Thermal Modeling and Temperature Prediction Using Least Square Model Averaging with Model Screening

Shengyang Xu, Yingbo Hua and Sheldon X.-D. Tan
Department of Electrical Engineering, University of California, Riverside, CA, 92521

***Abstract*—We consider thermal modeling and temperature prediction for a multicore microprocessor system given multiple power consumption and temperature observation sequences of finite length. Unlike conventional bottom-up approaches, a Linear Time-Invariant (LTI) Multiple-Input-Multiple-Output (MIMO) black-box model is adopted with power consumptions as inputs and temperature observations as outputs so that no details of the physical properties of specific multicore systems are required and no extremely complicated computations are needed. Along with model parameter estimation, model order selection based on the Akaike Information Criterion (AIC) is also useful for identifying the thermal model and obtaining accurate temperature prediction. However, the uncertainty inherited in model order selection is a problem which can cause the corresponding inference to be overly optimistic or misleading. As an alternative, model averaging mitigates this uncertainty by combining a set of candidate models with different weights. In this work, we propose Least Square (LS) based model averaging algorithm with model screening. The number of candidate models is first reduced by choosing those with the smallest AICs to improve the modeling accuracy. Then, we apply smoothed AIC (S-AIC) to provide the weights associated with each selected models and establish a combined estimate of model parameters for temperature prediction. Simulation results show that the proposed algorithm yields less temperature prediction errors than the traditional LS algorithm based on model order selection.**

I. Introduction

As more high-performance cores are integrated on a chip of shrinking size in multicore microprocessor architecture, its power density is expected to increase drastically, which could lead to rapidly increasing chip temperatures and associated significant adverse effects on chip packaging and cooling costs, circuit reliability, leakage power [1] [2]. To alleviate such problems, much attention is needed for dynamic thermal management where accurate temperature prediction plays a key role. A critical component for dynamic thermal management is thermal modeling which describes the relationship between input powers and output temperatures.

Thermal modeling can be done via bottom-up or black-box approaches. The bottom-up approaches are based on the physical properties of the circuits and devices in a microprocessor [2] [3] and can be too costly for accurate thermal modeling of a very large scale system. The black-box approaches, which generate thermal models only based on the observed external signals such as powers and temperatures, have attracted much interest recently. In [4], the authors use an AutoRegressive and Moving Average (ARMA) thermal model and assume that the running workload is a stationary stochastic process. Although a hypothesis detection is provided to update the model, the unreliability of the predicted temperatures between two updated models threats the reliability and performance of the system. In [5], the authors considered a black-box MIMO LTI thermal model which treats the powers to all cores as inputs and the temperatures from all cores as outputs. The thermal characteristic of a multicore system is captured by the impulse response matrix solved by ThermPOF. For this method, only step input functions are allowed to train the model, and only one core can be excited at a time. This requirement makes ThermPOF only useful off-line. In [6], the authors considered a state-space thermal modeling method called ThermSID which does not have the limitation for ThermPOF.

In this paper, we present a simple and perhaps the most basic approach to Black-Box MIMO Thermal Modeling. We treat a multicore microprocessor system as a deterministic MIMO thermal system whose inputs are the powers consumed by the cores and whose outputs are the temperatures at certain predetermined points. The input-output relationship is simply governed by a classical deterministic MIMO ARMA model. The parameters of the deterministic MIMO ARMA model can be determined by using the classical Least Square (LS) with model order selection where its uncertainty is neglected after the final model is chosen [7]. To overcome the uncertainty inherited in model order selection, a model averaging algorithm based on least square is proposed with model screening. We show that this algorithm can provide better prediction performance compared with model order selection.

The rest of the paper is organized as follows. In Section II, we describe the details of the deterministic MIMO ARMA model. In Section III, we present the classical LS algorithm and AIC rule. In Section IV, a LS based model averaging algorithm with model screening is proposed. Section V illustrates the performance of LS and the proposed algorithm.

