_{l_k} & -b_1 & \cdots & -b_{l_k}\end{bmatrix}^T, & \boldsymbol{x}\in\mathbb{R}^N \end{cases} \tag{6}$$

$$\boldsymbol{z}_i \triangleq \begin{bmatrix}1 & \exp(j\omega_i) & \cdots & \exp(j\omega_i(N-1))\end{bmatrix}^T \tag{7}$$

$$\boldsymbol{Z}_k \triangleq \begin{cases} \begin{bmatrix}\boldsymbol{z}_1 & \cdots & \boldsymbol{z}_{l_k}\end{bmatrix}, & \boldsymbol{x}\in\mathbb{C}^N \\ \begin{bmatrix}\mathrm{Re}(\boldsymbol{z}_1) & \cdots & \mathrm{Re}(\boldsymbol{z}_{l_k}) & \mathrm{Im}(\boldsymbol{z}_1) & \cdots & \mathrm{Im}(\boldsymbol{z}_{l_k})\end{bmatrix}, & \boldsymbol{x}\in\mathbb{R}^N \end{cases}$$

The notation $(\cdot)^T$ denotes transposition, and $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ take the real and imaginary part, respectively, of a complex-valued number.

2.1.2. The Prior

As the dimension of the vector $\boldsymbol{\alpha}_k$ of linear parameters varies between models, a proper prior distribution must be assigned on it [17]. For regression models, the Zellner's g-prior given by [22]

$$p(\boldsymbol{\alpha}_k|\boldsymbol{\omega}_k,\sigma^2,g,\mathcal{M}_k) = \mathcal{N}_r(\boldsymbol{\alpha}_k;\boldsymbol{0},g\sigma^2[\boldsymbol{Z}_k^H\boldsymbol{Z}_k]^{-1}) \tag{8}$$

has been widely adopted since it leads to analytically tractable marginal likelihoods and is easy to understand and interpret [23,24]. For mathematical convenience, the hyper-g prior given by [23]

$$p(g) = (\delta-1)(1+g)^{-\delta}\,, \quad \delta>1 \tag{9}$$

is assigned to g. The hyperparameter δ should be selected in the interval $1<\delta\leq 2$ [23]. Besides having some desirable analytical properties, $p(g)$ reduces to the Jeffreys' prior and the reference prior when $\delta=1$ [25]. However, since this prior is improper, it can only be used when the prior probability of the all-noise model is zero. When this is the case, the lp-BIC has no user-defined parameters. The noise variance σ^2 is a common parameter and has the same meaning in all models and can therefore be given an improper prior [16, 17]. We therefore use Jeffreys' prior $p(\sigma^2)=(\sigma^2)^{-1}$ which is scale invariant. That is, it includes the same prior knowledge whether we parametrise the model in terms of the noise variance σ^2, the standard deviation σ, or the precision parameter $\lambda=\sigma^{-2}$. For the frequencies, we assume the uniform prior

$$p(\boldsymbol{\omega}_k|\mathcal{M}_k) = \frac{1}{W_k}\prod_{i=1}^{l_k}\mathbb{I}_{\Omega_i}(\omega_i) \tag{10}$$

where $\mathbb{I}_{\Omega_i}(\omega_i)$ is the indicator function on the interval $\Omega_i\subseteq[0,2\pi)$, and W_k is the normalisation constant. The overall frequency parameter space is therefore $\boldsymbol{\Omega}_k=\Omega_1\times\Omega_2\times\cdots\times\Omega_{l_k}$. Finally, the prior on the models is also a uniform prior of the form $p(\mathcal{M}_k)=K^{-1}\mathbb{I}_{\mathcal{K}}(k)$ where $\mathcal{K}=\{1,\ldots,K\}$. However, another prior can easily be used in our framework (see (11) below).

