

# SYNTHESIS AND IDENTIFICATION OF HIDDEN MARKOV MODELS BASED ON A NOVEL STATISTICAL TECHNIQUE IN CONDITION MONITORING<sup>1</sup>

Lev S. Kuravsky, Sergey N. Baranov, Grigory A. Yuryev

Moscow State University of Psychology and Education, 29 Sretenka Street,  
127051 Moscow, Russia

Russian Aviation Co, Research Group, 33/50 Klara Tsetkin Street,  
125130 Moscow, Russia  
l.s.kuravsky@gmail.com

## Abstract

Under consideration are technologies for synthesis and identification of trained hidden Markov models based on a novel statistical technique and applied in condition monitoring. The approach can be used for both continuous- and discrete-time models of technical and non-technical systems. An initial rough model results from statistical multivariate analysis of observed data or their analysis by means of Kohonen self-organizing feature maps. Then it suffers proper corrections. Histograms of observed frequencies of being in different system states after the given exploitation periods are employed to train the constructed models. Free network parameters are identified by the chi-square minimum method. The synthesis may be carried out in case of some uncertainties: absence of complete information on system states and their connections is assumed. Undefined observed cases are partially identified by the classification propagation method with the aid of either self-organizing feature maps or cluster analysis. Connections between different system states are determined according to contiguity of their clusters or areas of winning units on the topological maps and statistical goodness-of-fit tests. In case of models represented by discrete-time structures (Markov chains) conversion to continuous-time trained structures (Markov networks) is performed to obtain more reliable results with the following return to discrete time and identification of transition probabilities. The methods presented are available for specialists responsible for maintenance of technical products. Obtained Markov models are useful for diagnostics, pattern recognition, structure service lives forecasting and establishing the justified routine maintenance. An example of synthesis and identification of Markov models representing fatigue failures of an aircraft air-intake panel is given for illustration.

---

**Keywords:** Markov networks, Markov chains, hidden Markov models, model synthesis, model identification, condition monitoring

## 1. INTRODUCTION

Jet engine noise, pressure fluctuations in turbulent boundary layer and other high-level acoustic loads result in fatigue failures of aircraft structures. These problems have been attracting attention of researchers since the 1950s after numerous acoustic fatigue failures of aircraft resulted from air-speed increase and changed engine types. In recent years, the interest to this problem has greatly

---

<sup>1</sup> This work was presented at the International Conference on Condition Monitoring and Machinery Failure Prevention Technologies in 2010. Published with permission of the British Institute of Non-Destructive Testing.

increased owing to a new generation of supersonic airplanes and hypersonic flight vehicles developed.

Acoustic loads are the most dangerous for thin-walled aircraft structures. These loads are wide-band (up to 5000 Hz) random process, with level varying from 145 dB to 170 dB in different points of aircraft surface.

In practice, destructiveness of these structures can be estimated by the changes of distributed structure stiffness, which is, in its turn, recognized by the qualitative changes of normalized spectral structure characteristics measured in checkpoints. Normalizing makes it possible to analyze only qualitative response and not to take into account the level of load. Such an approach based on the estimation of averaged structure properties seems more promising than the direct search for separate cracks<sup>[6,18]</sup>, which are not always accessible for direct observation and hard to forecast because of considerable dispersion in their evolution. The use of secondary characteristics (such as spectra) instead of time-domain realizations as check data is caused by the following facts:

- Spectra keep a sufficient amount of useful information about processes under study
- Spectra need less memory in digital representation
- Spectra may be easily and quickly computed with controlled accuracy.

Structure diagnostics and service life forecasting are two important technical problems to be solved during condition monitoring of aircraft structures suffered fatigue failures. Their solution is necessary for both planning maintenance works and estimating destruction level. The results obtained give the opportunity to cut down expenses for aircraft maintenance, raise the reliability of destruction monitoring, simplify the routine maintenance and derive its more justified schedules. Technology for monitoring of fatigue failures based on capabilities of neural<sup>2</sup> and discriminant networks was considered in papers and books<sup>[1,9-12,15]</sup>.

As it was shown in the book<sup>[5]</sup>, probabilistic models of cumulative damage are more preferable for forecasting in the discussed area than the deterministic ones. A relevant approach founded on Markov process theory was presented in works<sup>[1,9-10,12,14-15]</sup>.

According to this approach, forecasting is based on accumulated observations. Probability dynamics is described by continuous time, discrete state Markov processes. The given damage types are considered as separate discrete states in which the analyzed system has some probability to find itself. In due course transitions between the states are the case. It is assumed that state-to-state transitions (corresponding to each branch of the graph) meet the properties of Poisson flows of events. Time-domain dynamics of state probabilities is described by Kolmogorov set of ordinary differential equations.

Transition flow rates are free model parameters. Comparison of observed and expected histograms which present observed damage distributions at given time points, is used to identify them. The technique to find independent parameters as those minimizing a certain goodness-of-fit measure is used. According to the given approach, this measure is minimized at the specified time points, in which observed data are available. The employed procedure of parameter estimation as well as its software implementation were described in books<sup>[1,15]</sup>. In fact, an inverse problem is solved, viz.: coefficients of differential equations are determined by using the given solution characteristics. Obtained values of free parameters are considered as fatigue failure characteristics which have become apparent during observations.

It was shown that Markov models under the given inverse problem formulation may be regarded as specialized neural networks. The technique for their synthesis to select the most suitable network structure was developed on the base of statistical criteria<sup>[1,13,15]</sup>. Proposed in papers<sup>[12,14]</sup> was further development of this approach, the concept of multifactor Markov networks to represent subtle features of cumulative damage development and improve concordance between observed and predicted behavior. The papers mentioned present new features of these trained structures, including their 3-D applications, and show advantages of probabilistic predictions obtained with the aid of Markov networks to improve the damage diagnostics quality. Markov models in question make it

---

<sup>2</sup> Traditional neural networks with supervised learning (perceptrons, etc.) as well as Kohonen self-organizing feature maps were in use.

possible both to forecast probabilities of damage types and identify them on the base of observed system characteristics. Thereby they fix up typical condition monitoring problems solution for the structures under study.

Taking into account indirect observations for the system states, Markov networks under the given problem formulation can be considered as a special case of continuous-time hidden Markov models, which use measured spectral characteristics in checkpoints as observed parameters and are obviously regarded as more difficult instruments for practical applications than the traditional discrete time ones.

As for the structure service life forecasting in case of an identified Markov network, damage state probabilities as functions of time are obtained by means of integrating Kolmogorov set of differential equations. This operation is not difficult and is beyond the application field of hidden Markov models.

In its turn, diagnostics of the technical systems under consideration, which suffer damages of different types, is carried out in case of uncertainties when neither possible defect types nor connections between them are known beforehand. There is no information about the damage framework of such systems since:

- Some defects may not be identified easily and quickly (particularly, when it is difficult to get access to a structure part or find cracks because of their small size)
- Defects are usually not monitored directly and may be revealed by implication only via the system characteristics which are available for measurements (for example, spectra in checkpoints)
- Strict reasons for damage discrimination are frequently absent
- Experimental estimation of mutual connections between different defect types is a labor-intensive and long process, as a rule.

Two principal problems are usually to be solved to support structure diagnostics of the given ambiguous systems represented by the aforementioned probabilistic models:

- Given the parameters of Markov model in use and a particular sequence of observed spectral characteristics in structure checkpoints, find the sequence of different damages represented by model states, which is most likely occurred (identification problem)
- Given observed spectral characteristic sequences, find the most probable Markov model for representation of damage framework including the set of states, the structure of transitions between them and quantitative parameters of these transitions (synthesis problem).

Thus, the problem of constructing hidden Markov models best conformed with observations in case of uncertainties has been arisen. Presented in this paper are technologies for synthesis and identification of Markov models based on a novel statistical identification technique which can be applied for both continuous- and discrete-time models.

The given approach is intended for solution of both problems specified before. When solving the synthesis problem it determines a set of relevant states, their connections and optimal values of free model parameters for Markov networks. Some features of this technique were described in papers<sup>[12-13]</sup> and books<sup>[1, 15]</sup>. If discrete-time structures (Markov chains) fit an application problem better (for example, in case of discrete-time output of observed spectral characteristics with certain sampling interval), conversion from an initial continuous-time structure to the corresponding discrete-time one is carried out with the aid of a suggested special procedure.

Owing to universality and flexibility, the presented way of solution has obvious advantages over the methods used for maximum likelihood estimation of parameters for traditional hidden Markov models represented by Markov chains given a dataset of observed output is available. Until recently, no tractable algorithm has been known to solve such a problem, but for the Markov chains a local maximum likelihood can be efficiently derived by using Baum-Welch or Baldi-Chauvin algorithms<sup>[2,19]</sup>. It should be noted that the presented way of solution calculates not only optimal values of free model parameters like Baum-Welch and other algorithms but, in addition to this, a set of relevant states and their connections. It is significant that traditional algorithms for hidden Markov

models are unable to solve the applied technical problem considered here when uncertainties with damage types, model states and their connections are the case or the number of sampling points is too small.

