Computational Modeling of Multivariable Non-Stationary Time Series in the State Space by the AOKI_VAR Algorithm.

Johanna Tobar*, Celso Bottura**, Mateus Giesbrechtr***

***Abstract*—In this work we propose an iterative e algorithm for identification in the state space of multivariable non-stationary discrete time series considering and implementing a computational process which we will call AOKI_VAR.**
The proposed algorithm is based on an algorithm proposed by Masanao Aoki for computational modeling of time series.
A modeling example is presented as well as discussions on validation, prediction and modeling of time series.

***Index Terms*—Non-stationary time series, computational modeling, non stationary stochastic process, time variant identification.**

I. Introduction

Focused on our primary objective, computational modeling of multivariable non-stationary time series data, we initially do a brief study of the theoretical foundations for discrete time series state space identification. A structure to be used in the resolution of this kind of problem is also proposed, as well as a discussion on validation, modeling and prediction of time series is made.

In this article, we treat non-stationary time series as a set of time invariant models, that is to say, the system matrices A_k, K_k, C_k, supposedly present small time variations, "small changes" meaning that every array changes slowly, allowing us to generate an iterative algorithm for the proposed objective. For this study we modify the algorithm proposed by AOKI [2] for time series state space modeling and call it AOKI_VAR Algorithm. Finally we test our algorithm using a benchmark.

II. FUNDAMENTALS

The computational modeling of data is a fundamental problem in nearly every scientific discipline; particularly in engineering and economics, the multivariate input and output data are called vector signals or time series and their analysis generally serve to at least one of the following possible purposes:

- Modeling of signals that need to be recognized or retrieved by a process of analysis in applications such as communications or economic time series predictions.
- Since in the signals, the information stored correspond to the dynamical systems that produce them, or to the dynamical systems that could predict hypothetical data, the utilization of the signals to determine the unknown model parameters of the system model will allow the realization of these tasks in the state space.

In general, the computational modeling of data in the state space for linear dynamical systems with multiple inputs and multiple outputs (MIMO) from the input and output measurements in noisy environments, is a central problem in multivariable modeling of time series, signal processing, identification, analysis and control systems design, learning and inteligent systems analysis and design.

A. NON-STATIONARY VECTOR TIME SERIES MODELING

A time series is a set of vector observations y_k, at each specific time $k = \{0, 1, 2, ...\}$. The multivariable discrete time series model can be seen as the information generation system that transforms past and present signals in to future observations. The states collect the information contained in the signals and transmit it through the dynamic model of the series to generate new usable signal information.

A non-stationary discrete time series can be represented by $Y^T = \{y_1, y_2, ..., y_T\}$. There is a serial dependency relationship between these observations. The data modeling problem is to describe mathematically the properties of these non-stationary stochastic vectors.
One classical possibility for this treatment is through the decomposition of the time series, y_k, in their basic movements such as: trend, T; cyclic movements, C; seasonal movements, S and irregular or random movements, I. Thus y_k can be decomposed as:

$$y_k = (T + C + S + I)_k \quad ou \quad y_k = (TCSI)_k \tag{1}$$

Usually the trend can be estimated, for example, using graphics where the trend is drawn, or by the method of least squares, or by the method of the semimedias.
In this study we concentrate in non-stationary time series data modeling, validation and prediction. The main objective of our work is to determine the state space data model, including its order and its matrices for diferent time instants, based on an algorithm that explores the properties of subspaces for data supplied by multivariable stochastic time series.

A time series model of the observed data $\{y_k\}$ is a specification of the joint probability distribution of the stochastic process sequence the time series realizes. In this study we only consider second order stochastic process, so only means and covariances are accounted for.

State University of Campinas, Campinas - Brazil
*(e-mail: jou_tb@hotmail.com)
** (e-mail: bottura@dmcsi.fee.unicamp.br)
***(e-mail: mateus.giesbrecht@gmail.com)

(Revised online publication: 27 November 2010)

B. MODELING OF STATIONARY STOCHASTIC TIME SERIES IN THE STATE SPACE.

Stationary multivariable time series in the state space can be modeled by a discrete stochastic multivariable linear time invariant system [13] :

$$\left\{ \begin{array}{c} x_{k+1} = Ax_k + v_k \\ y_k = Cx_k + w_k \end{array} \right. \quad (2)$$

where the terms v_k, w_k are respectively the state and output noises in the time series due to its stochastic nature. These terms may be considered as inputs no one has any control on them.