II. Proposed Thermal Model

Denote all powers consumed by the system at time n by the input vector $\mathbf{u}(n) \in \mathbb{R}^{M\times 1}$ and the temperatures at certain points in the system at time n by the output vector $\mathbf{y}_a(n) \in \mathbb{R}^{N\times 1}$. Here, the notation $\mathbb{R}^{l\times k}$ means the set of all $l \times k$ real matrices. The number of inputs, M, may or may not be the same as the number of outputs, N, although one can

choose $M = N$ which is the number of cores in the system. Let b be an environment temperature of the system. Then the deviation of $\mathbf{y}_a(n)$ from b is $\mathbf{y}(n) = \mathbf{y}_a(n) - b\mathbf{1}_N$ where $\mathbf{1}_N = [1, \cdots, 1]^T \in \mathbb{R}^{N\times 1}$.

Under a causality condition, we can treat $\mathbf{y}(n)$ as a function of $\mathbf{y}(n-1), \cdots, \mathbf{y}(n-p)$ and $\mathbf{u}(n), \cdots, \mathbf{u}(n-q)$ plus noise, i.e.,

$$\mathbf{y}(n) = \mathbf{f}\left(\mathbf{y}(n-1), \cdots, \mathbf{y}(n-p), \mathbf{u}(n), \cdots, \mathbf{u}(n-q)\right) + \mathbf{e}(n) \tag{1}$$

where $\mathbf{e}(n)$ is the noise or model error, p is the regression order with respect to the output, and q is the regression order with respect to the input. Since $\mathbf{y}(n)$ is the deviation of the real temperature from the environment temperature, the function $\mathbf{f}$ should be such that $\mathbf{f}(0, \cdots, 0) = 0$.

In general, the function $\mathbf{f}$ can be nonlinear and time-varying. But within a limited dynamic range and a limited time window, it is reasonable to choose $\mathbf{f}$ to be linear and time-invariant. In this case, we can replace (1) by its first order Taylor's series expansion as follows

$$\mathbf{y}(n) = \sum_{i=1}^{p} \mathbf{A}_i^T \mathbf{y}(n-i) + \sum_{i=0}^{q} \mathbf{B}_i^T \mathbf{u}(n-i) + \mathbf{e}(n) \tag{2}$$

which is the MIMO ARMA model. The coefficient matrices $\mathbf{A}_i \in \mathbb{R}^{N\times N}$ and $\mathbf{B}_i \in \mathbb{R}^{N\times M}$ may change from one multicore system to another, from one environment to another. The determination of these matrices is known as system identification.

Unlike [4] where the powers are assumed to be unknown but stationary random processes, we treat the powers as deterministic and known quantities. Our assumption is more realistic and allows a more reliable estimation of the ARMA model parameters.

A more compact form of (2) is

$$\begin{aligned} \mathbf{y}(n) &= \mathbf{A}^T \mathbf{y}_c(n) + \mathbf{B}^T \mathbf{u}_c(n) + \mathbf{e}(n) \\ &= \mathbf{W}^T \mathbf{z}(n) + \mathbf{e}(n) \end{aligned} \tag{3}$$

where $\mathbf{z}(n) = \left[\mathbf{y}_c^T(n), \mathbf{u}_c^T(n)\right]^T \in \mathbb{R}^{(pN+(q+1)M)\times 1}$, $\mathbf{y}_c(n) = \left[\mathbf{y}^T(n-1), \cdots, \mathbf{y}^T(n-p)\right]^T$, $\mathbf{u}_c(n) = \left[\mathbf{u}^T(n), \cdots, \mathbf{u}^T(n-q)\right]^T$, $\mathbf{W} = \left[\mathbf{A}^T, \mathbf{B}^T\right]^T$, $\mathbf{A} = \left[\mathbf{A}_1^T, ..., \mathbf{A}_p^T\right]^T$ and $\mathbf{B} = \left[\mathbf{B}_0^T, ..., \mathbf{B}_q^T\right]^T$.

If the system has a weak non-linearity, we can also let the function $\mathbf{f}$ to be approximated by its second-order Taylor series expansion. However, our simulation based real data shows that the thermal behavior of a multicore microprocessor is described better by the linear MIMO ARMA model (3) than by its nonlinear extension which we will not further consider.

III. Least Square (LS) Algorithm with Model Order Selection

Given the thermal model (3), an algorithm is needed to estimate the matrix $\mathbf{W}$ by using all available $\mathbf{y}(n)$ and $\mathbf{z}(n)$. Notice that $\mathbf{z}(n)$ is determined by $\mathbf{u}(n), \cdots, \mathbf{u}(n-q)$ and $\mathbf{y}(n-1), \cdots, \mathbf{y}(n-p)$. In this section, we first assume that p and q have been chosen and then introduce Akaike Information Criterion (AIC), a widely used model order selection rule [8] that will also be used later for model averaging.

