Mechanics Transport Communications

Academic journal

ISSN 1312-3823
issue 3, 2009
article № 0445
http://www.mtc-aj.com

DATA DEPENDENT SYSTEMS AND THEIR USE IN DYNAMIC SYSTEMS IDENTIFICATION AND MODELLING

Bohuš Leitner
Bohus.Leitner@fsi.uniza.sk

Faculty of Special Engineering, University of Žilina, Department of Technical Sciences and Informatics, ul.1.mája 32, 010 26 Žilina
SLOVAKIA

Abstract: *The paper deals with the possibilities of using the data dependent systems mathematical apparatus to identification and modelling of easy stochastic dynamic systems, especially mechanical ones. Its purpose is to briefly characterise fundamental terms and equations of mathematical apparatus of time series and to describe algorithm of a statistically adequate discrete model of stochastically dynamic system.*
Key words**:** *Data Dependent System, Dynamic System, Identification, Modelling, ARMA models, Criterion of Model Adequacy.*

INTRODUCTION

One of possible way of complex systems analysis without loss of accuracy and without necessity of complicated mathematical apparatus utilisation is observation of system during its working and utilisation of produced data to its analyses. Such analysed system we call Data Dependent Systems (DDS). It means that we do not need to know anything about composition of system and all analysis and conclusions are made just based on observed system output.

Data Dependent Systems are represented by sets of continuous output signals, which discretized with uniform sampling interval, gets a serie of data (values) which forms a base for description and analysis of investigated system. The main aim is to get possibility of system behaviour forecasting and eventually influencing of its behaviour not to get system in any unwanted state. In the first chapter of the paper the principal terms, relationships and characteristics of time series theory are introduced so as-some coherence to the better known apparatus of autocorrelation functions and power spectral density. The next chapter deals with an algorithm of a statistically adequate order of given time series and with conditions of tests of adequate ARMA model order.

PRINCIPAL CONCEPTIONS AND RELATIONSHIPS

For description of DDS, one can use advantageously ARMA models (Auto-Regressive Moving Average Models). Their advantages are mainly precisely formulated statistically criteria and relatively simple mathematical apparatus of statistically regressive analysis and testing of statistical hypothesis. Their disadvantages are complications of identification procedures and time consuming computer calculations despite relatively simple mathematical apparatus. Therefore, they are suitable just for off-line dynamic systems identification and modelling and their main application – forecasting of dynamic system behaviour is suitable just by stationary systems, which do not change their parameters with time [1].

The principal idea of an autoregressive expression of a discrete stationary stochastic process (Fig.1) is to express dependence of each immediate value of process X_t not as a function of time but as a function of former values [1,2,3].

Therefore, the simplest type of dependence will be the linear dependence of immediate values X_t on immediate previous values X_{t-1} is

(1) $$X_t = a_1 . X_{t-1} + \varepsilon_t$$

which describes a so called **autoregressive model of 1st order**, next signed **AR(1).**

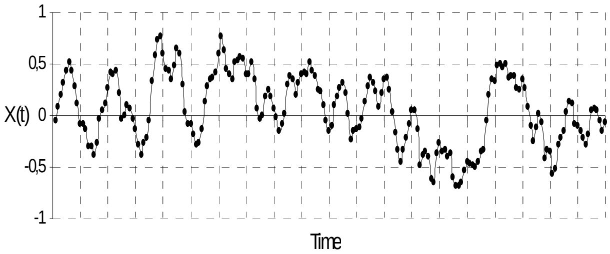

Fig.1. An example of discrete stochastic process

Really, if one tries to express dependence (Fig.1) graphically jets clear linear trend (Fig.2). It means that measure a_1 of dependence of immediate values X_t proceeding ones can be determined using a linear least square procedure which minimises the sum of deviation squares.