2.2. Bayesian Model Comparison

From Bayes' theorem, the posterior distribution on the models has the probability mass function (pmf)

$$p(\mathcal{M}_k|\boldsymbol{x}) = \frac{\mathrm{BF}[\mathcal{M}_k;\mathcal{M}_b]p(\mathcal{M}_k)}{\sum_{i=1}^K \mathrm{BF}[\mathcal{M}_i;\mathcal{M}_b]p(\mathcal{M}_i)} \tag{11}$$

where $\mathcal{M}_b$ is some base model, all other models are compared against, and the Bayes' factor is given by

$$\mathrm{BF}[\mathcal{M}_j;\mathcal{M}_i] = \frac{p(\boldsymbol{x}|\mathcal{M}_j)}{p(\boldsymbol{x}|\mathcal{M}_i)} \triangleq \frac{m_j(\boldsymbol{x})}{m_i(\boldsymbol{x})}\,. \tag{12}$$

The function $m_k(\boldsymbol{x})$ is an unnormalised marginal likelihood whose normalisation constant must be the same for all models. Working with $m_k(\boldsymbol{x})$ rather than the normalised marginal likelihood $p(\boldsymbol{x}|\mathcal{M}_k)$ is simpler. Moreover, $p(\boldsymbol{x}|\mathcal{M}_k)$ does not even exist if an improper prior such as the Jeffreys' prior on the noise variance is used. Given g and $\boldsymbol{\omega}_k$, the marginal likelihood is given by

$$\begin{aligned} p(\boldsymbol{x}|\boldsymbol{\omega}_k,g,\mathcal{M}_k) = \int_0^\infty\int_{\mathcal{A}_k} & p(\boldsymbol{x}|\boldsymbol{\alpha}_k,\boldsymbol{\omega}_k,\sigma^2,\mathcal{M}_k) \\ & \times p(\boldsymbol{\alpha}_k|\boldsymbol{\omega}_k,\sigma^2,g,\mathcal{M}_k)p(\sigma^2)d\boldsymbol{\alpha}_k d\sigma^2 \end{aligned} \tag{13}$$

where $\mathcal{A}_k$ is either the k-dimensional set of real- or complex-valued numbers. By performing the integration in (13), it can be shown that

$$p(\boldsymbol{x}|\boldsymbol{\omega}_k,g,\mathcal{M}_k) \propto m_k(\boldsymbol{x}|\boldsymbol{\omega}_k,g) = \frac{m_N(\boldsymbol{x})}{(1+g)^{l_k}}\left(\frac{\hat{\sigma}_N^2}{\hat{\sigma}_k^2(\boldsymbol{\omega}_k,g)}\right)^{N/r}$$

where we have defined

$$\hat{\sigma}_k^2(\boldsymbol{\omega}_k,g) \triangleq \frac{\boldsymbol{x}^H(\boldsymbol{I}_N-\frac{g}{1+g}\boldsymbol{P}_k)\boldsymbol{x}}{N} = \hat{\sigma}_N^2\left(1-\frac{g}{1+g}R_k^2\right)$$

$$R_k^2 \triangleq \frac{\boldsymbol{x}^H\boldsymbol{P}_k\boldsymbol{x}}{\boldsymbol{x}^H\boldsymbol{x}} \tag{14}$$

$$m_N(\boldsymbol{x}) \triangleq \Gamma(N/r)(N\pi\hat{\sigma}_N^2)^{-N/r}\,. \tag{15}$$

The matrix $\boldsymbol{P}_k$ is the orthogonal projection matrix for the column space of $\boldsymbol{Z}_k$, and $\hat{\sigma}_k^2(\boldsymbol{\omega}_k,g)$ is asymptotically equal to the ML estimate of the noise variance in the limit $\hat{\sigma}_{\mathrm{ML}}^2=\lim_{g\to\infty}\hat{\sigma}_k^2(\hat{\boldsymbol{\omega}}_k,g)$

