Kołowrocki Krzysztof

Soszyńska Joanna
Maritime University, Gdynia, Poland

Methods and algorithms for evaluating unknown parameters of operation processes of complex technical systems

Keywords

system operation process, semi-Markov model, sojourn time, distribution, estimation, maritime transport

Abstract

The paper objectives are to present the methods and tools useful in the statistical identifying unknown parameters of the operation models of complex technical systems and to apply them in the maritime industry. There are presented statistical methods of determining unknown parameters of the semi-markov model of the complex technical system operation processes. There is also presented the chi-square goodness-of-fit test applied to verifying the distributions of the system operation process conditional sojourn times in the particular operation states. Applications of these tools to identifying and predicting the operation characteristics of a ferry operating at the Baltic Sea waters are presented as well.

1. Introduction

Many real transportation systems belong to the class of complex systems. It is concerned with the large numbers of components and subsystems they are built and with their operating complexity. Modeling the complicated system operation processes, is difficult because of the large number of the operation states, impossibility of their precise defining and because of the impossibility of the exact describing the transitions between these states. The changes of the operation states of the system operations processes cause the changes of these systems reliability structures and also the changes of their components reliability functions [2]. The models of various multistate complex systems are considered in [1]. The general joint models linking these system reliability models with the models of their operation processes [5], allowing us for the evaluation of the reliability and safety of the complex technical systems in variable operations conditions, are constructed in [2].

In order to be able to apply these general models practically in the evaluation and prediction of the reliability of real complex technical it is necessary to elaborate the statistical methods concerned with determining the unknown parameters of the proposed models, namely the probabilities of the initials system operation states, the probabilities of transitions between the system operation states and the distributions of the sojourn times of the system operation process in the particular operation states and also the unknown parameters of the conditional multistate reliability functions of the system components in various operation states. It is also necessary the elaborating the methods of testing the hypotheses concerned with the conditional sojourn times of the system operations process in particular operations states and the hypotheses concerned with the conditional multistate reliability functions of the system components in the system various operation states. The model of the operation process of the complex technical system with the distinguished their operation states is proposed in [5]. The semi-markov process is used to construct a general probabilistic model of the considered complex industrial system operation process. To construct this model there were defined the vector of the probabilities of the system initial operation states, the matrix of the probabilities of transitions between the operation states, the matrix of the distribution functions and the matrix of the density functions of the conditional sojourn times in the particular operation states. To describe the system operation process conditional sojourn times in the particular operation states the uniform distribution, the triangular distribution, the double trapezium distribution, the quasi-trapezium distribution, the exponential distribution, the Weibull's distribution, the normal distribution and the chimney distribution are suggested in [5]. In this paper, the formulae estimating unknown parameters of these distributions are given and the chi-square test is applied to verifying the hypotheses about these distributions validity. Moreover these tools are applied to unknown parameters estimation and characteristics prediction of the Stena Baltica ferry operation process.

2. Identification of the operation process of the complex technical system

2.1. Estimation of unknown parameters of the semi-markov model of the operation process

We assume, similarly as in [2], [5] that a system during its operation at the fixed moment t, $t \in <0,+\infty>$, may be in one of v, $v \in N$, different operations states z_b, b = 1,2, ..., v. Next, we mark by $Z(t)$, $t \in <0,+\infty>$, the system operation process, that is a function of a continuous variable t, taking discrete values in the set $Z = \{z_1, z_2, ... z_v\}$ of the operation states. We assume a semi-markov model [2], [4], [5] of the system operation process $Z(t)$ and we mark by θ_{bl} its random conditional sojourn times at the operation states z_b when its next operation state is z_l, $b, l = 1,2,...,v$, $b \neq l$.

Under these assumption, the operation process may be described by the vector $[p_b(0)]_{1\times v}$ of probabilities of the system operation process staying in particular operations states at the initial moment t = 0, the matrix $[p_{bl}(t)]_{v\times v}$ of the probabilities of the system operation process transitions between the operation states and the matrix $[H_{bl}(t)]_{v\times v}$ of the distribution functions of the conditional sojourn times θ_{bl} of the system operation process at the operation states or equivalently by the matrix $[h_{bl}(t)]_{v\times v}$ of the distribution functions of the conditional sojourn times θ_{bl} of the system operation process at the operation states.

To estimate the unknown parameters of the system operations process, firstly during the experiment, we should collect necessary statistical data performing the following steps:

i) to analyze the system operation process and either to fix or to define its following general parameters:

- the number of the operation states o the system operation process v,
- the operation states o the system operation process z_1, z_2, …, z_v,
- the duration time of the experiment Θ;

ii) to fix and to collect the following statistical data necessary to evaluating the probabilities of the initial states of the system operations process:

- the number of the investigated (observed) realizations of the system operation process $n(0)$,
- the numbers of staying of the operation process respectively in the operations states z_1, z_2, …, z_v, at the initials moment $t = 0$ of all $n(0)$ observed realizations of the system operation process $n_1(0)$, $n_2(0)$, …, $n_v(0)$;

iii) to fix and to collect the following statistical data necessary to evaluating the probabilities of transitions between the system operation states:

- the numbers n_{bl}, b, l = 1,2,...,v, $b \neq l$, of the transitions of the system operation process from the operation state z_b to the operation state z_l during all observed realizations of the system operation process;
- the numbers n_b, b = 1,2,...,v, of departures of the system operation process from the operation states z_b;

iv) to fix and to collect the following statistical data necessary to evaluating the unknown parameters of the distributions of the conditional sojourn times of the system operation process in the particular operation states:

- the realizations θ_{bl}^k, k = 1,2, …, n_{bl}, for each b, l = 1,2,...,v, $b \neq l$ of the conditional sojourn times θ_{bl} of the system operations process at the operation state z_b when the next transition is to the operation state z_l during the observation time;

