

Construction of Predictive Models of Systems with Incomplete Observation Matrix

Irina Serysheva
Irkutsk National Research Technical University
 Irkutsk, Russia
 sia_cyber@mail.ru

Yuri Khrustalev
Irkutsk National Research Technical University
 Irkutsk, Russia
 khrustalev@istu.irk.ru

Abstract – The formalized technique of processing the measurement information obtained in the subsystem of internal comparisons of the standard based on the results of indirect measurements performed in the process of the functioning of the group standard of time and frequencies is offered. The program modules realizing a technique allow to solve a problem of full automation of process of creation of models of time series according to empirical data and to lower an error of reproduction of units of time and frequency by atomic clock to 30%.

Keywords – group standards of time and frequency (ensemble of atomic clocks), predictive models of time series, filtration of anomalous measurements, regression analysis, extremum of function of many variables.

I. INTRODUCTION

The accuracy of the estimation of the state of various technical systems can be significantly improved through the use of predictive models describing the dynamics of the object. Typical examples of dynamic stochastic systems are group standards of physical quantities, in particular time and frequency standards. The measuring systems implemented in such standards, as a rule, are based on use of the "each with reference" scheme and are underdetermined systems since they are described by a system of the linear equations with an incomplete observation matrix.

Typical examples of dynamic stochastic systems are group standards of physical quantities, in particular time and frequency standards. The measuring systems implemented in such standards, as a rule, are based on use of the "each with reference" scheme and are underdetermined systems since they are described by a system of the linear equations with an incomplete observation matrix.

The solution of tasks in the field of defenses and safety of the country, in radar, in space geodesy and cartography, in automated control systems for various devices in energy industry, on transport, etc. directly depends on the accuracy of measurement of time and frequency. Besides, the role of time-and-frequency measurements in fundamental researches has increased significantly.

The predictive models based on use of properties stationary (difference-stationary) stochastic processes, in the systems of processing of time-and-frequency information began to appear since the 1970s. A great contribution to the study of such systems made Bauch A., Breakiron L. A., Greenhall C. A. [1, 2], Jones R. H. [3], Tryon P. V. [3], Panfilo G.[4], Percival D. B. [5], Filimonov B. P., Khrustalev

Yu. P. [6], Tolstikov A. S. According to the results of researches executed in various organizations, use of the predictive models allows to increase the accuracy of state estimation of standards of time and frequency (ensemble of atomic clock) by 20–30% [6, 7, 8].

Practically all modern algorithms of estimation of a state vector of ensemble of the atomic clock relying on use of the predictive models use linear models. The prediction of frequency value of clock is calculated based on division of the frequency change process into two components: determined – a linear frequency trend and stochastic – described by the first-order moving average process [9, 10, 11].

Authors of article developed an algorithm [6, 8, 12], the using more general class of mathematical models: the models describing the higher-order deterministic trends (nonlinear frequency trends); dynamic stochastic models - autoregressive moving average models (ARMA(p, q)), where p is the autoregression order, q is the order of the moving-average.

The widespread use of algorithms that use the ARMA model in the activities of time services is hampered by the absence of a formalized technique of creation such models on experimental data, due to the following reasons:

- Lack of initial time series (empirical data) of models, necessary for creation of ARMA models, since the measurements which are carried out in atomic clock – indirect (measured by the frequency differences between the reference generator and all other generators included in the atomic clock).
- The creation of the predictive models is possible only for stationary processes. At practice in the time series of estimates of relative deviations of frequency of the hydrogen standards, which form the basis for secondary and working frequency and time standards in Russia, there are non-stationary components: anomalous measurements, jumps of frequency, the deterministic polynomial trends.
- The current methods of creation of APMA models, in particular, the most widely used and implemented in various application packages, for example, in STATISTICA [13], the Box-Jenkins method [14] cannot be fully formalized, have interactive character and demand involvement of the highly qualified specialists having wide experience in creation of ARMA models.

Work purpose: to increase the accuracy of state estimation of time and frequency standards based on the use of the predictive models describing the processes of frequency change of the hydrogen standards, which are a part of group frequency and time standard.

Achievement of this goal is associated with the solution of the following tasks:

- To develop a method of increase in efficiency of estimates of the state vector of the time and frequency standard, i.e. to reduce the bias of estimates and to increase their noise immunity.
- To develop the formalized technique of creation of the predictive ARMA models.
- To develop a technique of identification of non-stationary components which can be present in the time series, representing realization of processes of frequency change of hydrogen standards.
- To create the software realizing the developed techniques and to experimentally investigate quality of state estimations of a group standard got by means of the developed package of application programs

II. PROBLEM STATEMENT OF STATE ESTIMATION OF DYNAMIC STOCHASTIC SYSTEMS

As group standards function continuously, and processes of change of frequency of generators are accidental, it is possible to consider standards as dynamic stochastic systems, and the problem of processing of measured data, to consider as a problem of estimating the state vector of the dynamic systems from the results of the measurements.

