# A NEW APPROACH TO ORDER SELECTION AND PARAMETRIC SPECTRUM ESTIMATION

*Soosan Beheshti*

Electrical and Computer Engineering Department, Ryerson University, Toronto, Canada
soosan@ee.ryerson.ca

**ABSTRACT**

One important challenge in parametric spectrum estimation is the choice of an optimum number of parameters. In this paper, we provide a new method of order selection for the spectrum density estimation denoted by data model error whiteness (DMEW) criterion. Unlike the existing methods, this approach estimates the parameters and selects the optimum order simultaneously. We demonstrate the advantages of the new method over the existing approaches.

## 1. INTRODUCTION

We consider the power spectral density estimation of a wide-sense stationary (WSS) random process and the problem of order selection for parametric model-based spectrum estimations. We provide a new criterion for comparison of models of different order. In the existing parametric estimation methods, such as Yule-Walker and covariance methods, the data modeling error plays a very important role [4, 5]. The new approach examines whiteness of this error. Calculation of the new criterion is similar to the calculation of mean square error in data denoising which is presented in [1].

## 2. ORDER SELECTION AND PARAMETRIC SPECTRUM ESTIMATION (OSPSE)

Consider a wide sense stationary (WSS) random process $Y$ which is generated by the following minimum phase rational system

$$Y(e^{j\omega}) = \frac{\sum_{k=0}^{q} b_q(k)e^{-jk\omega}}{1-\sum_{k=1}^{p} a_p(k)e^{-jk\omega}} W(e^{j\omega}). \tag{1}$$

Finite length, $N$, sample of this random process, $y^N = [y[1], y[2], \cdots, y[N]]^T$, is available. Parametric spectrum estimation (PSE) method uses the available data, $y^N$, to estimate the random process and its power spectral density $P_y(e^{j\omega})$ as follows:

$$\hat{Y}(e^{j\omega}) = \frac{\sum_{k=0}^{q} \hat{b}_q(k)e^{-jk\omega}}{1-\sum_{k=1}^{p} \hat{a}_p(k)e^{-jk\omega}} W(e^{j\omega}), \tag{2}$$

$$\hat{P}_y(e^{j\omega}) = \frac{|\sum_{k=0}^{q} \hat{b}_q(k)e^{-jk\omega}|^2}{|1-\sum_{k=1}^{p} \hat{a}_p(k)e^{-jk\omega}|^2}. \tag{3}$$

where $w[n]$ is a zero-mean white noise with unit variance. For given orders $p$ and $q$, PSE uses the observed data $y^N$ to find estimates $\hat{A}_p = [\hat{a}_p(1), \hat{a}_p(2)\ , \cdots\ , \hat{a}_p(p)]$ and $\hat{B}_q = [\hat{b}_q(1), \hat{b}_q(2), \cdots, \hat{b}_q(q)]$. However, if the correct orders are not known, an order selection and parametric spectrum estimation (OSPSE) method selects the optimum orders $p_{opt}$ and $q_{opt}$, and provides the coefficient estimates, $\hat{A}_{p_{opt}}$ and $\hat{B}_{p_{opt}}$.

### 2.1. All-pole Modeling and OSPSE

Random process $Y$ in (1) can be modeled with the following all-pole structure

$$Y(e^{j\omega}) = \frac{b^*}{1-\sum_{k=1}^{p^*} a^*_{p^*}(k)e^{-jk\omega}} W(e^{j\omega}). \tag{4}$$

If the true model in (1) is such that $q = 0$, then $p^* = p$, $a^*_{p^*} = a_p$ and $b^* = b_q(0)$. If the true model has any zeros involved, then $p^*$ in (4) is infinite.