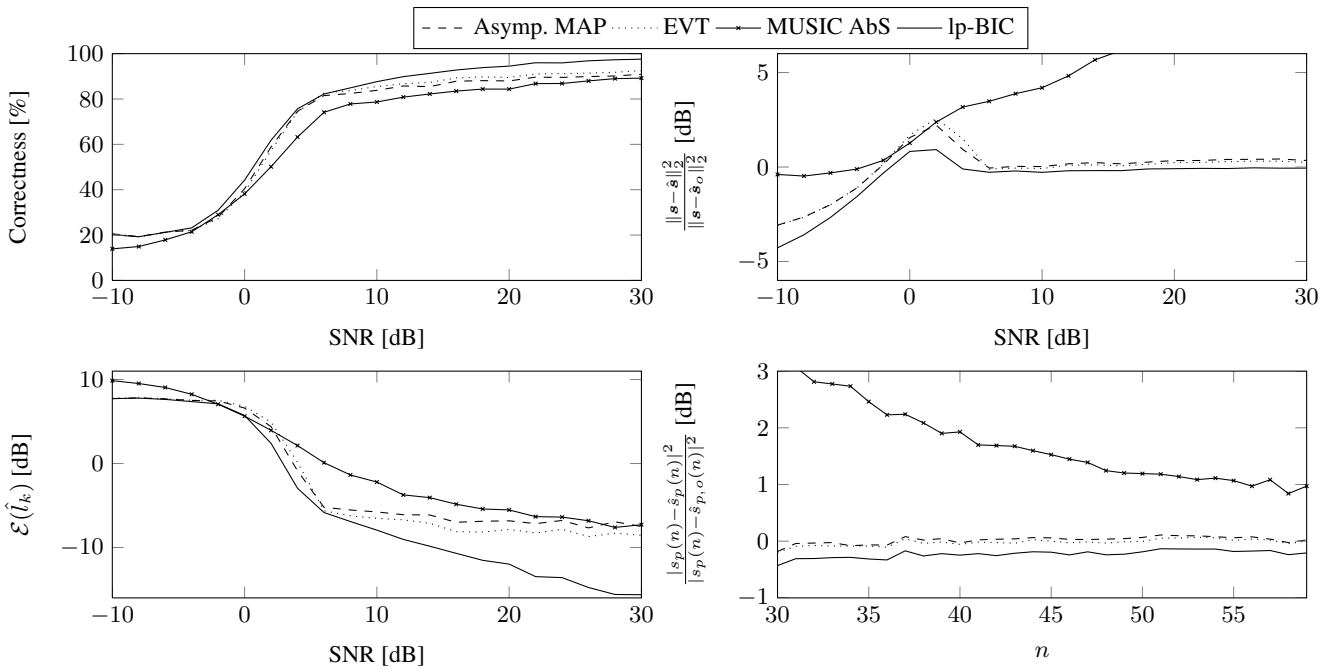

Fig. 1. Model detection and prediction performance of various model selection methods for white Gaussian noise.

with $\hat{\boldsymbol{\omega}}_k$ being the ML estimate of the frequency parameters. The estimate $\hat{\sigma}_N^2$ is the estimated noise variance of the null model $\mathcal{M}_N$ which is the all-noise model ($l_k = 0$) and has the unnormalised marginal likelihood $m_N(\boldsymbol{x})$. If we select the null model to be the base model, the Bayes' factor given $\boldsymbol{\omega}_k$ and g is thus

$$\mathrm{BF}[\mathcal{M}_k;\mathcal{M}_N|\boldsymbol{\omega}_k,g] = \frac{\left[\hat{\sigma}_N^2/\hat{\sigma}_k^2(\boldsymbol{\omega}_k,g)\right]^{N/r}}{(1+g)^{l_k}} \,. \tag{16}$$

To find the Bayes' factor in (12) for $\mathcal{M}_i = \mathcal{M}_N$, we have to multiply (16) with the priors on $\boldsymbol{\omega}_k$ and g and integrate this product over these parameters. Unfortunately, this cannot be done analytically, and we, therefore, evaluate the integral using the Laplace approximation. As in [26], the integral can also be evaluated using Markov chain Monte Carlo methods, but this is much more computationally intensive and typically requires the specification of many user-defined parameters.