After collecting the above statistical data it is possible to estimate the unknown parameters of the system operation process performing the following steps:

i) to determine the vector

$$[p(0)] = [p_1(0), p_2(0), \ldots, p_v(0)],$$

of the realizations of the probabilities $p_b(0)$, $b = 1,2,...,v$, of the initial states of the system operation process, according to the formula

$$p_b(0) = \frac{n_b(0)}{n(0)} \text{ for } b = 1,2,...,v,$$

where $n(0) = \sum_{b=1}^{\nu} n_b(0),$ is the number of the realizations of the system operation process starting at the initial moment $t = 0$;

ii) to determine the matrix

$$[p_{bl}] = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1\nu} \\ p_{21} & p_{22} & \cdots & p_{2\nu} \\ \cdots \\ p_{\nu 1} & p_{\nu 2} & \cdots & p_{\nu\nu} \end{bmatrix},$$

of the realizations of the probabilities p_{bl}, $b, l = 1,2,...,\nu,$ of the system operations process transitions from the operations state z_b to the operations state z_l during the experiment time Θ, according to the formula

$$p_{bl} = \frac{n_{bl}}{n_b} \text{ for } b, l = 1,2,...,\nu,\ b \neq l,$$

$$p_{bb} = 0 \text{ for } b = 1,2,...,\nu,$$

where $n_b = \sum_{b \neq l}^{\nu} n_{bl}$, $b = 1,2,...,\nu,$ is the realization of the total number of the system operations process departures from the operations state z_b during the experiment time Θ;

iii) to determine the following empirical characteristics of the realizations of the conditional sojourn time of the system operation process in the particular operation states:

- the realizations of the mean values $\bar{\theta}_{bl}$ of the conditional sojourn times θ_{bl} of the system operations process at the operations state $H_{bl}(t)$ when the next transition is to the operation state θ_{bl}, according to the formula

$$\bar{\theta}_{bl} = \frac{1}{n_{bl}} \sum_{k=1}^{n_{bl}} \theta_{bl}^k \ b, l = 1,2,...,\nu,\ b \neq l,$$

- the number θ_{bl}^k, of the disjoint intervals $k = 1,2,...,n_{bl},$, θ_{bl}, that include the realizations θ_{bl}^k, $k = 1,2,...,n_{bl}$, of the conditional sojourn times θ_{bl} at the operation state z_b when the next transition is to the operation state z_l, according to the formula

$$\bar{r} \cong \sqrt{n_{bl}},$$

- the length d of the intervals $I_j = < a_{bl}^j, b_{bl}^j)$, $j = 1,2,...,\bar{r}$, according to the formula

$$d = \frac{\bar{R}}{\bar{r} - 1}, \text{ where } \bar{R} = \max_{1 \le k \le n_{bl}} \theta_{bl}^k - \min_{1 \le k \le n_{bl}} \theta_{bl}^k,$$

- the ends a_{bl}^j, b_{bl}^j, of the intervals $I_j = < a_{bl}^j, b_{bl}^j)$, $j = 1,2,...,\bar{r}$, according to the formulae

$$a_{bl}^1 = \min_{1 \le k \le n_{bl}} \theta_{bl}^K - \frac{d}{2},\ b_{bl}^j = a_{bl}^1 + jd,$$
$$j = 1,2,...,\bar{r},$$

$$a_{bl}^j = b_{bl}^{j-1},\quad j = 2,3,...,\bar{r},$$

in the way such that

$$I_1 \cup I_2 \cup ... \cup I_{\bar{r}} = < a_{bl}^1, b_{bl}^{\bar{r}}),$$

and

$$I_i \cap I_j = \emptyset \text{ for all } i \neq j,\ i, j \in \{1,2,...,\bar{r}\},$$

- the numbers n_{bl}^j of the realizations θ_{bl}^k in particular intervals I_j, $j = 1,2,...,\bar{r}$, according to the formula

$$n_{bl}^j = \#\{k : \theta_{bl}^k \in I_j,\ k \in \{1,2,...,n_{bl}\}\},$$
$$j = 1,2,...,\bar{r},$$

where $\sum_{j=1}^{\bar{r}} n_{bl}^j = n_{bl}$, whereas the symbol # means the number of elements of the set;

iv) to estimate the parameters of the distributions of the conditional sojourn times of the system operation process in the particular operation states for the following distinguished distributions respectively in the following way:

- the uniform distribution with a density function

$$h_{bl}(t) = \begin{cases} 0, & t < x_{bl} \\ \dfrac{1}{y_{bl} - x_{bl}}, & x_{bl} \le t \le y_{bl} \\ 0, & t > y_{bl}, \end{cases}$$

where $0 \le x_{bl} < y_{bl} < +\infty,$

the estimates of the unknown parameters of this distribution are:

$$x_{bl} = a_{bl}^{1},\ y_{bl} = x_{bl} + \bar{r}d\ ;$$

- the triangular distribution with a density function

$$h_{bl}(t) = \begin{cases} 0, & t < x_{bl} \\ \dfrac{2}{y_{bl} - x_{bl}} \dfrac{t - x_{bl}}{z_{bl} - x_{bl}}, & x_{bl} \le t \le z_{bl} \\ \dfrac{2}{y_{bl} - x_{bl}} \dfrac{y_{bl} - t}{y_{bl} - z_{bl}}, & z_{bl} \le t \le y_{bl} \\ 0, & t > y_{bl}, \end{cases}$$

where $0 \le x_{bl} < z_{bl} < y_{bl} < +\infty$,
the estimates of the unknown parameters of this distribution are:

$$x_{bl} = a_{bl}^{1},\ y_{bl} = x_{bl} + \bar{r}d\ ,\quad j = 1,2,...,\bar{\bar{r}};$$