For increase in accuracy and reliability of reproduction of units of time and frequency, the atomic clock is realized in the form of group standards, which include n quantum standards. In modern standards, hydrogen standards of frequency of the new generation are used as keepers of the time scale and stable frequency sources, the frequency instability of which on daily intervals of averaging is $3 \cdot 10^{-16}$ [15]. One of them, as a rule, possessing the best metrological characteristics, is chosen as a reference. In the course of functioning of a standard periodic measurements of a difference of frequencies of the reference generator with each of other $n-1$ of generators are made (actually the differences of phases of periodic signals have been measured, however taking the first differences from the received empirical series of phase measurements allows to pass to series of relative frequencies differences easily [10]). Based on these results, corrections to the readings of the reference clock are formed.

Data processing is carried out in two modes: static (the mode of accumulation of data), dynamic (at the rate of arrival of measurement information).

Let's understand as the state of the standard at the moment of time s ($s = 1, 2, \dots, N$) y_i^s – values of a relative deviation of frequency of the i th generator from the attributed value of the frequency for all n generators which are a part of a group standard ($(i = 1, 2, \dots, n)$; N is the length of the time series which is used at a stage of static data processing). The

measurement results z_i^s are the frequency differences of the reference and i th generators. In the work we will consider data processing, received on daily intervals, we neglect the measurement errors due to their smallness [10].

The matrix equation describing the measuring system of the group standard has the form

$$\mathbf{Z}^s = \mathbf{H}^s \cdot \mathbf{Y}^s \quad (1)$$

where $\mathbf{H}^s = \begin{bmatrix} 1 & -1 & 0 & K & 0 \\ 1 & 0 & -1 & K & 0 \\ K & K & K & K & K \\ 1 & 0 & 0 & K & -1 \end{bmatrix}$ is the matrix of

measurements for the case when we consider the reference generator as number one, which does not change the generality of reasoning; $\mathbf{Y}^{sT} = [y_1^s, y_2^s, \dots, y_n^s]$ is the state vector of standard; $\mathbf{Z}^{sT} = [z_1^s, z_2^s, \dots, z_{n-1}^s]$ is the vector of measurements.

Since the number of results of measurements is one less than the number of generators of the group standard, the rank of the \mathbf{H}^s matrix is $n - 1$, and the system is underdetermined.

It is convenient to introduce an virtual measurement into \mathbf{Z}^s vector: $z_1^s = y_1^s - y_1^s = 0$. An additional (top) row consisting of zeros is entered into the \mathbf{H}^s matrix. The matrix rank in this case does not change. Thus, we deal with the linear underdetermined dynamic system the problem of state estimation which consists in finding of the solution of the matrix equation (1). In the general case, system (1) has the infinite number of solutions.

We impose an additional restriction on the vector of the obtained solutions $\hat{\mathbf{Y}}$ in order to obtain the only decision: the norm of vector $\hat{\mathbf{Y}}$ has to be minimum. The vector $\hat{\mathbf{Y}}$ is found by means of a pseudoinverse matrix \mathbf{H}^+ in this case. If at data processing, received on daily intervals, the measurements error can be neglected, it is enough to find assessment of one component of the vector $\hat{\mathbf{Y}}$ (as a rule, it is frequency estimate of the reference generator). The estimate of the relative deviation of the reference generator coincides with arithmetic mean of the results of the measurements

$$\hat{y}_1^s = \frac{1}{n} \sum_{i=1}^n z_i^s \quad (2)$$

This estimate will be a basis for finding of preliminary estimates \hat{y}_i^s which we will receive directly from the equations of a measuring system $z_i^s = y_1^s - y_i^s$, by substitution instead of the value y_1^s of its estimates found by formula (2).

Estimates of the relative deviation of the frequency of the reference generator \hat{y}_1^s , and, consequently, estimates \hat{y}_i^s , are estimates of the method of least squares. They are not noiseproof and have a bias as all least squares method estimations.

The method of increase in efficiency of estimates of the state of time and frequency standards of based on combination α -truncated and jackknife estimates is offered in article [16]. We experimentally showed that the combined α -truncated jackknife estimates allow to reduce the bias of estimates and to filter outliers (to 20%) simultaneously. It is possible to apply these estimates if a group standard includes at least five generators. At present in practice this requirement is not fulfilled, therefore, in general, we propose to improve the quality of estimates using the dynamic properties of the object – the group standard of time and frequency.

III. STATE ESTIMATION OF STANDARD OF TIME AND FREQUENCY TAKING INTO ACCOUNT DYNAMIC PROPERTIES

Kalman obtained the general solution of the task of obtaining optimal estimates of the state vector of dynamic systems based on the use of dynamic properties (Kalman filter). One of the filter options proposed in [6] is an algorithm that uses predictions of generator frequency values, calculated based on the ARMA models, constructed at a stage of static data processing. It is proposed to take into account the dynamic properties of an object when calculating estimates using predictive models. In this case, the frequency estimate of the reference generator will be found as [6]

$$\hat{y}_1^s = \sum_{i=1}^n g_i \left(z_i^s + \hat{y}_i^s(1) \right) \quad (3)$$

where $\hat{y}_i^s(1)$ is the prediction of the frequency of the i th generator, calculated on the preceding cycle, obtained on the basis of mathematical models constructed on empirical time series; $g_i=1/n$ is the weight if the i th measurement.