2.2.1. The Laplace-BIC (lp-BIC)

Since the marginal posterior pdf over g is not symmetric and in order to avoid edge effect near $g = 0$, the re-parametrisation $\tau = \ln g$ with the Jacobian $dg/d\tau = \exp(\tau)$ is made [23]. This re-parametrisation suggests that the posterior distribution over g is approximately a log-normal distribution. Thus, if we define the integrand

$$q(\boldsymbol{\omega}_k,\tau) \triangleq \mathrm{BF}[\mathcal{M}_k;\mathcal{M}_N|\boldsymbol{\omega}_k,\exp(\tau)]p(\boldsymbol{\omega_k}|\mathcal{M}_k)p(\tau)\,, \tag{17}$$

the Laplace approximation to the Bayes' factor in (12) is [16]

$$\mathrm{BF}[\mathcal{M}_k;\mathcal{M}_N] = \int_{\boldsymbol{\Omega}_k}\int_{-\infty}^{\infty} q(\boldsymbol{\omega}_k,\tau)d\tau d\boldsymbol{\omega}_k \tag{18}$$

$$\approx \mathrm{BF}[\mathcal{M}_k;\mathcal{M}_N|\hat{\boldsymbol{\omega}}_k^{\mathrm{MAP}},\hat{g}]\frac{\hat{g}(\delta-1)(2\pi)^{(l_k+1)/2}}{(1+\hat{g})^{\delta}W_k} \times\sqrt{\gamma(\hat{g}|\hat{\boldsymbol{\omega}}_k^{\mathrm{MAP}})}\left|-\boldsymbol{H}(\hat{\boldsymbol{\omega}}_k^{\mathrm{MAP}}|\hat{g})\right|^{-1/2}\Gamma(l_k+1) \tag{19}$$

where we have defined

$$\hat{g} = -(\beta_\tau + \sqrt{\beta_\tau^2 - 4\alpha_\tau})/(2\alpha_\tau) \tag{20}$$

$$\alpha_\tau = (1-R_k^2)(1-l_k-\delta) \tag{21}$$

$$\beta_\tau = (N/r-1)R_k^2 + (2-l_k-\delta) \tag{22}$$

$$\hat{\boldsymbol{\omega}}_k^{\mathrm{MAP}} = \underset{\boldsymbol{\omega}_k\in\boldsymbol{\Omega}_k}{\arg\max}\,\boldsymbol{x}^H\boldsymbol{P}_k\boldsymbol{x} \tag{23}$$

$$\gamma(\hat{g}|\hat{\boldsymbol{\omega}}_k^{\mathrm{MAP}}) = \frac{r}{\hat{g}}\left[\frac{N(1-R_k^2)}{[1+\hat{g}(1-R_k^2)]^2} - \frac{N-rl_k-r\delta}{(1+\hat{g})^2}\right]^{-1} \tag{24}$$

$$\boldsymbol{H}(\boldsymbol{\omega}_k|g) = \frac{\partial^2\ln q(\boldsymbol{\omega}_k,\tau)}{\partial\boldsymbol{\omega}_k\partial\boldsymbol{\omega}_k^T}\,. \tag{25}$$