Solution of the identification problem requires to find the most probable damage sequence over all the possible ones, with the probabilities of certain observed spectra belonging to given states in use. In case of Markov chains, the solution of interest can be found by Viterbi algorithm<sup>[20]</sup> but, taking into account a relatively low number of model states in many application problems, exhaustive search algorithms are also acceptable for solution in case of both continuous-time and discrete-time models.

In spite of the technology in question is considered here in application to a particular technical problem, the presented mathematical tools are not fastened to the given field of use and can be exploited for solution of many other technical and non-technical tasks. In particular, the conversion of discrete-time models to continuous-time models with their following identification and return to the initial structure is highly promising allowing for greater capabilities for model synthesis in continuous-time domain.

## 2. SOLUTION TECHNOLOGY

### 2.1. Synthesis problem

The synthesis problem solution includes the following stages:

- 1) Reduction of input data dimension to remove the redundant information
- 2) Determination of a set of states for Markov network via either Kohonen classification with the aid of Kohonen self-organizing feature maps (Kohonen maps) or multivariate clusterization procedure
- 3) Determination of initial distribution of connections between the states according to either contiguity of activated parts of Kohonen maps or multidimensional scaling
- 4) Adjustment of connections (removing redundant and statistically insignificant connections) by means of successive identification of transition flow rates by the chi-square minimum method
- 5) Final identification of network parameters by the chi-square minimum method.

#### 2.1.1. *Input data*

As is shown in paper<sup>[9]</sup>, it is convenient to estimate the level of fatigue failure for the structures suffered damages via response power spectral densities measured in checkpoints. These spectra are further considered as initial analyzed data. It is assumed that some representative sample of spectra resulted from either structure tests or observations during operation is available, with a damage type being specified for only certain of the spectra. Power spectral densities are regarded to be calculated with the accuracy sufficient for qualitative recognition<sup>3</sup>.

#### 2.1.2. *Reduction of input data dimension: Synthesis stage 1*

In case of observed data represented by spectral characteristics measured in structure checkpoints it is expedient to bring analyzed variables to conformity with frequency ranges taken with a certain step. Values of these variables are averaged normalized spectral characteristics for the corresponding frequency ranges. A large number of input variables, which is typical for this representation, may complicate the input sample analysis: in particular, the number of observed data necessary for qualitative network synthesis may become too great. Solution of this problem by worsening the spectral resolution capability is unacceptable as the useful information is lost. It is possible to improve the analysis quality only if the input data dimension is eliminated at the expense of redundant or insignificant information.

---

<sup>3</sup> For example, accuracy of digital spectral analysis is discussed in books<sup>[3,17]</sup>.

The ways of this problem solution are considered in paper<sup>[9]</sup> and books<sup>[1,15]</sup>. The method of principal components<sup>[16]</sup> is the most suitable among the permissible variants since it is the simplest and does not require any special assumptions on the initial data. The given approach makes the results of analysis independent of the spectrum evaluation features.

### ***2.1.3. Determination of a set of states for Markov network: Synthesis stage 2***

A set of Markov network states is determined by classifying the damage types: its own state corresponds to each type. As the damages for input cases are assumed to be partially known, it is necessary to propagate the given classification to the spectra which have no assigned damages to. Two alternative instruments can be used to determine the model states: Kohonen maps or multivariate clusterization.

The classification propagation method based on Kohonen maps contains the following steps:

- 1) Select Kohonen map<sup>[8]</sup>, where the number of elements exceeds considerably the expected number of damage types.
- 2) Train the network by using input cases for which damages are known. Mark the neurons of the topological map, which won for one damage type only. (Labels correspond to damage types.)
- 3) Input the remaining part of the sample to the network. Attribute the damage labels to the input cases, for which the marked neurons win.
- 4) If a subset of input cases, for which damages are known, is extended, then go to Step 2 else stop the computation.

If the sample of input cases still contains elements with unknown damages, the following actions are possible:

- Provide the network with a special state **Unknown** for unknown damages and attribute unclassified elements to this type (see Section 3);
- Undertake the structure inspection and classify the remaining elements of the sample (if their number is not too great).

It is expedient to carry out clusterization in four steps:

- 1) Tree clustering to get hierarchical tree plot and estimate reasonable number of clusters which are used later to form the model states<sup>4</sup> (in case of input data discussed before, the Ward's method with Euclidean distances is recommended)
- 2) K-means clustering for the number of clusters determined at the previous step to find out the content of each cluster
- 3) Assigning proper labels to clusters which contain the predominant number of marked elements of a given damage type
- 4) Meshing clusters with identical labels to higher-level clusters corresponding to a given damage type (optional).

### ***2.1.4. Determination of connections between the states of Markov network: Synthesis stage 3***

To determine initial distribution of connections between Markov network states, information on contiguity of activated parts of Kohonen maps or analysis of multidimensional scaling results can be used.

In the first case the associative properties of Kohonen maps are used, viz.: similar cases activate the groups of closely set neurons. Connections between the states are predicted according to contiguity of activated areas of the marked topological map. Each of these areas corresponds to a fixed damage type (Figure 4).

---

<sup>4</sup> It is recommended to determine this number of clusters as the number of hierarchical tree branch intersections by the cross-section through one of the longest alienation distance intervals in this tree. Selection of this interval is conditioned by availability of certain label dominance in each cluster.

The formal rule to find out contiguity of the areas is not unique. One of the convenient ways is to count minimal distances between pairs of marked neurons in Manhattan metric. Used for comparison are minimal distances between two differently marked neurons:

$$M = \min_{L(\alpha) \neq L(\beta)} (|i_\alpha - i_\beta| + |j_\alpha - j_\beta|),$$

where  $L(\alpha)$  and  $L(\beta)$  are labels of neurons with the  $\alpha$  and  $\beta$  indices; indices  $\alpha$  and  $\beta$  take all the values meeting the indicated condition;  $i_\alpha$  and  $i_\beta$  are numbers of rows, in which the neurons with indices  $\alpha$  and  $\beta$  are disposed at the topological map;  $j_\alpha$  and  $j_\beta$  are corresponding numbers of columns. Those and only those pairs of states  $p$  and  $q$  are supposed to be adjacent, for which the following condition is the case:

$$\min_{\substack{L(\alpha)=p \\ L(\beta)=q}} (|i_\alpha - i_\beta| + |j_\alpha - j_\beta|) \leq M + k,$$

where indices  $\alpha$  and  $\beta$  take on all the values that meet the indicated conditions;  $k \geq 0$  is the integer-valued parameter (in standard conditions  $k$  may be equal to zero).

Multidimensional scaling is applied to a matrix of Euclidean distances between the centers of clusters which correspond to different damage types and were determined at the previous step. As a result, these centers are usually inserted into two- or three-dimensional space. Possibility of transitions between the model states is derived from the values of distances between them. Thereby, the hypothesis of transition possibility dependence on distance in the obtained space is used.

When forming an initial Markov model structure, each state is connected with the states which centers are at the minimal distance from its center or at the distances exceeding the minimal one less than the given percent (in practice the 50%-limit may be recommended). The cluster center corresponding to the normal system operation is represented by the initial state. Directions of transitions along state connections are determined according to proximity to the initial state, viz.: transition from one state to another is possible if they are connected and the minimal number of transitions from the initial state to the first one is less than the analogous/similar number of transitions to the second one. If direction is ambiguous, both variants of transitions are attributed to the corresponding connection.

The structure obtained as a result of any given procedure takes into account redundant and statistically insignificant connections, which are revealed and removed at the next stage of calculations.

### 2.1.5. Adjustment of connections: Synthesis stage 4

In training the network, its free parameters are selected to get the best correspondence between observed and expected histograms of being in the system states at the given time points. Pearson statistic

$$\sum_{k=0}^n \frac{(F_k - p_k N)^2}{p_k N}$$

is used as a goodness-of-fit measure, where  $F_k$  is observed frequency of getting into the  $k$ -th system state at a given time point,  $p_k N$  is corresponding expected frequency,  $p_k$  is expected probability of being in the  $k$ -th state,  $N$  is the sample size,  $n$  is the number of model states. The smaller the value of this statistic the better the correspondence between observed and expected results. Under some conditions<sup>[7]</sup>, it is distributed asymptotically according to a chi-square distribution. Expected state probabilities are obtained by means of numerical integrating Kolmogorov set of differential equations.

