Technical University of Denmark

DTU

Prediction of Repair Work Duration for Gas Transport Systems Based on Small Data Samples

Lesnykh, Valery; Litvin, Yuri; Kozin, Igor

DTU Library
Technical Information Center of Denmark

Prediction of repair work duration for gas transport systems based on small data samples

Valery Lesnykh[1,2], Yuriy Litvin[1,2] and Igor Kozine[3]

Abstract

Prediction of the duration of a repair and maintenance project of a gas transport system is an important part of planning activities. Time overruns may lead to lost profit, damage of operator's image, penalties to repair and service companies, and other negative consequences. Numerous sources of uncertainties during the repair and maintenance work result in the uncertainty of how long this work will be performed. Our experience in planning this work suggests that departing from the deterministic point of view and accepting the point that the project duration is a random value is a constructive step towards the preparedness to contingencies, defining liability, and an explicit presentation of risks the operator faces. To support this approach, one needs to construct probability distributions of the durations of the projects. The samples of observed data are usually scarce, which complicates the problem. We suggest using a bootstrap resampling procedure to overcome this difficulty. Gram-Charlier functions and order statistics are employed to approximate the probability distributions. It is demonstrated how to derive a probability distribution of a separate repair project and a larger project consisting of a number of concurrently running subprojects. As soon as this problem is solved, guidance is provided on how to decide about what duration should define the deadline for completion of the whole work. A simple example is provided.

Keywords
Time to repair, gas transport, bootstrap, Gram-Charlier function, order statistics

Introduction

Gas transport systems (GTS) are complex networks of pipelines involving a series of processes and arrays of physical facilities that are distributed over large territories and that require constant support of their work capacity as well as preventive maintenance and repair works.[1,2] Timely performed preventive maintenance and repair works increase the reliability of gas supply and create added value for business. On the opposite, interruption of gas supply to customers because of failures or planned maintenance and repair works may lead to lost profit and limited opportunities to redistribute gas flows. This is why the owners and operators of the system strive, on the one hand, to decrease the disconnection time of some customers from the network and, on the other hand, to provide high quality of maintenance and repair within as short periods of time as possible.

Each maintenance and repair work is a rather complex project that depends on numerous factors and the duration of these activities is difficult to predict precisely. Exceeding the time planned is a persistent issue that should be properly addressed by an improved ability to predict it, including the use of adequate mathematical models. Our modelling approach is to regard time to repair (TTR) as a random variable dependent on many parameters

[1] Science Research Institute of Economics and Management in Gas Industry, Moscow, Russia
[2] National Research University "High School of Economics", Energy Institute, Moscow, Russia
[3] Technical University of Denmark, Department of Management Engineering, Copenhagen

characterizing the involved subsystems. A characteristic feature of the repair and maintenance process is that it is usually performed concurrently on many stretches and subsystems by several repair teams. This reduces the duration of TTR but makes the assessment of the disconnection time of a GTS' subsystem more difficult. Figure 1 depicts a typical GTS' subsystem that is disconnected from the network during maintenance and repair.
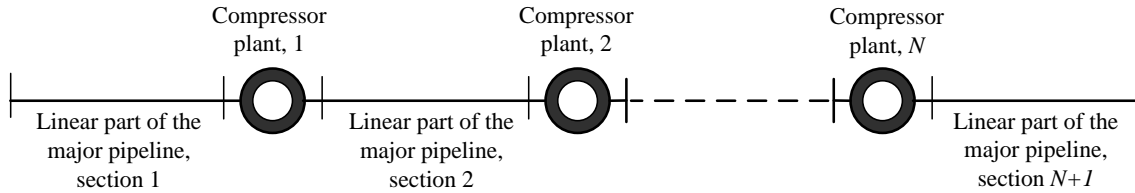

Figure 1. A typical GTS subsystem that can be disconnected to be repaired or maintained

The focus of this paper is the prediction of a possible increase of a GTS' subsystem disconnection time compared to the time planned to perform maintenance and repair work. For this purpose a probability distribution function of the duration of repair and maintenance works is assessed. This knowledge allows the operator to define penalties that can be imposed to the repair organization if it delays the completion of the work. The scale of the penalties is balanced against losses that can be incurred from the excess downtime.

Repair and maintenance works at a GTS' subsystem can be planned as any other project in accordance with known project management methodologies (see, for example [3]). Along with the uniqueness related to geographic and natural conditions, a majority of GTS' repair projects have a number of common characteristics (Table I).[4,5] Such typification of the project characteristics is possible because of the use of similar components and units in the subsystems of the GTS. The typified structure allows the operator to specify common requirements to the repair project management and use accumulated statistical data from previous projects to better predict the TTR and manage new projects.[5]

Table I. Main common characteristics of a standard repair project

#	Characteristics of a standard project
1	Similar work structure corresponding to the accepted repair technology
2	Durations and man-hours of the phases of a repair project that can be linearly scaled up or scaled down
3	Similar resource types to complete works at similar project phases
4	Opportunity to accumulate statistical data for their future use in planned assessments and monitoring

As the duration of each concurrent repair or maintenance project is a random variable, the total duration of the whole project is also a random variable determined by the duration of the longest subproject. Thus, the assessment of the probability distribution of the maximal time among the concurrent repair activities becomes the primary objective of the mathematical modelling. After the probability distribution has been assessed, other probability measures can be derived and conclusions made on the duration of the TTR and probabilities associated with time overruns.