The factor $\Gamma(l_k+1)$ in (19) accounts for that the integrand $q(\boldsymbol{\omega}_k,\tau)$ is invariant to permutations of the frequency parameters. Therefore, the Laplace approximation is made to the $l_k!$ identical peaks of $q(\boldsymbol{\omega}_k,\tau)$. The MAP estimate $\hat{\boldsymbol{\omega}}_k^{\mathrm{MAP}}$ is the same as the ML estimate of the frequency parameters from (4). Finding an analytical expression for the Hessian $\boldsymbol{H}(\boldsymbol{\omega}_k|g)$ is a tedious task, and we refer the interested reader to [16] for a derivation. When the signal-to-noise ratio (SNR) is high enough, enough data is available, and the frequencies are well-separated (relative to N), it can be approximated by

$$\boldsymbol{H}(\hat{\boldsymbol{\omega}}_k^{\mathrm{MAP}}|\hat{g}) \approx -\frac{N^3\hat{g}}{6r^2\hat{\sigma}_k^2(\hat{\boldsymbol{\omega}}_k^{\mathrm{MAP}},g)(1+\hat{g})}\,\mathrm{diag}(|\hat{\boldsymbol{\alpha}}_k|^2) \tag{26}$$

where $\hat{\boldsymbol{\alpha}}_k$ is the ML estimate of the linear parameters. Inserting the Bayes' factor in (19) into (11) allows us to select the most likely model and to compute the posterior probabilities of all models. Moreover, note the following.

1. For $\boldsymbol{g} \to \infty$, it is interesting to note that $-\boldsymbol{H}^{-1}(\hat{\boldsymbol{\omega}}_k^{\mathrm{MAP}}|\hat{g})$ is approximately identical to the asymptotic Cramér-Rao bound of the frequency parameters with the true values of the noise

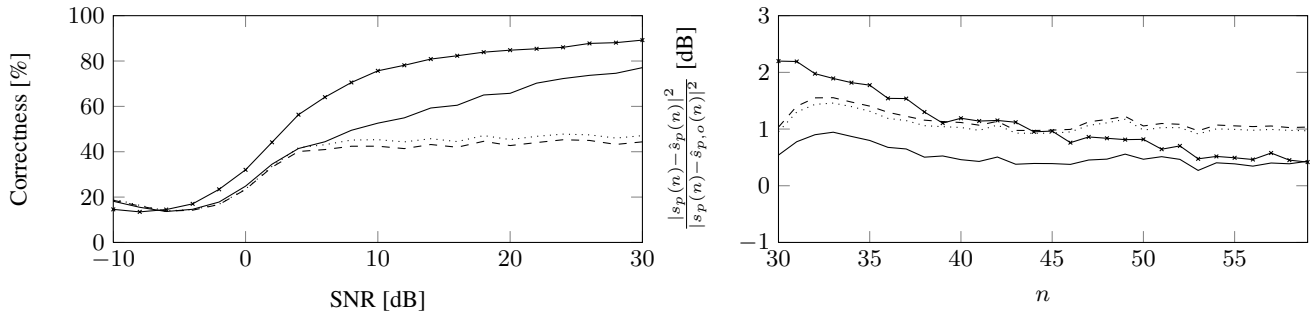

Fig. 2. Model detection and prediction performance for mildly coloured noise (see Fig. 1 for legends).

variance and amplitudes replaced by their ML estimates. In our experience, the approximation in (26) works very well in practice and not much is gained by using the exact expression for the Hessian.

2. The Bayes' factor in (19) has only the single user parameter δ. As discussed in [23], it should be selected in the interval $1 < \delta \leq 2$, and we usually select it as $3/2$.
3. If we select $g = N$, assume that $N \gg rl_k$, only make the Laplace approximation to a single mode of $\boldsymbol{\omega}_k$, and neglect all first-order terms, the resulting model selection rule is basically the same as the asymptotic MAP method.

3. SIMULATIONS

We here present some typical simulation results for the subset of the model selection methods, we believe have the best performance, but we also encourage the interested reader to try other configurations and methods by modifying the simulation code which is available at `http://kom.aau.dk/~jkn/publications/publications.php`. Specifically, we compare the performance of the popular asymptotic MAP criterion [8, 9] (see also [11, 15]), the recently proposed EVT [10], the angle between subpaces (MUSIC AbS) method [1], and the lp-BIC method proposed in this paper. To keep the plots simple, we do not show the performance of the AIC [3], the naive MDL [5], the GAIC [4], and the ESTER method [13] since we have found that their performance is generally worse than that of the other methods.