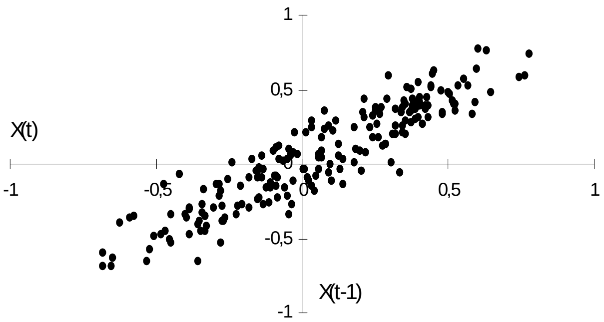

Fig.2. The course of dependence of Xt =f(Xt – 1)

This approach can be generalized on dependence of immediate X_t values on former *n* values, which can be described as **autoregressive model of n-th orders - AR(n)** as

(2) $$X_t = a_1 . X_{t-1} + a_2 . X_{t-2} + ... + a_n . X_{t-n} + \varepsilon_t$$

The basic presumption of adequacy of AR (n) model is the independence of stochastic values ε_t, which must form an independent series. If this presumption does not apply it means that εt depends on ε_{t-1}, ε_{t-2}, ε_{t-3} ... etc. The pure autoregressive models change to **Autoregressive Moving Average Models - ARMA**, generally of (n, m) order. With use of ARMA models one can express more complex types of internal dependencies, and as it will be shown further, their parameters have a very narrow dependence with the physical principal of followed processes [2,3,4].

A general type of ARMA dependence is a model of n-th order in an autoregressive part and (n-1)-th order in moving average part – **ARMA (n, n-1)** described by

(3) $$X_t - a_1 . X_{t-1} - a_2 . X_{t-2} - ... - a_n . X_{t-n} = \varepsilon_t - b_1 . \varepsilon_{t-1} - b_2 . \varepsilon_{t-2} - ... - b_{n-1} . \varepsilon_{t-n+1}$$

one can suppose, that residual deviations ε_t are of normal distribution of probability with zero mean and dispersion of σ_{e2}, $\left(\varepsilon_t \approx N\left(0, \sigma_\varepsilon^2\right)\right)$.

The basic characteristics of ARMA models are **impulse response function** – so called Greens function which can express conditions of stability of models and **inverse function** describing dynamics of models by expression if influence of former values of the process on the present ones [2]. To express it more simply we can introduce a back shift operator B in general mode defined as

(4) $$B.X_t = X_{t-1} \quad \text{or} \quad B^j . X_t = X_{t-j}$$

and using this the general ARMA (n, n-1) model form from equation (3) gets the form

(5) $$\left(1 - a_1 . B - a_2 . B^2 - ... - a_n . B^n\right) . X_t = \left(1 - b_1 . B - b_2 . B^2 - ... - b_{n-1} . B^{n-1}\right) . \varepsilon_t$$

Greens function G_J of difference equation (5) can be used to express values of X_t as a linear combination of stochastic values of ε_t as

(6) $$X_t = \sum_{j=0}^{\infty} G_j . \varepsilon_{t-j} = \left(\sum_{j=0}^{\infty} G_j . B^j\right) . \varepsilon_t = \left(G_0 + G_1 . B + G_2 . B^2 + ...\right) . \varepsilon_t$$

Similarly can be expressed the value of X_t as a linear combination of former values. Function of coefficients I_j in this expression is called „inverse function“ which is defined as

(7) $$X_t = \sum_{j=1}^{\infty}(I_j.X_{t-j}) + \varepsilon_t = I_1.X_{t-1} + I_2.X_{t-2} + ... + I_q.X_{t-q} + \varepsilon_t$$ or

(8) $$\varepsilon_t = \left(1 - I_1.B - I_2.B^2 - ...\right).X_t$$

The condition of stability of ARMA (n, n-1) model is generally in form $|\lambda_k| < 1$, for k = 1,2,...,n where λ_k are roots of characteristic equation on left-hand side of equation (3) in form of