- the double trapezium distribution with a density function

$$h_{bl}(t) = \begin{cases} 0, & t < x_{bl} \\ q_{bl} + \dfrac{C_{bl} - q_{bl}}{z_{bl} - x_{bl}}(t - x_{bl}), & x_{bl} \le t \le z_{bl} \\ w_{bl} + \dfrac{C_{bl} - w_{bl}}{y_{bl} - z_{bl}}(y_{bl} - t), & z_{bl} \le t \le y_{bl} \\ 0, & t > y_{bl}, \end{cases}$$

where

$$C_{bl} = \frac{2 - q_{bl}(z_{bl} - x_{bl}) - w_{bl}(y_{bl} - z_{bl})}{y_{bl} - x_{bl}},$$

$$0 \le x_{bl} < z_{bl} < y_{bl} < +\infty,\ 0 \le q_{bl} < +\infty,$$
$$0 \le w_{bl} < +\infty,$$
$$0 \le q_{bl}(z_{bl} - x_{bl}) + w_{bl}(y_{bl} - z_{bl}) \le 2;$$

the estimates of the unknown parameters of this distribution are:

$$x_{bl} = a_{bl}^{1},\ y_{bl} = x_{bl} + \bar{r}d\ ,\ q_{bl} = \frac{n_{bl}^{1}}{n_{bl}d},$$

$$w_{bl} = \frac{n_{bl}^{\bar{r}}}{n_{bl}d},\ z_{bl} = \bar{\theta}_{bl},$$

- the quasi-trapezium distribution with a density function

$$h_{bl}(t) =$$

$$\begin{cases} 0, & t < x_{bl} \\ q_{bl} + \dfrac{A_{bl} - q_{bl}}{z_{bl}^{1} - x_{bl}}(t - x_{bl}), & x_{bl} \le t \le z_{bl}^{1} \\ A_{bl}, & z_{bl}^{1} \le t \le z_{bl}^{2} \\ w_{bl} + \dfrac{A_{bl} - w_{bl}}{y_{bl} - z_{bl}^{2}}(y_{bl} - t), & z_{bl}^{2} \le t \le y_{bl} \\ 0, & t > y_{bl}, \end{cases}$$

where $A_{bl} = \dfrac{2 - q_{bl}(z_{bl}^{1} - x_{bl}) - w_{bl}(y_{bl} - z_{bl}^{2})}{z_{bl}^{2} - z_{bl}^{1} + y_{bl} - x_{bl}}$,

$$0 \le x_{bl} < z_{bl}^{1} \le z_{bl}^{2} < y_{bl} < +\infty\ ,\ 0 \le q_{bl} < +\infty,$$
$$0 \le w_{bl} < +\infty,$$
$$0 \le q_{bl}(z_{bl}^{1} - x_{bl}) + w_{bl}(z_{bl}^{2} - y_{bl}) \le 2\ ,$$

the estimates of the unknown parameters of this distribution are:

$$x_{bl} = a_{bl}^{1},\ y_{bl} = x_{bl} + \bar{r}d,\ q_{bl} = \frac{n_{bl}^{1}}{n_{bl}d},$$

$$w_{bl} = \frac{n_{bl}^{\bar{r}}}{n_{bl}d},\ z_{bl}^{1} = \bar{\theta}_{bl}^{1},\ z_{bl}^{2} = \bar{\theta}_{bl}^{2},$$

where

$$\bar{\theta}_{bl}^{1} = \frac{1}{n_{(me)}} \sum_{j=1}^{n(me)} \theta_{bl}^{j}\ ,$$

$$\bar{\theta}_{bl}^{2} = \frac{1}{n_{bl} - n_{(me)}} \sum_{j=n(me)+1}^{n_{bl}} \theta_{bl}^{j}\ ,\ n_{(me)} = [\frac{n_{bl} + 1}{2}];$$

- the exponential distribution with a density function

$$h_{bl}(t) = \begin{cases} 0, & t < x_{bl}, x_{bl} \ge 0, \\ \alpha_{bl} \exp[-\alpha_{bl}(t - x_{bl})], & t \ge x_{bl}, \end{cases}$$

where $0 \le \alpha_{bl} < +\infty,\ 0 \le x_{bl} = a_{bl}^{1},$,
the estimates of the unknown parameters of this distribution are:

$$x_{bl} = a_{bl}^1,\ \alpha_{bl} = \frac{1}{\bar{\theta}_{bl} - x_{bl}},$$

- the Weibull's distribution with a density function

$$h_{bl}(t) =$$

$$\begin{cases} 0, & t < x_{bl}, x_{bl} \geq 0, \\ \alpha_{bl}\beta_{bl}(t - x_{bl})^{\beta bl - 1} \exp[-\alpha_{bl}(t - x_{bl})^{\beta bl}], & t \geq x_{bl}, \end{cases}$$

where $0 \leq \alpha_{bl} < +\infty$, $0 \leq \beta_{bl} < +\infty$, $0 \leq x_{bl} = a_{bl}^1,$, the estimates of the unknown parameters of this distribution are:

$$x_{bl} = a_{bl}^1,\ \alpha_{bl} = \frac{n_{bl}}{\sum_{j=1}^{n_{bl}} (\theta_{bl}^j)^{\beta bl}},$$

$$\alpha_{bl} = \frac{\frac{n_{bl}}{\beta_{bl}} + \sum_{j=1}^{nbl} \ln(\theta_{bl}^j - x_{bl})}{\sum_{j=1}^{nbl} (\theta_{bl}^j)^{\hat{\beta} bl} \ln(\theta_{bl}^j - x_{bl})};$$

- the normal distribution with a density function

$$h_{bl}(t) = \frac{1}{\sigma_{bl}\sqrt{2\pi}} \exp[-\frac{(t - m_{bl})^2}{2\sigma_{bl}{}^2}],$$
$$t \in (-\infty, \infty),$$

where $-\infty < m_{bl} < +\infty$, $0 \leq \sigma_{bl} < +\infty$, the estimates of the unknown parameters of this distribution are:

$$m_{bl} = \bar{\theta}_{bl},\ \sigma_{bl}{}^2 = \bar{\sigma}_{bl}^2 = \frac{1}{n_{bl}} \sum_{j=1}^{nbl} (\theta_{bl}^j - m_{bl})^2,$$

- the chimney distribution with a density function

$$h_{bl}(t) = \begin{cases} 0, & t < x_{bl} \\ \frac{a_{bl}}{z_{bl}^1 - x_{bl}}, & x_{bl} \leq t \leq z_{bl}^1 \\ \frac{c_{bl}}{z_{bl}^2 - z_{bl}^1}, & z_{bl}^1 \leq t \leq z_{bl}^2 \\ \frac{d_{bl}}{y_{bl} - z_{bl}^2}, & z_{bl}^2 \leq t \leq y_{bl} \\ 0, & t > y_{bl}, \end{cases}$$

where $0 \leq x_{bl} \leq z_{bl}^1 \leq z_{bl}^2 \leq y_{bl} < +\infty$, $0 \leq q_{bl} < +\infty$, $0 \leq w_{bl} < +\infty$, $a_{bl} \geq 0$, $c_{bl} \geq 0$, $d_{bl} \geq 0$, $a_{bl} + c_{bl} + d_{bl} = 1$.