The perturbation vectors $v_k \in \mathbb{R}^n$ and $w_k \in \mathbb{R}^l$ are white noise stochastic processes sequences of zero mean and covariance matrices represented by:

$$E\left[\left(\begin{array}{c} v_k \\ w_k \end{array}\right)\left(\begin{array}{cc} v_s^T & w_s^T \end{array}\right)\right] = \begin{array}{cc} \left[\begin{array}{cc} Q & S \\ S^T & R \end{array}\right] & k = s \\ 0 & k \neq s \end{array} \quad (3)$$

where E is the mathematical expectance operator.
Defining the innovation vector e_k, with $E\left[e_k\right] = 0\ \forall k$, $e_k \in \mathcal{R}^l$ serially uncorrelated, stationary stochastic process in the weak sense, with covariance matrix $\Delta = E\left(e_k e_k^T\right)$, we can also represent the time series by the state space innovation form:

$$\left\{ \begin{array}{c} x_{k+1} = Ax_k + Ke_k \\ y_k = Cx_k + e_k \end{array} \right. \quad (4)$$

where $x_k \in \mathcal{R}^n$ is the state vector stochastic process stationary in the weak sense, where K is a constant matrix. The perturbation vector is $\left[\begin{array}{c} Ke_k \\ e_k \end{array}\right]$ and the covariance matrix is:

$$E\left\{\left[\begin{array}{c} Ke_k \\ e_k \end{array}\right]\left[\begin{array}{cc} e_k K^T & e_k^T \end{array}\right]\right\} =$$

$$\left[\begin{array}{cc} KE(e_k e_k^T)K^T & KE(e_k e_k^T) \\ E(e_k e_k^T)K^T & E(e_k e_k^T) \end{array}\right] = \quad (5)$$

$$\left[\begin{array}{cc} K\triangle K^T & K\triangle \\ \triangle K^T & \triangle \end{array}\right]$$

For analogy (3) and (5) are equal.

For the model described in the equation (4), we define the extended observability matrix by:

$$\mathcal{O} = \left[\begin{array}{cccccc} C^T & A^T C^T & \left(A^T\right)^2 C^T & \cdots & \left(A^T\right)^{n-1} C^T & \cdots \end{array}\right]^T$$

and the extended reachability matrix by

$$\Omega = \left[\begin{array}{cccccc} M & AM & (A)^2 M & \cdots & (A)^{n-1} M & \cdots \end{array}\right]$$

where the covariance matrix $M = E\left(x_{k+1} y_k^T\right)$, is given by:

$$M = A\Pi C^T + K\Delta$$

where $\Pi = E\left(x_k x_k^T\right)$:

$$\Pi = A\Pi A^T + K\Delta K^T$$

and Λ_o is the covariance matrix of the output stochastic process $\{y_k\}$:

$$\Lambda_o = C\Pi C^T + \Delta$$

Assuming $\Delta > 0$, Δ and K can be expressed as:

$$\triangle = \Lambda_o - C\Pi C^T$$

$$K = (M - A\Pi C^T)\Delta^{-1}$$

From these equations we have:

$$\Pi = A\Pi A^T + \left(M - A\Pi C^T\right)\left(\Lambda_o - C\Pi C^T\right)^{-1}\left(M - A\Pi C^T\right)^T \quad (6)$$

The stochastic realization problem for time series can be expressed in the following steps:

1)Determine the matrices Λ_o, A, M and C that represent a model for the covariances sequence Λ_i of a set of outputs y_k, assuming that the Hankel matrix of the covariances can be factored, as $H = \mathcal{O}\Omega$.

2)Solve the Riccati equation (6), for the state covariance Π.

3)Calculate Δ and K from $\Lambda_o, A, M, C\, and\, \Pi$.

