ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS DE MATHEMATICA
Volume 8, 2004

Regression model based on Markov chain theory for composite fatigue curve approximation

MARTINSH KLEINHOF, YURI PARAMONOV AND ALEXANDRA PARAMONOVA

ABSTRACT. Model of fatigue damage accumulation, which is founded on the bases of Markov chain theory, is used as nonlinear regression model for processing of S-N fatigue curve data of laminate. A simple method of approximate estimation of model parameters is suggested. A numerical example is given. Some parameters of this fatigue curve model can be considered as local static strength distribution parameters. This model can be used for prediction of fatigue curve changes as consequence of the static strength distribution function parameter changes and for prediction of the cumulative distribution function of fatigue life for arbitrary stress cycles consequence.

1. Introduction

Every year the use of composite material in aircraft structure increases. To provide reliability of flight we should study the fatigue phenomenon of this material. One of the main quantitative characteristics of this phenomenon is fatigue or S-N curve. It is a plot (often log-log plot) of cycling (maximum or amplitude) stress S versus the median (or mean) fatigue life N, which is expressed in cycles to failure. An extension of this concept is the p-quantile S-N curves, also called S-N-P curves, a generalization that relates the p-quantile of fatigue life to the applied stress. There are many offers for description of S-N-P curves. Wide discussion on this topic is presented in Pascual and Meeker (1999). Short review of this problem is given by Paramonova, Kleinhof and Paramonov (2002a). We will not repeat it because this paper is in some way a development of the latter. This paper is devoted mainly to processing of a dataset of fatigue tests of 125 carbon-fiber laminate specimens. The dataset forms a table of 125 lines. On every

Received October 5, 2003.
2000 *Mathematics Subject Classification.* 60J10, 60J02, 62P30.
Key words and phrases. Strength, fatigue life, composite.

line, corresponding to the fatigue test of one specimen, we see four numbers: i, S_i, T_i, A_i. These are: (1) order number of a specimen, (2) stress level, (3) test time and (4) result of the test (A_i is equal to 1, if fatigue test is finished by the failure of specimen, and A_i is equal to 0, if the test is finished without failure of specimen (right censored observation)). In the considered dataset there were 5 different stress levels and 25 specimens were tested at every stress level. The purpose of this paper is to get estimates of parameters of the fatigue damage accumulation model, based on the Markov chain theory (Paramonova, Kleinhof and Paramonov, 2002a), which can be used for description of this dataset. It was shown already in the previous work of the authors that, although the considered model is too simple and does not provide numerical coincidence with experimental fatigue test data, nevertheless it allows to get a formula for S-N-P fatigue curve by the use of the static strength distribution parameters and some additional parameters, which have some 'physical' interpretation. In this paper we consider an inverse problem: by the use of this model, which can be considered now as a nonlinear regression analysis model, we try to get estimates of local static strength distribution parameters. The likelihood of these estimates and the likelihood of "theoretical" and experimental fatigue curves can be considered as a proof of likelihood of the studied model. The model discussed in Paramonova, Kleinhof and Paramonov (2002a) has 6 parameters altogether. In this paper approximately for the same level of precision of fatigue curve description we have used only 4 parameters. It means significant decrease of the difficulty of statistical analysis.

The reminder of the discussed model ideas is given in Section 2. A method for model parameters estimation is discussed in Section 3 and a numerical example is given in Section 4.

2. Cumulative damage model based on the Markov chains theory

We consider the process of fatigue damage accumulation as Markov chain with the following matrix of transition probabilities

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{23} & \cdots & p_{1r} & p_{1(r+1)} \\ 0 & p_{22} & p_{23} & p_{24} & \cdots & p_{2r} & p_{2(r+1)} \\ 0 & 0 & p_{33} & p_{34} & \cdots & p_{3r} & p_{3(r+1)} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & p_{rr} & p_{r(r+1)} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

The model corresponds to a MC (Markov Chain) with one $(r+1)$-th absorbing state and r nonrecurrent states.