The estimates of the unknown parameters of this distribution are:

$$x_{bl} = a_{bl}^1,\ y_{bl} = x_{bl} + \bar{r}d,$$

and moreover, if

$$\hat{n}_{bl} = \max_{1 \leq j \leq \bar{r}} \{n_{bl}^j\},\ i = j,\ \text{where}\ j \in \{12,..., \bar{r}\},$$

is such a number of the interval for which $n_{bl}^j = \hat{n}_{bl}$, then:

for $i = 1$

$$z_{bl}^1 = x_{bl} + (i-1)d,\ z_{bl}^2 = x_{bl} + id,$$

$$a_{bl} = 0,\ c_{bl} = \frac{n_{bl}^i}{n_{bl}},\ d_{bl} = \frac{n_{bl}^{i+1} + ... + n_{bl}^{\bar{r}}}{n_{bl}},$$

when $n_{bl}^{i+1} = 0$ or $n_{bl}^{i+1} \neq 0$ and $\frac{n_{bl}^i}{n_{bl}^{i+1}} \geq 3$,

$$z_{bl}^1 = x_{bl} + (i-1)d,\ z_{bl}^2 = x_{bl} + (i+1)d,$$

$$a_{bl} = 0,\ c_{bl} = \frac{n_{bl}^i + n_l^{i+1}}{n_{bl}},\ d_{bl} = \frac{n_{bl}^{i+2} + ... + n_{bl}^{\bar{r}}}{n_{bl}},$$

when $n_{bl}^{i+1} \neq 0$ and $\frac{n_{bl}^i}{n_{bl}^{i+1}} < 3$,

for $i = 2,3,..., \bar{r} - 1$

$$z_{bl}^1 = x_{bl} + (i-1)d,\ z_{bl}^2 = x_{bl} + id,$$

$$a_{bl} = \frac{n_{bl}^1 + ... + n_{bl}^{i-1}}{n_{bl}},\ c_{bl} = \frac{n_{bl}^i}{n_{bl}},\ d_{bl} = \frac{n_{bl}^{i+1} + ... + n_{bl}^{\bar{r}}}{n_{bl}},$$

when $n_{bl}^{i-1} = 0$ or $n_{bl}^{i-1} \neq 0$ and $\frac{n_{bl}^i}{n_{bl}^{i-1}} \geq 3$ and

$$z_{bl}^1 = x_{bl} + (i-1)d,\ z_{bl}^2 = x_{bl} + (i+1)d,$$

$$a_{bl} = \frac{n_{bl}^1 + ... + n_{bl}^{i-1}}{n_{bl}},$$

$$c_{bl} = \frac{n_{bl}^{i} + n_{l}^{i+1}}{n_{bl}},\ d_{bl} = \frac{n_{bl}^{i+2} + ... + n_{bl}^{\bar{r}}}{n_{bl}},$$

when $n_{bl}^{i-1} = 0$ or $n_{bl}^{i-1} \neq 0$ and $\frac{n_{bl}^{i}}{n_{bl}^{i-1}} \geq 3$ and

$$z_{bl}^{1} = x_{bl} + (i-2)d,\ z_{bl}^{2} = x_{bl} + id,$$

$$a_{bl} = \frac{n_{bl}^{1} + ... + n_{bl}^{i-2}}{n_{bl}},$$

$$c_{bl} = \frac{n_{bl}^{i-1} + n_{l}^{i}}{n_{bl}},\ d_{bl} = \frac{n_{bl}^{i+1} + ... + n_{bl}^{\bar{r}}}{n_{bl}},$$

when $n_{bl}^{i-1} \neq 0$ and $\frac{n_{bl}^{i}}{n_{bl}^{i-1}} < 3$ and when $n_{bl}^{i+1} = 0$ or $n_{bl}^{i+1} \neq 0$ and $\frac{n_{bl}^{i}}{n_{bl}^{i+1}} \geq 3,$

$$z_{bl}^{1} = x_{bl} + (i-2)d,\ z_{bl}^{2} = x_{bl} + (i+1)d,$$

$$a_{bl} = \frac{n_{bl}^{1} + ... + n_{bl}^{i-1}}{n_{bl}},$$

$$c_{bl} = \frac{n_{bl}^{i-1} + n_{bl}^{i} + n_{l}^{i+1}}{n_{bl}},\ d_{bl} = \frac{n_{bl}^{i+2} + ... + n_{bl}^{\bar{r}}}{n_{bl}},$$

when $n_{bl}^{i-1} \neq 0$ and $\frac{n_{bl}^{i}}{n_{bl}^{i-1}} < 3$ and when $n_{bl}^{i+1} \neq 0$ and $\frac{n_{bl}^{i}}{n_{bl}^{i+1}} < 3,$

for $i = \bar{r}$

$$z_{bl}^{1} = x_{bl} + (i-1)d,\ z_{bl}^{2} = x_{bl} + id,$$

$$a_{bl} = \frac{n_{bl}^{1} + ... + n_{bl}^{i-1}}{n_{bl}},\ c_{bl} = \frac{n_{bl}^{i}}{n_{bl}},\ d_{bl} = 0,$$

when $n_{bl}^{i-1} = 0$ or $n_{bl}^{i-1} \neq 0$ and $\frac{n_{bl}^{i}}{n_{bl}^{i-1}} \geq 3$,

$$z_{bl}^{1} = x_{bl} + (i-2)d,\ z_{bl}^{2} = x_{bl} + id,$$

$$a_{bl} = \frac{n_{bl}^{1} + ... + n_{bl}^{i-2}}{n_{bl}},\ c_{bl} = \frac{n_{bl}^{i-1} + n_{bl}^{i}}{n_{bl}},\ d_{bl} = 0,$$

when $n_{bl}^{i-1} \neq 0$ and $\frac{n_{bl}^{i}}{n_{bl}^{i-1}} < 3$.