The given technique of identifying free parameters is called the chi-square minimum method. For the problems under consideration, it yields estimations which are close to the ones obtained

by the maximum likelihood method<sup>[7]</sup>. It was proved that, under some general conditions, Pearson statistic values derived by substituting the indicated solutions were described asymptotically by the chi-square distribution with  $n-s-1$  degrees of freedom, where  $s$  is the number of free parameters to be determined, with the calculated values of free parameters converging in probability to the desired solution when increasing the sample size<sup>[7]</sup>. These facts allow to use the given statistic for testing the hypothesis of agreement between observation data and Markov network forecast.

The procedure for finding optimal values of free parameters, which is based on electronic spreadsheet capabilities, is considered in works<sup>[1,9,15]</sup>. It was also shown there how to optimize the model, using statistical goodness-of-fit tests.

Adjustment of connections in Markov networks (viz.: removing indirect and statistically insignificant connections) is carried out on the basis of the same approach. A network, in which all the transition flow rates are free parameters, will be called the complete network. Networks, in which some conditions are imposed on free parameters (for example, the conditions of equality to zero), will be called the reduced networks. The hypothesis of agreement between obtained observation data and a complete Markov network forecast will be denoted as  $H_c$ .

If there are no reasons to reject hypothesis  $H_c$ , each network connection tests of the significance are carried out in turn. For this purpose, free parameters are evaluated by the chi-square minimum method under the condition of flow rate equality, which corresponds to the connection tested, to zero. Obtained value  $X_r^2$  of Pearson statistic is compared with the analogous characteristic  $X_c^2$  for the complete network. Since difference  $X_r^2 - X_c^2$  is asymptotically distributed as a chi-square with 1 degree of freedom<sup>[4]</sup>, this statistic is used for testing null hypothesis  $H_r$  of the agreement between the observation data and the reduced network forecast against the alternative hypothesis  $H_c$ . If hypothesis  $H_r$  is not rejected at the given significance level, the corresponding connection is marked as a candidate for withdrawal from the network.

It is expedient to refine the network by iterations, using a sufficiently great value for the significance level, at which a connection is removed ( $p=0.5$  is recommended), and a small value for the significance level at which a connection remains in the network ( $p=0.01$  is recommended). The result obtained after removing connections at the current stage of calculations is considered as a complete network at the following iteration. Under such conditions, only *a fortiori* redundant connections are withdrawn from the network after each iteration ( $p>0.5$ ). Solution with regard to questionable connections ( $0.01 \leq p \leq 0.5$ ) is postponed. As a rule, significance of questionable connections is clarified step-by-step. This fact is caused by simplified network under study.

### 2.1.6. Final identification of network parameters: Synthesis stage 5

At this stage, free parameters of the obtained network are identified by using the chi-square minimum method. The goodness-of-fit for observed data and forecast is estimated by the minimal value of Pearson statistic. Number of degrees of freedom for the chi-square goodness-of-fit test is determined as a difference between the number of independent observed statistics  $F_k$  and the number of parameters to be found.

## 2.2. Identification problem

The number of sampling instants  $q$ , at which spectral characteristics in structure checkpoints are estimated, should be selected beforehand for the identification problem solution. The identification procedure includes the following steps:

- 1) Determination of all the model state sequences  $S_i = \{s_{i1}, \dots, s_{iq}\}$ ,  $i=1, \dots, r$ , which are possible for developing a system under study during a given number of sampling instants
- 2) Estimation of occurrence probabilities  $P(S_i)$  for every sequence  $S_i$ ,  $i=1, \dots, r$  indicated at the previous step via the product of probabilities of transitions between the model states during intervals limited by the given sampling instants in the following way:

$P(S_i) = \prod_{u=1}^{q-1} p_{s_i, i, u, u+1}$ , where

$p_{s_i, u, u+1}$  is probability of transition from state  $s_{iu}$  occupied at instant  $u$  to state  $s_{i, u+1}$  occupied at instant  $u+1$ . These transition probabilities can be calculated by using either Markov chains or Markov networks<sup>5</sup>

3) Estimation of occurrence probabilities  $P_{iz}$  of observed sequence  $Z = \{z_1, \dots, z_q\}$  of spectral characteristics in structure checkpoints for state sequences  $S_i, i=1, \dots, r$ , in the following way:

$$P_{iz} = P(S_i) \prod_{j=1}^q p_{z_i, j}$$
, where  $p_{z_i, j}$  is probability of acquisition of observed characteristic  $z_j$  being at state  $s_{ij}$ <sup>6</sup>.

4) Selection of the most probable state sequence  $S_{max} \in \{S_i\}_{i=1, \dots, r}$  corresponding to the maximal probability  $P_{z, max} = \max_i \{P_{iz}\}_{i=1, \dots, r}$ .

When sampling estimations of distribution parameters for the distances of observed data corresponding to each model state from state centroids<sup>7</sup> of these data are available, it is expedient to replace quantities  $p_{z_i, j}$  by probability density functions  $f_{i, j}(x_j)$  of Euclidean distances  $x_j$  of characteristics  $z_j$  from the observed data centroids for corresponding states  $s_{ij}$ <sup>8</sup>. It is the distance parameter that determines estimations of occurrence probabilities for observed characteristics in this case.

It is necessary to note that Markov chains are more convenient for identification than Markov networks since they ensure simpler calculations. In its turn, Markov networks are more convenient for synthesis of hidden models owing to their greater flexibility and capabilities of problem solution when only a few observation instants are available and model states are unknown a priori.

### 2.3. Conversions from continuous-time models to discrete-time models

Conversions from a continuous-time model to the corresponding discrete-time one and vice versa are carried out if this is expedient for the application problem solution. Conversion to a continuous-time model is in fact determined by a proper mapping of a sequence of Markov chain discrete steps onto Markov network continuous-time scale. In case of inverse conversion, the following procedure is suggested:

- 1) Selection of a sampling interval  $\tau$  on the continuous-time scale, which corresponds to the discrete step of Markov chain
- 2) Integration of Kolmogorov set of differential equations describing the state probability dynamics in an initial Markov network to calculate, as a result, the sequence of probability distribution vectors  $P_i (i=0, 1, \dots, l)$  corresponding to a given sequence of  $l+1$  discrete time points taken with the selected sampling interval  $\tau$
- 3) Construction of transition probability matrix  $A = ||a_{ij}||$  of order  $n \times n$ , where  $n$  is the number of states of Markov chain under consideration which represent this chain dynamics with the aid of free variables denoting transition probabilities and, if necessary, analytical expressions assembled from these variables
- 4) Numerical solution of multivariate optimization problem with the following conditions to estimate required values of free variables which matrix  $A$  are composed of:

<sup>5</sup> In case of Markov chains such a probability equals to the probability of transition between the indicated states. In case of Markov networks it is determined by integrating Kolmogorov set of differential equations, with the previous state at the previous instant having unity probability and the other ones – zero probabilities.

<sup>6</sup> It is expedient to calculate distances  $x_j$  for representation of input data with reduced dimension (see Step 1 of the synthesis problem solution) to avoid influence of redundant information.

<sup>7</sup> These distributions are assumed to be normal according to the central limit theorem of the probability theory and the structure of Euclidean distances.

<sup>8</sup> In case of normal distance distribution, positive arguments are used for calculating the mentioned probability density functions.



- least squares criterion to be minimized is  $\sum_{i=0}^{l-1} (\mathbf{P}_{i+1} - \mathbf{A}\mathbf{P}_i)^T (\mathbf{P}_{i+1} - \mathbf{A}\mathbf{P}_i)$
- $0 \leq a_{ij} \leq 1$  for all  $i, j$
- $\sum_{i=1}^n a_{ij} = 1$  for all  $j$ .

Obtained values of the specified free variables yield the sought transition probability matrix  $\mathbf{A}$  that determines state changes of Markov chain in question:  $\mathbf{P}_{i+1} = \mathbf{A}\mathbf{P}_i$ ,  $i=0, 1, \dots, l$ . Statistical significance of relevant residual error can be evaluated by means of obviously modified Pearson statistic (see Section 2.1.5).