From (4), the random process in time domain has the following structure

$$y[n] - \sum_{k=1}^{p^*} a^*_{p^*}(k) y[n-k] = b^* w[n]. \tag{5}$$

where the coefficients in $A_{p^*}$ represent impulse response $h^*$ with the following z-transform

$$h^*(z) = \sum_{k=1}^{p^*} a^*_{p^*}(k) z^{-k}. \tag{6}$$

Figure 1 shows the considered all-pole model structure and Figure 2 shows the estimated model for a given order $p$ with $\hat{h}_{A_p}$, the impulse response estimate, and $y_{\hat{A}_p}$, the data estimate

$$h_{\hat{A}_p} = \sum_{k=1}^{p} \hat{a}_p(k) z^{-k}, \quad y_{\hat{A}_p} = h_{\hat{A}_p} * y, \tag{7}$$

where '*' denotes the convolution operator. In this modeling, the data modeling error (DME) is

$$e_{\hat{A}_p}[n] = y[n] - y_{\hat{A}_p}[n]. \tag{8}$$

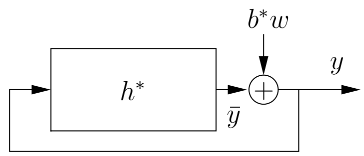


**Fig. 1**. Desired all-pole model.

An OSPSE method compares models of different order and chooses the optimum order $p_{opt}$ and the corresponding coefficients $\hat{b}_{p_{opt}}$ and $\hat{A}_{p_{opt}}$ (or equivalently $\hat{h}_{p_{opt}}$).

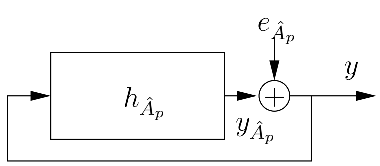


**Fig. 2**. Estimated all-pole structure for a given order $p$.

## 3. EXISTING OSPSE METHODS

The existing OSPSE methods separate the modeling to the following two steps:

### 3.1. Data Modeling Error (DME) and PSE Step

To calculate parameter estimates for a given order $p$, well known PSE methods such as Yule-walker or autocorrelation method, Burg's method, covariance method, and modified covariance method can be used. In all these methods, the desired solution is

$$A_p^* = \arg\min_{A_p} E(|e_{A_p}[n]|^2), |b_p^*|^2 = E(|e_{A_p^*}[n]|^2). \tag{9}$$

The importance of minimizing this error is rooted in the fact that if the order of the true model is $p$, then minimizing $E(|e_{A_p}[n]|^2)$ provides the correct parameters. However, in practical application $E(|e_{A_p}[n]|^2)$ is not available. In some PSE methods, such as Yule-Walker, the estimate of $A_p^*$ is calculated by using the empirical estimate of the autocorrelation function. Other approaches, such as covariance method, focus on minimizing the empirical estimate of $E(|e_{A_p}[n]|^2)$ and the estimate of the optimum coefficient is

$$\hat{A}_p = \arg\min_{A_p} \frac{1}{N-p} \sum_{n=1+p}^{N} |e_{A_p}[n]|^2. \tag{10}$$

In this approach, as the data length increases, convergence of the parameter estimates to the desired parameters in (9) is usually faster than that of the Yule-Walker approach.

### 3.2. Order Selection Step

The mean-square error of the modeling error is

$$m(p, y^N) = \frac{1}{N-p} \sum_{n=1+p}^{N} |e_{\hat{A}_p}[n]|^2. \tag{11}$$

This modeling error is an important element in comparison of model sets of different order. To compensate for the fact that this is a decreasing function of $p$, the order selection approaches use an additional term and provide criteria with the following structure

$$C(p, y^N) = m(p, y^N) + p\frac{f(N)}{N}, \tag{12}$$

where $f(N)$ is a function of data length $N$ and the penalty term $p\frac{f(N)}{N}$ is an increasing function of $p$. The optimum order minimizes the desired criterion

$$p_{opt} = \arg\min_{p} C(p, y^N). \tag{13}$$

Two well known OEPSE methods are Akaike information criterion (AIC), with $f(N) = 2$, and two-stage minimum description length (MDL) approach with $f(N) = log(N)$ [6, 7].