The cumulative distribution function (cdf) of time to absorption for this process is given by the equality

$$F_T(t) = p_{1,(r+1)}(t), \quad t = 1, 2, 3, ...,$$

where $p_{1,(r+1)}(t)$ is the element of the first row and $(r+1)$-th column of matrix $P(t) = P^t$. It can be defined also in the following way:

$$F_T(t) = aP^t b, \tag{1}$$

where $a = (100...0)$ is a row vector, $b = (000...1)'$ is a column vector. It is worth to notice, that product $P^t b$ represents a column vector of cumulative distribution functions of times to absorption, components of which correspond to different initial states of MC $:(F_T^{(1)}(t),\ F_T^{(2)}(t),\ ...,\ F_T^{(r)}(t))'$. In general case it can be used to get cdf when the probability distribution on initial states of MC, $\pi = (\pi_1, \pi_2, ..., \pi_r, \pi_{r+1})$, is known

$$F_T(t) = \pi P^t b = \pi(F_T^{(1)}(t), F_T^{(2)}(t), F_T^{(r)}(t)).$$

This possibility should be considered as some reserve, which can be used to take into account some specific features of specific composite structure, induced by some specific technology. Now we do not use this possibility, because we deliberately try to decrease the number of parameters of the considered model. Moreover, in this paper we consider Simple Markov Chain Model of Fatigue Life (SMCMFL) of composite material. This is the model, for which only transitions to the nearest 'senior' states are allowed and

$$P = \begin{bmatrix} q_1 & p_1 & 0 & & & \dots & 0 \\ 0 & q_2 & p_2 & 0 & & \dots & 0 \\ 0 & 0 & q_3 & p_3 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & & & \dots & 0 & q_r & p_r \\ 0 & & & \dots & 0 & 0 & 1 \end{bmatrix}, \ q_1 = 1 - p_i,\ i = 1, ..., r.$$

The main characteristics of this type MC are well known. Time to failure (time to absorption) $T = X_1 + X_2 + ... + X_r$, where X_i (time spent in ith state), $i = 1, ..., r$, are independent random variables. Random variable X_i has geometric distribution with pmf (probability mass function)

$$P(X_i = n) = (1-p)^{n-1} p_i,$$

mean

$$E(X_i) = 1/p_i$$

and variance

$$V(X_i) = (1 - p_i)/p_i^2.$$

It seems that, in fact, we could make all calculations without using results from Markov chain theory. But this theory gives us very simple formula (1) for cdf calculation. Moreover this theory gives us similar formula for cdf of any program of loading. The matrix of transition probabilities corresponds to specific mode of loading in one cycle, which in the considered case is defined by specific initial nominal stress level (in the first cycle).

The probability generating function for the random variable T (which can be used to obtain pmf of T) is defined by formula

$$G_T(z) = \sum_{i=0}^{\infty} p_T(i) \cdot z^i = \prod_{i=1}^{\infty} \frac{zp_i}{1 - z(1 - p_i)}.$$

The pmf can be calculated also by the formula

$$p_T(t) = F_T(t) - F_T(t-1).$$

If we assume, that one step in MC corresponds to k_M cycles in fatigue test, then for calculation of expected value and variance of T we should use formulae

$$E(T) = k_M \sum_{i=1}^{r} 1/p_i, \tag{2}$$

$$V(T) = k_M^2 \sum_{i=1}^{r} (1 - p_i)/p_i^2. \tag{3}$$

But the main and most difficult problem is to connect these probabilities, p_i, $i = 1, ..., r$, with the parameter of composite material component static strength distribution and applied stress level in such a way that we can get the fatigue curve equation. The simplest suggestion, which is studied in this paper, is to assume that in one step of Markov process only one parallel structural item (for example, strand) can fail. If we have $(R - i)$ still alive parallel structural items and the same cdf $F(s)$ for every item, then the fracture probability of at least one item is equal

$$p_i = 1 - (1 - F(s_i))^{R-i},$$

where R is initial number of items, i is the number of items, which are failed already, s_i is the corresponding stress applied uniformly to all $(R-i)$ items. We suppose also that

$$s_i = \frac{SR - S_f i}{R - i} = \frac{S(1 - S_f i/SR)}{1 - i/R},$$

where S is the initial stress (force) in every item (at the beginning of the test), S_f is the stress (force) which an already failed item can carry (because at least in the beginning of damage accumulation the rupture of fibers can be in different cross sections).