In terms of the covariances, the Markov parameters of the system can be represented as:

$$\Lambda_i = \left\{ \begin{array}{cc} C\Pi C^T + R & i = 0 \\ M^T \left(A^T\right)^{-i-1} C^T & i < 0 \\ CA^{i-1} M & i \geq 1 \end{array} \right. \quad (7)$$

In other words:

$$\left\{ \begin{array}{c} R = \Lambda_0 - C\Pi C^T \\ Q = \Pi - A\Pi A^T \\ S = M - A\Pi C \end{array} \right. \quad (8)$$

For better understanding, see [**?**], [**?**], [5], [13].

C. STATE SPACE REPRESENTATION OF A NON-STATIONARY NOISE .

Consider $v(k)$ a stochastic process with zero mean and covariance matrix:

$$E_v\left(k_2, k_1\right) = V(k_1)\delta(k_2 - k_1)$$

where $V(k) \geq 0$ is its intensity.

In the case where $V(k)$ is a constant V, the process is stationary.

On the basis of a white noise e_k we generate the non-stationary noise $e_{j,k}$, that we call innovation noise, by:

$$\left\{ \begin{array}{c} z_{k+1} = A_k z_k + K_k e_k \\ e_{j,k} = C_k z_k + e_k \end{array} \right. \quad (9)$$

where A_k, C_k, K_k are time variant.

D. NON-STATIONARY TIME SERIES MODELING IN THE STATE SPACE: AOKI_VAR ALGORITHM.

The algorithm we propose is defined initially for a number of n intervals of experimentation for the signal identification. Thus the AOKI_VAR algorithm here proposed will be evaluated a number n of times whith T samples for a window, to determine n sets of systems matrices for each experiment j.

Let L_j be a specific integer for the j^{th} experimentation interval I_j, given by:

$$I_j = [k_j - L_j,\ k_j + L_j] \tag{10}$$

with $L_j = v * \nabla$ and $\nabla = S * \nabla_j$, where v and S are adequately fixed integers and ∇_j is the j^{th} sampling period.

In the proposed algorithm we will treat the innovation model of the non-stationary time series as a set of time invariant innovation models. Hence the modeling of the non-stationary time series will consist in a set of n stationary models that will describe the system for the proposed algorithm. As the innovation vector covariance matrix is time variant, $cov\, e_k = \Delta_k$.

To interpret the identification of the parameters of the problem, we have sets of indices j, k that tell us the input sample for the k^{th}instant of time for the j^{th}experimental interval for system (11). Therefore we can annotate that $j \in [j_0, j_0 + n - 1]$ and $k \in [k_0, k_0 + T - 1]$ where j_0 is the first interval of experimentation, k_0 is the first instant of time, n is the total number of experiments or tests and T is the number of samples for a simple experiment.

The noisy non-stationary linear system is represented by the following state space equations:

$$\begin{cases} x_{j,k+1} = A_{j,k}x_{j,k} + Ke_{j,k} \\ y_{j,k} = C_{j,k}x_{j,k} + e_{j,k} \end{cases} \tag{11}$$

where $T \geq n$, where $e_{j,k}$ is a stochastic process variant over time generated by a white noise.

The problem is to determine the state space description

$$\begin{bmatrix} x_{j,k+1} \\ y_{j,k} \end{bmatrix} = \begin{bmatrix} A_{j,k} & k_{j,k} \\ C_{j,k} & 1 \end{bmatrix} \begin{bmatrix} x_{j,k} \\ e_{j,k} \end{bmatrix} \tag{12}$$

based on the following associated output data sequences:

$$Y_{j,k} = \begin{bmatrix} y_{j_0,k_0} & y_{j_0,k_{0+1}} & \cdots & y_{j_0,k_0+T-1} \\ y_{j_{0+1},k_0} & y_{j_{0+1},k_{0+1}} & \cdots & y_{j_0+1,k_0+T-1} \\ \vdots & \vdots & \cdots & \vdots \\ y_{j_{0+n-1},k_0} & y_{j_{0+1},k_0} & \cdots & y_{j_0+n,k_0+T-1} \end{bmatrix}$$

The matrix $Y_{j,k}$ represents the set of n intervals of experimentation, allowing us to develop general expressions able to relate the outputs starting at an initial time k_0 and establishing an adequate experiment in window j, for a discrete time-varying system.