(9) $$\lambda^n - a_1.\lambda^{n-1} - a_2.\lambda^{n-2} - ... - a_n = 0 .$$

Similarly we can defined a condition of invertibility as $|\nu_k| < 1$, for k = 1,2,...,n-1 where ν_k are roots of characteristic equation on right-hand side of relationship (3) in form of

(10) $$\nu^{n-1} - b_1.\nu^{n-2} - ... - b_{n-1} = 0$$

Using Greens function one can developed an implicit expression of discrete values of autocorrelation function (ACF), which get the form [1, 2, 5] of

(11) $$\begin{aligned} R_0 &= a_1.R_1 + a_2.R_2 + ... + a_n.R_n + \\ &\left(1 - b_1.G_1 - b_2.G_2 - ... - b_{n-1}.G_{n-1}\right).\sigma_\varepsilon^2 \\ R_1 &= a_1.R_0 + a_2.R_1 + ... + a_n.R_{n-1} + \\ &\left(-b_1 - b_2.G_2 - ... - b_{n-1}.G_{n-1}\right).\sigma_\varepsilon^2 \\ &\vdots \\ R_{n-1} &= a_1.R_{n-2} + a_2.R_{n-3} + ... + a_n.R_1 - b_{n-1}.\sigma_\varepsilon^2 \\ R_k &= a_1.R_{k-1} + a_2.R_{k-2} + ... + a_n.R_{k-n} \\ &for\ k \ge n \end{aligned}$$

Power spectral density (PSD) can be determined using Fourier transform of ACF or in a simpler way directly [1,2] from the formula

(12) $$S(\omega) = \sigma_\varepsilon^2 . \frac{\left(\varepsilon^{i.(n-1).\omega} - b_1.\varepsilon^{i.(n-2).\omega} - ... - b_{n-1}\right)^2}{\left(\varepsilon^{i.n.\omega} - a_1.\varepsilon^{i.(n-1).\omega} - ... - a_n\right)^2} .$$

which holds for $\omega \in \left[(-\pi/\Delta t) \le \omega \le (+\pi/\Delta t)\right]$, where Δt is a sampling interval. Better estimate of power spectra through the whole frequency band one can get from continues model as it is shown in [1].

ALGORITHM OF STATISTICALLY ADEQUATE DISCRETE MODEL DETERMINATION

As it was indicated in the former chapter, the aim of identification is to determine order n of statistically adequate model ARMA (n, n-1), coefficients on left and right-hand side ($a_1, a_2 \ldots, a_n, b_1, b_2 \ldots, b_{n-1}$) and sum of squares of residual deviations $\sum \varepsilon_t^2$ (or their dispersion σ_ε^2).

Because of the necessity of recurrent determination of deviations ε_t from the start, the result is from the point of view of coefficients non-linear. So that it necessary to apply a non-linear least square procedure looking in interactive steps for the minimum of sum of squares.

A special case of identification therefore forms pure autoregressive models of n-th orders AR(n) because for determination of its parameters a linear least squares procedure is suitable. Searched vector of coefficients $\mathbf{a} = [a_1, a_2, ..., a_n]^T$ of AR(n) model one gets as a solution of matrix equation

(13) $$\mathbf{S}.\mathbf{a} = \mathbf{T}$$

where $\mathbf{S} = \mathbf{X}^\mathbf{T}.\mathbf{X}$ and $\mathbf{T} = \mathbf{X}^\mathbf{T}.\mathbf{Y}$, where

(14) $$\mathbf{X} = \begin{bmatrix} X_n, & X_{n-1}, & \cdots & X_1 \\ X_{n+1}, & X_n, & \cdots & X_2 \\ \vdots & & & \\ X_{N-1}, & X_{N-2}, & \cdots & X_{N-n} \end{bmatrix}$$