## 4. A NEW OSPSE METHOD: DME WHITENESS (DMEW) CRITERION

As it was mentioned in the previous section, DME ($e_{A_p}[n]$ in (8)) plays an important role in the existing OSPSE approaches (11,12). In these approaches the goal is to minimize the mean-square of this error for any model of order $p$. Here, we present a new OSPSE approach based on comparison of the desired structure in Figure 1 and the model of order $p$ in Figure 2. To make the structure of the model estimate as close as possible to the desired model, it is not necessarily needed to make the variance of $e_{A_p}[n]$ small, but rather, it is needed for this error to have the properties of a white noise similar to $b^*w[n]$. Therefore, to choose the optimum parameter for a given $p$ and evaluate the estimated models of different order, the "whiteness" of the DME needs to be examined. The more colored $e_{A_p}[n]$ is, the farther it is from the desired white noise $b^*w[n]$. The following lemma shows the connection between the DME and the desired white noise:
*lemma:* The DME, defined in (8)(shown in Figure 2), is

$$e_{A_p} = b^*w + y * (h^* - h_{A_p}). \tag{14}$$

*Proof:*

$$e_{A_p} = y - y * h_{A_p} = y * (1 - h_{A_p}) \tag{15}$$

$$= b^*w * \tfrac{1}{1-h^*} * (1 - h_{A_p}) = b^*w + b^*w * \tfrac{h^* - h_{A_p}}{1-h^*} \tag{16}$$

$$= b^*w + y * (h^* - h_{A_p}). \qquad \diamondsuit$$

The lemma provides useful information about the colored error DME. The behavior of $y*(h^* - h_{A_p})$ represents the closeness of DME to the desired white noise $b^*w[n]$

$$v_{A_p}[n] = e_{A_p}[n] - b^*w[n] = y*(h^* - h_{A_p}). \quad (17)$$

The smaller this random variable is, the closer our model is to the desired model in Figure 1. Therefore, $E(|v_{A_p}[n]|^2)$ can be used to evaluate the whiteness error. However, note that with only the available finite data, calculation of this value is impossible. Similar to the existing PSE methods, we suggest using the empirical estimate of this error

$$\mathrm{WNME}(p, A_p, y^N) = \frac{1}{N}\sum_{n=1}^{N} ||v_{A_p}[n]||^2, \quad (18)$$

which is defined as *white noise modeling error*. Minimizing this error over all the orders and the parameters provides the optimum order and coefficients estimates simultaneously. Nevertheless, the minimization can be provided in two steps. For a fixed p, minimizing WNME provides $\hat{A}_p$. Interestingly, solving this minimization is identical to calculation of parameter estimates in the available PSE covariance method

$$\hat{A}_p = \arg\min_{A_p} \mathrm{WNME}(p, A_p, y^N). \quad (19)$$

Minimum of the WNME error, given order $p$, is denoted as data modeling error whiteness (DMEW) criterion

$$\mathrm{DMEW}(p, y^N) = \mathrm{WNME}(p, \hat{A}_p, y^N). \quad (20)$$

The optimum order minimizes the desired criterion

$$p_{opt} = \arg\min_{p} \mathrm{DMEW}(p, y^N). \quad (21)$$

### 4.1. Calculation of DMEW Criterion

In order to examine the whiteness of the available DME in (8), the unavailable error $v_{\hat{A}_p}[n]$ is needed. In DMEW, the desired criterion in (20), which is the mean-square error of $v_{\hat{A}_p}[n]$, is estimated using the available mean-square error of DME, $m(p, y^N)$ in (11). While in the existing approaches, to define a criterion, $m(p, y^N)$ is used as an additive term, in DMEW approach the same quantity is used in a novel approach to probabilistically validate proper bounds on the desired criterion

$$L(p, y^n, p_1, p_2) \le \mathrm{DMEW}(p, y^N) \le U(p, y^N, p_1, p_2) \quad (22)$$