For achieving our objetive we utilize an iteractive algorithm structure. This kind of structure is described in Figure 1, and is based in [14], where representations for iterative and recursive data processing structures are presented. The recursive algorithm is obtained by working serially with the data, a sample at a time, using a recursion. On the other hand, the iteractive algorithm utilizes the *en bloc* method of analysis, where a single estimate is obtained by operating over the entire set of data in one operation. A sequence of these *en bloc* operations characterizes the iteractive process; the *en bloc or batch* solution can be considered as a single iteration on the data.

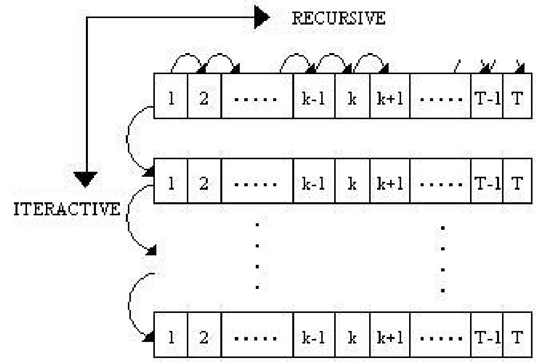

Figura 1. Recursive and Iteractive

To model the time variant series on the state space we can apply an iterative algorithm assuming small variations on the systems matrices in a predefined interval of operation. To the data we apply this iteractive scheme, that we call the AOKI_VAR algorithm, that is summarized in this section and where the following operations are made:

1)Determine the signal $y_{j;k} \equiv \bar{y}$, generating the matrices $H^A, H^M, H^C, H, Y_-, Y_+$

$$Y_- = \begin{bmatrix} \bar{y}_1 & \bar{y}_2 & \bar{y}_3 & \cdots & \bar{y}_{N-1} \\ 0 & \bar{y}_1 & \bar{y}_2 & \cdots & \bar{y}_{N-2} \\ 0 & 0 & \bar{y}_1 & \cdots & \bar{y}_{N-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \bar{y}_{N-k-1} & \bar{y}_{N-k} \end{bmatrix}$$

$$Y_+ = \begin{bmatrix} \bar{y}_2 & \bar{y}_3 & \bar{y}_4 & \cdots & \bar{y}_N \\ \bar{y}_3 & \bar{y}_4 & \bar{y}_5 & \cdots & 0 \\ \bar{y}_4 & \bar{y}_5 & \bar{y}_6 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \bar{y}_{j+1} & \bar{y}_{j+2} & \bar{y}_{j+3} & \cdots & 0 \end{bmatrix}$$

$$H = \frac{Y_+ Y_-^T}{N} = \begin{bmatrix} \Lambda_1 & \Lambda_2 & \cdots & \Lambda_k \\ \Lambda_2 & \Lambda_3 & \cdots & \Lambda_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_j & \Lambda_{j+1} & \cdots & \Lambda_{j+k} \end{bmatrix}$$

$$H^A = \begin{bmatrix} \Lambda_2 & \Lambda_3 & \cdots & \Lambda_{k+1} \\ \Lambda_3 & \Lambda_4 & \cdots & \Lambda_{k+2} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_{j+1} & \Lambda_{j+2} & \cdots & \Lambda_{j+k+1} \end{bmatrix}$$

$$H^M = \begin{bmatrix} \Lambda_1 \\ \Lambda_2 \\ \vdots \\ \Lambda_j \end{bmatrix}$$

$$H^C = \begin{bmatrix} \Lambda_1 & \Lambda_2 & \cdots & \Lambda_k \end{bmatrix}$$

2)Determine the singular value decomposition of the Hankel matrix of covariances

$$H = U\sum^{1/2}\sum^{1/2}V^T$$

3)Calculate the matrices $A_{j,k}, C_{j,k}, K_{j,k}$ and $\triangle_{j,k}$.
4)Update: iterate on the algorithm as described above, to obtain the state and covariance matrices for $y_{j,k}$ for each instant k of time and time interval j.
5)Validate.

