Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®
Signal Processing 86 (2006) 2426–2429
SIGNAL PROCESSING
www.elsevier.com/locate/sigpro

A comparison of multivariate autoregressive estimators

Alois Schlögl*

Institute for Human-Computer Interfaces, Graz University of Technology, Krenngasse 37, A-8010 Graz, Austria

Received 17 January 2005; received in revised form 16 September 2005; accepted 14 November 2005
Available online 27 December 2005

Abstract

Recently, a new estimator—Arfit—for multivariate (vector) autoregressive (MVAR) parameters has been proposed. Several other MVAR estimators (e.g. Levinson recursion, Burg-type Nuttall–Strand, etc.) were already well known in the field of signal processing.

The various MVAR estimators have been implemented for Octave and Matlab. A method based on cross-validation and bootstrapping has been developed for comparing the various estimators. Thousand realizations of a MVAR(6)-process with 5 channels and a length of 1000 samples were generated. Each realization was separated into training and a test period. The training period was used to estimate the MVAR-parameters with each algorithm; the testing period was used to probe the accuracy of the estimates.

For large sample sizes, the Burg-type algorithm and Arfit yielded similar results, the multivariate Levinson method was worse. For small sample sizes, the Burg-type Nuttall–Strand method was significantly better than multivariate Levinson, the Arfit estimates performed worst.

In summary, the Nuttall–Strand method (multivariate Burg) for estimating MVAR parameters yielded the best results. The implementation of the algorithms for Octave and Matlab has been made available on the world wide web.

Keywords: Stochastic signal processing; Parametric modeling; Bootstrapping; Cross-validation; Stationary multivariate spectral analysis; Coherence; Directed transfer function; Causality; Information flow

1. Introduction

Several multivariate autoregressive (MVAR) estimators, like the multichannel Levinson recursion, the Burg-type Nuttall–Strand method, or the Vieira–Morf method have been known for more than 25 years [2]. Marple and Nuttall [3] compared three multichannel spectral estimates and recommended the Nuttall–Strand method. De Hoon et al. [1] and Salau [7] investigated the post-sampling accuracy of various MVAR estimators; he recommended the use of a Burg-type algorithm.

Recently, a new method, "Algorithm 808: Arfit" has been published [6,9] and the software implementation has been made available in Matlab. In this work, the MVAR estimators of [2,4,5,10] have been implemented for the use in Octave (see http://www.octave.org) and Matlab (see http://www.mathworks.com) [8]. Moreover, a method for comparing the different MVAR estimators has been developed.

Since all MVAR estimation algorithms are based on the principle of minimizing the prediction error,

*Tel.: +43 316 873 5319; fax: +43 316 873 5320.
E-mail address: alois.schloegl@tugraz.at.

0165-1684/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2005.11.007

we expect only small, if any, differences. Theoretical approaches, like Cramer–Rao bounds, do not take into account differences in the implementation of an algorithm. Therefore, a numerical method using cross-validation and bootstrapping has been developed for identifying subtle differences between the various MVAR estimators.

2. Method

2.1. Theory

An M-variate autoregressive process Y_k can be described by Eq. (1), $\boldsymbol{A}_i$, are $M \times M$ MVAR parameter matrices, p is the order of the MVAR-model and X_k is the M-variate innovation process with zero mean and covariance $\mathrm{var}\{X_k\}$. Schneider and Neumaier [9] also incorporated an intercept vector w, which "is included to allow nonzero mean of the time series". However, without loss of generality we can make Y_k a zero mean process by removing the mean. The relationship between the mean, μ_Y, and the intercept, w, is $w = (\boldsymbol{I} - \boldsymbol{A}_1 - \cdots - \boldsymbol{A}_p)\mu_Y$.

$$Y_k = \boldsymbol{A}_1 Y_{k-1} + \boldsymbol{A}_2 Y_{k-2} + \cdots + \boldsymbol{A}_p Y_{k-p} + w + X_k. \quad (1)$$

Eq. (1) can be seen as linear forward prediction $\sum_i(\boldsymbol{A}_i Y_{k-i})$ with some noise X_k added. However, the "noise" is important; if X_k were zero all the time, then the MVAR process Y_k would also be zero. For this reason, X_k is also called the innovation process.

The equation can be also seen as a model of how the observed values Y_k have been generated. Estimation algorithms are applied to a limited number of observed samples and provide estimates $\hat{\boldsymbol{A}}_i$ of the true underlying model parameter $\boldsymbol{A}_i$. Naturally, the estimates will never be the exact parameters. We will compare several of these estimation algorithms.