Let $\mathbf{y}(n)$ and $\mathbf{z}(n)$ be available for $n = n_1, \cdots, n_2$. We can determine $\mathbf{W}$ by minimizing the following cost

$$\begin{aligned} J^{LS}(\mathbf{W}) &= \sum_{i=n_1}^{n_2} \|\mathbf{y}(i) - \mathbf{W}^T \mathbf{z}(i)\|^2 \\ &= \sum_{i=n_1}^{n_2} \mathrm{Tr}\left(\left[\mathbf{y}(i) - \mathbf{W}^T \mathbf{z}(i)\right]\left[\mathbf{y}(i) - \mathbf{W}^T \mathbf{z}(i)\right]^T\right) \end{aligned}$$

which is known as the LS problem. Tr denotes the trace. The LS problem minimizes the model mismatch error $\mathbf{e}(n)$ in (2). The solution to this problem can be found by setting the gradient of $J^{LS}(\mathbf{W})$ with respect to $\mathbf{W}$ to zero. To find the gradient, we first take the differential of $J^{LS}(\mathbf{W})$ and following the fact that if $\partial J = \mathrm{Tr}(A\partial X)$ then $\frac{\partial J}{\partial X} = A^T$, we have

$$\frac{\partial J^{LS}(\mathbf{W})}{\partial \mathbf{W}} = -2 \sum_{i=n_1}^{n_2} \mathbf{z}(i)\left[\mathbf{y}(i) - \mathbf{W}^T \mathbf{z}(i)\right]^T \tag{4}$$

Setting the gradient to zero leads to

$$\mathbf{Z}\mathbf{W} = \mathbf{Y} \tag{5}$$

where $\mathbf{Z} = \sum_{i=n_1}^{n_2} \mathbf{z}(i)\mathbf{z}^T(i)$ and $\mathbf{Y} = \sum_{i=n_1}^{n_2} \mathbf{z}(i)\mathbf{y}^T(i)$. If $\mathbf{Z}$ is invertible which is typically the case when $n_2 - n_1 + 1$ is no less than the dimension $pN + (q+1)M$ of $\mathbf{z}(n)$, then the LS solution is

$$\hat{\mathbf{W}} = \mathbf{Z}^{-1}\mathbf{Y} \tag{6}$$

The above expression is also known as the batch LS algorithm where $\mathbf{Z}$, $\mathbf{Y}$ and $\mathbf{Z}^{-1}\mathbf{Y}$ are determined by using all available data at once. This algorithm has the complexity in the order of $(pN + (q+1)M)^3$ multiplications.

To derive AIC rule, we assume that the elements of $\mathbf{e}(n)$ are i.i.d zero mean Gaussian with variance σ^2 for simplicity. Following [8] we have

$$\begin{aligned} \mathrm{AIC}(p,q) &= -2\ln p\big(\mathbf{y}(n_1), \ldots, \mathbf{y}(n_2); \hat{\mathbf{W}}(p,q)\big) + 2l \\ &= N(n_2 - n_1 + 1)\left(\ln 2\pi + 1 + \ln \hat{\sigma}^2(p,q)\right) + 2l \end{aligned} \tag{7}$$

where $p\big(\mathbf{y}(n_1), \ldots, \mathbf{y}(n_2); \hat{\mathbf{W}}(p,q)\big)$ is the probability density function of $\mathbf{y}(n_1), \ldots, \mathbf{y}(n_2)$ given LS estimate $\hat{\mathbf{W}}(p,q)$, $l = pN + (q+1)M$, and

$$\hat{\sigma}^2(p,q) = \frac{\sum_{i=n_1}^{n_2} \|\mathbf{y}(i) - \hat{\mathbf{W}}(p,q)^T \mathbf{z}(i)\|^2}{N(n_2 - n_1 + 1)} \tag{8}$$

Then, the model orders are chosen as the ones minimizing the above AIC among a set of candidate model orders. So far, we are able to obtain the LS estimate of model parameters according to AIC rule from a set of model order candidates. Note that other model order selection rules can also be employed such as Bayesian Information Criterion (BIC) and Generalized cross-validatory KL approach (GIC) [8].