The errors of estimates (2) and (3) respectively are

$$\varepsilon_1 \left(\hat{y}_1^s \right) = \frac{1}{n} \sum_{i=1}^n y_i^s \quad (4)$$

$$\varepsilon_2 \left(\hat{y}_1^s \right) = \frac{1}{n} \sum_{i=1}^n \left(y_i^s - \hat{y}_i^s(1) \right) \quad (5)$$

The estimate (5) is unbiased, and estimate (4) is asymptotically unbiased.

Thus, it is necessary to consider expedient in data processing systems obtained in the course of the functioning of group standards, to rely on state estimation algorithms using the dynamic properties of frequency variation processes.

Methods of construction of ARMA models based on the separation of procedures for identifying the structure of models and fitting of parameters of model (the structure of

which is previously defined) to empirical time series are widely used and are implemented in software currently [13]. Such approach demands participation of the researcher having serious experience in field of the statistical analysis of time series, and is not amenable to formalization. The point is that the problem of identification of structure of model is solved in the Box-Jenkins technique on the basis of the visual analysis of empirical autocorrelation and partial autocorrelation functions and their "similarity" to theoretical analogs of these functions for this or that structure of random processes [14].

We offer alternative approach, which consists in refusal of interactive procedures, namely, from a stage of identification of structure of model and use of a method of simple search of all possible structures of the predictive models [12].

IV. PROPOSED TECHNIQUE OF CONSTRUCTING PREDICTIVE MODELS

The offered procedure of static data processing consists of the following stages:

- Filtering outliers from measurement series.
- Obtaining initial estimates of values of frequency of the hydrogen standards which are a part of a group standard.
- Construction of ARMA models by stationary components of time series with simultaneous obtaining of refined estimates of the state vector of the standard.
- Checking of adequacy of the received models, based on the analysis of time series of residuals.

A. Outliers filtration

For filtration of outliers in the measurement series the α -truncated estimators with unknown percent of anomalous measurements are used. The traditional procedure of rejection of outliers is complemented with introduction on each step of the procedure of check of a hypothesis: whether the variance of the remaining part of the sample decreased significantly after the next pair of supposed outliers was removed. Process stops if the sample variance practically does not change (all outliers are filtered).

B. Formation of series of preliminary estimates

On the basis of the time series of measurements obtained as a result of filtration, we form series of preliminary estimates \hat{y}_i^s . The obtained preliminary estimates of the method of least squares are not suitable for construction ARMA models, since they may contain nonstationary components.

C. Extracting of the stationary component of the time series

Time series, describing the processes of frequency change of hydrogen standards may contain nonstationarities of two types: step-functions describing frequency jumps and deterministic (polynomial) trends.

1) *Technique of identification of step functions*

In this work it is offered to define jump as "outlier" in the series of first-order differences. The first step in identifying the step-function is to obtain a series of first-order differences, from which the robust estimate of the standard

deviation is found: $\hat{\sigma}_i = \frac{\text{med}\left\{\left|y_i^s - \text{med}\left(y_i^s\right)\right|\right\}}{0,6745}$, where

$\text{med}\left(y_i^s\right)$ operator of calculation of a median of a series y_i^s .

The value of a series of first-order differences is considered abnormal (i.e., the original series contains a "frequency jump") if it oversteps the bounds $\pm 6\hat{\sigma}_i$. A step function describing the jumps is constructed for each of the generators and is removed from the corresponding time series.

2) *Identification of polynomial trends*

After removing the step functions from the series of preliminary estimates, it is necessary to remove polynomial trends, if any. As a rule, there are no trends higher than the second order in the series describing the processes of changing the frequency of hydrogen generators [8, 10, 11].

Identification procedure of such trends is a task of the regression analysis, is rather fully described in various publications and does not require special clarification. For definition of an order of a polynomial (linear or quadratic) the procedure of comparison of residual variances is used. If the residual variance calculated for a second-order polynomial is significantly less than the residual variance of a first-order polynomial, a quadratic trend takes place. The corresponding component is removed from the initial time series describing the process of frequency change. The remainder of the time series represents stationary process.

D. *Formalized procedure of constructing predictive models based on empirical data*

The technique of creation of a system of automatic constructing of the predictive models is based on relatively small orders of autoregression and moving average. By results of the researches executed recently the maximum order of autoregression is three, order of moving averages is two [6, 8]. As a result number of possible structures which need to be analyzed in the course of automatic creation of models is eleven [12]. The procedure for fitting parameters of the ARMA models is based on minimization of the sum of squares of deviations of measurement results from their predictions. The authors proved a theorem [12] that allows us to consider a task of process of adjustment of parameters of model to empirical data as a problem of search of an unconditional extremum that it significantly reduces time of the solution of this task. For the solution of a problem of search of an unconditional extremum the method of conjugate gradients is used, which has a high speed.

