Cause and effect analysis of failures at Russian nuclear power plants between years 1992 and 2018

A I Ivanov¹, M Yu Egorov¹, A S Stepashkina¹ and A S Kritchenkov²,³,⁴

¹St. Petersburg State University of Aerospace Instrumentation, Russia, 190000, St.Peterburg, Bol’shaya Morskaya str., 67
²Peoples’ Friendship University of Russia (RUDN University), Russia, 117198, Moscow, Miklukho-Maklaya str. 6
³Institute of Macromolecular Compounds of the Russian Academy of Sciences, Russia, 199004, St.Petersburg, Bolshoy pr. 31
⁴Saint-Petersburg National Research University of Information Technologies, Mechanics, and Optics, Russia, 197101, Kronverksky pr., 49

Abstract. Multiple systems determine safe operation of every nuclear power plant. Elements of these system may periodically fail due to some reasons. This article contains annual failure statistics analysis of Russian nuclear power plants between years 1992 and 2018. It was also assessed the effectiveness of nuclear power plant safety improvement measures, aimed on failure elimination.

1. Introduction
Term reliability means ability to keep execution of its functions in time at all modes and conditions of storage, transportation, operation, and maintenance [1]. In the context of nuclear power plants (NPPs), terms reliability and safety are interrelated because a failure (means the lack of reliability) of the plant system’ element may lead to harmful conditions for people and nature [2].

Within the operation of NPP, operating organization JSC Rosenergoatom collects data including type, frequency, and cause of failures, deviations from normality in order to improve safety. In Russia, it was decided to classify all deviations by:

- its influence on safety. In compliance with [3], it was decided in Russia to use the International Nuclear and Radiological Event Scale (INES), assessing all system deviations from normality by its impact on safety with respect to radiation using the eight levels scale. The scale includes seven levels with impact on safety where 7th level is maximal, first level is minimal. The scale also includes level named “zero” or “below scale” having no impact on safety though used for statistical purposes;
- cause of the deviation which can be subdivided into:
  - common cause failures (CCFs), which may occur “as a result of a single failure or human error or internal or external impact (event) or due to other reason” [1] (clause 46).
  - independent (single) failures, having no influence on other failures.
Reliability of Russian NPP elements is the subject of continuous analysis which is based on national operational experience and failure study of other countries. Due to this fact, a recent complex analysis of the reliability of elements of Russian NPPs shows that all failures caused by internal reasons such as fire, flood or were hidden defects. Notably that “none of failures have common cause” [4]. On the one hand, this circumstance witnesses in favor of the method. On the other hand, failures of NPP elements still happen periodically. Therefore, actual efficiency of the reliability improvement measures is the subject of further analysis and discussion.

2. Purpose of the research

Actual study aims to assess statistics of annual failures of Russian nuclear power plants elements, having significant impact on unit’s safety.

Annual failures spread between years 1992 and 2018 inclusive is shown on the f. The dataset was collected from annual reports of JSC Rosenergoatom which were published at [5] and also presented at [6]. It is remarkable that within the period from 1992 to 2001 the failures frequency has decreased in 2.5 times (from 197 to 79 events per year respectively). In the consequent period, from 2002 to 2018, the number of failures has stabilized at the level of approximately 40–50 events per year.

![Figure 1](image)

**Figure 1.** Spread of annual number of failures, occurred at Russian nuclear power plant units between years 1992 and 2018.

Using the least square approach, it is possible to derive the mathematical model from the data pool. Since the data has negative curvilinear distribution, application of the binominal model of third extent rather than the simple linear model for better precision. The equation of 3-d extent binominal model is:

\[ y = a \cdot x^3 + b \cdot x^2 + c \cdot x + d. \]

According to the approach, sum of squares of deviations between actual values of the dependent variables and the values predicted by the model. This could be transformed into following equation:

\[ \sum_{i=1}^{n} \delta^2 = \sum_{i=1}^{n} (y_i - (a \cdot x_i^3 + b \cdot x_i^2 + c \cdot x_i + d))^2. \]  \( (1) \)