2.2. Identification of distributions of conditional sojourn times in operation states

To formulate and next to verify the non-parametric hypothesis concerning the form of the distribution function $H_{bl}(t)$ of the system conditional sojourn time θ_{bl} at the operation state z_b when the next transition is to the operation state z_l, on the basis of its realizations θ_{bl}^{k}, $k = 1,2,...,n_{bl}$, it is necessary to proceed according to the following scheme:
- to construct and to plot the realization of the histogram of the system conditional sojourn time θ_{bl} at the operation states, defined by the following formula

$$\bar{h}_{n_{bl}}(t) = \frac{n_{bl}^{j}}{n_{bl}} \text{ for } t \in I_j,$$

- to analyze the realization of the histogram, comparing it with the graphs of the density functions $h_{bl}(t)$ of the previously distinguished distributions, to select one of them and to formulate the null hypothesis H_0 and the alternative hypothesis H_A, concerning the unknown form of the distribution function $H_{bl}(t)$ of the conditional sojourn time θ_{bl} in the following form:

H_0 : The system conditional sojourn time θ_{bl} at the operation state z_b when the next transition is to the operations state z_l, has the distribution function $H_{bl}(t)$,

H_A : The system conditional sojourn time θ_{bl} at the operation state z_b when the next transition is to the operations state z_l, has the distribution function different from $H_{bl}(t)$,

- to join each of the intervals I_j that has the number n_{bl}^{j} of realizations is less than 4 either with the neighbour interval I_{j+1} or with the neighbour interval I_{j-1} this way that the numbers of realizations in all intervals are not less than 4,
- to fix a new number of intervals $\bar{\bar{r}}$,

- to determine new intervals $\bar{I}_j = <\bar{a}^j_{bl}, \bar{b}^j_{bl})$, $j = 1,2,..,\bar{\bar{r}}$,
- to fix the numbers $\bar{n}^j_{bl}$ of realizations in new intervals $\bar{I}_j$, $j = 1,2,..,\bar{\bar{r}}$,
- to calculate the hypothetical probabilities that the variable θ_{bl} takes values from the interval $\bar{I}_j$, under the assumption that the hypothesis H_0 is true, i.e. the probabilities

$$p_j = P(\theta_{bl} \in \bar{I}_j) = P(\bar{a}^j_{bl} \le \theta_{bl} < \bar{b}^j_{bl})$$

$$= H_{bl}(\bar{b}^j_{bl}) - H_{bl}(\bar{a}^j_{bl}),\ j = 1,2,...,\bar{\bar{r}},$$

where $H_{bl}(\bar{b}^j_{bl})$ and $H_{bl}(\bar{a}^j_{bl})$ are the values of the distribution function $H_{bl}(t)$ of the random variable θ_{bl} defined in the null hypothesis H_0,
- to calculate the realization of the χ^2(*chi-square*)-Pearson's statistics $U_{n_{bl}}$, according to the formula

$$u_{n_{bl}} = \sum_{j=1}^{\bar{\bar{r}}} \frac{(\bar{n}^j_{bl} - n_{bl}\, p_j)^2}{n_{bl}\, p_j},$$

- to assume the significance level α ($\alpha = 0.01$, $\alpha = 0.02$, $\alpha = 0.05$ or $\alpha = 0.10$) of the test,
- to fix the number $\bar{\bar{r}} - l - 1$ of degrees of freedom, substituting for l for the distinguished distributions respectively the following values: $l = 0$ for the uniform, triangular, double trapezium, quasi-trapezium and chimney distributions, $l = 1$ for the exponential distribution, $l = 2$ for the Weibull's and normal distributions,
- to read from the Tables of the χ^2 – Pearson's distribution the value u_α for the fixed values of the significance level α and the number of degrees of freedom $\bar{\bar{r}} - l - 1$ such that the following equality holds

$$P(U_{n_{bl}} > u_\alpha) = 1 - \alpha,$$

and next to determine the critical domain in the form of the interval $(u_\alpha, +\infty)$ and the acceptance domain in the form of the interval $< 0, u_\alpha >$,
- to compare the obtained value $u_{n_{bl}}$ of the realization of the statistics $U_{n_{bl}}$ with the red from the Tables critical value u_α of the chi-square random variable and to verify previously formulated the null hypothesis H_0 in the following way: if the value $u_{n_{bl}}$ does not belong to the critical domain, i.e. when $u_{n_{bl}} \le u_\alpha$, then we do not reject the hypothesis H_0, otherwise if the value $u_{n_{bl}}$ belongs to the critical domain, i.e. when $u_{n_{bl}} > u_\alpha$, then we reject the hypothesis H_0 in favor of the hypothesis H_A.

3. Application in maritime transport

3.1. The Stena Baltica ferry operation process and its statistical identification

Taking into account the operation process of the considered ferry we distinguish the following as its eighteen operation states:

- an operation state z_1 – loading at Gdynia Port,
- an operation state z_2 – unmooring operations at Gdynia Port,
- an operation state z_3 – leaving Gdynia Port and navigation to "GD" buoy,
- an operation state z_4 – navigation at restricted waters from "GD" buoy to the end of Traffic Separation Scheme,
- an operation state z_5 – navigation at open waters from the end of Traffic Separation Scheme to "Angoring" buoy,
- an operation state z_6 – navigation at restricted waters from "Angoring" buoy to "Verko" Berth at Karlskrona,
- an operation state z_7 – mooring operations at Karlskrona Port,
- an operation state z_8 – unloading at Karlskrona Port,
- an operation state z_9 – loading at Karlskrona Port,
- an operation state z_{10} – unmooring operations at Karlskrone Port,
- an operation state z_{11} – ship turning at Karlskrone Port,
- an operation state z_{12} – leaving Karlskrone Port and navigation at restricted waters to "Angoring" buoy,
- an operation state z_{13} – navigation at open waters from "Angoring" buoy to the entering Traffic Separation Scheme,
- an operation state z_{14} – navigation at restricted waters from the entering Traffic Separation Scheme to "GD" buoy,
- an operation state z_{15} – navigation from "GD" buoy to turning area,

- an operation state z_{16} – ship turning at Gdynia Port,
- an operation state z_{17} – mooring operations at Gdynia Port,
- an operation state z_{18} – unloading at Gdynia Port.