3.1. Monte Carlo Simulation for White Gaussian Noise

In the first simulation, the model detection performance, the de-noising performance, and the prediction performance were evaluated for white Gaussian noise (WGN) using a Monte Carlo simulation consisting of 5000 runs at SNRs from -10 dB to 30 dB in steps of 2 dB. In each run, $N = 30$ data points were generated from a complex-valued sinusoidal model. As recommend in [27], most parameters were generated at random in each run. The model order was generated uniformly from the set $\{1, 2, \ldots, 5\}$, the amplitudes were set to ones, and the phases and frequencies were generated uniformly from the interval $[0; 2\pi)$. For all of the model selection methods, the set of candidate model orders were $l_k \in \{1, \ldots, 8\}$. The all-noise model ($l_k = 0$) was not included since the MUSIC AbS method is unable to handle such a model. The ML estimates of the frequency parameters were found using the RELAX algorithm [4]. The subvector length of the MUSIC AbS method were set to the typical value of $N/2$, and the hyperparameter δ of the lp-BIC was set to 3/2.

Fig. 1 shows the results. In the upper left plot, the percentage of correctly detected model orders is shown. Clearly, the lp-BIC outperformed the other methods with the EVT being slightly better than the asymptotic MAP method and much better than the MUSIC AbS method. The same trend is also seen in the lower left plot where the mean-squared error (MSE) of the detected model order given by $\mathcal{E}(\hat{l}_k) = (l_k - \hat{l}_k)^2$ is shown. In the upper right plot, the de-noising performance is shown. To facilitate a better comparison, we have shown the results relative to the performance of an Oracle who knew the true model order. Thus, if a curve is below the 0 dB line, the corresponding method performed better than the Oracle. The lp-BIC method performed slightly better than the EVT and the asymptotic MAP method, which performed equally well, and much better than the MUSIC AbS method. The same trend is also found in the lower right plot where the prediction performance relative to the Oracle's performance is shown at an SNR of 6 dB. Although not presented here, we observed similar results for white Laplacian noise.

3.2. Monte Carlo Simulations for Mildly Coloured Noise

Using the same simulation setup as above, we also ran a Monte Carlo simulation for mildly coloured noise. Specifically, each noise vector was a realisation from a first-order and complex-valued autoregressive process whose parameter was drawn uniformly from the interior of circular region with a radius of 0.95. In Fig. 2, the model detection and prediction performance are shown. The performance of lp-BIC and in particular the MUSIC AbS method were less affected by the violated WGN assumption than the asymptotic MAP criterion and the EVT. For model detection, the MUSIC AbS method performed better than the lp-BIC whereas the lp-BIC outperformed the MUSIC AbS method in terms of prediction performance.

4. CONCLUSION

We have here presented a new model selection and comparison method for independent sinusoidal models. The method, the lp-BIC, was derived in a full Bayesian framework in which the popular asymptotic MAP criterion can also be derived using a number of asymptotic assumptions. Although the lp-BIC is not as simple as the asymptotic MAP criterion, its computational complexity is of the same order. Through simulations, we demonstrated that the lp-BIC outperformed the asymptotic MAP criterion and other state of the art model selection methods in terms of model order detection, de-noising, and prediction performance. The lp-BIC was also demonstrated to be much more robust to coloured noise than other information criteria such as the asymptotic MAP criterion.

5. REFERENCES

[1] M. G. Christensen, A. Jakobsson, and S. H. Jensen, "Sinusoidal order estimation using angles between subspaces," *EURASIP J. on Advances in Signal Process.*, vol. 2009, pp. 1–11, Nov. 2009.

[2] S. L. Marple, Jr., "Computing the discrete-time "analytic" signal via FFT," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2600–2603, Sept. 1999.

[3] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, Dec. 1974.