E. VALIDATION, MODELING AND PREDICTION BY THE AOKI_VAR ALGORITHM

The validation of a model can be defined as the demonstration of its accuracy for a particular application. In this sense, accuracy is the absence of systematic and random errors. The time series model validation requires:

- The confirmation of the model (that is, that it proves to be credible and admissible)
- The verification of the model (that is, that it proves to be true).

Our proposal of validation for the AOKI_VAR algorithm has the following scheme, Figure 2:

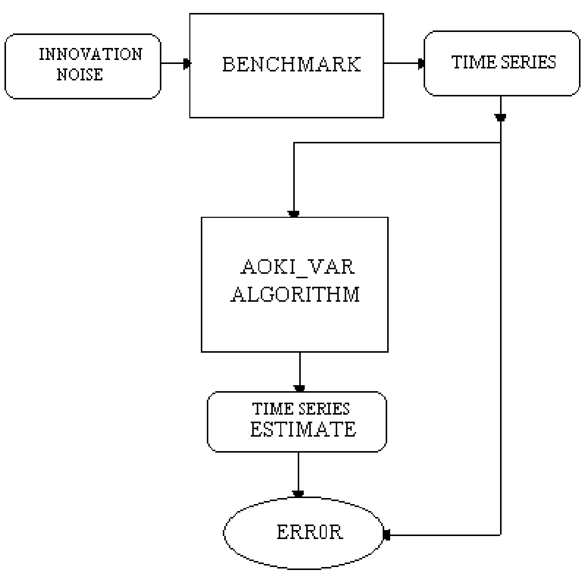

Figura 2. AOKI_VAR ALGORITHM VALIDATION

To confirm and to verify the quality of the proposed AOKI_VAR algorithm we implemented the following benchmark for the time variant system:

$$\begin{cases} X_{k+1} = A_k x_k + K_k e_k \\ y_k = C_k x_k + e_k \end{cases} \quad (13)$$

where

$$A_k = \begin{bmatrix} -0{,}3 & 0 \\ 0 & a_k \end{bmatrix} \quad (14)$$

with

$$a_k = -\frac{1}{3} - \frac{1}{10} sin(\frac{2\pi k}{400})$$

The remaining matrices are considered constants:

$$K_k = \begin{bmatrix} 5/4 & 5/4 \\ 7/10 & 7/10 \end{bmatrix} \quad (15)$$

$$C = \begin{bmatrix} 5/4 & -7/10 \\ 5/4 & -7/10 \end{bmatrix} \quad (16)$$

Using the procedure presented in II-D, the input to the system (11) is a random signal that changes for each iteraction of the algorithm according to (9).

For $j = 1$, the results for the validation phase of the proposed AOKI_VAR algorithm are presented in Figures 3 to 5 for $k = 1$.

The matrices $\hat{A}$, $\hat{B}$,$\hat{K}$ and $\hat{M}$ and the Markov parameters are also presented, as well as in Table 1,$\triangle$ (the noise covariance matrix) and $\hat{\triangle}$ (the estimated noise covariance matrix), for this case.