As a result of work of an algorithm of creation of the predictive models for the studied series eleven models are formed, each of which corresponds to a specific structure of process. Models are ordered by increasing residual variance. By the Fisher criterion the hypothesis of equality of residual variances is checked. If for some group of models the

residual variances differ statistically not significantly, the simplest model is preferred. The simplest models include autoregression models and moving average models having the minimum order. The autoregression model is more preferable.

V. EXPERIMENTAL CHECK OF OPERABILITY OF THE OFFERED TECHNIQUE

The operability of the proposed technique was checked by methods of statistical modeling: 1. Stationary time series were generated. 2. The nonstationarities of the three types listed above was superimposed on the series 3. Times series of measurements that implement the "each with reference" scheme, were simulated. 4. The resulting series were processed in accordance with the described technique. We implemented the software module of extracting the stationary component in MathCAD 15.0 and the module of automatic creation of the predictive models in the Python 3.6. The quality of the realized algorithms and the offered techniques was evaluated by comparing the time series of estimates obtained as a result of the work and the "true" (generated) series.

A. *Extracting of the stationary component*

As an illustrative example, we present the results of a computational experiment for the identification of a stationary component from the series containing nonstationarity of all three types simultaneously. As the "true" series in the described experiment, we used the series generated on the autoregression models of first order. A trend was superimposed on each time series simulating the frequency change processes in four hydrogen generators that make up the group standard of time and frequency. The autoregression coefficients φ_i , the standard deviation of white noise σ_i and other parameters of models are described in Table 1.

We added no more than two percent of frequency jumps to each series, the amplitude and moments of occurrence of which are random and generated as random numbers from a sample subject to the normal law of probability distribution with zero mean and variance exceeding twenty times the maximum of the residual variance of stationary components. "True" series containing trends and frequency jumps are shown in Fig. 1b. Subtracting the values \hat{y}_i^s termwise from the series \hat{y}_1^s , we create series of "measurements". Similarly, the above procedure of generating frequency jumps (but with a variance in excess of a thousand times), adding three percent of the outliers we generate the "contaminated" measurements. The obtained series of measurements containing nonstationarities of all three types are shown in Figure 1b. The results of work of the stationary component extraction algorithms are presented in Fig. 1.

TABLE I. PARAMETERS OF THE INVESTIGATED TIME SERIES

No. of series	φ_i	Trend	σ_i
1	0.3	$y(t) = 0.01t + 0.3$	0.1
2	0.6	$y(t) = 0.001t^2 - 0.12t + 2$	0.2
3	-0.4	$y(t) = 0.03t - 1.0$	0.3
4	0.8	$y(t) = -0.0015t^2 + 0.1t - 2$	0.4