Let us consider the case when cdf $F(s)$ has location and scale parameters:

$$F(s) = F_0((g(s) - \theta_0)/\theta_1), \tag{4}$$

where $g(\cdot)$ is some known function, $F_0(\cdot)$ is some known cdf. Later on we shall use normal cdf and $g(s) = log(s)$. Now the considered model has parameter η with 6 components: $\eta = (\theta_0, \theta_1, r, R, k_M, S_f)$. They have the following interpretation:

- θ_0, θ_1 are parameters of cdf of strength of composite item (strand or fiber); for example, if $g(s) = s$ (normal distribution of strength) then θ_0 is the expected value and θ_1 is the standard deviation of the item strength;
- R is the number of composite items in critical volume, failure of which corresponds to total failure of specimen;
- r is a critical number of failed elements inside of this critical volume, corresponding to failure of this volume; the ratio r/R defines the part of the cross section area of critical volume, the destruction of which we consider as failure of specimen; the variance and the coefficient of variation of fatigue life is mainly defined by the value r;
- k_M is the number of cycles corresponding to one step in MC;
- S_f is residual strength of failed item (it depends on the orientation and number of layers, the characteristics of matrix, ...).

From now on we shall use $F_T(t; S; \eta)$ as a specific notation of cdf of the random variable T instead of more general notation, $F_T(t)$.

3. Estimation of parameters of model

Formulae (1), (2), (3) can be used in both directions for calculating mean and p-quantile fatigue curves, if parameters are known, or for nonlinear regression analysis for the model parameters estimation, if fatigue life dataset is known. The mean and S-N-P fatigue curves are defined by the formulae

$$E(T(S_j)) = k_M \sum_{i=1}^{r} 1/p_i(S_j, \eta), \quad t_p(S_j) = F_T^{-1}(p; S_j, \eta)$$

where $E(T(S_j))$, $t_p(S_j)$ are the mean value and p-quantile of the fatigue life for stress S_j. The parameters of the model can be estimated by the use of Method of Moments (MM), Least Square Method (LSM) or Maximum Likelihood Method (MLM), which is preferable. For the profound investigation of this model the nonlinear regression procedure of SAS system can be recommended. But in any case it is a difficult problem to find 6 unknown parameters. We limit ourselves to an approximate solution of this problem. First of all we put: $k_M = 1$, $S_f = 0$, then we get approximate estimates of the remaining parameters: r, R, θ_0 and θ_1, and, finally, for fixed approximate estimates of parameters r, R we can find estimates of θ_0 and θ_1 by the use of MLM.

Approximate estimate of parameter r can be found, if we assume, that when

$$p_1 = p_2 = ... = p_r = p,$$

then

$$E(T) \cong \frac{r}{p}; \quad V(T) \cong \frac{r}{p^2};$$

coefficient of variation

$$C_V = \sqrt{V(T)}/E(T) \cong 1/\sqrt{r}.$$

An approximate preliminary estimate of parameter r is defined by the formula

$$\hat{r} \cong [1/(\hat{C}_V)^2] + 1,$$

where $\hat{C}_V$ is an estimate of C_V, $[x]$ is the integer part of x. The value $E(T) \cong \frac{r}{p}$ is very large $(10^5 - 10^7)$, r is small enough (see Section 4), so the value of p is very small and $F(s)$ is very small too. All this gives us an idea to make the following assumption (which is used only for preliminary approximate estimation of parameters):

$$p_i = cF_0((g(S) - \theta_0)/\theta_1)$$

$$for \quad all \quad i = 1, 2, ..., r,$$

where c is some constant. Then we have the following approximate formula

$$E(T(S)) = \frac{D_f}{F_0((g(S) - \theta_0)/\theta_1)}, \tag{5}$$

where $D_f = r/c$.