This type of problem statement is well-known and different solutions have been suggested to solve it. For example, Gnedenko, et al.[6] and O'Connor and Kleyner[7] suggest a number of mathematical models that allow the assessment of this and other reliability measures of technical systems. It should be noted that the main objective of these models is to determine an optimal moment for a subsystem disconnection and for the start of repair work. Jacobson[8] describes a widely-used approach to the planning of repair work duration based on deterministic and predefined performance times of repair teams and other types of resources used under repair. An obvious advantage of this approach is its simplicity, while the ignorance of the stochastic nature of these activities may results in predictions that deviate substantially from those observed in reality.

There are other quantitative and qualitative approaches to assessing the risk of exceeding the planned times to project completion. To name some, these are the project evaluation and review technique (PERT), the earned value methodology, Monte Carlo simulation, and various stochastic network models used for planning and managing project.[3,9,10] These methods help decision makers to choose an action path taking into account possible consequences and risks, and, in particular, the risk of time overruns and resources needed to reduce this time given it has been exceeded. In fact these models help also understand the value of taking risk and justify the "risk appetite"[11] the managers would like accept when managing GTS repair projects.

Significant help in the risk assessment can be obtained from various mathematical models of project' processes.[10] However, the problem of having reliable data as inputs for these models is common; and usually only small samples of data are available on which the assessments of inputs are based. A majority of the models require the following information as input:

- detailed project schedule,

- various statistics including ranges of possible deviations from the planned project durations, and

- probability distributions of the execution times of some critical tasks performed in the projects.

Indeed, the analysts usually have to work with limited information about the execution times of repair projects. In this paper it is assumed that scarce samples of TTRs are available and the only valid assumption about the form of the probability distribution is that it is unimodal, meaning that the probability density function has a single peak. The assumption of the unimodality is well supported by observation.[5]

The objective of this paper is to develop a method to set up the time planned for performing standard repair projects and to assess the probability of exceeding this time based on small

samples of available data. The assessment method of our choice is based on the bootstrap resampling procedure called Gram-Charlier method, and we suggest using order statistics to fulfil the objective.

Assessment of point and interval statistics

The first step of the approach is the assessment of the mean, standard deviation, and other statistics of the TTR of a GTS' subsystem as well as the confidence intervals. These assessments are supposed to be done assuming that only small samples of TTRs are available and the form of the probability distribution is unimodal.

Let n values of TTRs are available, $t_1, t_2, \dots t_n$, and without loss in generality they are ordered so that $t_1 \leq t_2, \dots, \leq t_n$. The small sample size and lack of information about the distribution function type do not allow deriving reliable probabilistic quantities and statistics by the use of classical methods of mathematical statistics. Assessments made directly on small samples lead to the bias of the statistics point estimator when compared to the assessments of the statistics made on large samples. This makes the assessments unstable and unreproducible when another small sample is used.

There are a number of alternative approaches to making the assessments of probabilities and probabilistic quantities based on scarce data and partial information. Notable are statistical models using Bayesian statistics and the theories of imprecise probabilities. The former requires the specification of prior distributions that can then be updated through the Bayes formula by observations. The latter can produce measures of chance or uncertainty based on scarce and partial data, though without sharp numerical probabilities (can be, for example, interval-valued).

There are simple reasons for deselecting these two for our purpose. Bayesian assessments are heavily dependent on the choice of prior distributions that is subjective and in general varying among subjects. Despite the updating rule (Bayes theorem) will produce posteriors that are closer to the true value, small samples cannot guarantee fast convergence to it. Thus, the Bayesian inference suffers the same problem as classical methods of mathematical statistics if samples are small: instability and irreproducibility. The use of non-informative priors is an option to avoid subjective judgements. Nevertheless, the posteriors are heavily influenced by the priors if the samples are small and can be far from the true value. The need to choose conjugate probability distributions for the updating contributes to the uncertainty.

The theories of imprecise probabilities can indeed produce stable and reproducible lower and upper bounds of probabilistic quantities. For example, multinomial and the beta-Bernoulli imprecise models[12-15] can be used as models of statistical inference to derive the bounds for probabilities given small samples. To avoid being dependent on subjective priors, we may assuming complete ignorance as the prior state of knowledge. In case we are compelled to employ the prior model that is the vacuous prior probability (the lower bound is equal to 0, while the upper is equal to 1). In this case, the bounds will be stable and reproducible. The major problems with the models of imprecise probabilities is that the interpretation of the bounds is rather problematic and the bounds are often too broad to be practical. To our best knowledge, the only available interpretation of the bounds is behavioural in a form of betting rates[15], which is problematic to employ in order to translate into the probabilities of exceeding the bounds.