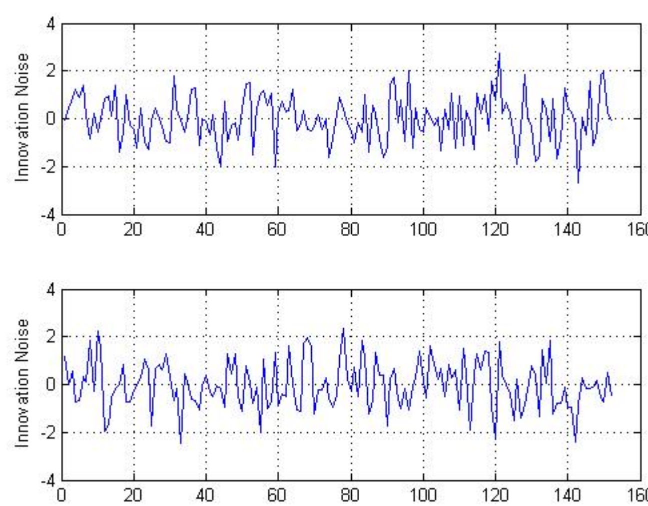

Figura 3. Innovation Noise for $k = 1$ and $j = 1$

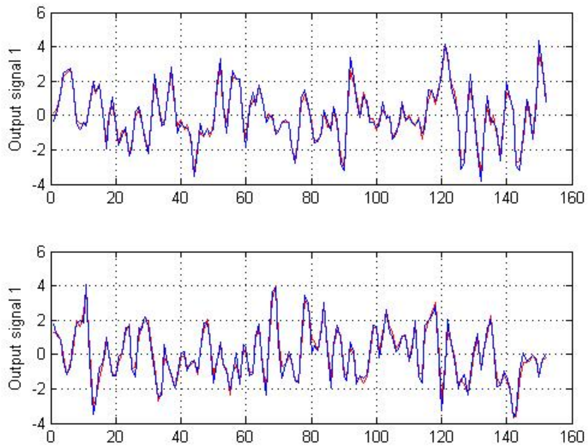

Figura 4. Outputs and outputs estimates for $k = 1$ and $j = 1$

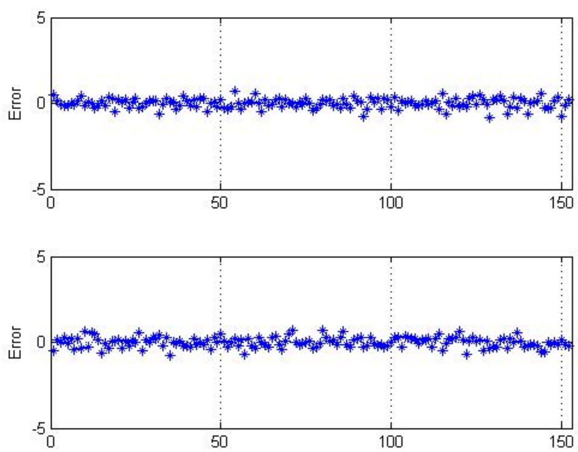

Figura 5. Error for $k = 1$ and $j = 1$

Table I
COVARIANCES MATRICES FOR $j = 1$ AND $k = 1$

Δ	$\begin{matrix} 0.9633 & -0.0856 \\ -0.0856 & 1.0092 \end{matrix}$
$\hat{\Delta}$	$\begin{matrix} 0.9176 & -0.0801 \\ -0.0799 & 1.2387 \end{matrix}$

$$\hat{A} = \begin{bmatrix} 0{,}174 & -0{,}21 \\ 0{,}216 & -1{,}02 \end{bmatrix}$$

$$\hat{K} = \begin{bmatrix} -1{,}02 & -0{,}93 \\ -0{,}17 & -0{,}18 \end{bmatrix}$$

$$\hat{C} = \begin{bmatrix} -1{,}10 & -0{,}05 \\ -1{,}15 & -0{,}04 \end{bmatrix}$$

$$\hat{M} = \begin{bmatrix} -1{,}1 & -1{,}15 \\ 0{,}05 & 0{,}045 \end{bmatrix}$$

Two ways of studying time series are here briefly considered: state space modelling and prediction. In our context, in the analysis of time series one tries to mainly determine the model order and the model structure (systems and covariances matrices) that generated the time series. Thus modeling of time series, or computational modeling of time series, or identification of time series in the state space involves finding a time series model that represents the time series characteristics: order and structure.

Assuming that the AOKI_VAR time series model has already been validated, it can be used for modeling any time series, as shown in Figure 7.

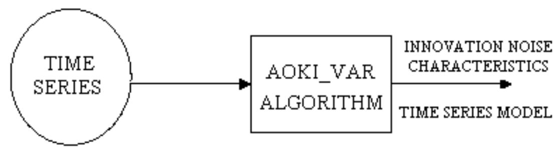

Figura 6. AOKI_ VAR MODELING

Based in Figure 6, we model the non-stationary time series presented in Figure 7 and obtain the matrices $\hat{A}$, $\hat{B}$,$\hat{K}$ and $\hat{M}$ for $j = 1$ and $k = 1$:

$$\hat{A} = \begin{bmatrix} 0{,}0337 & -0{,}1260 \\ -0{,}1260 & -0{,}5975 \end{bmatrix}$$

$$\hat{K} = \begin{bmatrix} -0{,}3476 & 1{,}0199 \\ 1{,}0102 & 1{,}0278 \end{bmatrix}$$

$$\hat{C} = \begin{bmatrix} -8{,}6150 & -0{,}0367 \\ 0{,}0234 & -0{,}0475 \end{bmatrix}$$

$$\hat{M} = \begin{bmatrix} -4{,}1824 & -1{,}1741 \\ 0{,}0418 & 0{,}0501 \end{bmatrix}$$