Once some estimates $\hat{\boldsymbol{A}}_i$ of the true parameters are available, we can rewrite (1) and replace the model parameters by its estimates. Instead of the innovation process X_k we get the prediction error E_k. In case where the estimates are "exact", the prediction error would be identical to the innovation process; in general, this is not the case.

$$E_k = Y_k - \hat{\boldsymbol{A}}_1 Y_{k-1} - \hat{\boldsymbol{A}}_2 Y_{k-2} - \cdots - \hat{\boldsymbol{A}}_p Y_{k-p}. \quad (2)$$

The mean squared prediction error will be, due to the estimation error, larger than the mean square of the innovation process (3). The difference of the variances is only caused by the estimation error $(\hat{\boldsymbol{A}}_i - \boldsymbol{A}_i)$. Hence, the smaller the variance of the prediction error the more accurate are the estimates.

$$\mathrm{var}\{E_k\} = \mathrm{var}\{X_k\} + E\left\langle \left[\sum_i (\hat{\boldsymbol{A}}_i - \boldsymbol{A}_i) Y_{k-i}\right] \times \left[\sum_i (\hat{\boldsymbol{A}}_i - \boldsymbol{A}_i) Y_{k-i}\right]^{\mathrm{T}} \right\rangle. \quad (3)$$

In practice, the variance of the prediction error, $\mathrm{var}\{E_k\}$, is larger than the variance of the innovation process $\mathrm{var}\{X_k\}$. Only when the estimates are correlated with the testing samples (e.g. the same samples were used for the estimation) might $\mathrm{var}\{E_k\}$ be smaller than $\mathrm{var}\{X_k\}$; this case is also known as "over-fitting". We will prevent over-fitting by using distinct data sets for estimating and for testing of the MVAR estimates.

2.2. Simulation

From a multichannel EEG recording 5-variate AR(6) parameters were obtained; these MVAR parameters were used to generate a thousand realizations; the intercept was $w = 0$; the covariance of the innovation process was the identity matrix, i.e. $\mathrm{var}\{X_k\} = \boldsymbol{I}$, and the mean square of the simulated signal $\{Y_k\}$ was MSY = 36. The MVAR parameters of the simulated process are available from Schlögl [8]. Each realization had a length of 1000 samples, and was then separated into training and testing segments. The training samples were used to estimate the MVAR parameters. From each realization, N samples were used (with $N = 40$, 50, 60, 70, 100, and 400) to obtain the MVAR-estimate.

The prediction error (2) was calculated from the test segment (500 samples) of each realization, and the total mean squared prediction error (MSE) (across all channels) was obtained from each testing period (500 samples). Non-overlapping training and testing periods ensured that over-fitting was avoided. The calculations were performed on a PC with Linux-OS and Matlab 5.3.0.10183 (R11).

Various MVAR estimators were tested. The following five algorithms were the most interesting candidates; and were, therefore, investigated in more detail.

(i) Arfit, it estimates also an "intercept" vector.
(ii) Arfit 0, assuming a zero intercept.

(iii) Multichannel Levinson algorithm with the correlation function estimation method; it is also called the multichannel Yule–Walker method.
(iv) Multivariate Burg-type algorithm (Nuttall–Strand) with biased estimation of covariance matrices.
(v) Multivariate Burg-type algorithm (Nuttall–Strand) with unbiased estimation of co-variances.

The estimation algorithms (i) and (ii) are available from Schneider and Neumaier [9], algorithms (iii)–(v) are implemented in the function MVAR.M of the TSA-toolbox [8].

3. Results

The mean and the standard error of the mean were obtained from the MSE of a thousand realizations. The results for different sample numbers N and for different algorithms are shown in Fig. 1. All results showed a MSE which is smaller than the MSY = 36; this indicates that stable MVAR estimates (all poles inside unit circle) have been obtained.

We also notice the prediction error became smaller with increasing sample numbers N (Fig. 1). This is common knowledge and can also be theoretically proven. It accounts for the fact that, for a fixed number of parameters, more data provide more accurate model estimates. On the other hand, Arfit contains additionally the intercept parameters w and should, therefore, perform worse than Arfit0. Also this expectation is confirmed by our experimental results (see Table 1 and Fig. 1).

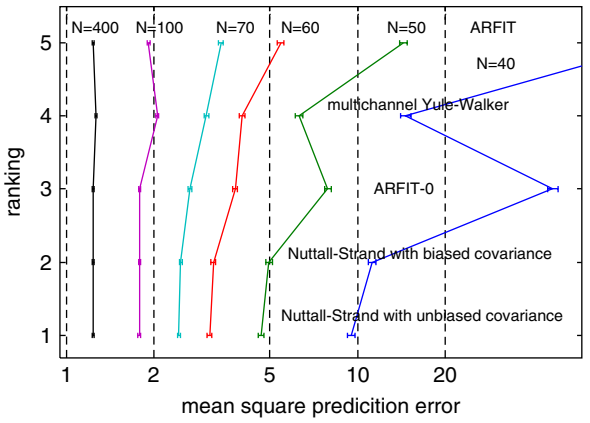

Fig. 1. Mean and twice the standard error of the total mean squared error from a thousand realizations of a MVAR(6) process with 5 channels; the total mean square of the MVAR-process was approx. 36. The x-scale shows the MSE, the y-axis shows the ranking of the 5 algorithms. The number of the samples N used for estimation was varied between $N = 40$ and 400. If the number of samples was sufficiently large ($N \geqslant 70$), the variance of the prediction error became smaller than 10% (= 3.6) of the signal variance.