As it has been stated in the introduction, the method of our choice is bootstrapping[16,17] that is based on bootstrap resampling procedures. Bootstrapping was suggested by Efron and used first as a method of statistics bias estimation based on bootstrap samples.[18] The main idea of bootstrapping lies in the fact that an analysed sample has the necessary information about a true distribution of a random variable. Generating a set of bootstrap samples of large sizes, say, 5 000 - 10 000 realisations in each set, and estimating statistics on this basis, will make them stable, independent of any priors and the bounds of confidence intervals interpretable in the conventional way. In fact, in many cases results obtained with classical methods of mathematical statistics and bootstrapping coincide.[18,19]

The 'technical' advantages of using bootstrap procedures are the following:

1. Bias correction of random variable statistics by bootstrapping: in some particular cases it is possible to derive analytical expressions for bias correction.[20]

2. Derivation of interval statistics $\left[\overline{\theta}^* - w \cdot \overline{\sigma}^* ; \overline{\theta}^* + w \cdot \overline{\sigma}^* \right]$, where the notation has the following meaning:

$$\overline{\theta}^* = \frac{1}{m}\sum_{i=1}^m \theta_i^* \tag{1}$$

is an average of m statistics θ_i^* each of which is obtained on a bootstrap sample;

$$\overline{\sigma}^* = [\frac{1}{m-1}\sum_{i=1}^m (\theta_i^* - \overline{\theta}^*)^2]^{0,5} \tag{2}$$

is the statistics' standard deviation; and w is a multiplier defining the breadth of the interval. The asterisk notation "*" indicates that the statistics are obtained on bootstrap samples.

3. Construction of confidence intervals (CIs) without having to introduce the assumption of a normal distribution: a set of θ_i^* values (i=1, …, m) arranged in the ascending order allows calculating bootstrap percentiles of statistics' distribution.[21] In practice, the use of normal distributions for constructing CIs may lead to significant errors even though there is an insignificant deviation from this assumption.[22]

4. The ability to prove statistical hypotheses.[23]

The bootstrap algorithm for the calculation of point-valued and interval-valued statistics consists in the following steps:

1. Given a small sample of TTRs $\overline{t} = (t_1, t_2, t_3, …, t_n)$, bootstrap samples $\overline{t}_i^* = (t_{i1}^*, t_{i2}^*, …, t_{in}^*)$, $i = 1, …, m$, of the same size n are generated (m is usually in the range [5 000–10 000]). This is done by random sampling of the variables t_k, $k = 1, …, n$ from the original sample $\overline{t}$ with further replacement. It should be noted that bootstrap samples are adjusted to the distribution shape that agrees with the empirical distribution of the original sample. Though, in some cases the bootstrap distribution mean may be biased relative to the original distribution.[24]

2. The statistics of interest are calculated for each bootstrap sample as arithmetic averages of the corresponding functions calculated based on the sampled random values. For example, mean TTR (MTTR) is calculated in the following way:

$$M_i^* = \frac{1}{n}\sum_{j=1}^{n} t_{ij}^*, \qquad i = 1, \dots, m;$$

while the standard deviation of the TTR, σ_i^*, is defined by the formula:

$$\sigma_i^* = [\frac{1}{n-1}\sum_{j=1}^{n}(t_{ij}^* - M_i^*)^2]^{0.5}$$

The moments of higher orders can be calculated in a similar way as arithmetic averages.

3. By plugging the values M_i^* and σ_i^*, $i = 1, \dots, m$ into formula (1) and (2), the bias-corrected statistics $\bar{M}^*$ and $\bar{\sigma}^*$ are obtained.

4. The CIs of bootstrap statistics are calculated. If one wants to have a 95% CI, 2.5% and 97.5% bootstrap percentiles are calculated using the values of the statistics such, for example, as MTTRs M_i^*, $i = 1, \dots, m$ obtained at the previous step. Thus, the 2.5% bootstrap percentile becomes the lower bound of the 95% confidence interval, while the 97.5% bootstrap percentile becomes the upper bound of it.[25]

The bootstrap percentiles method is not the only one that can be used for the construction of CIs. Shitikov and Rosenberg[26] compare seven different approaches to the determination of CIs, including jackknife, percentiles, and the main intervals method. The conclusion was that CI values calculated with all these seven methods are close to each other. The percentiles method has been chosen in this paper because of its simplicity.

It is important to stress that the above algorithm produces bias-corrected random variable statistics which can be used at the next stage of the method described in the following section in the form of the parameters for Gram-Charlier approximate series.

Assessment of the probability of exceeding the planned duration of a repair project

The main objective of this stage is to assess the probability of exceeding the planned duration of a repair project at a separate GTS subsystem. This can be done by generating the statistics shown in the previous section and then by approximating the unknown distribution function of the TTR.

There exists a great variety of approaches used for the approximation of probability distributions of random variables, nevertheless, Gram-Charlier series has become a widely-used method.[25] There exists two types of Gram-Charlier series: *A* and *B* Gram-Charlier series. It is appropriate to use Gram–Charlier A series to construct distributions that are close to normal, however, that are not. It has been proven that the TTR a subsystem of the GTS is governed by such a distribution that is close to normal.[5,25] Gram–Charlier B series is usually used to construct distributions that are close to a Poisson distribution.[25] Further in this paper only Gram–Charlier A series will be in focus.