I order to estimate unknown coefficients \( a, b, c, d \), the equation (1) could be partially differentiated with respect to each coefficient. Then, equalize each equation to zero \( \frac{df}{da} = 0, \frac{df}{db} = 0, \frac{df}{dc} = 0, \frac{df}{dd} = 0 \), and move all its multiples with \( y_i \) to the right hand side:
\[
\begin{aligned}
&\left\{\begin{array}{l}
a \cdot \sum_{i=1}^{n} x_i^6 + b \cdot \sum_{i=1}^{n} x_i^5 + c \cdot \sum_{i=1}^{n} x_i^4 + d \cdot \sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} y_i \cdot x_i^3 \\
a \cdot \sum_{i=1}^{n} x_i^5 + b \cdot \sum_{i=1}^{n} x_i^4 + c \cdot \sum_{i=1}^{n} x_i^3 + d \cdot \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i \cdot x_i^2 \\
a \cdot \sum_{i=1}^{n} x_i^4 + b \cdot \sum_{i=1}^{n} x_i^3 + c \cdot \sum_{i=1}^{n} x_i^2 + d \cdot \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \cdot x_i \\
a \cdot \sum_{i=1}^{n} x_i^3 + b \cdot \sum_{i=1}^{n} x_i^2 + c \cdot \sum_{i=1}^{n} x_i + d \cdot N = \sum_{i=1}^{n} y_i
\end{array}\right.
\]

(2)

Using Cramer approach, it is possible to solve the equation (2) and write the model equation with coefficients:

\[
y = -0.022x^3 + 1.297x^2 - 25.427x + 202.286.
\]

Precision of the model \(r^2\) could be determined from the following equation:

\[
r^2 = \frac{\sum_{j=1}^{N} (y_j - \bar{y})^2}{\sum_{j=1}^{N} (y_j - \bar{y})^2} = \frac{42270.90}{45134.07} = 0.9366,
\]

where

\[
\sum_{j=1}^{N} (y_j - \bar{y})^2 - \text{sum of squares of deviations between predicted values and average value of the data set;}
\]

\[
\sum_{j=1}^{N} (y_j - \bar{y})^2 - \text{sum of squares of deviations between actual values and average value of the data set.}
\]

Though the model precision is relatively close to 1 which means high quality of the model, a residual analysis will be conducted in order to estimate the prediction errors effect. Figure 2 illustrates the distribution of residuals by time.

![Figure 2. Residuals plot.](image)

Since no direct correlation observed between values, the Durbin-Watson statistic will be used to detect presence of autocorrelation among neighboring residuals. The value of Durbin-Watson criteria \(dw\) could be determined from the following equation:

\[
dw = \frac{\sum_{i=2}^{n} (e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2} = \frac{3755.5}{4543.5} = 1.2,
\]

where

\[e_i - \text{residual at the time } i.\]

Then, the \(dw\) value should be compared with critical values of Durbin-Watson statistic taken for following parameters: one variable \(Y\), number of observations in the data set \(N=27\), and the significance level \(\alpha = 0.05\%\). Since \(dw < dw_{1.32}\), is observed a week positive autocorrelation between neighboring residuals which, however, will be neglected in this study due to its low influence.

Let’s analyze how the annual failures number, a discrete random variable, is spreading over time. For this reason, following approach will be used:
• it will be calculated the mean, standard deviation;
• five basic characteristics, namely $x_{\text{min}}, Q_1, \text{Median}, Q_3, x_{\text{max}}$, will be represented on the box diagram. Since the population size is small (number of observations $N = 27$ observations), it is possible to analyze the whole population without taking sample. The mean $\mu$ could be found from the following equation (3):
\[
\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{1750}{27} = 64.81 \text{ (failures per year)}
\]
where $X_i$ – n-observation of the variable $X$, $\sum_{i=1}^{N} X_i$ – sum of variables of the population.

It is possible to estimate the standard deviation $\sigma$ of the population using following equation (4):
\[
\sigma = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}} = \sqrt{\frac{45134.07}{27}} = 41.66 \text{ (failures per year)}.
\]

So, the value of annual failures is fluctuating around 64.81 failures per year and deviates by 41.66 from the mean in most.

Using five basic characteristics, namely $x_{\text{min}}, Q_1, \text{Median}, Q_3, x_{\text{max}}$, it is possible to drove the box diagram. Minimum value $x_{\text{min}} = 29$ failures per year; maximum value $x_{\text{max}} = 197$ failures per year. On first, let’s estimate indexes of the values using following equations:
\[
\text{index } Q_1 = \frac{N+1}{4} = \frac{28}{4} = 7
\]
\[
\text{index } \text{Median} = \frac{2(N+1)}{4} = \frac{56}{4} = 14
\]
\[
\text{index } Q_3 = \frac{3(N+1)}{4} = \frac{84}{4} = 21.
\]

According the empiric rule [5], if an estimated index is integer, then value of each characteristics is equal to that of element of the population respectively. Thus, $Q_1 = 38, \text{Median} = 44, \text{ and } Q_3 = 83$ failures per year. Figure 3 illustrates spread of five basic characteristics ascending by value. Analyzing the boxplot, it can be concluded that the spread is asymmetric and right-skewed.