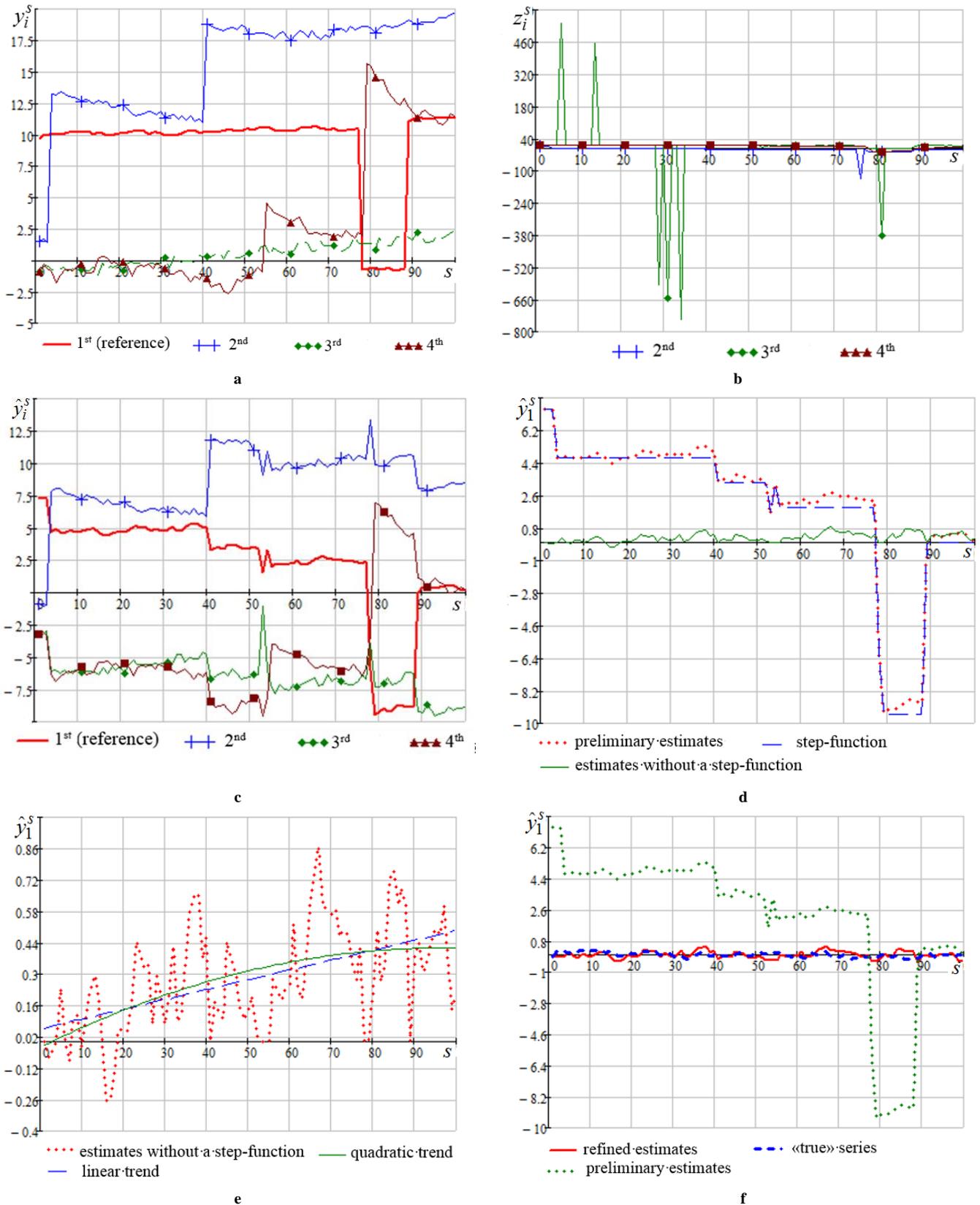


Fig. 1. The results of the extraction of the stationary component: a - "true" series with trends and frequency jumps; b - measurement series with outliers; c - series of preliminary estimates without outliers; d - preliminary estimates of the first (reference) generator (dotted line), step function (dashed line), a series of estimates without a step function (solid line); e - a series of estimates of the reference generator without a step function (dotted line), linear (solid line) and quadratic (dashed line) trends (significant linear regression); f - "true" series (dashed line), a series of initial estimates (dotted line), a stationary component (solid line) for the reference generator

As can be seen from the fig.1, the developed algorithm successfully filters outliers, allows to identify and eliminate step functions and polynomial trends from the series and to obtain the stationary components necessary for construction of predictive models at the following stage of static data processing.

B. Construction of autoregressive moving average models

Results of work of the program of automatic construction of the ARMA models are presented in table 2 where the structure and parameters of basic model, and the list of the predictive models with the corresponding indicators and estimates of their parameters are specified. The results are presented for three classes of processes: a second-order moving average process ($q = 2$); the process of autoregression of the third order ($p = 3$); mixed process of autoregression moving average ($p = 1, q = 1$). The resulting range of models corresponding to each of the basic models is ordered by increasing values of the residual variance.

From the results of the experiments, it follows that the models parameters obtained by the simple enumeration method, with the structure corresponding, structure of models on the basis of which the time series were generated, practically coincide with those specified during generation. For example, for model of moving average of the second order the received coefficients [$\theta_1=0.297, \theta_2=-0.133$] coincides with coefficients of initial model [$\theta_1=0.3, \theta_2=-0.1$]. It is necessary to keep in mind, however, that different dynamic stochastic models can correspond to one time series. Especially if the time series are realization of the mixed processes (i.e. process of autoregressive moving average). Therefore, when modeling data processing processes, one may encounter a situation when time series were generated with at some values of coefficients of autoregression and moving average, and in the process of automatic construction of models, the estimates of these parameters, which are a little differing from initial ones, are obtained. Such models are equivalent from the point of view of predictive ability (the parameters of these models lie on one level line, i.e. have the equal sums of squares of forecasts errors [14]). Hit in this or that point of lines of equal level depends on the choice of initial values of a vector of parameters. In the given experiments process began from a zero point as it belongs to the region of permissible values, i.e. meets requirements of stationarity and reversibility of models [14].

The values of the Fisher F-test show that the difference between the obtained models in terms of predictive ability, characterized by residual variance, is statistically insignificant (the hypothesis of equality of variances at a significance level of 0.05 is not rejected). Therefore, it is the most expedient to use model of autoregression of the first order for all ranks as the having the simplest structure among equivalent from the point of view of predicting abilities of models.

C. Checking the adequacy of the received models

Practically all approaches related to checking the adequacy of predictive models are based on the analysis of time series, which are forecast errors. In this case, the following requirements are imposed on the time series of residuals [14, 13, 17]: forecast errors have to submit to the

normal law of distribution of probabilities with zero mean; residuals (forecast errors) should not be correlated.

We checked adequacy of the received predictive models in STATISTICA [13].