(15) $$\mathbf{Y} = \begin{bmatrix} X_{n+1} \\ X_{n+2} \\ \vdots \\ X_N \end{bmatrix}$$

and N is length of time series. Searched dispersion σ_ε^2 can be found as

(16) $$\sigma_\varepsilon^2 = \frac{1}{N-n}.\sum_{t=n+1}^{N}\left(X_t - a_1.X_{t-1} - ... - a_n.X_{t-n}\right)^2 = \frac{1}{N-n}\sum_{t=n+1}^{N}\varepsilon_t^2$$

Further steps in an identification procedure (determination of an optimum order n of AR(n) model) is similar to that of general ARMA(n,n-1) model, which is shown in the next or in [1,2].

Determination of starting guess of parameters

The main problem in solution of effective identification algorithm of ARMA (n, n-1) model is the guessing of $(a_1, a_2, \ldots a_n)^{(0)}$, $(b_1, b_2, \ldots b_{n-1}$

$)^{(0)}$ values to ensure convergence of used iterative method and not too big a number of steps.

Far best of tested methods [2] is the one based on expression of inverse function. These formulas are then linear for each type of ARMA (n, n-1) model. Using general ARMA (n, n-1) in operator form (4) and for ε_t's giving such a definition of inverse function in form (8), comparing values by same powers of operators one gets system of equations as

$$\begin{aligned} a_1 &= b_1 + I_1 \\ a_2 &= b_2 - b_1.I_1 + I_2 \\ a_3 &= b_3 - b_1.I_2 + b_2.I_1 + I_3 \\ &\vdots \\ a_j &= b_j - b_1.I_{j-1} + b_2.I_{j-2} + \ldots + b_{j-1}.I_1 + I_j \end{aligned} \tag{17}$$

which holds for each j, knowing, that $b_j = 0$ for (j > n-1) and $a_j = 0$ for (j > n) for ARMA (n, n-1) model. Then for (j > n) is

$$\left(1 - b_1.B - b_2.B^2 - \ldots - b_{n-1}.B^{n-1}\right).I_j = 0 . \tag{18}$$

I mean that parameters a_j and b_j one can jet knowing values of inverse functions I_j. To solve the pure autoregressive model AR (p) can be used for a system of equations (13) is $I_j = a_j$ for $j=1, 2\ldots p$ and $I_j = 0$ for $(j > p)$. Procedure of starting guess of parameters a_j, b_j of general ARMA (n, n-1) model is then as follows:

1. *Solutions of parameters a_j of an AR(p) model where p=2.n-1 and values of inverse functions I_j simultaneously.*
2. *Determination of moving average parameters b_j by solving set of linear equations using equation (13) for j = n+1, n+2, ... , 2n-1.*
3. *Determination of autoregressive parameters a_j using set of equations (13) and then parameters b_j.*

Criterion of ARMA (n, n-1) model adequacy

To judge adequacy of chooses order n of ARMA (n, n-1) model (eventually AR(n) model) the procedure **of statistical hypothesis testing** was used. The test of model adequacy is in a principle the test of statistical independence of deviations ε_t. The standard F-test of statistical significance of two sample differences was chosen and this test was modified as a test of statistically significant decrease of the sum of squares. If one has to specify when by a regressive model with r parameters can be s of them counted as zero having N observations the test criterion becomes form as

$$F = \frac{\dfrac{A_0 - A_1}{s}}{\dfrac{A_1}{N - r}} \tag{19}$$

where A_1 is sum of squares of higher order model, A_0 sum of squares of initial model. Resulting value of F is to be compared to value of F_{crit} found in the table of critical values of F – distribution for $(N - r)$ and s degrees of freedom and chosen level of probability (mostly 95 %). When we get that $F > F_{krit}\left(s, N - r\right)_{0,95}$ then decrease the sum of squares during a change to higher order model was statistically significant, therefore, the initial model was not suitable. On the other hand, if $F < F_{krit}\left(s, N - r\right)_{0,95}$, the initial model was statistically adequate then an increase of its order does not make sense [1,4].