IV. Least Square Model Averaging with Model Screening

Note that we assume that the model orders are chosen as hopefully the "best" ones by model order selection in the last section. However, it has been recognized that the uncertainty inherited in model order selection is ignored once the "best" model is finalized and the corresponding inference is conducted according to the conditional statistical characteristics of parameter estimates associated with the chosen model [7]. In case that the selected model deviates from the true model, relying on this only model could result in overly optimistic or misleading inference. On the other hand, model averaging, as will be described below, alleviates the strong dependence on only one selected model through combining estimated model parameters across different model candidates. In fact, by averaging the inherited uncertainty resulted from model order selection is incorporated rather than neglected.

Unlike in the last section we consider a sequence of different models indexed by $k = 1, \ldots, K$, where the kth model employs p_k and q_k as its output and input model orders. Following (1) those models can be expressed as

$$\mathbf{y}(n) = \mathbf{f}_k\left(\mathbf{y}(n-1), \cdots, \mathbf{y}(n-p_k), \mathbf{u}(n), \cdots, \mathbf{u}(n-q_k)\right) + \mathbf{e}_k(n)$$

Similar to (3) the compact approximation of the first order Taylor's expansion is given by

$$\mathbf{y}(n) = \mathbf{W}_k^T \mathbf{z}_k(n) + \mathbf{e}_k(n) \tag{9}$$

where $\mathbf{z}_k(n) = \left[\mathbf{y}_{kc}^T(n), \mathbf{u}_{kc}^T(n)\right]^T \in \mathbb{R}^{(p_k N + (q_k+1)M)\times 1}$, $\mathbf{y}_{kc}(n) = \left[\mathbf{y}^T(n-1), \cdots, \mathbf{y}^T(n-p_k)\right]^T$, $\mathbf{u}_{kc}(n) = \left[\mathbf{u}^T(n), \cdots, \mathbf{u}^T(n-q_k)\right]^T$, $\mathbf{W}_k = [\mathbf{A}_k^T, \mathbf{B}_k^T]^T$, $\mathbf{A}_k = [\mathbf{A}_{k1}^T, ..., \mathbf{A}_{kp_k}^T]^T$ and $\mathbf{B}_k = [\mathbf{B}_{k0}^T, ..., \mathbf{B}_{kq_k}^T]^T$.

In order to combine different models, we define $\mathbf{w} = [w_1, \ldots, w_K]$ as the weight vector where $0 \le w_k \le 1$ and $\sum_{k=1}^K w_k = 1$. The kth element of $\mathbf{w}$ is assigned to the kth model and the new estimated model parameter is given by

$$\hat{\mathbf{W}}_{MA} = \sum_{k=1}^{K} w_k \left[\hat{\mathbf{A}}_k^T \quad \mathbf{0} \quad \hat{\mathbf{B}}_k^T \quad \mathbf{0}\right]^T \tag{10}$$

where $\hat{\mathbf{W}}_{MA}$ is a $p_{\max}N + (q_{\max}+1)M$ by N parameter matrix, $p_{\max}$ and $q_{\max}$ are the maximum among p_k and q_k, $\hat{\mathbf{A}}_k$ and $\hat{\mathbf{B}}_k$ result from the LS estimate of $\hat{\mathbf{W}}_k$ by (6) based on the kth model and the zero elements in $\hat{\mathbf{W}}_{MA}$ are assigned to those data that don't appear in models with smaller p_k, q_k but do exist in larger orders to keep this formulation consistent. Since the LS algorithm provides the solution of $\hat{\mathbf{W}}_k$, the only concern left is how to determine the weight vector $\mathbf{w}$.

For this problem, we first compute the AIC_k for each model as in (7) and then choose a number of $K' < K$ models with the smallest AICs to screen out those less matching candidates. Let Γ denote the set of selected model candidates. Then, smoothed Akaike Information Criterion (S-AIC), a simplified form of Akaike model averaging [7], is adopted where the weight associated with the kth model in Γ is obtained as

$$w_k = \frac{\exp\left(-\frac{1}{2}\mathrm{AIC}_k\right)}{\sum_{k\in\Gamma}\exp\left(-\frac{1}{2}\mathrm{AIC}_k\right)}, \quad k \in \Gamma \tag{11}$$

The proposed least square model averaging with model screening algorithm is illustrated in Algorithm 1.