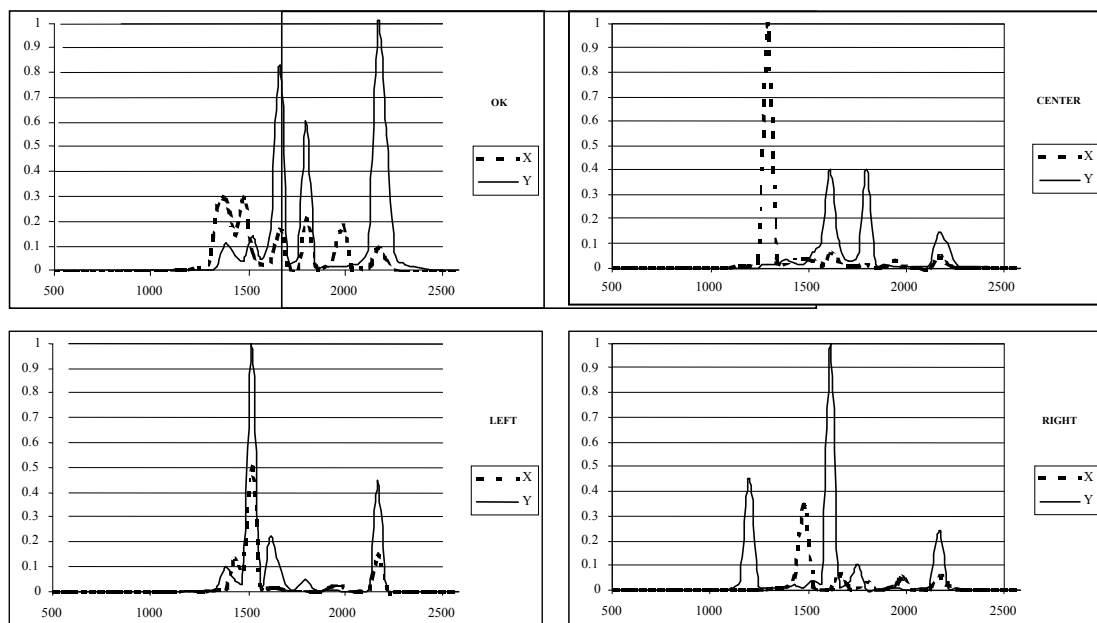
### 3. APPLICATION EXAMPLE

The proposed technique features are shown by the example of synthesis and identification of Markov models representing fatigue failures of an aircraft air-intake panel. Its characteristics and damage types are presented in works<sup>[1,10]</sup>.

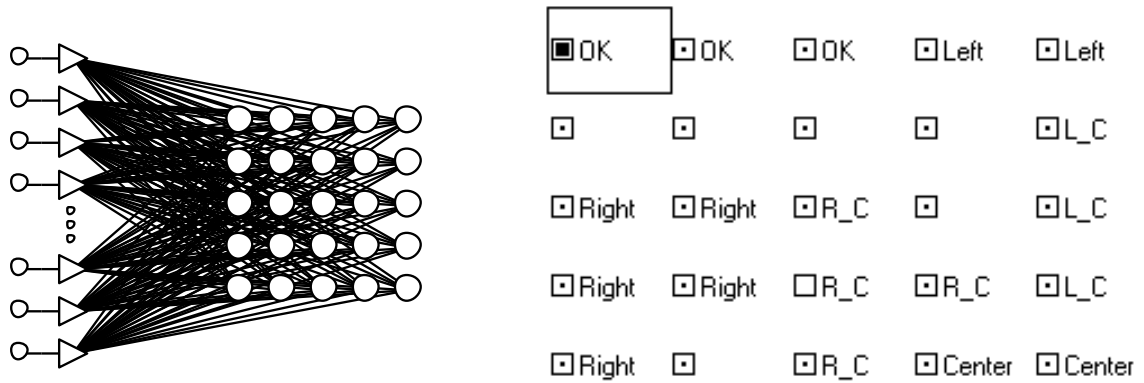
Observation data were represented by a sample of 144 pairs of stress power spectral densities measured at two checkpoints in the range from 500 to 2500 Hz. Some patterns of these characteristics are given in Figure 1. Damage types were supposed to be known beforehand for 25 spectral pairs only. Spectra for the test sample of air-intakes were estimated after 1000 and 2000 exploitation hours. Six damage types were identified. Each observed case contained average values of normalized spectral densities in 45 frequency ranges of 45-Hz width. Thus, the initial set of data to be analyzed included 90 input variables.

After reducing the problem dimension by the principal component method, the number of input data was diminished up to 10 variables, which were used in the following analysis (10 first principal components explained more than 80% of observed variance).

The method of classification propagation made it possible to expand the identified part of the sample up to 115 elements by two iterations. The Kohonen map composed of 25 units was used (Figure 2). For five pairs of spectra, the damage types were determined by the additional inspection. 24 pairs of spectra remained unclassified. The Kohonen maps work was simulated with the aid of the STATISTICA Neural Networks software package.

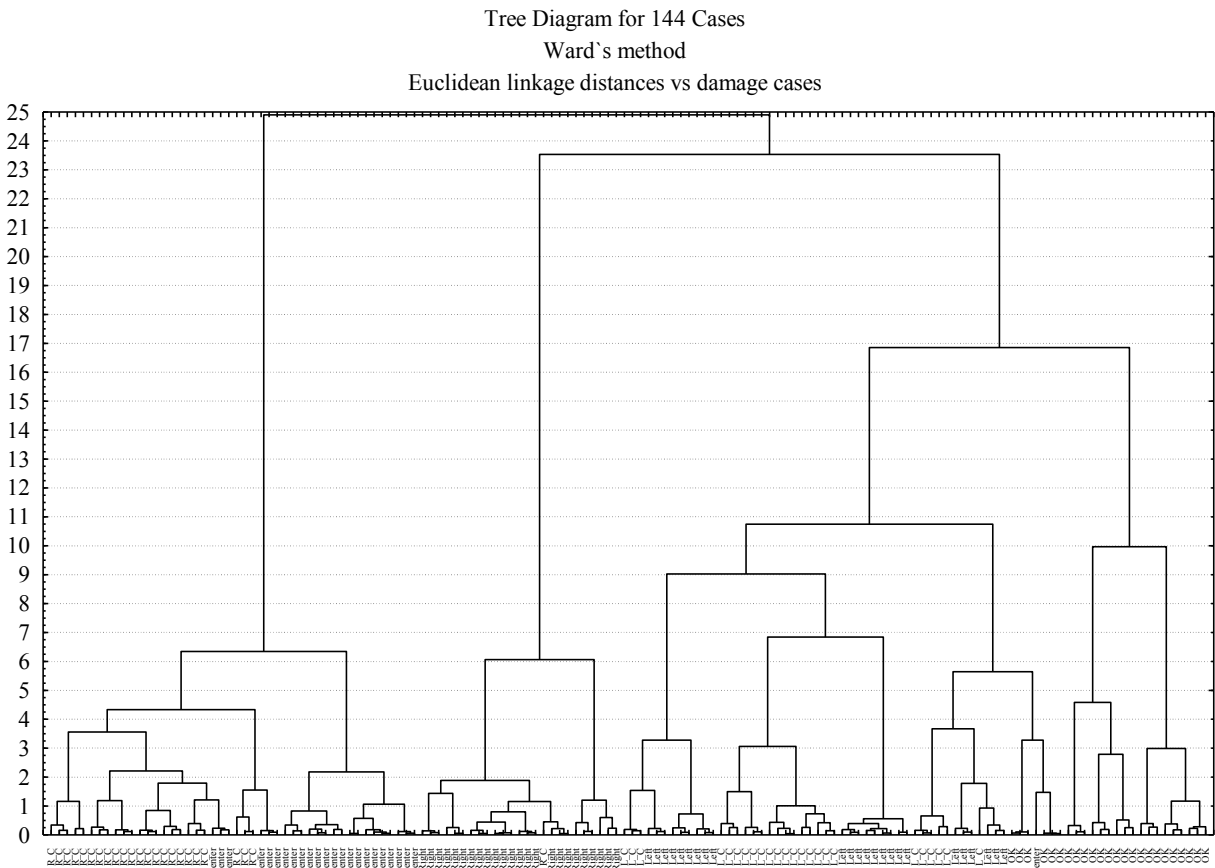


**Figure 1.** Patterns of normalized stress spectra for the recognized structure states at checkpoints X and Y.



**Figure 2.** Kohonen map (5x5 units) for classification propagation. The total view and marked topological map at the last algorithm step are shown.