At fixed D_f we get the following linear regression model

$$y_i = F_0^{-1}(D_f/T(S_i))$$

$$= -\theta_0/\theta_1 + (1/\theta_1)g(S_i) = \beta_0 + \beta_1 x_i, \quad i = 1, 2, ..., n. \tag{6}$$

The value of $T(S_i)$ we get from our dataset. Parameters β_0 and β_1 of this model can be estimated by some statistical program of linear regression analysis at every fixed value of parameter D_f. It is not a complicated problem to find one nonlinear parameter D_f. Remind, that in the considered numerical example we have 5 stress levels and 25 observations for every stress level. We have got the following estimates for θ_0 and θ_1:

$$\hat{\theta}_1 = 1/\hat{\beta}_1, \quad \hat{\theta}_0 = -\hat{\beta}_0/\hat{\beta}_1.$$

An estimate of the parameter R can be obtained after estimation of the ratio $\rho = r/R$. Remind, that this ratio defines the part of the cross section area, the destruction of which we consider as total failure of specimens. In the Daniels's model of static strength (Paramonov and Kleinhof, 2000) this value corresponds to the value of $F(x^*)$, where x^* satisfies

$$x^*(1 - F(x^*)) = \max_x x(1 - F(x)).$$

We can estimate this value, using the estimates of θ_0 and θ_1. So we have

$$\hat{\rho} = F(x^*),$$

$$\hat{R} =]1/((\hat{C}_V)^2 \hat{\rho})[\ +1. \tag{7}$$

Now we have preliminary approximate estimates of all four parameters θ_0 and θ_1, r and R. Using the fixed estimates of r and R, the more precise estimates of θ_0 and θ_1 can be found by the use of MLM. For the probability mass function we have now the following formula

$$p_T(t; S; \eta) = F_T(t; S; \eta) - F_T(t-1; S; \eta).$$

For calculation of cdf we should know P^t. So we try to find an approximation for cdf. It appears, that lognormal approximation is appropriate:

$$F_T(t; S; \eta) \cong \Phi\Big(\frac{log(t) - \theta_{0LT}}{\theta_{1LT}}\Big),$$

where θ_{0LT}, θ_{1LT} are such, that we have the same mean and standard deviation

$$\theta_{0LT} = log(E(T)) - (log(C_V^2 + 1))/2,$$

$$\theta_{1LT} = (log(C_V^2 + 1))^{1/2}.$$

Now we have the maximum likelihood function in logarithm scale

$$l(\eta) = ln(L(\eta)),$$

where $L(\eta) = \prod_{i=1}^{n} f_i(1 - F_i)^{1-A_i}$, f_i, F_i are the probability density function and the cumulative distribution function of the random variable T (for fixed η and S). Remind, that A_i is equal to 1, if fatigue test is finished by the failure of specimens, and A_i is equal to 0, if the time of test is limited (right censored observation). After obtaining the MLM estimates of the parameters θ_0 and θ_1 at the preliminary estimated r and R we can get more precise estimates of r and R at the next step and etc. The convergence conditions of this process have to be examined in future.

Until now we have considered mainly uniform load-sharing system of isolated parallel items loaded by tension. But we want to apply this model to the more complex structure, for which fracture of longitudinal items means the failure of specimen as a whole.