Algorithm of optimum model determination

Using former shown dependencies and formulas one can describe an optimum ARMA (n, n-1) model getting algorithm in words as follows [5]:

1. *Calculation of ARMA (2n, 2n-1) model parameters for n=1 and its sum of squares* $A_0 = \sum \varepsilon_t^2$.
2. *Increase of order* n $\rightarrow$ (n+1) *and calculation of model parameters and sum of squares A_1.*
3. *Testing of statistically significance of sum of squares decrease $\Delta A = A_0 - A_1$. In case that decrease is statistically significant, go to 2. In the other case the former model was statistical adequate.*
4. *Test of a_{2n}, b_{2n-1} parameters if their value is near zero or if their internal of confidence contains zero. When not ARMA (2n, 2n-1) model is suitable.*
5. *When a_{2n}, b_{2n-1} are zeros or near zero, calculation of ARMA (2n-1, 2n-2) model parameters.*
6. *Testing of moving average parameters b_j of ARMA (2n-1, 2n-2) model. If some of them are near zero, create of ARMA (2n-1,m) model for m < 2n-2 and calculation of its parameters eventually of pure AR (2n-1) model.*

Shown algorithm was used by developing of program ARMAGET (developed on authors department). It contains users menu, which apart from basic functions with file, configurations, work with windows and help functions. Basic is two submenus – submenu of “Simulation” and submenu of “Identification” (Fig.3).

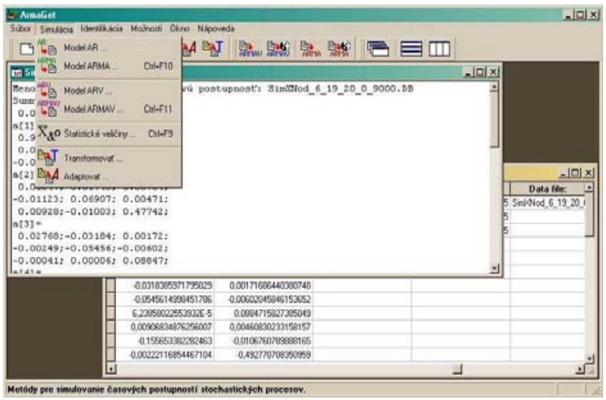

Fig.3 : The Environment of ARMAGET software

Item “*Simulation*” enables adjustment and conversion of incompatible input files of time series to compatible ones and simulation (generation) of time series basing on given AR or ARMA models order and parameters with possibilities of mean and dispersion selection of simulated serie.

The heart of the program is submenu “*Identification*”, by means, which is possible to make selection of the identification method and way of chosen time series, whereupon it is possible to use either adaptive algorithm of time series identification or make identification using non-linear least squares method. Identification by means of higher presented non-linear (respectively for AR models – linear) least square method is available in item *Identification* and its sub-menu *NLINLS*. Here are four options. First two- *Model AR – after orders* and *Model AR-complete calculation* give as results of identification AR mode.

Next item – *Model ARMA- after orders* gives coefficients of beforehand selected order of ARMA (n, n-1) models determination. It means, that we it is necessary beforehand to determine required order (known number of coefficients) of autoregressive part - a_k and moving average part - b_k which principally determine number of former values the calculated value depends on. The initial guess is of coefficients of ARMA (n, n-1) model. To coefficients of moving average part are assigned value of zero and coefficients of autoregressive part obtained by application of linear least square method. Simultaneously the sum of squares of deviations value expressing deviation of theoretical model from real model is calculated. Then the proper iterative calculation follows, which outputs are the coefficients of model (Fig.4).