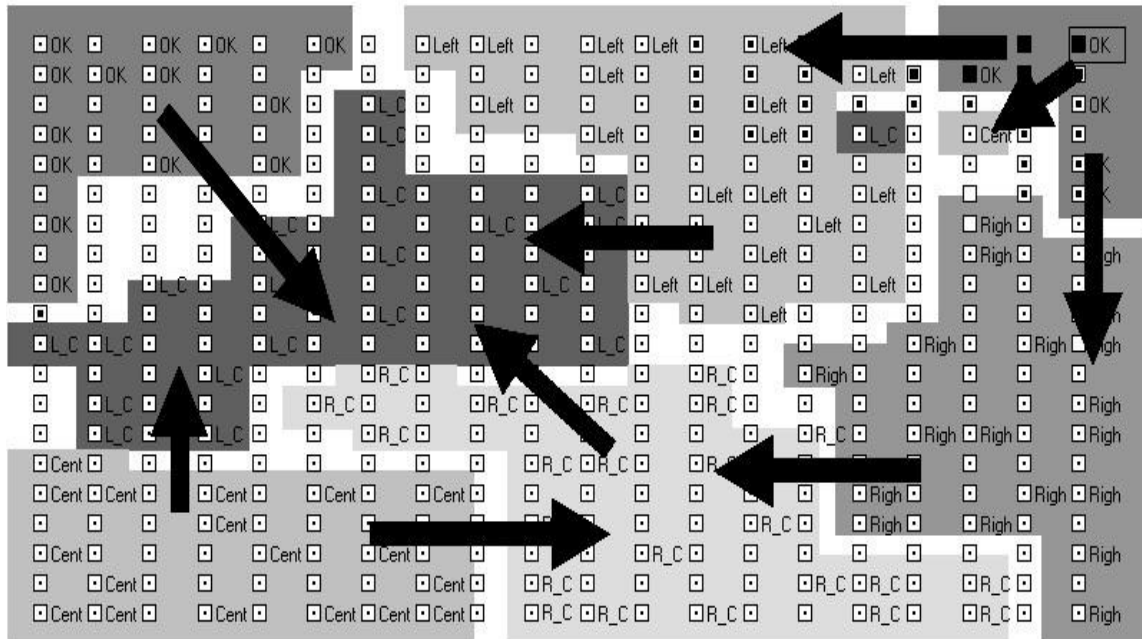
The **OK**, **Center**, **Left**, **Right**, **L\_C** (**Left+Center**), **R\_C** (**Right+Center**) states, which correspond to identified damages, as well as the state **Unknown** for unknown damages, proved to be necessary for Markov network. To simplify results interpretation, network construction is carried out, firstly, for the six recognizable states, to which the **Unknown** state is then added.



**Figure 3.** Hierarchical tree plot obtained after tree clustering with the aid of Ward method with Euclidean distances.

Multivariate clusterization procedure yielded the hierarchical tree plot presented in Figure 3<sup>9</sup>. Taking into account the comments given in Section 2.1.3 reasonable initial clusters forming the model states were estimated with the aid of a tree cross-section at the 5-unit distance. After refining cluster contents via K-means clustering, assigning predominant labels to obtained sets and meshing clusters with identical labels, six higher-level clusters corresponding to the aforementioned damage types (**OK**, **Center**, **Left**, **Right**, **L\_C** and **R\_C**) were derived. Slight mismatches in distributions of cases over damage clusters are permissible if they do not cause qualitative distortions in a Markov network set of states and their reciprocal connections.

Structure of connections between the recognizable states of Markov network was determined with the aid of Kohonen map consisting of 400 units (Figure 4).



**Figure 4.** The marked topological map (20×20 units) used to determine connections between Markov network states. Revealed connections are denoted by arrows.

The shortest distances between neurons of the marked areas are presented in Table 1.

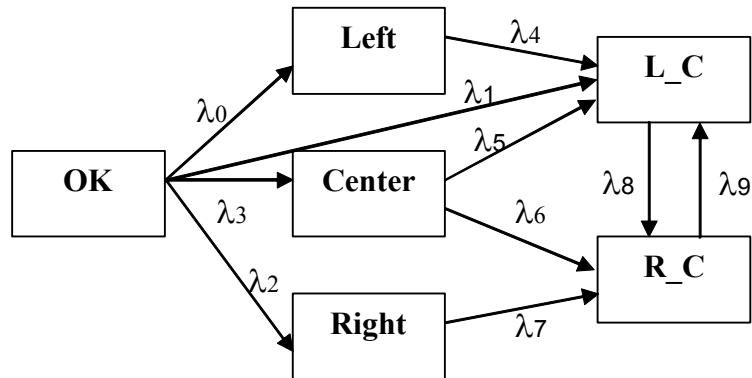
**Table 1.** The shortest distances between neurons of the marked areas of the topological map ( $M=2$ ).

	<b>OK</b>	<b>Center</b>	<b>Right</b>	<b>Left</b>	<b>R_C</b>	<b>L_C</b>
<b>OK</b>	0	2	2	2	9	2
<b>Center</b>	2	0	3	4	2	2
<b>Right</b>	2	3	0	3	2	5
<b>Left</b>	2	4	3	0	3	2
<b>R_C</b>	9	2	2	3	0	2
<b>L_C</b>	2	2	5	2	2	0

Pairs of states, for which the shortest distance was equal to 2, were supposed to be adjacent ( $k=0$ ). To ascertain directions of transitions between the adjacent states, we took into consideration that the **OK** state was initial and transitions from this state to the other ones were nonreversible as well as

<sup>9</sup> Calculations were carried out with the aid of the STATISTICA software package.

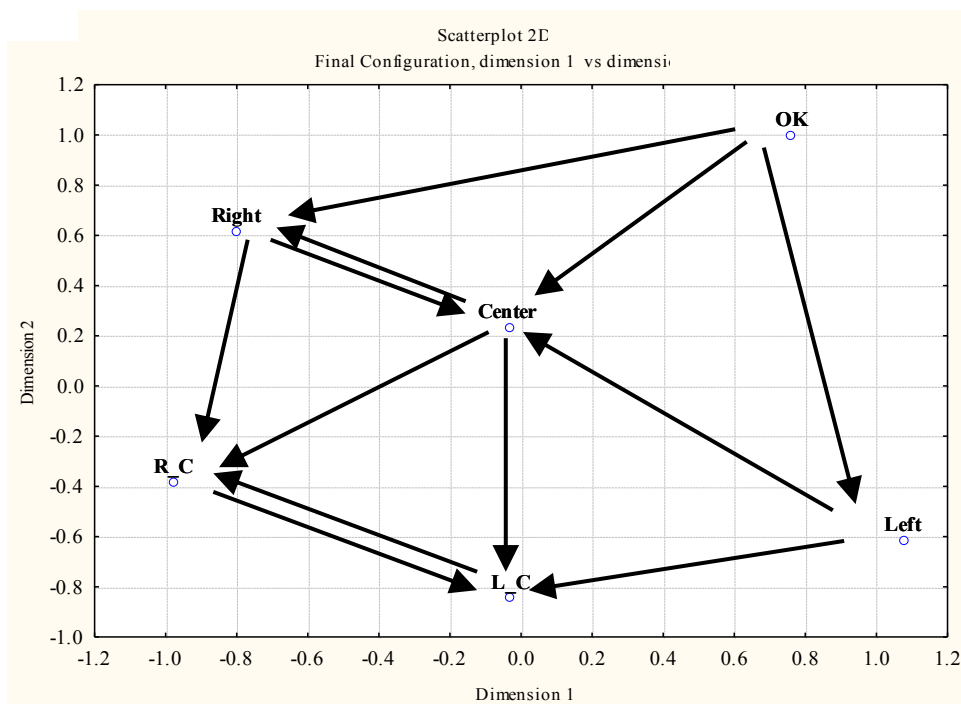
the facts that the simple damages (**Center**, **Right** and **Left**) preceded more complex ones (**R\_C** and **L\_C**), in which they were included as components. The obtained Markov network with redundant connections is presented in Figure 5.



**Figure 5.** Markov network with redundant connections.

Taking into account the contiguity of activated areas at the topological map, the network formally included connections between the states **R\_C** and **L\_C** at a given synthesis step in spite of the fact that none of them could be predecessor of another. Record of these connections did not influence the final result since they would be removed later as redundant.

Alternative way of revealing state connections was based on the multidimensional scaling and yielded the 2-D diagram shown in Figure 6<sup>10</sup>. The arrows denote connections determined in accordance with the rules specified in Section 2.1.4. The structure obtained agrees substantially with the results calculated by means of Kohonen maps.



**Figure 6.** Redundant state connections resulted from multidimensional scaling.

<sup>10</sup> Calculations were carried out with the aid of STATISTICA software package.

Redundant connections were revealed by comparison of Pearson statistic values obtained by the chi-square minimum method for the complete network shown in Figure 5 and ten reduced networks. In each of these networks, one of  $\lambda_i$  ( $i=0, 1, \dots, 9$ ) parameters was supposed to be equal to zero at significance level 0.5. Estimates resulted from the first iteration of network adjustment are shown in Table 2.