In Table II the estimated noise covariance matrix for $j = 1$ and $k = 1$ is presented.

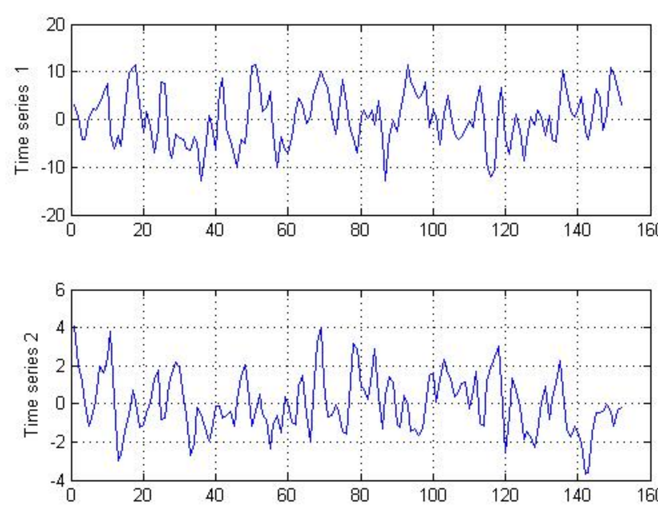

Figura 7. Time Series

Table II
COVARIANCE MATRIX

$\hat{\Delta}$	14.0581	-0.3927
	-0.3835	1.0264

Once a state space model for a time series is validated, it can be used for prediction. In Figure 8 we present a proposal for prediction of future values of the time series based on the model obtained by the AOKI_VAR Algorithm. The results of the prediction by the AOKI_VAR algorithm are presented in Figure 10, and in Table III. In Figure 9 the innovations noises for the validated model, for $j = 1$ and $k = 1$, are presented.