In general, Gram–Charlier A series includes an infinite number of terms and has the following probability density function (pdf):

$$f_A(x) = \varphi(x) - \frac{r_3}{6}\cdot\varphi^{(3)}(x) + \frac{r_4 - 3}{24}\cdot\varphi^{(4)}(x) - \frac{r_5 - 10r_3}{120}\cdot\varphi^{(5)}(x)$$

$$+ \frac{r_6 - 15r_4 + 30}{720} \cdot \varphi^{(6)}(x) - \cdots,$$

where

$f_A(x)$ is the value of pdf at point x;

$\varphi(x)$ is the standard normal distribution $(\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}})$;

r_q is the q–th moment of the random variable;

$\varphi^{(q)}(x)$ is the derivative of q-order of the standard normal distribution:

$$\varphi^{(q)}(x) = (-1)^q \cdot H_q(x) \cdot \varphi(x),$$

where $H_q(x)$ is the Chebyshev–Hermite polynomial. The first five Chebyshev–Hermite polynomials are of the following form:

$H_0(x) = 1,$

$H_1(x) = 1,$

$H_2(x) = (x^2 - 1),$

$H_3(x) = (x^3 - 3 \cdot x),$

$H_4(x) = (x^4 - 6 \cdot x^2 + 3).$

For a majority of applications it is enough to use only three first terms of Gram–Charlier approximating series: [25]

$$f_A(x) = \varphi(x) - \frac{r_3}{6} \cdot \varphi^{(3)}(x) + \frac{r_4 - 3}{24} \cdot \varphi^{(4)}. \tag{3}$$

Thus, by using this series it is possible to construct distribution functions that are close to the normal distribution though having a non-zero skewness and kurtosis.

The cumulative probability distribution function $F_A(x)$ of Gram–Charlier series is the following:

$$F_A(x) = F(x) - \frac{r_3}{6} \cdot F^{(3)}(x) + \frac{r_4 - 3}{24} \cdot F^{(4)}(x), \tag{4}$$

where

$F_A(x)$ is the value of the cumulative distribution function at point x;

$F(x)$ is the standard normal cumulative distribution function; and

$F^{(q)}(x)$ is the derivative of q-order of $F(x)$.

Formula (3) and (4) allow constructing better approximations of unknown distributions in comparison with asymptotic normal distributions when only small samples are available. It is important to note that unknown values in (3) and (4), r_3 and r_4, are bias-corrected statistics which can be obtained with the help of the bootstrap procedure described above.

As soon as the probability distribution of the TTR is known and a deadline for completing the repair work is provided, the probability of exceeding the deadline can be easily

computed. To do so, a quantile for the TTR can be computed and then used to set up the deadline for repair work. It provides also the risk of exceeding it. If a 95% quantile is chosen, which is often the case, the probability of exceeding it amounts to 0.05.

Thus, by using the suggested type of approximate Gram-Charlier A series, it is possible to construct the probability distribution function of TTR and to assess the risk of completing the repair works beyond the planned time.

Assessment of the probability of exceeding the planned duration of multiple repair projects

The probability assessment algorithm described above makes it possible to predict an increase in the TTR for a repair and maintenance project that is considered indivisible into subprojects possibly run in parallel. The problem becomes significantly complicated in the case when a GTS fragment consists of several subsystems and TTRs of all the subsystems are governed by different probability distributions. In this case, the total TTR of the GTS fragment is determined by the longest TTR among the subprojects. The assessment of the total TTR is a nontrivial problem which can be solved with the methods of order statistics.[27]

Let a GTS fragment subjected to repair works consist of $N+1$ consecutive linear parts of the pipeline and N compressor plants (Figure 1). A TTR of a subsystem is considered independent of any other subsystem's TTR. $F_i(t)$ is the cumulative distribution function of the TTR of an i-th subsystem ($i = 1, ..., 2N + 1$), while $f_i(t)$ is the pdf. Let t_i stand for the TTR of an i-th subsystem. If all t_i are arranged in the ascending order $t_1 \leq t_2 \leq \cdots \leq t_{2N+1}$, then t_{2N+1} is the value that determines the total downtime of the GTS fragment. That is $TTR_{GTS} = max_i(t_i), \ i = 1, ..., 2N + 1$

If all t_i are identically distributed random variables, the problem of determining the distribution of $max_i(t_i)$ is rather simple and its solution can be found in a number of textbooks (see, for example [27]). However, this problem becomes more difficult to solve if each t_i is a non-identically distributed random variable. A solution to this problem is provided by Balakrishnan.[28] The pdf of t_i , $i = 1, ..., 2N + 1$ is defined as follows:

$$f_i(t) = \frac{1}{(i-1)!(2N+1-i)!} \cdot Per(A_{2N+1}), \tag{5}$$

where $per (A_{2N+1})$ is the permanent of matrix A_{2N+1}:[29]

$$A_{2N+1} = \begin{pmatrix} F_1(t) & \dots & F_{2N+1}(t) \\ f_1(t) & \dots & f_{2N+1}(t) \end{pmatrix} \begin{matrix} \} \, 2N \\ \} \, 1 \end{matrix} \tag{6}$$