Figure 3. Failures between years 1992 and 2018 boxplot

3. Hypothesis
Basing on the analysis of the failure spread over time, it can be supposed that certain type of failures of NPP’ elements peaked before 1992 year. Following that event, a course of specific corrective measures was undertaken which resulted in steady decrease of the number of failures until 2001. The course, however, did not affect other type of failures. If it is true, the number of failures starting from 2002 will be normally distributed. It means the population features must conform that of normal distribution:

• minimum and maximum of the population are in the range $\mu \pm 3\sigma$;
• the interquartile range value does not exceed $1.33\sigma$;
• normalized values lies reasonably close to the straight line and show systematic pattern on the quantile plot.
Population of failures within the period from 2002 to 2018 year inclusive (population A) which contains \( N_A = 17 \) elements is shown on the figure 4. As it could be seen from the graph, the value is fluctuating around approximately 40 failures per year.

Let's analyze how the annual failures value is spreading between years 2002 and 2018 using similar approach, used for the whole period analysis. Using equations (3) and (4) for estimation of the mean \( \mu_A \) and the standard deviation \( \sigma_A \) of the population respectively, it can be concluded that the value of failures is distributed around 39.82 ± 4.83 failures per year. Substituting these numbers into the following equation, it is possible to assess boundaries of the theoretically normally distributed value from the mean:

\[
(X_{norm, min}; X_{norm, max}) = \mu_A \pm 3 \cdot \sigma_A = 39.82 \pm 3 \cdot 4.83 = (25.34; 54.49).
\]

**Figure 4.** Spread of annual number of failures between years 2002 and 2018.

It can be noted that minimum and maximum values of the population which are 29 and 47 respectively, does not exceed the boundaries and, therefore, first condition of the theoretical distribution is fulfilled. Assessing values of the median, first and third quartile of the population A from the equations (5), (6) and (7), yields 38, 40 and 42 failures per year respectively. So the interquartile range \( IQR \) could be found from the following equation:

\[
IQR = Q_3 - Q_1 = 42 - 38 = 4.
\]

As it could be seen, \( IQR < 1.33\sigma_A = 1.33 \cdot 4.83 = 6.42 \) that witnesses about fulfillment of the second condition.

In order to estimate normality of the population A, quantiles of the normal distribution must be computed using following equation:

\[
q_i = \frac{i}{N_A + 1},
\]

where

\( q_i \) – \( i \)-th quantile of the normal distribution having \( \mu = 0 \) with the standard deviation \( \sigma = 1 \).

Figure 5 represents the normal quantile plot of the population A where each \( x \) is an original population value and \( y \) is the corresponding \( z \) score of computed quantile which was taken from the standard normal distribution values table [5, p.1240].
As it might be seen from the graph provided above, the normalized points are evenly distributed across the straight line. This circumstance witnesses about fulfillment of third requirement of conformity.

4. Conclusion

- Basing on the findings, it can be stated that corrective measures, undertaken for elimination of the failures among elements of Russian nuclear power plants, have had specific character.
- Within the period from 1992 to 2001, the speed of annual failure emergence has descending trend.
- From 2002 to 2018, the failure frequency has stabilized at the level of 39.82 ± 4.83 failures per year.

References

[1] Interstate council for standardization, metrology and certification (ISC) Industrial product dependability. General principles. Terms and definitions Interstate standard GOST 27.002-15 Moscow: Standartinform 2016 p 22
[2] Institut Mashinovedeniya AN SSSR Industrial product dependability. General principles. Terms and definitions. Interstate standard GOST 27.002-89 Moscow: Gosstandart SSSR 1990 p 13
[3] International Atomic Energy Agency (IAEA) International Nuclear and Radiological Event Scale. User's manual Vienna 2013 p 218
[4] Tokmachev G 2011 Issues related to selection and treatment of common mode failure data Nuclear and Radiation Safety 4(62) 29-39
[5] Levine D M, Stephan D, Krebbiel T C and Berenson M L 2005 Statistics for managers using Microsoft Excel ed 4th Klushin D A, (Moscow: Williams Publishing House) p 186
[6] Sturges H 1926 J. Amer. Statist. Assoc. 21 65-6