where $p_1$ and $p_2$ are defined as validation and confidence probabilities. By using the available DME and with confidence probability, $p_2$, the estimate of mean and variance of random variable $\mathrm{WNME}(p, \hat{A}_p, Y^N)$ can be provided. Then, the one sample of this random variable, $\mathrm{DMWE}(p, y^N)$, is probabilistically validated. In application, the provided worst case bound $U(p, y^N, p_1, p_2)$ is used for comparison of models of different order. The choice of $p_1$and $p_2$ are up to the user. The higher these probabilities are, the higher is our confidence on the bounds and the farther are the bounds from each other. Detailed calculation of these bounds using the DME is identical to calculation of reconstruction error in a new denoising approach which is presented in [1]. The calculation is also similar to calculation of minimum description complexity (MDC) for linear models, where MDC is a new information theoretic approach for statistical system modeling and is provided in [2, 3].

In [2, 3], it is shown that the existing order selection methods of form (12) are special cases of closed forms of form $U(y^N, p, p_1, p_2)$ with particular choices of the confidence and validation probabilities. For example, AIC is the same as $U(y^N, p, 0, 0)$ where $p_1 = p_2 = 0$. This explains the inconsistency of AIC which results over-modeling in some applications, i.e., the reality that AIC usually chooses more parameters than the true $p^*$. The simulations in the following section demonstrates the advantages of the DMEW approach over AIC and MDL approaches.

## 5. SIMULATION

Consider an autoregressive (AR) process of order $p^* = 14$ with the following coefficients[1]

$$a^*_{p^*} = [-0.0791\ -0.1070\ 0.0790\ -0.9214\ -0.2228\ 0.0472\ -0.1920\ 0.1822\ -0.2144\ 0.0152\ -0.0646\ 0.4186\ 0.1321\ -0.1142]$$

The goal is to use the available data of length $N = 140$ to provide the all-pole estimate of this model. Three order selection methods AIC, MDL, and the new DMEW criterion are used to choose the optimum order from $1 \le p \le 30$. The covariance method is used for calculation of the coefficients for each $p$. Throughout these simulations, the confidence and validation probabilities for DMEW approach are both 0.99. Figure 3 shows the true power spectrum and the results of different approaches. AIC estimate has 18 poles, MDL estimate has 12 poles and both methods are in dotted-dashed lines. The dashed line is the DMEW estimate. The new method has chosen the correct order and has 14 poles. The associated WNME error (18) of AIC, MDL and DMEW estimates are .1, .056 and .03 respectively, which shows that the error in the new approaches is whiter than the corresponding error from the other two methods. Figure 4 provides an interesting insight into the optimum order choice for different data lengths. The results are shown for data length ranging from 60 to 250. For any order selection method and a data length $N$, the figure shows two quantities, the average and

[1] The filter that generates the data has the following poles $.7$, $.35$, $.7e^{j(.75\pi)}$, $.7e^{j(-.75\pi)}$, $-.6$, $.8$, $.98e^{j(2\pi)}$, $.98e^{j(-.2\pi)}$, $.98e^{j(3\pi)}$, $.98e^{j(-.3\pi)}$.

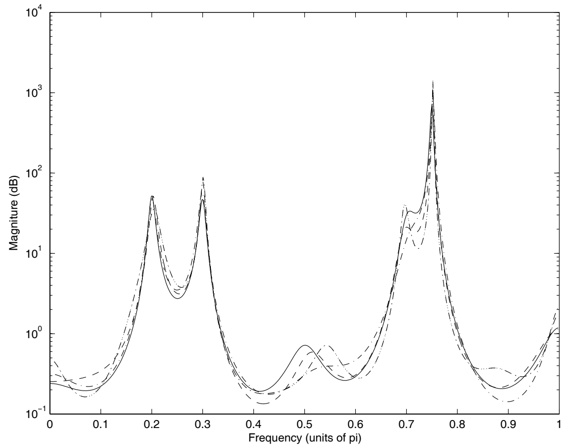


**Fig. 3**. Solid line is the desired spectrum. Estimated spectrums are: Dotted-dashed lines (AIC and MDL) and dashed line (DMEW). AIC and MDL's optimum orders are 18 and 12 poles. DMEW chooses $p = 14$ which is the correct order.