There are several approaches to simplifying the computation of the permanent of a matrix. We use Ryser's formula[30]:

$$Per(A_{2N+1}) = \sum S(A_{2N+1}) - \sum S(A_{2N+1}^1) + \sum S(A_{2N+1}^2) + \dots + (-1)^{2N} \sum S(A_{2N+1}^{2N}).$$

Where

$$\sum S(A_{2N+1}) = \sum_{i=1}^{2N+1} F_i(t)^{2N} \times \sum_{i=1}^{2N+1} f_i(t),$$

and $S\left(A_{2N+1}^i\right)$ is a matrix obtained from A in which the i-th column is populated by zeros; while $\sum S\left(A_{2N+1}^i\right)$ is the value obtained as the product of the summed up components in each row (an example is below).

Examples of applications of (5) and (6) can be found in Litvin.[31,32] These formulae allow assessing the distribution of a process duration with any finite number of simultaneously executed repair works characterized by different distributions.

If all TTRs of the parallel works are assumed to have the same distribution, e.g. repair works of n similar compressor plants, then the pdf, $f_{n:n}(t)$, and cumulative distribution function, $F_{n:n}(t)$, of the whole duration of the repair project are derived, correspondingly, as follows:

$$f_{n:n}(t) = n \cdot [F(t)]^{n-1} \cdot f(t),$$

$$F_{n:n}(t) = [F(t)]^n.$$

As an example, let us consider a subsystem consisting of two linear parts and a compressor plant with cumulative distribution functions $F_1(t)$, $F_2(t)$ and $F_3(t)$ and pdfs $f_1(t)$, $f_2(t)$ and $f_3(t)$. In this case matrix (10) has the following form:

$$A_3(t) = \begin{pmatrix} F_1(t) & F_2(t) & F_3(t) \\ F_1(t) & F_2(t) & F_3(t) \\ f_1(t) & f_2(t) & f_3(t) \end{pmatrix}.$$

By using Ryser's formula[30], which significantly simplifies the process of the permanents' computation, one can get the following:

$$\sum S(A_3) = [F_1(t) + F_2(t) + F_3(t)]^2 [f_1(t) + f_2(t) + f_3(t)].$$

$$\sum S(A_3^1) = [F_2(t) + F_3(t)]^2 [f_2(t) + f_3(t)] + [F_1(t) + F_3(t)]^2 [f_1(t) + f_3(t)]$$

$$+ [F_1(t) + F_2(t)]^2 [f_1(t) + f_2(t)]$$

$$\sum S(A_3^2) = F_1(t)F_1(t)f_1(t) + F_2(t)F_2(t)f_2(t) + F_3(t)F_3(t)f_3(t).$$

Finally,

$$Per(A_3) = \sum S(A_3) - \sum S(A_3^1) + \sum S(A_3^2) =$$

$$2[F_2(t)F_3(t)f_1(t) + F_1(t)F_2(t)f_3(t) + F_1(t)F_3(t)f_2(t)].$$

$$f_{3:3}(t) = F_2(t)F_3(t)f_1(t) + F_1(t)F_2(t)f_3(t) + F_1(t)F_3(t)f_2(t). \tag{7}$$

This last formula is the pdf of the total TTR for the GTS fragment consisting of the three subsystems. It should be noted that this expression can be also derived in the following way:

$$f_{3:3}(t) = \frac{d}{dt}\left(\prod_{i=1}^{3} F_i(t)\right).$$

To define the pdfs of the TTRs for each subsystem separately formula (5) should be used.

Expression (7) allows the analyst to evaluate the pdf of the TTR of a subsystem of a GTS consisting of three units. As the distribution function is known, the assessment of the quantiles becomes a rather straightforward exercise. Knowing the quantiles will provide input to deciding on how long time the repair project should last and what is the probability of that the work will indeed be performed within this interval. It is clear that the probability of exceeding the deadline becomes known as well given the quantiles.

An example

As above, consider a part of a GTS consisting of two linear sections of a pipeline and a compressor plant. Table II contains 15 fictitious values of the TTRs for each subsystem that presumably have been collected in the past for similar subsystems. The TTRs are ordered in the ascending order.

Table II. TTRs the three subsystems

Project	Factual repair project duration, days														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Major pipeline section 1	20	21	22	23	25	29	30	34	35	42	43	43	55	61	70
Compressor plant 1	7	9	10	10	10	11	11	12	13	17	18	18	21	21	24
Major pipeline section 2	33	40	41	42	43	47	54	56	56	58	58	58	63	68	69

To calculate the statistics for these projects durations we use the bootstrap procedure that produces *m=10 000* samples of TTRs for each GTS subsystem. Each bootstrap sample consists of 15 TTRs bootstrapped from the original sample by the use of Monte Carlo simulation. By applying formulae (1) and (2) and the bootstrap percentiles method we obtain bias-corrected statistics collected in Table III. This table includes both the statistics needed to apply the Gram-Charlier method and other statistics that can be used for risk assessment.