TABLE II. THE RESULTS OF THE WORK OF THE SOFTWARE AUTOMATIC CONSTRUCTION OF PREDICTIVE MODELS

Model class	Estimates of coefficients	Estimates of residual variance	F-calculated	F-critical
ARMA(0,2), $N=1000, y_0 = 0.5, \sigma = 1, [\theta_1=0.3, \theta_2=-0.1]$				
(3, 1)	[-0.873, -0.135, 0.0829, -0.5805]	1.049	1.000	1.1099
(3, 2)	[-0.82, -0.075, 0.10, -0.527, 0.044]	1.050	1.001	1.1099
(3, 0)	[-0.297, 0.036, 0.077]	1.051	1.002	1.1099
(0, 2)	[0.297, -0.133]	1.052	1.002	1.1099
(2, 2)	[-0.198, -0.156, 0.099, -0.228]	1.052	1.003	1.1099
(1, 2)	[0.025, 0.321, -0.14]	1.053	1.003	1.1099
(1, 0)	[-0.300]	1.055	1.006	1.1098
(2, 1)	[0.396, 0.235, 0.688]	1.055	1.006	1.1099
(2, 0)	[-0.296, 0.013]	1.056	1.007	1.1099
(1, 1)	[-0.326, -0.029]	1.056	1.007	1.1099
(0, 1)	[0.263]	1.068	1.018	1.1098
ARMA(3,0), $N=1000, y_0 = 0.5, \sigma = 1, [\varphi_1=0.5, \varphi_2=0.3, \varphi_3=-0.2]$				
(3, 0)	[0.515, 0.294, -0.146]	1.003	1.000	1.110
(2, 2)	[0.305, 0.24, -0.209, -0.167]	1.004	1.000	1.110
(3, 1)	[0.586, 0.26, -0.162, 0.073]	1.005	1.001	1.110
(3, 2)	[0.422, 0.227, -0.056, -0.091, -0.121]	1.005	1.001	1.110
(1, 2)	[0.598, 0.080, -0.242]	1.007	1.003	1.110
(2, 1)	[0.126, 0.452, -0.372]	1.011	1.007	1.110
(2, 0)	[0.482, 0.224]	1.025	1.021	1.110
(1, 1)	[0.765, 0.232]	1.043	1.039	1.110
(0, 2)	[-0.479, -0.432]	1.075	1.071	1.110
(1, 0)	[0.621]	1.078	1.073	1.110
(0, 1)	[-0.403]	1.337	1.332	1.110
ARMA(1,1), $N=1000, y_0 = 0.5, \sigma = 1, [\varphi_1=0.5, \theta_1=0.2]$				
(2, 2)	[-0.374, 0.55, -0.656, 0.253]	0.930	1.000	1.110
(3, 2)	[-0.301, 0.585, -0.026, -0.584, 0.312]	0.931	1.001	1.110
(2, 0)	[0.280, 0.115]	0.932	1.003	1.110
(1, 1)	[0.607, 0.325]	0.933	1.003	1.110
(2, 1)	[0.431, 0.067, 0.152]	0.933	1.003	1.110
(1, 2)	[0.555, 0.277, -0.029]	0.933	1.004	1.110
(3, 0)	[0.279, 0.112, 0.010]	0.933	1.004	1.110
(3, 1)	[0.637, 0.010, -0.026, 0.359]	0.934	1.004	1.110
(0, 2)	[-0.276, -0.159]	0.941	1.012	1.110
(1, 0)	[0.317]	0.944	1.015	1.110
(0, 1)	[-0.249]	0.968	1.041	1.110

Autocorrelation and partial autocorrelation functions of the original generated series for model of moving average of the second order with coefficients $\theta_1 = 0.3$, $\theta_2 = -0.1$ (from Table 2) are presented in Fig. 2. A visual analysis of these plots confirms that the process of structural identification is largely subjective, as is the Box-Jenkins technique in general, since in the example given, the structure can be identified ambiguously (we can assume the structures ($p = 1; q = 0$) or ($p = 0; q = 2$)).

The parameters of the predictive models for this case and the estimation of residual variances (Table 3) obtained as a result of the regression analysis in STATISTICA absolutely coincide with the corresponding characteristics of the models obtained in the developed software (Table 2). This confirms the validity of the models received by means of the developed software of automatic construction of predictive models.

Testing the hypothesis about the normal distribution of the prediction error series is not rejected by any of the tests performed in STATISTICA for either the ARMA(1, 0) or ARMA(0, 2) models, which is confirmed by the values of various criteria from Table 3.