**Table 2. Statistical estimates resulted from the first iteration of adjustment of Markov network with redundant connections.**

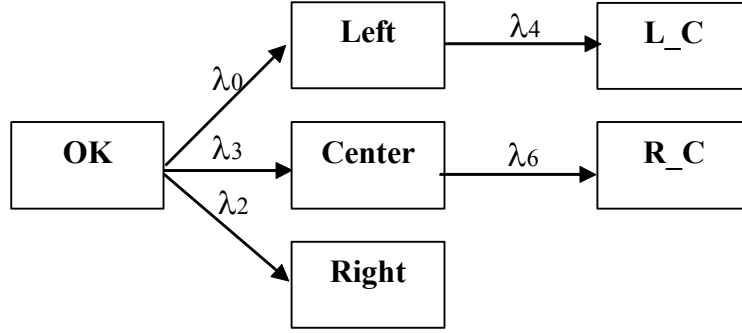
No	Network	Chi-square	Chi-square difference for the complete and reduced networks	$p$ -value for the difference (for 1 d.o.f.)
1	Complete (2 d.o.f.)	0.026	-	-
2	Reduced ( $\lambda_0=0$ )	457.580	457.555	0.000
3	Reduced ( $\lambda_1=0$ )	0.026	0.000	<b>1.000</b>
4	Reduced ( $\lambda_2=0$ )	710.305	710.279	0.000
5	Reduced ( $\lambda_3=0$ )	718.422	718.396	0.000
6	Reduced ( $\lambda_4=0$ )	2.978	2.953	0.086
7	Reduced ( $\lambda_5=0$ )	0.052	0.026	<b>0.871</b>
8	Reduced ( $\lambda_6=0$ )	2.104	2.078	0.149
9	Reduced ( $\lambda_7=0$ )	0.097	0.072	<b>0.789</b>
10	Reduced ( $\lambda_8=0$ )	0.026	0.000	<b>1.000</b>
11	Reduced ( $\lambda_9=0$ )	0.062	0.037	<b>0.848</b>

One can see from Table 2 that connections with flow rates  $\lambda_1$ ,  $\lambda_5$ ,  $\lambda_7$ ,  $\lambda_8$  and  $\lambda_9$  are redundant and may be removed from the network ( $p > 0.5$ ); connections with flow rates  $\lambda_0$ ,  $\lambda_2$  and  $\lambda_3$  are highly significant and must be retained ( $p < 0.01$ ); connections with flow rates  $\lambda_4$ ,  $\lambda_6$  are supposed to be questionable ( $0.01 \leq p \leq 0.5$ ) and must be studied at the following iteration.

The second iteration of adjustment showed that connections with flow rates  $\lambda_4$  and  $\lambda_6$  are highly significant and must be retained (see Table 3). Markov network obtained is presented in Figure 7.

**Table 3. Statistical estimates resulted from the second iteration of adjustment of the Markov network with redundant connections.**

No	Network	Chi-square	Chi-square difference for the complete and reduced networks	$p$ -value for the difference (for 1 d.o.f.)
1	Complete	0.276	-	-
2	Reduced ( $\lambda_4=0$ )	911.668	911.392	0.000
3	Reduced ( $\lambda_6=0$ )	1086.616	1086.340	0.000



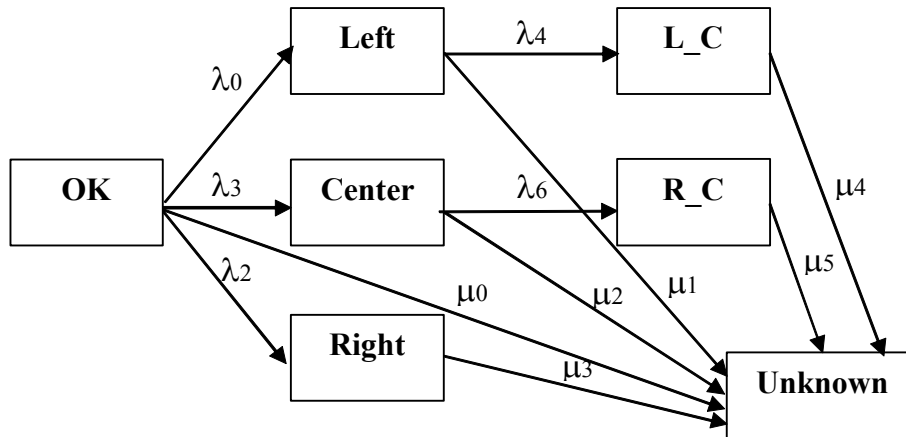
**Figure 7. Markov network adjusted.**

The **Unknown** state was inserted in the network according to the above-mentioned technique, with the exception of adjacency study with the aid of Kohonen map. A network with redundant connections contained all the possible ways of transition to the **Unknown** state (Figure 8). Removing redundant connections by two iterations of adjustment resulted in a final network presented in Figure 9 (only connection with flow rate  $\mu_5$  turned out to be significant). Kolmogorov set of differential equations describing time-domain dynamics of state probabilities  $p_{ok}$ ,  $p_{left}$ ,  $p_{center}$ ,  $p_{right}$ ,  $p_{l-c}$ ,  $p_{r-c}$  and  $p_{unknown}$ , is

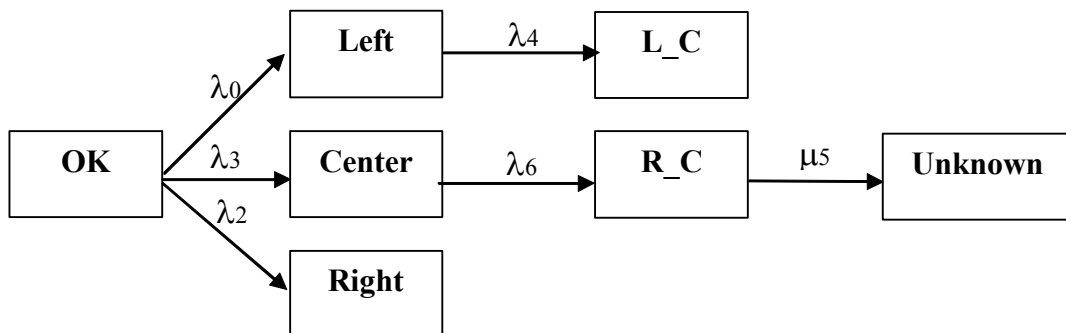
$$\begin{aligned}\frac{dp_{ok}}{dt} &= -(\lambda_0 + \lambda_2 + \lambda_3)p_{ok}, \\ \frac{dp_{left}}{dt} &= \lambda_0 p_{ok} - \lambda_4 p_{left}, \\ \frac{dp_{center}}{dt} &= \lambda_3 p_{ok} - \lambda_6 p_{center}, \\ \frac{dp_{right}}{dt} &= \lambda_2 p_{ok}, \\ \frac{dp_{l-c}}{dt} &= \lambda_4 p_{left}, \\ \frac{dp_{r-c}}{dt} &= \lambda_6 p_{center} - \mu_5 p_{r-c}, \\ \frac{dp_{unknown}}{dt} &= \mu_5 p_{r-c}.\end{aligned}$$

The following initial conditions are given for integration:

$$p_{ok}(0)=1, p_{left}(0)=p_{center}(0)=p_{right}(0)=p_{l-c}(0)=p_{r-c}(0)=p_{unknown}(0)=0.$$



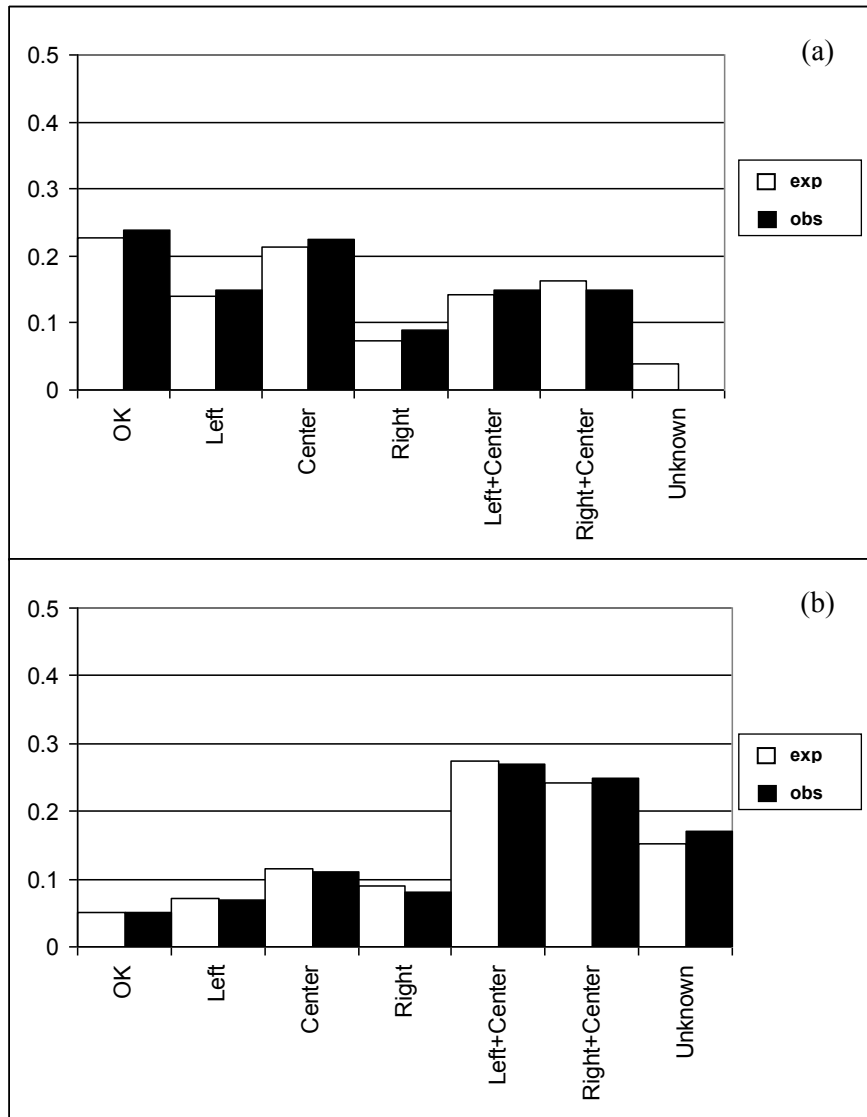
**Figure 8.** Markov network with the Unknown state and redundant connections.