[4] J. Li and P. Stoica, "Efficient mixed-spectrum estimation with applications to target feature extraction," *IEEE Trans. Signal Process.*, vol. 44, no. 2, pp. 281–295, Feb. 1996.

[5] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, Sep. 1978.

[6] J. Rissanen, "Estimation of structure by minimum description length," *Circuits, Systems, and Signal Process.*, vol. 1, no. 3, pp. 395–406, 1982.

[7] G. Schwarz, "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, Mar. 1978.

[8] P. M. Djuric, "A model selection rule for sinusoids in white Gaussian noise," *IEEE Trans. Signal Process.*, vol. 44, no. 7, pp. 1744–1751, Jul. 1996.

[9] P. M. Djuric, "Asymptotic MAP criteria for model selection," *IEEE Trans. Signal Process.*, vol. 46, no. 10, pp. 2726–2735, Oct. 1998.

[10] B. Nadler and A. Kontorovich, "Model selection for sinusoids in noise: Statistical analysis and a new penalty term," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1333–1345, Apr. 2011.

[11] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*, Englewood Cliffs, NJ, USA: Prentice Hall, May 2005.

[12] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 387–392, Apr. 1985.

[13] R. Badeau, B. David, and G. Richard, "A new perturbation analysis for signal enumeration in rotational invariance techniques," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 450–458, Feb. 2006.

[14] P. Stoica and Y. Selén, "Model-order selection: a review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, July 2004.

[15] L. Kavalieris and E. J. Hannan, "Determining the number of terms in a trigonometric regression," *J. of Time Series Analysis*, vol. 15, no. 6, pp. 613–625, Nov. 1994.

[16] J. K. Nielsen, M. G. Christensen, A. T. Cemgil, and S. H. Jensen, "Bayesian model comparison with the g-prior," *IEEE Trans. Signal Process.*, vol. 62, no. 1, pp. 225–238, 2014.

[17] J. O. Berger and L. R. Pericchi, "Objective Bayesian methods for model selection: Introduction and comparison," *Institute of Mathematical Statistics Lecture Notes – Monograph Series*, vol. 38, pp. 135–207, 2001.

[18] G. L. Bretthorst, *Bayesian Spectrum Analysis and Parameter Estimation*, Springer-Verlag, Berlin Heidelberg, 1988.

[19] E. T. Jaynes, "Bayesian spectrum and chirp analysis," in *Maximum Entropy and Bayesian Spectral Analysis and Estimation Problems*, C. R. Smith and G. J. Erickson, Eds., pp. 1–37. AA Dordrecht, Holland: D. Reidel Publishing Company, 1987.

[20] P. Stoica and P. Babu, "The Gaussian data assumption leads to the largest Cramér-Rao bound," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 132–133, May 2011.

[21] K. Kim and G. Shevlyakov, "Why Gaussianity?," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 102–113, Mar. 2008.

[22] A. Zellner, "On assessing prior distributions and Bayesian regression analysis with g-prior distributions," in *Bayesian Inference and Decision Techniques*. New York, NY, USA: Elsevier, 1986.

[23] F. Liang, R. Paulo, G. Molina, M. A. Clyde, and J. O. Berger, "Mixtures of g priors for Bayesian variable selection," *J. Amer. Statistical Assoc.*, vol. 103, pp. 410–423, Mar. 2008.

[24] J. K. Nielsen, M. G. Christensen, and S. H. Jensen, "Bayesian model comparison and the BIC for regression models," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2013, pp. 6362–6366.

[25] R. Guo and P. L. Speckman, "Bayes factor consistency in linear models," in *The 2009 International Workshop on Objective Bayes Methodology*, Jun. 2009.

[26] C. Andrieu and A. Doucet, "Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2667–2676, 1999.

[27] P. Stoica, Y. Selén, and J. Li, "Multi-model approach to model selection," *Digital Signal Process.*, vol. 14, no. 5, pp. 399–412, Sep. 2004.