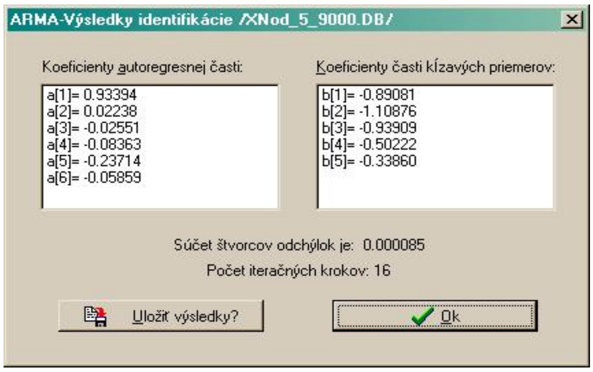

Fig.4 : Results of identification

The last important item is *Model ARMA-complete calculation*, which aim is to find an optimal ARMA (n, n-1) model. This model is the best describe of stochastic system, which output is a time serie. Because, in most cases we don’t know optimal order of model, beforehand it is necessary to determine by an iterative procedure an optimal order of model for system description.

CONCLUSIONS

Method for determination of adequate autoregressive models of stochastic systems and developed relationships between parameters of discrete and continuous models shown in former chapters were successfully verified on case of identification of mechanical dynamic system of machine tool during cutting with its application by adaptive geometrical control in real time [3] and by analysis of feed-back system (machine tool – cutting) [4]. Shown connections can be utilized by determination of modal characteristics of mechanical dynamics systems too.

There is the mostly used procedure of experimental investigation of dynamic characteristics of different systems and structures at present the application of dynamic compliance matrices determination and analysis of structure modes.

REFERENCES:

[1] Leitner, B.: Stochastic Time Series and Possibilities of Their Use for selected Modal Characteristics Determination of Mechanical Structures. In: Scientific Papers of Žilina University "Communications", Žilina 2001. ISSN 1335 – 4205.

[2] Beňo, B.: Stochastické metódy identifikácie dynamických systémov dopravných a stavebných strojov. [PhD Thesis]. ŽU, Žilina 2003, 100 p.

[3] Broersen, P.M.T., Bos, R.: Time-series analysis if data are randomly missing, TU Delft digital repository, IEEE, 2006, Netherlands, web: http://repository.tudelft.nl/file/379514/370833.

[4] D. Dottori, T. Ocaktan, M. Picchio, R. Staub, Multivariate time series analyisis and its applications, web: http://www.core.ucl.ac.be/~laurent/ts/pdf/VARMA.pdf, 2005.

[5] Leitner, B.: Modelling and Simulation of Transport Machines Working Conditions by using of Autoregressive Models, In: Academic Journal "Mechanics, Transport, Communications", Issue 1/2007, Art. No. 0079, VTU of Todor Kableshkov, Sofia 2007, Bulgaria, 8 p., www.mtc-aj.com

Reviewer: Prof. Dr. Eng. Milan SÁGA., University of Žilina, Faculty of Mechanical Engineering, Department of Applied Mechanics.

ЗАВИСИМИ ОТ ДАННИ СИСТЕМИ И ПРИЛОЖЕНИЕТО ИМ В ИДЕНТИФИКАЦИЯТА И МОДЕЛИРАНЕТО НА ДИНАМИЧНИ СИСТЕМИ

Бохуш Лейтнер
Bohus.Leitner@fsi.uniza.sk

Факултет по специализирано инженерство, Университет в Жилина, катедра по технически науки и информатика, ул. "1 Май" 32, Жилина 010 26
СЛОВАКИЯ

Ключови думи: *Зависима от данни система, Динамична система, Идентификация, Моделиране, модели ARMA, Критерий за адекватност на модела*

Анотация*: Статията третира възможностите за приложението на математически апарат от зависими от данни системи при идентификацията и моделирането на удобни стохастични динамични системи и най-вече, на механични такива. Целта е да се представят накратко основните членове и уравнения на математически апарат за последователни измервания във времето и да се опише алгоритъм за статистически адекватен дискретен модел на стохастично динамична система.*