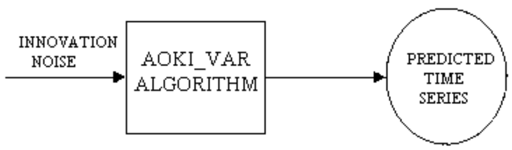

Figura 8. AOKI_VAR PREDICTION $j = 1$ and $k = 1$

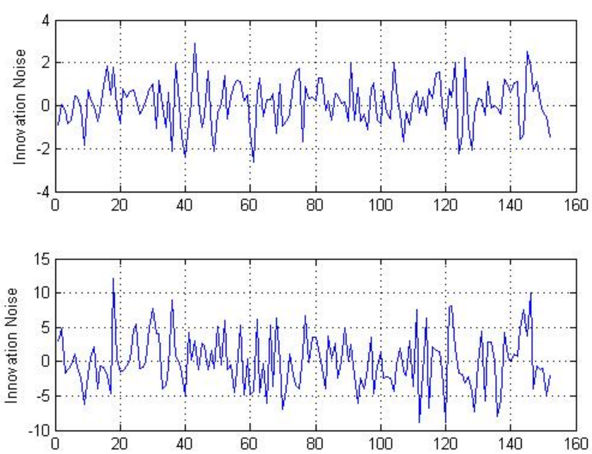

Figura 9. Innovations Noises $j = 1$ and $k = 1$

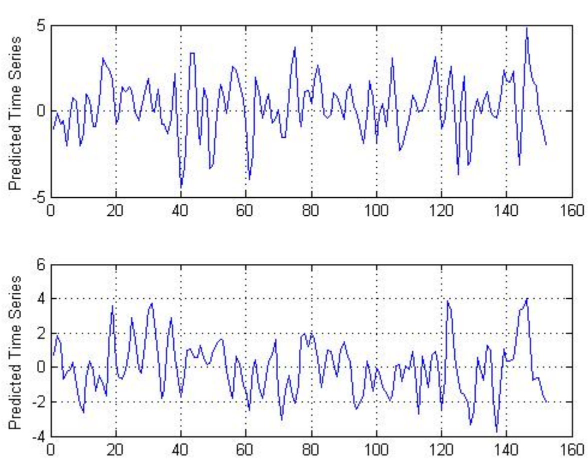

Figura 10. Predicted Time Series for $j = 1$ and $k = 1$

Table III
COVARIANCE MATRIX

$\triangle$	1.1097	0.8081
	0.8081	15.6649

III. CONCLUSIONS

In this study we formulated a computational procedure for state space modeling of multivariable non-stationary time series we called AOKI_VAR algorithm, based on the AOKI algorithm. We also discussed the validation, modeling and prediction procedures of multivariable non-stationary time series by the AOKI_VAR algorithm and an application using a proposed benchmark has been presented. The results can be considered good and the proposal useful under the considered hypothesis, but they should still be considered as preliminary due to the complexity of the problem considered.

REFERENCES

[1]Aoki, Masanao, "State Space Modeling of Time Series", Springer-Verlag, 1987.
[2]Barreto, G., "Modelagem Computacional Distribuída e Paralela de Sistemas e de Séries Temporais Multivariáveis no Espaço de Estado", PhD Thesis, Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas, UNICAMP, 2000
[3]Cáceres, Angel Fernando Torrico, "Identificação e Controle Estocásticos Decentralizados de Sistemas Interconectados Multivariáveis no Espaço de Estado", PhD Thesis, Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas, UNICAMP, 2005.
[4]Dewilde, Patrick and Van der Veen, Alle-Jan, "Time-Varying Systems and Computations", Kluwer Academic Publishers, 1998.
[5]Ohsumi, Akira and Kawano T., "Subspace Identification for a Class of Time-Varying Continuous-Time Stochastic Systems Via Distribution Based Approach", Proceedings 15th IFAC World Congress, Barcelona, July 2002.
[6]Shokoohi, S. and Silverman, L.M., "Identification and Model Reduction of Time-Varying Discrete-Time Systems", Automatica 23, pp. 509-522, 1987.
[7]Tamariz A.D.R, "Modelagem Computacional de Dados e Controle Inteligente no Espaço de Estados". PhD Thesis, LCSI/FEEC/UNICAMP, Campinas SP, Brasil, 2005.
[8]Tamariz A.D.R., C.P. Bottura, G. Barreto, "Iterative MOESP Type Algorithm for Discrete Time Variant System Identification", Proceedings of the 13th Mediterranean Conference on Control and Automation, (MED'2005), Limassol, Cyprus, June 27-29, 2005.
[1]Tamariz A.D.R., Bottura C.P. and Barreto G., "Algoritmo Iterativo do tipo MOESP para Identificação de sistemas discretos variantes no tempo - Parte I: Formulação", DINCON, 2003.
[10]Tamariz A.D.R., Bottura C.P. and Barreto G., "Algoritmo Iterativo do tipo MOESP para Identificação de sistemas discretos variantes no tempo - Parte II: Implementação e Experimentação", DINCON, 2003.
[11]Verhaegen, Michel and Deprettere, E., "A fast recursive MIMO State Space model Identification Algorithm", Proceedings of the 30th Conference on Decision and Control, Brighton, England, December 1991.
[12]Verhaegen, Michel and Dewilde, Patrick., "Subspace Model Identification - Part 1 : The outputerror state-space model identification class of algorithms", International Journal of Control, Volume 56, Number 5, 1187-1210, November, 1992.
[13]Verhaegen, Michel and Yu Xiaode, "A Class of Subspace Model Identification Algorithms to Identify Periodically and Arbitrarily Time-varying Systems", Automatica, Vol. 31, N0. 2, pp. 201-216, 1995.
[14]Young, Peter, Recursive Estimation and Time-Serires Analysis, Springer-Verlag, 1984.

(Revised online publication: 27 November 2010)