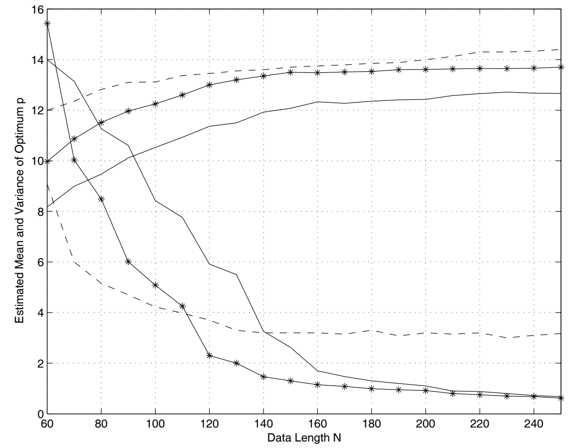


**Fig. 4**. Estimated mean and variance of optimum $p$ for data length $N$ between 60 and 250. AIC is the dashed line. MDL is the solid line and DMEW is the solid line with '*'.

variance of the optimum order for 300 trials. For example, for $N = 140$, AIC chooses 14 poles on average. However, AIC's variance is very high and of order 3. DMEW's average is also close to 14. However, the variance of this method is smaller and around one. MDL's average is 12 with variance of order one. As the figure shows, all methods choose orders less than the true order for data legnth $N$ less than 140. However, as the data length grows, AIC with its large variance, tends to overmodel and both MDL and DMEW provide the correct order since they are both consistent approaches. Nevertheless, as the figure shows, compare to DMEW, MDL tends to undermodel for the data length below 250.

## 6. CONCLUSION

We presented a new OSPSE approach. The new method focuses on calculation and optimization of a new criterion. In this criterion, the data modeling error (DME), that has an important role in the existing PSE approaches, plays a critical different role. While in the existing PSE approaches the goal is to minimize the DME, the new approach uses the DME to model the structure of the random process with the desired parametric model. Consequently, it is the whiteness of the DME, and not just its power, which plays an important role in the new OSPSE approach. The desired criterion is DME whiteness (DMEW) criterion which simultaneously selects the optimum order and the optimum parameters. This method is very different from the existing order selection approaches that use one of the PSE approaches for parameter estimation and a different criterion to compare these estimates for the order selection step. The approach shows many advantages over the existing methods and explains occasional under-modeling and over-modeling of important existing methods such as AIC and MDL. The consistent theory of DMEW introduces a fundamentally new approach to the OSPSE problem and the method promises to succeed in various areas of practical applications.

## 7. REFERENCES

[1] S. Beheshti and M.A. Dahleh, "A new information theoretic approach to signal denoising and best basis selection" *IEEE Trans. on Signal Processing,* Vol 53, No. 10, pp. 3613-3624, October 2005.

[2] S. Beheshti and M.A. Dahleh, "LTI systems, additive noise and order estimation", *Conference on Decision and Control,* Volume 6, pp.6491-6496, 2003.

[3] S. Beheshti and M.A. Dahleh, "A new information theoretic approach to order estimation problem", *Proceedings of the IFAC Symposium on System Identification,* 2003.

[4] M.H. Hayes. *Statistical Digital Signal Processing and Modeling*, Wiley and Sons, Inc 1996.

[5] J.G. Proakis, D.G. Manolakis. *Digital Signal Processing, Principles. Algorithms, and Applications*, Printice-Hall, Inc, 2002.

[6] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. on Automat. Contr.*, vol. 19, pp. 716-723, 1974.

[7] J. Rissanen, "Universal Coding, Information, Prediction, and Estimation," *IEEE Trans. Inform. Theory*, vol. 30, pp. 629-636, 1984.