To identify all parameters of Stena Baltica ferry operation process the statistical data about this process is needed. The statistical data that has been collected up to now is given in *Tables 1-7* in [6, Appendix 5A]. In the *Tables 1-7* there are presented the realizations θ_{bl}^k, $k = 1,2,...,42$, for each $b = 1,2,...,17$, $l = b+1$ and $b = 18$, $l = 1$ of the ship operation process conditional sojourn times θ_{bl}, $b = 1,2,...,17$, $l = b+1$ and $b = 18$, $l = 1$ in the state z_b while the next transition is the state z_l during the experiment time $\Theta = 42$ days.

These statistical data allow us, applying the methods and procedures given in the section 2, to formulate and to verify the hypotheses about the conditional distribution functions $H_{bl}(t)$ of the Stena Baltica ferry operation process sojourn times θ_{bl}, $b = 1,2,...,17$, $l = b+1$ and $b = 18$, $l = 1$ in the state z_b while the next transition is to the state z_l on the base of their realizations θ_{bl}^k, $k = 1,2,...,42$.

On the basis of the statistical data, given in the Appendix 5A in [5], the vector of the probabilities of the system initial operation states was evaluated in the following form

$$[p_b(0)] = [1, 0, 0, \ldots, 0, 0].$$

The matrix of the probabilities p_{bl} of transitions from the operation state z_b into the operation state z_l were evaluated as well. Their evaluation are given in the matrix below

$$[p_{bl}]_{18x18} = \begin{bmatrix} 0\,1\,0 \ldots 0\,0 \\ 0\,0\,1 \ldots 0\,0 \\ \ldots \\ 0\,0\,0 \ldots 0\,1 \\ 1\,0\,0 \ldots 0\,0 \end{bmatrix},$$

Next the matrix $[h_{bl}(t)]_{18x18}$ of conditional density functions of the system operation process $Z(t)$ conditional sojourn times θ_{bl} [6, Appendix 5A] were evaluated.

The results of the distributions unknown parameters estimation and the hypotheses testing are as follows:

- the conditional sojourn time θ_{12} have a triangular distribution with the density function

$$h_{12}(t) = \begin{cases} 0, & t < 7, \\ 0.00044t - 0.0031, & 7 \le t < 54, \\ 0.044 - 0.00043t; & 54 \le t < 103, \\ 0, & t \ge 103; \end{cases}$$

- the conditional sojourn time θ_{23} have an exponential distribution with the density function

$$h_{23}(t) = \begin{cases} 0, & t < 1.6 \\ 1.03\exp[-1.03(t-1.6)], & t \ge 1.6; \end{cases}$$

- the conditional sojourn time θ_{34} have a steep chimney distribution with the density function

$$h_{34}(t) = \begin{cases} 0, & t < 29, \\ 0.0278, & 29 \le t < 35, \\ 0.1984, & 35 \le t < 38, \\ 0.0266, & 38 \le t < 47, \\ 0, & t \ge 47; \end{cases}$$

- the conditional sojourn time θ_{45} have a chimney distribution with the density function

$$h_{45}(t) = \begin{cases} 0, & t < 41, \\ 0.0095, & 41 \le t < 46, \\ 0.0762, & 46 \le t < 56, \\ 0.0127, & 56 \le t < 71, \\ 0, & t \ge 71; \end{cases}$$

- the conditional sojourn time θ_{56} have a double trapezium distribution with the density function

$$h_{56}(t) = \begin{cases} 0, & t < 467.8, \\ -0.00004t + 0.0277, & 467.8 \le t \le 525.95, \\ -0.00006t + 0.0397, & 525.95 \le t \le 650.2, \\ 0, & t < 650.2; \end{cases}$$

- the conditional sojourn time θ_{67} have a double trapezium distribution with the density function

$$h_{67}(t) = \begin{cases} 0, & t < 31.9, \\ 0.0067t - 0.1747, & 31.9 \le t \le 37.17, \\ 0.0031t - 0.0395, & 37.7 \le t \le 45.1, \\ 0, & t > 45.1; \end{cases}$$

- the conditional sojourn time θ_{78} have a double trapezium distribution with the density function

$$h_{78}(t) = \begin{cases} 0, & t < 4.5, \\ -0.0183t + 0.2922, & 4.5 \le t \le 7.02, \\ -0.0069t + 0.2122, & 7.02 \le t \le 10.5, \\ 0, & t > 10.5; \end{cases}$$

- the conditional sojourn time θ_{89} have a triangular distribution with the density function

$$h_{89}(t) = \begin{cases} 0, & t < 0, \\ 0.0021t, & 0 \le t \le 21.4 \\ -0.002t + 0.087, & 21.4 \le t \le 44.4, \\ 0, & t > 44.4; \end{cases}$$

- the conditional sojourn time θ_{910} have a double trapezium distribution with the density function

$$h_{910}(t) = \begin{cases} 0, & t < 14.6, \\ 0.0001t + 0.0109, & 14.6 \le t \le 52.26, \\ 0.0062, & 52.2 \le t \le 127.4, \\ 0, & t > 127.4; \end{cases}$$

- the conditional sojourn time θ_{1011} have a double trapezium distribution with the density function

$$h_{1011}(t) = \begin{cases} 0, & t < 1.6, \\ -0.3071t + 1.0014, & 1.6 \le t \le 2.93, \\ 0.0398t - 0.0145, & 2.93 \le t \le 6.4, \\ 0, & t > 6.4; \end{cases}$$

- the conditional sojourn time θ_{1112} have a quasi-trapezium distribution with the density function

$$h_{1112}(t) = \begin{cases} 0, & t < 3.8, \\ -148.899t + 567.4863, & 3.8 \le t \le 3.81, \\ 0.181, & 3.81 \le t \le 4.48, \\ 0.3773t - 1.5094, & 4.48 \le t \le 6.2, \\ 0, & t > 6.2; \end{cases}$$