**Figure 9.** Markov network adjusted (with the Unknown state).

Construction (modification) of a network without adjacency study with the aid of Kohonen network is expedient if the number of free parameters to be found does not exceed the number of independent observed statistics  $F_k$  and Kohonen maps are definitely unable to determine the structure of connections (as is the case for the **Unknown** state).

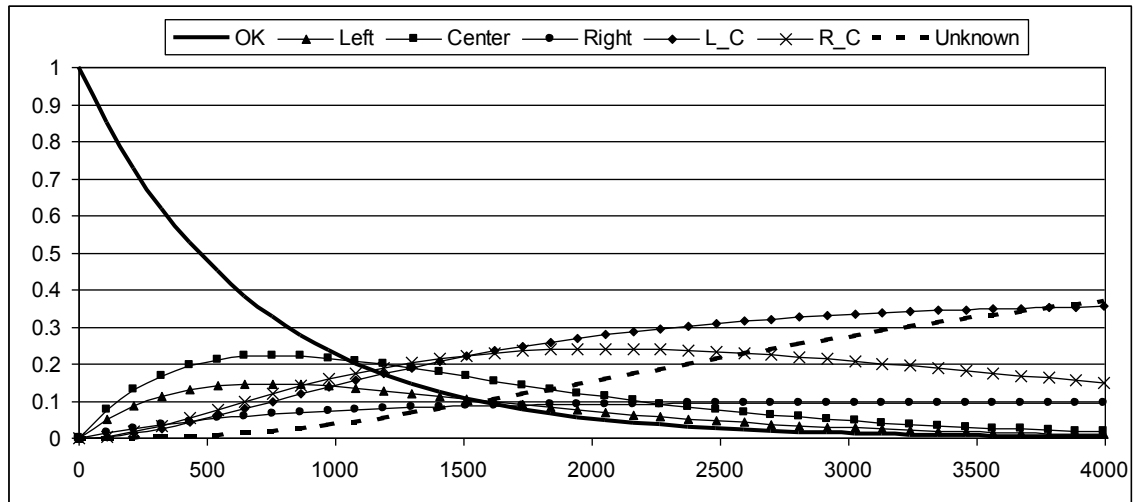
The synthesized network yields a good forecast: as the probability of exceeding the obtained value 3.396 of the chi-square statistic for six degrees of freedom ( $(7-1) \times 2 = 12$  independent statistics were used to find six free parameters) is 0.76, the chi-square goodness-of-fit test does not give grounds to consider the differences between observed and expected histograms (Figure 10) as statistically significant.



**Figure 10.** Expected and observed percentage of different damage types appearance after  
(a) 1000 service hours, (b) 2000 service hours.

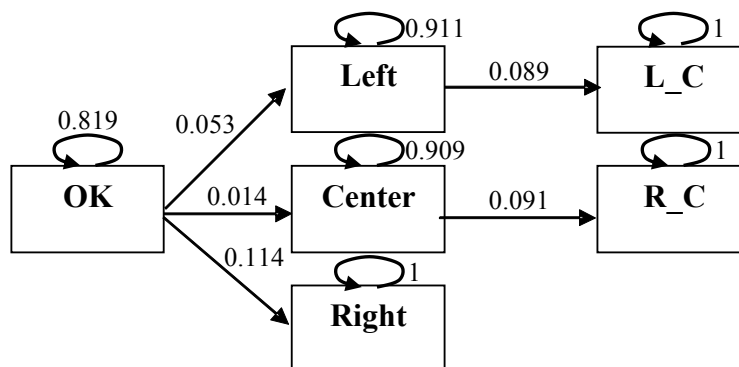
Probabilities of different system damages as functions of time are shown in Figure 11. The indicated curves make it possible to forecast the structure service life, establish the justified routine maintenance and facilitate diagnostics.





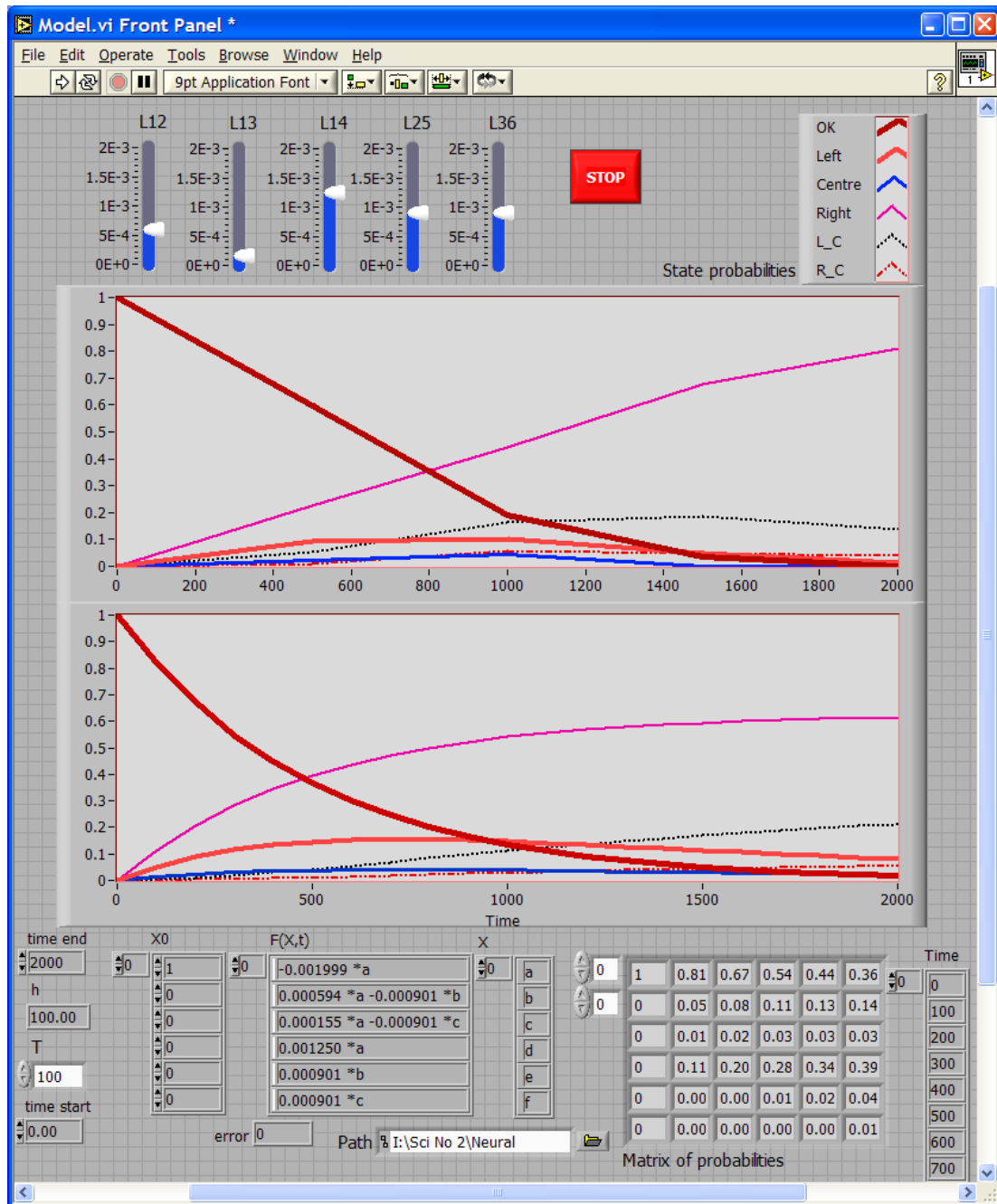
**Figure 11.** Probabilities of different system damages as functions of time (time is given in hours).

To get the simplest instrument for reproducing the dynamics of damage probabilities, Markov chain for engineering applications was generated from the network presented in Figure 7, by using the technique described in Section 2.3. 100-hour sampling intervals commensurable with 1000-hour damage estimation intervals under consideration were used. Numerical solution of the relevant conditional multivariate optimization problem was computed by means of a special macro implemented at the MS Excel spreadsheet. Residual error resulted from calculations with the sought transition probability matrix was negligible<sup>11</sup>. Markov chain obtained is given in Figure 12. Shown in Figure 13 for illustration is comparison of the integrated Markov network probability functions and simulation results for the corresponding Markov chain after 100 iterations. This simulation was carried out with the aid of the LabVIEW graphical programming environment.