Algorithm 1 LS Model Averaging with Model Screening

1) Given p_k, q_k for each model k, $k = 1, \ldots, K$, prepare data $\mathbf{Z}_k$ and $\mathbf{Y}_k$.
2) Estimate $\hat{\mathbf{W}}_k$ using LS algorithm of (6) for each candidate model and compute the estimate $\hat{\sigma_i}^2$ of σ_i^2 by (8).
3) Calculate AIC by (7) for each model k. Select $K' < K$ models with the smallest AICs and let Γ denote the set of these models.
4) Compute the K'-dimensional weight vector $\mathbf{w}$ by (11). Note that $\sum_{k\in\Gamma} w_k = 1$.
5) Combine the K' models as in (10) to obtain the LS model averaging estimate of model parameters.

V. Simulation Results

To verify the proposed thermal model and estimation algorithm, we use the real data for a quad-core system with one cache core obtained from Intel as shown in Fig. 1 and Fig. 2 where $M = 5$ and $N = 5$. The whole set of data is divided into two parts. Training part comprises the first 1000 samples for the estimation of model parameters and validation part comprises the second 1000 samples for measuring the accuracy of prediction performance. For the data we use, the sampling (or iteration) interval is uniform and equals 0.001s.

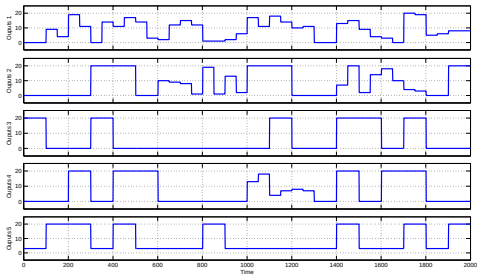

Fig. 1. Power inputs, $M = 5$.

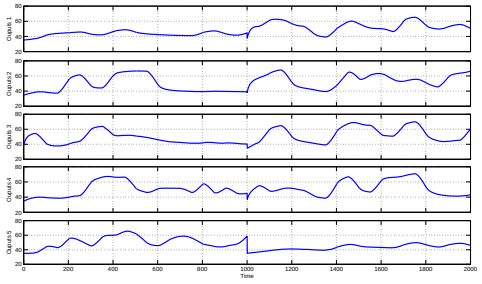

Fig. 2. Temperature outputs in Celsius, $N = 5$.

Our simulation results show that when we tried to estimate both b and $\mathbf{W}$ jointly in the LS fitting, we found that the LS

cost is almost invariant to a large range of b and $\mathbf{W}$. From a practical point of view, it is a good news that the MIMO ARMA modeling is not sensitive to b. We will choose $b = 30$ which is less than the minimum value in Fig. 2.

For simplicity, we consider a set of $K = 10$ models with model orders $p_k = q_k = k$, $k = 1, \ldots, 10$. For each model, the LS algorithm uses the first 1000 training data to determine a model matrix $\hat{\mathbf{W}}$ via the batch LS algorithm. The model mismatch error due to $\hat{\mathbf{W}}$ for the entire data is defined as $\hat{\mathbf{e}}(i) = \mathbf{y}(i) - \hat{\mathbf{W}}^T \mathbf{z}(i)$. The corresponding RMSE (root mean squared error) of the model mismatch for the training data is denoted as the estimation error by

$$e_T = \sqrt{\frac{\sum_{i=k+1}^{n_T} \|\hat{\mathbf{e}}(i)\|^2}{(n_T - k)N}} \tag{12}$$

where $n_T = 1000$. Note that from the definition of $\mathbf{z}(i)$ and the available data set, $\mathbf{z}(i)$ is available starting from $i = \max(p_k, q_k) + 1$. Similarly, the RMSE of the model mismatch due to $\hat{\mathbf{W}}$ for the validation data is defined as the prediction error by

$$e_V = \sqrt{\frac{\sum_{i=n_T+k+1}^{n_V} \|\hat{\mathbf{e}}(i)\|^2}{(n_V - n_T - k)N}} \tag{13}$$

where $n_V = 2000$.