4. Numerical example

In the numerical example we consider the problem of fitting the experimental data of fatigue test of laminate panel. The data was kindly given to the authors by professor W.Q. Meeker, who studied them in Pascual and Meeker (1999) and give the following description of these data: "the data come from 125 specimens in four-point out-of-plane bending tests of carbon eight-harness-satin/epoxy laminate. Fiber fracture and final specimen fracture occurred simultaneously. Thus, fatigue life is defined to be the number of cycles until specimen fracture. The dataset includes 10 right censored observations (known as "run outs" in the fatigue literature)". In the papers Paramonova, Kleinhof and Paramonov (2002a)-(2002c) we have considered already extreme values of fatigue lives for 5 stress levels, which we have got from the Figure 1 in Pascual and Meeker (1999). Now we have the original information, the same, on the bases of which the fatigue curve of this figure was made. As it was told already, this time we decrease the numbers of model parameters in order to increase the stability of estimates of others parameters. We put $k_M = 1$, $S_f = 0$. Then four main steps are made for estimation of the parameters θ_0, θ_1, r and R.

The first step. We make an additional assumption, that static strength of items has lognormal distribution (Paramonov and Kleinhof, 2000) : $F_0(\cdot)$ is cdf of the standard log-normal distribution, $g(s) = log(s)$. By the use of regression analysis (and by sequence of calculations for different D_f) approximate parameter estimates were found: $\hat{\theta}_0 = 7.6906$, $\hat{\theta}_1 = 0.3541$ and $\hat{D}_f = 0.0229$.

The second step. Estimation of r. For this purpose the estimate of the coefficient of variation $C_V = 0.5839$ was obtained for some middle stress, at which there was no censoring ($S = 340$ MPa). So $\hat{r} = 3$.

The third step. We have got an estimate $\hat{R} = 15$, because at the approximate estimates $\hat{\theta}_0$, $\hat{\theta}_1$ the value of $\hat{\rho} = F(x^*)$ appears to be equal to 0.2072 (later on, for final MLM estimates $\hat{\theta}_0$, $\hat{\theta}_1$, a new estimate of $\hat{\rho} = F(x^*)$ appears to be equal to 0.2022 and the final estimate $\hat{R} = 15$ did not change). Remind, that x^* is such, that:

$$x^*(1 - F(x^*)) = \max_x x(1 - F(x)).$$

Thr fourth step. Estimation of the parameters θ_0 and θ_1 by MLM appears to be very difficult problem. After detailed analysis the estimates were found: $\hat{\theta}_0 = 7.6461$, $\hat{\theta}_1 = 0.34471$.

Now we can check the validation of the log-normal approximation of $F_T(t; S; \eta)$. We consider two models:

$$F_T(t; S; \eta) \cong \Phi\Big(\frac{log(t) - \theta_{0LT}}{\theta_{1LT}}\Big)$$

and

$$F_T(t; S; \eta) \cong \Phi\Big(\frac{t - \theta_{0T}}{\theta_{1T}}\Big).$$

In the first case we should get a straight line

$$log(t) = \theta_{0LT} + \theta_{1LT}\Phi^{-1}(F_T(t; S; \eta)). \quad (8)$$

In the second case

$$t = \theta_{0T} + \theta_{1T}\Phi^{-1}(F_T(t; S; \eta)). \quad (9)$$

We calculate the cdf and pdf for stress level $S = 290.1$ MPa by formula (1). On Figure 1 the results of calculation are shown. We see that formula (8) gives nearly straight line, so the log-normal approximation of $F_T(t; S; \eta)$ is more appropriate than the normal approximation.