**Figure 12.** Markov chain with transition probabilities generated from Markov network.

<sup>11</sup> Pearson statistic yielded  $X^2 < 0.0001$  and  $p > 0.9999$ .



**Figure 13.** Comparison of the integrated Markov network probability functions (lower plot) and simulation results for the corresponding Markov chain (upper plot) after 100 iterations.

The identification problem was solved to get a sequence of different damages represented by model states, which is most likely occurred, for the sequence of three observed spectral characteristics in structure checkpoints taken with a 100-hour sampling interval. Possible state sequences for the adjusted model shown in the form of Markov network in Figure 7 and in the form of Markov chain in Figure 12 are presented in the second column of Table 4. Table 5 contains parameters of probability density functions for Euclidean distances of observed characteristics arising in case of each state from the corresponding centroids as well as Euclidean distances of observed characteristics from state centroids. To avoid influence of redundant information, the indicated distances were calculated for observed data representation with reduced dimension. According to the problem formulation the system under study starts from the **OK** state.

**Table 4.** Possible state sequences for the adjusted model and their probabilistic characteristics.

State sequence number	State sequence	Occurrence probability for state sequence	Probabilistic estimation of occurrence of observed sequence of spectral characteristics
1	{ <b>OK,OK,OK</b> }	0.671	0.000000064
2	{ <b>OK,OK,Left</b> }	0.043	0.129708896
3	{ <b>OK,OK,Center</b> }	0.011	0.000283591
4	{ <b>OK,OK,Right</b> }	0.093	0.000571344
5	{ <b>OK,Left,Left</b> }	0.048	0.374853359
6	{ <b>OK,Left,L_C</b> }	0.005	0.031801329
7	{ <b>OK,Center,Center</b> }	0.013	0.000007695
8	{ <b>OK,Center,R_C</b> }	0.001	0.000000002
9	{ <b>OK,Right,Right</b> }	0.114	0.000004562

**Table 5.** Parameters of normal probability density functions for Euclidean distances of observed characteristics arising in case of each state from the corresponding centroids and Euclidean distances of observed characteristics from state centroids.

	<b>OK</b>	<b>Left</b>	<b>Center</b>	<b>Right</b>	<b>L_C</b>	<b>R_C</b>
Mean	1,126	0,594	0,432	0,484	0,669	0,549
Standard deviation	0,049	0,253	0,303	0,224	0,252	0,193
Euclidean distance of the 1 <sup>st</sup> observed characteristic from state centroids	1,108	1,777	1,842	1,967	1,897	1,993
Euclidean distance of the 2 <sup>nd</sup> observed characteristic from state centroids	1,251	0,282	1,420	1,314	0,913	1,503
Euclidean distance of the 3 <sup>rd</sup> observed characteristic from state centroids	1,427	0,463	1,363	1,288	0,878	1,448

Comparison of values presented in the last column of Table 4 yields the following most likely occurred sequence of damages: {**OK,Left,Left**}.

## 4. MAIN RESULTS AND CONCLUSIONS

Methods of construction and analysis of hidden Markov models, which are in the best conformity with observations in case of uncertainties, have been developed. They are based on a novel statistical identification technique and can be applied to both continuous- and discrete-time models.

The model synthesis may be carried out in case of some uncertainties: absence of complete information on system states and their connections is assumed. When solving this problem, a set of relevant states, their connections and optimal values of free parameters of Markov networks are determined. An initial rough model results from either cluster analysis and multidimensional scaling of observed data or their analysis by means of Kohonen maps. Then it suffers proper corrections. Connections between different system states are determined according to contiguity of their clusters or areas of winning units on the topological maps and statistical goodness-of-fit tests. Undefined observed cases are partially identified by the method of classification propagation with the aid of either Kohonen maps or cluster analysis. Histograms of observed frequencies of being in different system states after the given exploitation periods are employed to train the constructed models. Free network parameters are identified by the method of chi-square minimum. In case of models represented by discrete-time structures conversion to continuous-time trained structures is performed to obtain more reliable results with the following return to discrete time and identification of transition probabilities.

Owing to universality and flexibility, the presented way of solution has obvious advantages over the methods used for maximum likelihood estimation of parameters for traditional hidden Markov chains. Moreover, traditional algorithms used for hidden Markov models are unable to solve the applied technical problem considered here when uncertainties with damage types, model states and their connections are the case or the number of sampling points is too small.

The methods suggested are available for specialists responsible for maintenance of technical products. However, these techniques are not fastened to the particular technical problem considered in this paper and can be applied to solve many other technical and non-technical tasks. The obtained Markov models are useful for diagnostics, pattern recognition, structure service lives forecasting and establishing the justified routine maintenance.

### **Acknowledgements**

This work is supported by grant 10-06-00423 from the Russian Foundation for Basic Research.

## REFERENCES

1. Baranov S.N. and Kuravsky L.S., “Acoustic vibrations: modeling, optimization and diagnostics”, 2nd Edition, enlarged, Moscow: RUSAVIA, 224 pp, 2006.
2. Baum L.E., Petrie T., Soules G., and Weiss N., “A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains”, *Ann. Math. Statist.*, Vol 41, No 1, pp 164–171, 1970.
3. Bendat J.S. and Piersol A.G., “Random data. Analysis and measurement procedures”, New York: John Wiley & Sons, 1986.
4. Bishop Y.M.M., Fienberg S.E., and Holland P.W., “Discrete multivariate analysis: Theory and practice”, Cambridge, MA: M. I. T. Press, 1975.
5. Bogdanoff J.L. and Kozin F., “Probabilistic Models of Cumulative Damage”, New York: John Wiley & Sons, 1985.
6. Brousset C. and Baudrillard G., “Neural network for automating diagnosis in aircraft inspection”, *Review of Progress in Quantitative Nondestructive Evaluation* (Ed. by D.O. Thompson and D.E. Chimenti), Plenum Press, New York, Vol 12, pp 797–802, 1993.
7. Cramer H., “Mathematical methods of statistics”, Princeton: Princeton University Press, 1946.
8. Kohonen T., “Self-organizing maps”, Heidelberg: Springer Verlag, 1995.
9. Kuravsky L.S. and Baranov S.N., “Application of neural networks for diagnostics and forecasting of fatigue failures of thin-walled structures”, *Neurocomputers: Design and Applications*, No 12, pp 47–63, 2001 (in Russian).
10. Kuravsky L.S. and Baranov S.N., “Condition monitoring of the structures suffered acoustic fatigue failure and forecasting their service life”, *Proc. Condition Monitoring 2003*, Oxford, United Kingdom, pp 256–279, July 2003.
11. Kuravsky L.S. and Baranov S.N., “Discriminant networks to solve diagnostics problems”, *Neurocomputers: Design and Applications*, No 8-9, pp 3–9, 2003 (in Russian).
12. Kuravsky L.S. and Baranov S.N., “Neural networks in fatigue damage recognition: diagnostics and statistical analysis”, *Proc. 11th International Congress on Sound and Vibration*, St.-Petersburg, Russia, pp 2929–2944, July 2004.
13. Kuravsky L.S. and Baranov S.N., “Synthesis of Markov networks for forecasting fatigue failures”, *Proc. Condition Monitoring 2003*, Oxford, United Kingdom, pp 76–91, July 2003.
14. Kuravsky L.S. and Baranov S.N., “The concept of multifactor Markov networks and its application to forecasting and diagnostics of technical systems”, *Proc. Condition Monitoring 2005*, Cambridge, United Kingdom, pp 111–117, July 2005.
15. Kuravsky L.S., Baranov S.N., and Malykh S.B., “Neural networks to solve the problems of forecasting, diagnostics and data analysis”, Moscow: RUSAVIA, 2003 (in Russian).
16. Lawley D.N. and Maxwell A.E., “Factor analysis as a statistical method”, London: Butterworths, 1963.
17. Marple S.L., Jr., “Digital spectral analysis with applications”, New Jersey: Prentice-Hall, 1987.
18. Pidaparti R.M.V. and Palakal M.J., “Neural network approach to fatigue-crack-growth predictions under aircraft spectrum loadings”, *Journal of Aircraft*, Vol 32, pp 825–831, 1995.
19. Rabiner L.R., “A tutorial on hidden Markov models and selected applications in speech recognition”, *Proc. IEEE*, Vol 77, No 2, pp 257–286, 1989.
20. Viterbi A.J., “Error bounds for convolutional codes and an asymptotically optimum decoding algorithm”, *IEEE Transactions on Information Theory*, Vol 13, No 2, pp.260–269, 1967.