Table III. Point and interval-valued statistics for GTS outages

Project	Statistics	Sign	Measurement unit	Point value	Bias-corrected point value	CI, 95%
Major	Mean	M_1	*days*	36.9	36.9	[29.8; 44.7]

Project	Statistics	Sign	Measurement unit	Point value	Bias-corrected point value	CI, 95%
pipeline section 1	Variance	D_1	$days^2$	218.4	203.2	[79.9; 365.9]
	Standard deviation	σ_1	$days$	14.5	13.8	[8.9; 19.1]
	Third main statistic	$r_{1,3}$	-	$64.8 \cdot 10^{-2}$	$57.8 \cdot 10^{-2}$	$[-5.6 \cdot 10^{-2};$ $149.9 \cdot 10^{-2}]$
	Forth main statistic	$r_{1,4}$	-	$23.1 \cdot 10^{-1}$	$23.5 \cdot 10^{-1}$	$[12.7 \cdot 10^{-1};$ $47.4 \cdot 10^{-1}]$
Compressor plant 1	Mean	M_2	$days$	14.2	14.2	[11.7; 16.9]
	Variance	D_2	$days^2$	25.7	23.8	[12.6; 39.7]
	Standard deviation	σ_2	$days$	5.0	4.8	[3.6; 6.3]
	Third main statistic	$r_{2,3}$	-	$40.6 \cdot 10^{-2}$	$39.1 \cdot 10^{-2}$	$[-34.4 \cdot 10^{-2};$ $127.2 \cdot 10^{-2}]$
	Forth main statistic	$r_{2,4}$	-	$18.1 \cdot 10^{-1}$	$19.4 \cdot 10^{-1}$	$[11.6 \cdot 10^{-1};$ $36.9 \cdot 10^{-1}]$
Major pipeline section 2	Mean	M_3	$days$	52.4	52.4	[47.1; 57.64]
	Variance	D_3	$days^2$	109.2	101.2	[55.0; 170.6]
	Standard deviation	σ_3	$days$	10.4	9.9	[7.4; 13.1]
	Third main statistic	$r_{3,3}$	-	$-13.5 \cdot 10^{-2}$	$-14.8 \cdot 10^{-2}$	$[-87.0 \cdot 10^{-2};$ $57.3 \cdot 10^{-2}]$
	Forth main statistic	$r_{3,4}$	-	$17.9 \cdot 10^{-1}$	$18.9 \cdot 10^{-1}$	$[12.4 \cdot 10^{-1};$ $29.8 \cdot 10^{-1}]$

The values in Table III are an informative input to decision making even though they are not further mathematically treated. The data demonstrate that the maximum mean repair project duration is attributable to the third subproject (M_3=52.4 *days*), while the mean duration of the second subproject is minimal (M_2=14.2 *days*). The duration of the third subproject dominates significantly the second one. On this ground the operator may decide to postpone the starting date for the second subproject, which will reduce the disconnection time of the customers fed from this section of the pipeline. By comparing the rations of the standard deviations and MTTRs we can conclude that the largest deviation of TTR from the mean value is attributable to the first project, which may be a warning to be more attentive to the causes that make it possible.

The use of the Gram-Charlier method allows deriving the pdf of the TTR for each repair subproject. This is done by plugging the statistics presented in Table IV into formula (3).

Gram-Charlier A series functions of all projects have the following forms:

- *for the major pipeline section 1*:

$$f_{1,A}(x) = f(x) - \frac{57.8 \cdot 10^{-2}}{6} \cdot (x^3 - 3x) \cdot f(x) + \frac{23.5 \cdot 10^{-1} - 3}{24} \cdot$$
$$(x^4 - 6x^2 + 3) \cdot f(x) \tag{8}$$

- *for the compressor plant 1*:

$$f_{2,A}(x) = f(x) - \frac{39.1 \cdot 10^{-2}}{6} \cdot (x^3 - 3x) \cdot f(x) + \frac{19.4 \cdot 10^{-1} - 3}{24} \cdot$$
$$\cdot (x^4 - 6x^2 + 3) \cdot f(x), \tag{9}$$

- *for the major pipeline section 2*:

$$f_{3,A}(x) = f(x) - \frac{-14.8 \cdot 10^{-2}}{6} \cdot (x^3 - 3x) \cdot f(x) + \frac{18.9 \cdot 10^{-1} - 3}{24} \cdot$$
$$\cdot (x^4 - 6x^2 + 3) \cdot f(x). \tag{10}$$