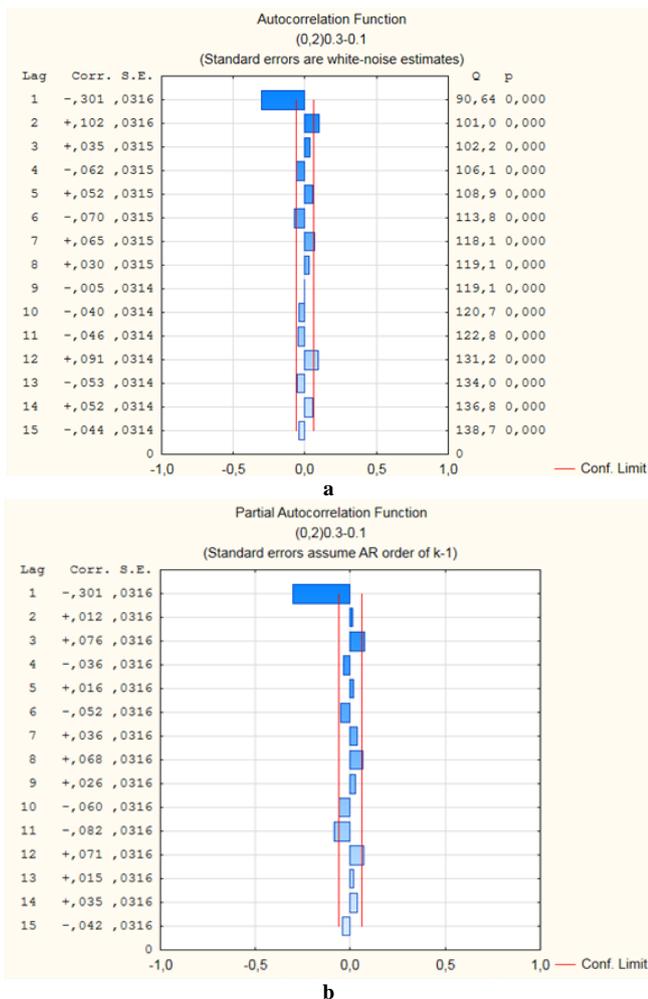


Fig. 2. Sample autocorrelation (a) and partial autocorrelation functions (b) of the second order moving average process

TABLE III. RESULTS OF THE REGRESSION ANALYSIS IN STATISTICA

Characteristics	ARMA(0, 2)	ARMA(1, 0)
Autoregression coefficient	-	$\phi_1 = -0.3$
Coefficients of moving average	$\theta_1 = 0.2973$ $\theta_2 = -0.1332$	-
Residual variance	1.0515	1.0550
Skewness	-0.0465	-0.0647
Kurtosis	-0.0173	-0.0077
Standard error of skewness	0.0773	0.0773
Standard error of kurtosis	-0.1545	-0.1545
p-value, by Shapiro-Wilk's W test ($p > 0.05$)	0.9091	0.9091
Chi-Square test	4.4331, $df=9$ (adjusted), $p=0.88067$	4.9771, $df=9$ (adjusted), $p=0.83630$
Kolmogorov-Smirnov	$d=0.01035$	$d=0.00797$
Box and Pierce test ($p > 0.05$)	40.49 < 42.56	40.49 < 41.34

We checked all initial models from Table 2 in a similar way. In each of the cases, the parameters of the models (the values of the autoregression coefficients and the moving average, the values of the residual variance) obtained with the help of the program module of constructing of models coincide with the parameters of the similar models built in STATISTICA. In addition, in all the above cases, the series of forecast errors obey the normal distribution law and are uncorrelated. In all cases, we can use the first-order autoregression model, as having the simplest structure among models equivalent in prediction ability.

The constructed models are a basis for obtaining the refined estimates of the relative deviations of the frequencies of the reference generator in the static mode, and are used at data processing in the dynamic mode.

The results of numerous computational experiments have confirmed the efficiency and operability of the methods of processing measurement information proposed above.

VI. CONCLUSION

The paper deals with the problems of constructing predictive models of technical systems with incomplete observation matrix. We consider group time and frequency standards as a typical example of such systems. Two classes of state estimation algorithms for such systems are considered: algorithms that use only the results of current measurements; algorithms based on the use of predictive models (taking account of the dynamics of the system). It is shown that the estimates of the state vector obtained by the algorithms of the second class are more accurate.

The paper describes a formalized technique of constructing predictive models of underdetermined systems. The proposed approach allows building predictive models of frequency change processes without the participation of highly qualified specialists in the field of time series analysis.

Methods of outliers filtering, identification of frequency jumps and deterministic trends have been developed and software realized, allowing extracting a stationary component of the studied random processes.

In this paper we consider the procedure of static data processing, since dynamic data processing is based on the use of predictive models built at the stage of static processing, and requires special consideration.

The operability of the offered technique and the developed software is confirmed by statistical modeling and processing of real data. We processed the data obtained in the course of the work of the secondary standard of time and frequency VET 1-5, functioning on the basis of the East Siberian branch of VNIIFTRI (Irkutsk) [18]. The results do not contradict the data obtained with use of results of external comparisons.

Because of the work, software was developed that allows automating the construction of predictive models. The use of predictive models in the evaluation systems of time and frequency standards (and other dynamic systems) can significantly reduce the algorithmic error of the measurement information processing. That allows increasing the accuracy of information-measuring systems.