Table 1
Ranking of algorithms

Rank	Algorithm	$\mu \pm \sigma_c$
1	Nuttall–Strand with unbiased covariance	2.4812 ± 0.0179
2	Nuttall–Strand with biased covariance	2.5197 ± 0.0183
3	Arfit0	2.7217 ± 0.0207
4	Multichannel Yule–Walker	3.1659 ± 0.0378
5	Arfit	3.5643 ± 0.0427

The total mean square prediction error (MSE) was calculated for one thousand realizations. The mean and the standard error of the MSE are depicted for a sample number of $N = 70$.

The algorithms were ranked according to average MSE based on the results with $N = 70$ (Table 1). The differences between Nuttall–Strand and Arfit0 were highly significant. For the case $N \leqslant 50$, Arfit0 became even worse than the multi-channel Yule–Walker method. For large number of samples ($N \geqslant 400$) only Yule–Walker seems to be slightly worse; the other methods did not show significant differences.

4. Discussion

Although, all MVAR estimators are based on the idea of minimizing the residual process, the presented method was able to identify subtle differences. The importance of good model estimates is also demonstrated; better estimates provide a smaller prediction error; the predictability is larger and the properties of the data are described more accurately. As a side effect, parameters derived from MVAR estimates (e.g. eigenmodes and associated damping periods, auto- and cross-spectra, coherence, directed transfer function, etc.) are also likely to be more accurate.

The results confirm the outcome of Marple and Nuttall [3] and Salau's [7]; i.e. the Burg-type algorithm (Nuttall–Strand) is better than the multichannel Yule–Walker method. When the number of observations was limited, the most accurate MVAR estimates were provided by the Nuttall–Strand method. When a sufficient number of samples were used for the estimates, the difference between

Nuttall–Strand and the Arfit was not significant. This result is in agreement with De-Hoon et al. [1], who showed for univariate AR parameters, that the Burg approach is superior to the Yule–Walker method.

The algorithm Arfit did not show any obvious advantage with respect to the Nuttall–Strand algorithm. But, if the number of observations was limited, the Nuttall–Strand method provided the most accurate estimates. Consequently, it is recommended to use the Nuttall–Strand method (a multivariate Burg-type algorithm) for estimating multivariate AR parameters.

References

[1] M.J.L. De-Hoon, T.H.J.J. Van-der-Hagen, H. Schoonewelle, H. Van-Dam, Why Yule–Walker should not be used for autoregressive modeling, Ann. Nucl. Energy 23 (15) (1996) 1219–1228.

[2] S.L. Marple, Digital Spectral Analysis with Applications, Prentice-Hall, Englewood Cliffs, NJ, 1987.

[3] S.L. Marple, A.H. Nuttall, Experimental comparison of three multichannel linear prediction spectral estimators, IEE Proc. F 130 (1983) 218–229.

[4] A.H. Nuttall, FORTRAN program for multivariate linear predictive spectral analysis, employing forward and backward averaging, Naval Underwater Systems Center Technical Report 5419, New London, Conn., 1976.

[5] A.H. Nuttall, Multivariate linear predictive spectral analysis employing weighted forward and backward averaging: a generalization of Burg's algorithm, Naval Underwater Systems Center Technical Report 5501, New London, Conn., 1976.

[6] A. Neumaier, T. Schneider, Estimation of parameters and eigenmodes of multivariate autoregressive models, ACM T. Math. Software 27 (March 2001) 27–57.

[7] M.O. Salau, Post-sample forecasting accuracy of estimators for vector autoregressive models, International Conference on Modelling and Simulation Proceedings, vol. 1, Univ. de Santiago de Compostela, Santiago de Compostela, 1999, pp. 369–379.

[8] A. Schlögl. The time series analysis toolbox for octave and matlab, available online at http://www.dpmi.tu-graz.ac.at/~schloegl/matlab/tsa/ and http://cvs.sourceforge.net/viewcvs.py/octave/octave-forge/extra/tsa/.

[9] T. Schneider, A. Neumaier, Algorithm 808: Arfit-a Matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models, ACM T. Math. Software 27 (March 2001) 58–65.

[10] O.N. Strand, Multichannel complex maximum entropy (autoregressive) spectral analysis, IEEE T. Automat. Contr. 22 (August 1977) 634–640.