- the conditional sojourn time θ_{1213} have a triangular distribution with the density function

$$h_{1213}(t) = \begin{cases} 0, & t < 18.7, \\ 0.025t - 0.4675, & 18.7 \le t \le 23.86, \\ -0.012t + 0.4116, & 23.86 \le t \le 34.3, \\ 0, & t > 34.3; \end{cases}$$

- the conditional sojourn time θ_{1314} have a chimney distribution with the density function

$$h_{1314}(t) = \begin{cases} 0, & t < 410, \\ 0.0017, & 410 \le t < 478, \\ 0.0189, & 478 \le t < 512, \\ 0.0024, & 512 \le t < 614, \\ 0, & t \ge 614; \end{cases}$$

- the conditional sojourn time θ_{1415} have a double trapezium distribution with the density function

$$h_{1415}(t) = \begin{cases} 0, & t < 36.8, \\ -0.0006t + 0.0518, & 36.8 \le t \le 50.14, \\ 0.0003t + 0.0062, & 50.14 \le t \le 75.2, \\ 0, & t > 75.2; \end{cases}$$

- the conditional sojourn time θ_{1516} have a chimney distribution with the density function

$$h_{1516}(t) = \begin{cases} 0, & t < 30, \\ 0.0317, & 30 \le t < 33, \\ 0.2698, & 33 \le t < 36, \\ 0.0084, & 36 \le t < 48, \\ 0, & t \ge 48; \end{cases}$$

- the conditional sojourn time θ_{1617} have a triangular distribution with the density function

$$h_{1617}(t) = \begin{cases} 0, & t < 2.7, \\ 0.305t - 0.823, & 2.7 \le t \le 4.52, \\ -0.313t + 1.9719, & 4.52 \le t \le 6.3, \\ 0, & t > 6.3; \end{cases}$$

- the conditional sojourn time θ_{1718} have a double trapezium distribution with the density function

$$h_{1718}(t) = \begin{cases} 0, & t < 2.3, \\ -0.1134t + 0.6707, & 2.3 \le t \le 5.62, \\ 0.0071t - 0.0063, & 5.62 \le t \le 10.7, \\ 0, & t > 10.7; \end{cases}$$

- the conditional sojourn time θ_{181} have a triangular distribution with the density function

$$h_{181}(t) = \begin{cases} 0, & t < 0, \\ 0.0023t, & 0 \le t \le 18.74, \\ -0.0016t + 0.0729, & 18.74 \le t \le 45.59, \\ 0, & t > 45.59. \end{cases}$$

3.2. The Stena Baltica ferry operation process prediction

On the basis of the previous section, the mean values $M_{bl} = E[\theta_{bl}]$, $b, l = 1,2,...,18$, $b \ne l$, (12) in [5] of the system operation process $Z(t)$ conditional sojourn times in particular operation states were determined and there are given by:

$$M_{12} = 54.33,\ M_{23} = 2.57,\ M_{34} = 36.57,$$
$$M_{45} = 52.5,\ M_{56} = 525.95,\ M_{67} = 37.16,$$
$$M_{78} = 7.02,\ M_{89} = 21.43,\ M_{910} = 53.69,$$
$$M_{1011} = 2.93,\ M_{1112} = 4.38,\ M_{1213} = 23.86,$$
$$M_{1314} = 509.69,\ M_{1415} = 50.14,\ M_{1516} = 34.28,$$
$$M_{1617} = 4.52,\ M_{1718} = 5.62,\ M_{181} = 18.74.$$

Hence, by (21) in [5], the unconditional mean sojourn time in the particular operation states are given by:

$$M_1 = E[\theta_1] = p_{12}M_{12} = 1 \cdot 54.33 = 54.33,$$
$$M_2 = E[\theta_2] = p_{23}M_{23} = 1 \cdot 2.57 = 2.57,$$
$$M_3 = E[\theta_3] = p_{34}M_{34} = 1 \cdot 36.57 = 36.57,$$
$$M_4 = E[\theta_4] = p_{45}M_{45} = 1 \cdot 52.5 = 52.5,$$
$$M_5 = E[\theta_5] = p_{56}M_{56} = 1 \cdot 525.95 = 525.95,$$
$$M_6 = E[\theta_6] = p_{67}M_{67} = 1 \cdot 37.16 = 37.16,$$
$$M_7 = E[\theta_7] = p_{78}M_{78} = 1 \cdot 7.02 = 7.02,$$
$$M_8 = E[\theta_8] = p_{89}M_{89} = 1 \cdot 21.43 = 21.43,$$
$$M_9 = E[\theta_9] = p_{910}M_{910} = 1 \cdot 53.69 = 53.69,$$
$$M_{10} = E[\theta_{10}] = p_{1011}M_{1011} = 1 \cdot 2.93 = 2.93,$$
$$M_{11} = E[\theta_{11}] = p_{1112}M_{1112} = 1 \cdot 4.38 = 4.38,$$
$$M_{12} = E[\theta_{12}] = p_{1213}M_{1213} = 1 \cdot 23.86 = 23.86,$$
$$M_{13} = E[\theta_{13}] = p_{1314}M_{1314} = 1 \cdot 509.69 = 509.69,$$
$$M_{14} = E[\theta_{14}] = p_{1415}M_{1415} = 1 \cdot 50.14 = 50.14,$$
$$M_{15} = E[\theta_{15}] = p_{1516}M_{1516} = 1 \cdot 34.28 = 34.28,$$
$$M_{16} = E[\theta_{16}] = p_{1617}M_{1617} = 1 \cdot 4.52 = 4.52,$$
$$M_{17} = E[\theta_{17}] = p_{1718}M_{1718} = 1 \cdot 5.62 = 5.62,$$
$$M_{18} = E[\theta_{18}] = p_{181}M_{181} = 1 \cdot 18.74 = 18.74.$$