To evaluate the performance of proposed algorithms, we choose the model orders corresponding to the 1st (model 4), 3rd (model 6) and 9th (model 2) smallest AICs. From Fig. 3 we can see that model order selection does have a significant impact on thermal modeling and temperature prediction due to different model mismatches under different chosen model orders as shown in Table I. However, model 4 with the smallest AIC has a larger prediction error than model 6 with the 3rd smallest AIC. In case that model 4 is used for future temperature prediction, it may not perform as well as model 6, which justifies the uncertainty from model order selection.

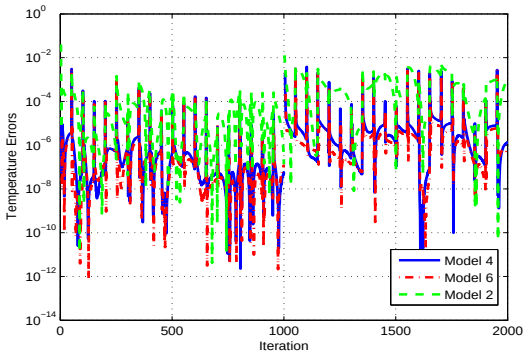

Fig. 3. Model-mismatch errors for output 3 of model 4, Model 6, and Model 2.

Although we know that for the current data, model 6 has better prediction performance, in general which model performs the best has to be decided after data validation. Thus, before any validation data are available, Algorithm 1 alleviates the uncertainty resulted from model order selection. Let assume $K' = 3$, then the K' smallest AICs have model index 4, 5, and 6, respectively, and their corresponding weights can be computed as $\mathbf{w} = [0.947, 0.052, 0.001]^T$. Since most model mismatches of the new parameters estimated by Algorithm 1 are so close to that of model 2 that they can not almost be distinguished, we just show its estimation and prediction errors in Table I. We can see that the prediction error from Algorithm 1 is 0.0057 which is smaller than that of model 2 with the smallest AIC.

TABLE I
e_T AND e_V FOR MODEL 4, MODEL 6, MODEL 2 AND LSMA.

	Model 4	Model 6	Model 2	LSMA
e_T	0.0026	0.0026	0.0065	0.0026
e_V	0.0061	0.0058	0.0226	0.0057

VI. CONCLUSION

In this paper we propose a black-box LTI MIMO thermal model for multicore microprocessor system. First, LS method combined with AIC model order selection rules are used to estimate the model parameters and predict model temperatures. Then, in order to mitigate the uncertainty inherited in model order selection, we introduce the idea of model averaging and propose LS based algorithm with model screening whose model weights can be computed by S-AIC. Simulation results show that the proposed algorithm has better prediction performance compared to LS method with AIC. Note that model averaging algorithm can also be extended to online estimation of model parameters by Recursive Least Square (RLS) method.

REFERENCES

[1] D. Brooks and M. Martonosi, "Dynamic thermal management for high-performance microprocessors," *In Proceedings of the 7th International Symposium on High-Performance Computer Architecture*, pp. 171–182, 2001.

[2] K. Skadron, M. R. Stan, W. Huang, S. Velusamy, K. Sankaranarayanan, and D. Tarjan, "Temperature aware microarchitecture," *Proc. Int. Symp. Comput. Architect.*, pp. 2–13, Jun. 2003.

[3] Y. Yang, Z. Gu, C. Zhu, R. P. Dick, and L. Shang, "ISAC: Integrated space-and-time-adaptive chip-package thermal analysis," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 26, no. 1, pp. 86–99, 2007.

[4] K. Coskun, T. S. Rosing, and K. C. Gross, "Utilizing predictors for efficient thermal management in multiprocessor socs," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 28, no. 10, pp. 1503–1516, Oct. 2009.

[5] D. Li, S. X.-D. Tan, E. H. Pacheco, and M. Tirumala, "Architecture-level thermal characterization for multicore microprocessors," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 17, no. 10, pp. 495–1507, Oct. 2009.

[6] T. J. Eguia, S. X.-D. Tan, R. Shen, E. H. Pacheco, and M. Tirumala, "General behavioral thermal modeling and characterization for multi-core microprocessor design," *Design, Automation and Test in Europe Conference and Exhibition (DATE)*, pp. 1136–1141, Mar. 2010.

[7] N. L. Hjort and G. Claeskens, "Frequentist model average estimators," *Journal of the American Statistical Association*, vol. 98, no. 464, pp. 879–899, Dec. 2003.

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