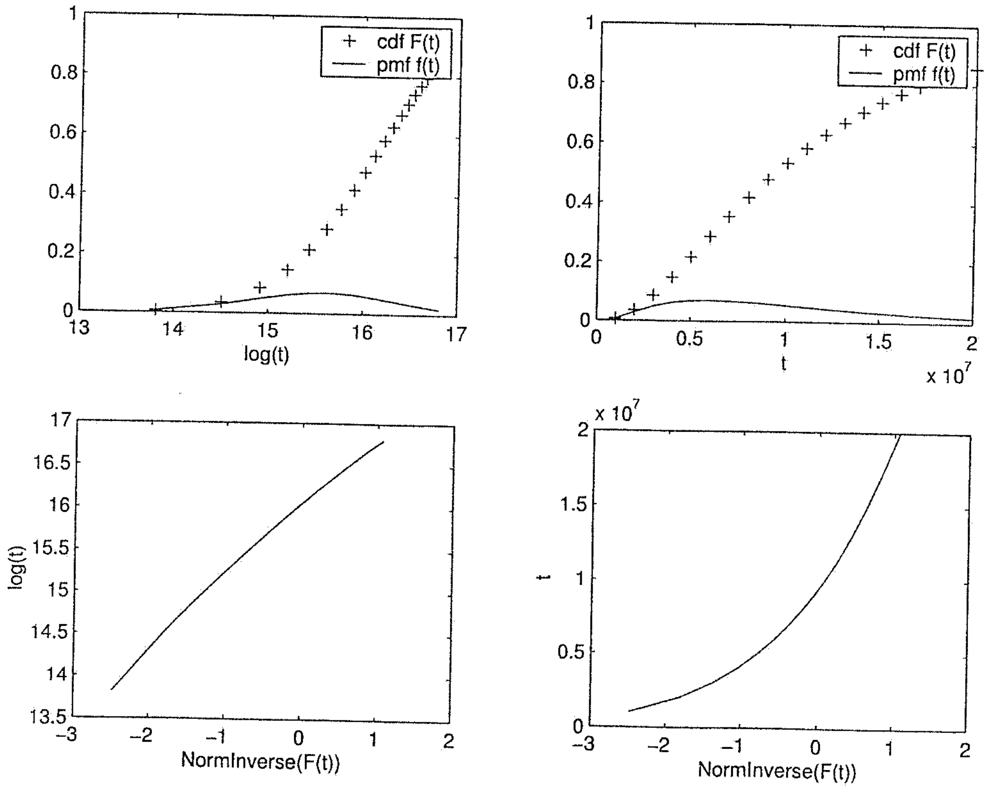

Figure 1 : Cumulative distribution (+) and probability mass functions (−) in the upper part. Functions $log(t) = \theta_{0LT} + \theta_{1LT}\Phi^{-1}(F_T(t; S; \eta))$ and $t = \theta_{0T} + \theta_{1T}\Phi^{-1}(F_T(t; S; \eta))$ in the lower part.

Finally we have got the fatigue curve. The experimental data (+) and results of calculations of the expected values of extreme order statistics (o) (minimum and maximum) are shown on Figure 2.