REFERENCES

- [1] C. A. Greenhall, "Reduced Kalman filters for clock ensembles," IEEE International Frequency Control Symposium and Exposition. USA, pp. 1–5, May 2011.
- [2] M. Suess and C. A. Greenhall, "Combined covariance reductions for Kalman filter composite clocks," Metrologia, vol. 49, pp. 588–596, 2012.
- [3] P. V. Tryon and R. H. Jones, "Estimation of parameters in models cesium beam atomic clocks," Journal of Research of the National of Standards, vol. 88, no. 1, pp. 3–16, 1983.
- [4] G. Panfilo, A. Harmegnies and L. Tisserand, "A new prediction algorithm for the generation of International Atomic Time," Metrologia, vol. 49, pp. 49–56, 2012.
- [5] D. B. Percival and K. L. Senior, "A wavelet-based multiscale ensemble time-scale algorithm," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 59, no. 3, pp. 510–522, 2012.
- [6] Yu. P. Khrustalev, "Statistical and dynamic processing of data obtained when handling time and frequency standards," Measurement Techniques, vol. 47, no. 6, pp. 555–561, 2004.
- [7] K. G. Mishagin, S. D. Podogova, I. N. Chernyshev, S. Yu. Medvedev, Algorithm for generating the output signal of a group frequency reference, Measurement Techniques, vol. 56, no. 8, pp. 887–893, 2013.
- [8] Yu. P. Khrustalev, V. M. Akulov, A. A. Ippolitov and L. N. Kurysheva, "Processing data obtained as a result of reciprocal measuring of secondary standard of time and frequency," Proceedings of Irkutsk State Technical University, no. 7, pp. 22–28, 2012 ["Obrabotka dannyh poluchennyh po rezul'tatam vzaimnyh izmerenij vtorichnogo ehtalona vremeni i chastoty," Vestnik Irkutskogo gosudarstvennogo tehniceskogo universiteta, no. 7, pp. 22–28, 2012] (in Russian).
- [9] J. Levine, "The statistical modeling of atomic clocks and the design of timescales," Review of scientific instruments, vol. 83, 021101, pp.1–28, 2012.
- [10] I.V. Bezmenov and I.Yu. Blinov, Theoretical bases of building of models for the description of modern time scales and frequency standards. Mendeleevo: VNIIFTRI Publ., 2015 [Teoreticheskie osnovy postroeniya modelej dlya opisaniya sovremennyh shkal vremeni i standartov chastoty. Mendeleevo: VNIIFTRI, 2015] (In Russian).
- [11] A. V. Zderiglazova, V. M. Akulov and L. N. Kurysheva, "Processing the data obtained by the operation results of the system for secondary clock time scale formation," in Proceedings of Irkutsk State Technical University, no. 2, pp. 23–29, 2012 ["Obrabotka dannykh, poluchennykh po rezul'tatam raboty kompleksa formirovaniya shkal vremeni vtorichnogo ehtalona vremeni i chastoty," Vestnik Irkutskogo gosudarstvennogo tehniceskogo universiteta, no. 2, pp. 23–29, 2012] (In Russian).
- [12] Yu.P. Khrustalev and I.A. Serysheva, "Automation of dynamic stochastic model creation process," in Proceedings of Irkutsk State Technical University, no. 9, pp. 95–103, 2017 ["Avtomatizaciya processa postroeniya dinamicheskikh stokhasticheskikh modelei," Vestnik Irkutskogo gosudarstvennogo tehniceskogo universiteta, no. 9, pp. 95–103, 2017] (In Russian).
- [13] V. P. Borovikov, Popular introduction to modern data analysis in STATISTICA system. Moscow: Goryachaya liniya – Telekom Publ., 2013) [Populyarnoe vvedenie v sovremennyj analiz dannyh v sisteme STATISTICA. Moskva: Goryachaya liniya – Telekom, 2013]. (In Russian).
- [14] G. E. P. Boks, G. M. Jenkins, G. C. Reinsel and G. M. Ljung, Time series analysis: forecasting and control Hoboken: John Wiley & Sons, Inc., 2016.
- [15] A. A. Belyaev, N. A. Demidov, S. Yu. Medvedev, A. V. Pastukhov, B. A. Sakharov and I. Yu. Blinov, "Apparatuses of hydrogen keepers of standards of units of time and frequency," Vestnik metrologa, no. 2, pp. 14–23, 2015 ["Apparatura vodorodnyh hranitelej ehtalonov edinic vremeni i chastoty," Vestnik metrologa, no. 2, pp. 14–23, 2015] (In Russian).
- [16] Yu. P. Khrustalev and I. A. Serysheva, "Increasing of robustness of estimators of the state of time and frequency standards," Measurement Techniques, vol. 57, no 5, pp. 519–525, 2014.
- [17] R. R. Wilcox, Introduction to robust estimation and hypothesis testing. Amsterdam: Elsevier, 2012.
- [18] I. A. Serysheva and Yu. P. Khrustalev, "The formalized technique of processing measuring information obtained in the process of functioning group standard of time and frequency," in Information and mathematical technologies in science and management, no. 4, pp. 163–172, 2018 ["Formalizovannaya metodika obrabotki izmeritel'noj informacii, poluchaemoj v processe funkcionirovaniya gruppovogo ehtalona vremeni i chastoty," Informacionnye i matematicheskie tekhnologii v nauke i upravlenii, no. 4, pp. 163–172, 2018] (In Russian).