The three pdfs, (8)-(10), are depicted in Figure 2. Based on the obtained results we can conclude, for example, the following. If the MTTR of section 1(36.9 days) is chosen to serve as the planned time to perform the repair and maintenance work for this section, in 30% of cases we can expect an increase of the time up to 44 days, and in 10% of cases this time can reach 58 days. In case the operator plans the completion of the work within 63 days, on average in 5% of cases this deadline will be exceeded. The decision maker has to accept some level of risk, which depends on the operators risk appetite. The decision on choosing the deadlines has to be made for each subproject. As soon as this is done, the probability of not completing the whole repair and maintenance project in time can be assessed. Alternatively, the decision maker can decide first on the level of residual risk (probability) the company can accept for the whole project; and then the durations of the subprojects can be determined, so that the total risk of the project is preserved.
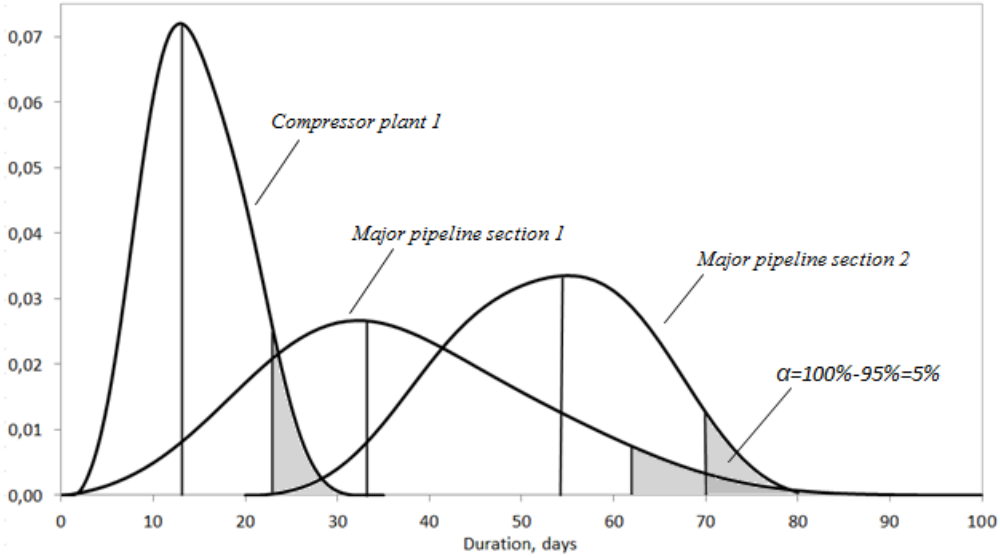
Figure 2. Pdfs of repair projects' durations

To calculate the pdf of the disconnection time for the whole project consisting of the three subprojects, Gram-Charlier functions (8)-(10) are plugged into formula (7). The resulting bulky function is plotted in Figure 3.
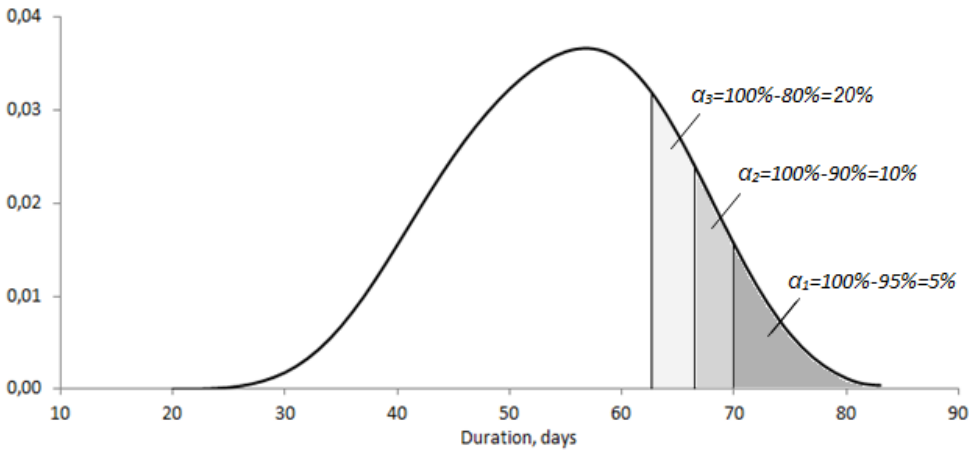

Figure 3. The pdf of the disconnection time for the GTS subsystem (section 1 – compressor – section 2)

The MTTR for the whole project consisting of the three subprojects can now be calculated and is equal to 57 days. On average, in 10% of cases the duration of the repair and maintenance project will exceed 67 days, while in 5% of cases this time will be greater than 70 days. 80%, 90%, and 95% quantiles for the durations of the subprojects and the whole project are given in Table IV.

Table IV. Quantiles of the three repair projects durations

Project	Distribution quantiles		
	$\alpha_3=80\%$	$\alpha_2=90\%$	$\alpha_1=95\%$
Major pipeline section 1	50 days	58 days	64 days
Compressor plant 1	19 days	21 days	23 days
Major pipeline section 2	62 days	66 days	69 days
Total GTS subsystem	**63 days**	**67 days**	**70 days**

Concluding note

The planning and execution of repair and maintenance projects are costly activities. The GTS operator strives to reduce these costs, justify penalties to repair companies in case of delays while attempting to shorten the times to do the work. Numerous sources of uncertainties during the repair and maintenance work result in the uncertainty of how long this work will be performed. While considering time to repair a random value, the operator often lacks data to reliably determine the probability distribution of this value. A bootstrap resampling procedure is the core of the approach of our choice that has been applied to solve the problem. It has demonstrated to be workable and at present is used by one of the major gas suppliers in the Russian Federation.