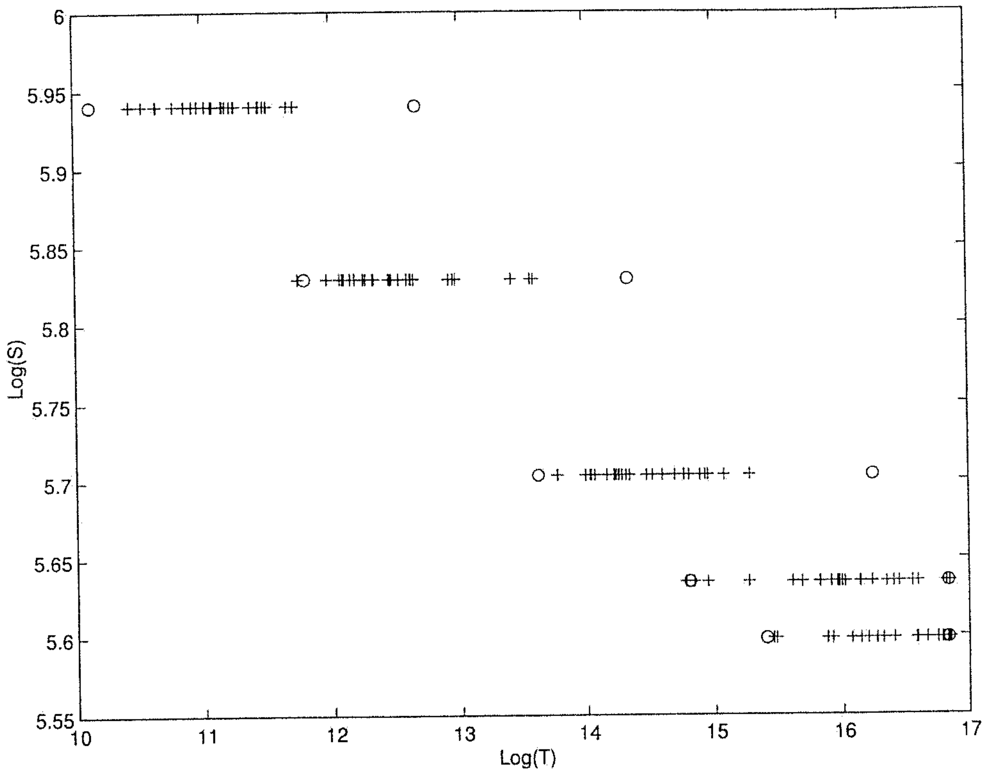

Figure 2: Dataset (+) and results of calculations of the expected values of extreme (minimum and maximum) order statistics (o).

5. Conclusions

1. *Simple Markov Chain Model of Fatigue Life* (SMCMFL) of composite material can be used as a nonlinear regression model for fatigue curve approximation. Processing a dataset of fatigue test of carbon-fiber laminate specimens we have got estimates of parameters, which can be interpreted as *equivalent local static strength distribution parameter estimates*. Probably, there is some discrepancy between these estimates ($\hat{\theta}_0 = 7.6461$, $\hat{\theta}_1 = 0.34471$) and real parameters of static strength distribution of carbon fibers, which are unknown for us. For example, in Paramonov and Kleinhof (2000) the following estimations of carbon fiber static strength distribution are given: $\theta_0 = 7.198$, $\theta_1 = 0.467$. This discrepancy can be explained by the difference of the original material, difference of load types (bending instead of tension), difference of "effective" length of fibers and so on.

2. This discrepancy can be used for description of specific features of a specific structure.

3. We do not need to estimate the static strength distribution parameters on the bases of fatigue data. We considered the likelihood of these estimates only as an argument, that considered SMCMFL of the composite material has right to exist and it can be used, for example, for forecasting of fatigue curve changes, when there are some changes in real static strength distribution parameters.

References

Paramonova, A. Yu., Kleinhof, M. A. and Paramonov, Yu. M. (2002a). Markov chains theory use for fatigue curve of composite material approximation. *Aviation* **6**, Technika, Vilnius, 103–108.

Paramonova, A. Yu, Kleinhof, M. A. and Paramonov, Yu. M. (2002b). Parameter estimation of Markov chain model of fatigue life of composite material. *Scientific Proceedings of Riga Technical University. Transport and Engineering: Transport, Aviation Transport* **6(8)**, 15–25.

Paramonova, A. Yu, Kleinhof, M. A. and Paramonov, Yu. M. (2002c). The use of Markov chains theory for approximation fatigue curve of composite material. In: *Proceedings of Third International Conference on Mathematical Methods in Reliability. Methodology and Practice. June 17-20*, NTNU, Trondheim, 509–512.

Paramonov, Yu. M. and Kleinhof, M. A. (2000). Simple statistical model of composite material strength. *Aviation* **5**, Technika, Vilnius, 128–133.

Pascual, F. G. and Meeker, W. Q. (1999). Estimating fatigue curves with the random fatigue-limit model. *Technometrics* **41**, 277–302.

RIGA TECHNICAL UNIVERSITY, KALKU 1, LV-1658 RIGA, LATVIA
E-mail address: `rauprm@junik.lv`