Since from the system of equations below (23) in [5] that takes the form

$$\begin{cases} [\pi_1, \pi_2, ..., \pi_{18}] = [\pi_1, \pi_2, ..., \pi_{18}][p_{bl}]_{18x18} \\ \pi_1 + \pi_2 + ... + \pi_{18} = 1, \end{cases}$$

we get

$$\pi_1 = \pi_2 = ... = \pi_{18} = 0.056,$$

then the limit values of the transient probabilities (the portions of time of a week, as the operation process is periodic) $p_b(t)$ at the operational states z_b, according to (22) in [5], are given by

$$p_1 = 0.037,\ p_2 = 0.002,\ p_3 = 0.025,$$
$$p_4 = 0.036,\ p_5 = 0.364,\ p_6 = 0.025,$$
$$p_7 = 0.005,\ p_8 = 0.014,\ p_9 = 0.037,$$
$$p_{10} = 0.002,\ p_{11} = 0.003,\ p_{12} = 0.017,$$
$$p_{13} = 0.354,\ p_{14} = 0.035,\ p_{15} = 0.024,$$
$$p_{16} = 0.003,\ p_{17} = 0.004,\ p_{18} = 0.013.$$

Hence by (26) in [5], the mean values of the system operation process total sojourn times $\hat{\theta}_b$ in the particular operation states z_b, for the operation time $\theta = 1$ month = 720 hours are approximately given by

$$E[\hat{\theta}_1] = p_1\theta = 26.64,\ E[\hat{\theta}_2] = p_2\theta = 1.44,$$
$$E[\hat{\theta}_3] = p_3\theta = 18.00,\ E[\hat{\theta}_4] = p_4\theta = 25.92,$$
$$E[\hat{\theta}_5] = p_5\theta = 262.08,\ E[\hat{\theta}_6] = p_6\theta = 18.00,$$
$$E[\hat{\theta}_7] = p_7\theta = 3.6,\ E[\hat{\theta}_8] = p_8\theta = 10.08,$$
$$E[\hat{\theta}_9] = p_9\theta = 26.64,\ E[\hat{\theta}_{10}] = p_{10}\theta = 1.44,$$
$$E[\hat{\theta}_{11}] = p_{11}\theta = 2.16,\ E[\hat{\theta}_{12}] = p_{12}\theta = 12.24,$$
$$E[\hat{\theta}_{13}] = p_{13}\theta = 25.88,\ E[\hat{\theta}_{14}] = p_{14}\theta = 25.20,$$
$$E[\hat{\theta}_{15}] = p_{15}\theta = 17.28,\ E[\hat{\theta}_{16}] = p_{16}\theta = 2.16,$$
$$E[\hat{\theta}_{17}] = p_{17}\theta = 2.88,\ E[\hat{\theta}_{18}] = p_{18}\theta = 9.36.$$

4. Conclusion

The statistical methods and algorithms for the unknown parameters of the operation process of complex technical systems in variable operation conditions are proposed. Next, these methods are applied to estimating the operation process of Stena Baltica ferry operating between Gdynia Port in Poland and Karsklone Port in Sweden. The proposed methods other very wide applications to port and shipyard transportation systems operation processes characteristics evaluation are obvious. The results are expected to be the basis to the reliability and safety of complex technical systems optimization and their operation processes effectiveness and cost analysis.

Acknowledgements

The paper describes part of the work in the Poland-Singapore Joint Research Project titled “Safety and Reliability of Complex Industrial Systems and Processes” supported by grants from the Poland’s Ministry of Science and Higher Education (MSHE grant No. 63/N-Singapore/2007/0) and the Agency for Science, Technology and Research of Singapore (A*STAR SERC grant No. 072 1340050).

References

[1] Blokus-Roszkowska, A., Guze, S., Kolowrocki, K., Kwiatuszewska-Sarnecka, B. & Soszynska, J. (2008). *Models of safety, reliability and availability evaluation of complex technical systems related to their operation processes*. WP 4 – Task 3.1 – 31.05.2008. Poland-Singapore Joint Research Project, 2007-2010.

[2] Blokus-Roszkowska, A., Guze, S., Kolowrocki, K., Kwiatuszewska-Sarnecka, B. & Soszynska, J. (2008). *Models of safety, reliability and availability evaluation of complex technical systems related to their operation processes*. WP 4 – Task 4.1 – 31.05.2008. Poland-Singapore Joint Research Project, 2007-2010.

[3] Blokus-Roszkowska, A. Guze, S. Kolowrocki, K., Kwiatuszewska-Sarnecka, B., Milczek B. & Soszynska, J. (2009). *Data Mining for Identification and Prediction of Safety and Reliability Characteristics of Complex Industrial Systems and Processes*. WP6 – Task 6.1. Methods-algorithms for evaluating unknown parameters of system operation processes, reliability and safety models. WP6 - Task 6.1.2 – 16.03.2009. Poland-Singapore Joint Research Project, 2007-2010.

[4] Grabski, F. (2002). *Semi-Markov Models of Systems Reliability and Operations Analysis*. Warsaw: Systems Research Institute, Polish Academy of Science.

[5] Kolowrocki, K. & Soszynska, J. (2008). *A general model of technical systems operation processes related to their environment and infrastructure*. WP 2 – Task 2.1 – 31.05.2008. Poland-Singapore Joint Research Project, 2007-2010.

[6] Soszynska, J., Kołowrocki, K., Kamiński, P., Jurdziński, M., Guze, S., Milczek, B., & Golik, P. (2009). *Data Mining for Identification and Prediction of Safety and Reliability Characteristics of Complex Industrial Systems and Processes*. WP6– Task 6.2. Preliminary safety and risk analysis and evaluation of a Stena Baltica ferry in variable operation conditions. Sub-Task 6.2.5 - Sub-Task 6.2.5 – 16.03.2009. Poland-Singapore Joint Research Project, 2007-2010.

[7] Kolowrocki, K. & Kwiatuszewska- Sarnecka, B. (2009). *Data Mining for Identification and Prediction of Safety and Reliability Characteristics of Complex Industrial Systems and Processes*. WP6 – Task 6.1. Methods-algorithms for evaluating unknown parameters of system operation processes, reliability and safety models – a preliminary study. WP6 - Task 6.1.1 – 16.03.2009. Poland-Singapore Joint Research Project, 2007-2010.