References

1. Mokhatab S, Poe AW and Mak JY. *Handbook of Natural Gas Transmission and Processing*. Oxford: Gulf Professional Publishing is an imprint of Elsevier, 2015

2. Alimov SV, Nefedov SV, Milko-Butovsky GA, et al. *Optimization of long-term planning for diagnostics and repair of gas main linear parts within the GTS. Technical Condition and Integrity Control System of the Gazprom JSC*. Proceedings of gas industry research and development 2013. UDC 622.691.4.053

3. Larson E and Gray C. *Project Management: The Managerial Process*. 5th ed. New York: McGraw-Hill Companies, 2011.

4. Reshetnikov AD. Methods and technological processes of repair of main gas pipe lines sited in marshy areas. *Doctoral dissertation in technical sciences*. Moscow 2006. (in Rissian).

5. Litvin Y. Project and portfolio management in major repairs and underwater pipe lines diagnostic company. Part 1. Functional-purpose approach. *Problems of economics and management of oil and gas complex* 2009; 12: 20-26. (in Russian)

6. Gnedenko BV, Belyayev YK and Solovyev AD. *Mathematical Methods of Reliability Theory*. New York: Academic Press 1969.

7. O'Connor PDT and Kleyner A. *Practical Reliability Engineering*. 5th ed. New Delphi: Wiley 2012.

8. Jacobson MO. *A Unified System of Planning Preventive Maintenance and Exploitation of the Equipment of Machinery Construction Plants*. Moscow: Izdatelstvo Mashinoctroenie 1967. (in Russian)

9. Fleming, CD, Koppelman JM. *Earned Value Project Management*. 2nd ed. Pennsylvania: Project Management Institute 2000.

10. Golenko-Ginzburg D. *Stochastic network models in innovative projecting. Vol. 1: Network projects with deterministic structure*: Monograph. - Yelm, WA, USA: Science Book Publishing House 2014..

11. COSO. *Enterprise Risk Management – Understanding and Communicating Risk Appetite*. COSO, 2012.

12. Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall.

13. Utkin, L. (2004). Probabilities of judgments provided by unknown experts by using the imprecise Dirichlet model. Risk, Decision and Policy, 9(4), 391–4000.

14. Utkin, L. (2006). Cautious reliability analysis of multi-state and continuum-state systems based on the imprecise Dirichlet model. International Journal of Reliability, Quality and Safety Engineering, 13(5), 433–453.

15. Walley, P. (1996). Inferences from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society*, *58*(1), 3–57.

16. Athreya KB. Bootstrap of the mean in the infinite variance case. *Ann. Stat*, 1987; 15(2):724-731.

17. DiCiccio TJ, Romano JP. A review of bootstrap confidence intervals (with discussions). *J. R. Stat. Soc.,* 1988; B.50:538-554.

18. Efron B. Bootstrap methods: Another look at jackknife. Ann. Stat, 1979; 7(1):1-26.

19. Verbec M. *A Guide to Modern Econometrics*, 2nd ed. New York: John Wiley& Sons 2004.

20. Steck H and Jaakkola TS. Bias-Corrected Bootstrap and Model Uncertainty. *Advances in Neural Information Processing Systems* 16, 2003.

21. Efron B and Tibshirani RJ. *An introduction to the bootstrap*. N.Y.: Chapman&Hall 1993.

22. Orlov AI. Nonparametric point and interval estimation characteristics of the distribution. *Zavodskaya Laboratoriya* 2004; 70(5): 65-70.

23. Simar L and Wilson W. *Statistical inference in non-parametric frontier models*. New York: Oxford University Press 2008.

24. Hesterberg T, Monaghan S, et al. *The Practice of Business Statistics. Chapter 18, Bootstrap methods and permutation tests*. USA: W. H. Freeman and Company 2003; p. 18.1-18.63.

25. Willemse I. *Statistical Methods and Calculation Skills*. 3rd ed. Cape Town: Juta & Co Ltd 2009 .

26. Shitikov VK and Rosenberg GS. *Randomisation in Bootstrap: Statistical Analysis in Biology and Ecology*. Toliatti: Izdatelstvovo Kasandra 2013. (in Russian)

27. Feller W. *An Introduction to Probability Theory and Its Applications*, V 2. John Wiley & Sons 1971.

28. Balakrishnan N. Permanents, order statistics, outliers, and robustness. *Rev. Mat. Comput*, 2007; 20(1):7-107.

29. Marcus M and Minc H. *Survey of matrix theory and matrix inequalities*. New York: Dover Publications 1964.

30. Ryser HJ. Combinatorial Mathematics. The Carus mathematical monographs series. Washington D.C.: Mathematical Association of America; 1963. 154 p.

31. Litvin Y. Applying order statistics in planning and risk assessment of projects repair. *Problems of economics and management of oil and gas complex*. 2014; 7: 24-28. (in Russian)

32. Litvin Y. Order statistics in estimation time of parallel execution of projects works with limited resources. *Problems of economics and management*. 2013; 11(27): 41-46. (in Russian)