

# INSERTION OF SYNTHETIC DATA INTO AN EMPIRICAL DATABASE OF SOME ALLOYS' CORROSION DEPTH AND THEIR INFLUENCE ON THE SELECTION OF THE BEST-FITTED CONTINUOUS DISTRIBUTIONS

Nataša Kovač<sup>1</sup>, PhD; Špiro Ivošević<sup>2</sup>, PhD

<sup>1</sup> University of Donja Gorica, Faculty of Applied Sciences,  
Podgorica, MONTENEGRO, natasa.kovac@udg.edu.me

<sup>2</sup> University of Montenegro, Faculty of Maritime Studies, Kotor,  
MONTENEGRO, spiroi@ucg.ac.me

**Abstract:** *The importance of studying corrosive processes is evident. However, researchers often face the problem of a lack of sufficient empirical data, especially in the case of the application of advanced modeling techniques in the field of artificial intelligence. In this paper, the technique of inserting synthetic data into empirical databases based on measuring the depth of corrosion caused by three different marine environments over samples of three different alloys after 12 and 18 months of exposure to the environment is applied. Empirical and extended databases were further used to analyze a linear model of corrosion depth based on the assumption that corrosion processes occur immediately after exposure to the effects of the marine environment. In each observed database, the best two-parameter, three-parameter, and multiparameter continuous distributions were selected by fitting. After a comparative presentation of the obtained results, the influence of the inserted synthetic data was detected.*

**Keywords:** *corrosion depth, empirical database, extended database, continuous distribution, distribution fitting*

## 1. INTRODUCTION

Since the appearance of metallic materials in various industries, corrosion has emerged as the dominant degradation factor. Corrosion occurs in various physical forms such as, general corrosion, pitting corrosion, fretting corrosion, galvanic or two-metal corrosion, crevice corrosion, intergranular corrosion, selective leaching or parting, erosion-corrosion, stress corrosion, etc., which affect the change of the chemical composition of metals and the degradation of their technical-technological and exploitation conditions [1]. Previous research has shown that damage caused by corrosion of all types of material is about 3-4% of the Gross National Income of developed countries [2, 3].

Due to the different forms of corrosion, as well as the great harmful effect that corrosion has on various metallic and non-metallic materials, numerous studies are conducted to discover the cause, corrosion intensity, and harmful consequences.

Although the number of authors believes that the corrosion process is unstable and time-dependent and that it has a constant velocity that can be expressed linearly a nonlinear model is more realistic in terms of describing the corrosion process [4]. For these reasons, numerous studies have shown that the corrosion process in different environmental conditions, taking a significant number of influential parameters that affect the course of corrosion over time can be described as linear or nonlinear models [5]. These models describe the loss of base metal thickness in mm of wear or as a percentage of wear concerning the built-in value [6, 7] or mass lost [8].

Special attention to the research of the corrosion process is paid to the influence of corrosion in new materials such as smart materials. These materials can change their shape, position, stiffness, natural frequency, and other mechanical characteristics when they expand to temperature, stress, moisture, pH, electric or magnetic fields. Between different smart materials as piezoelectric materials, self-healing materials, chromogenic materials, etc., Shape Memory materials (SMA) are more attractive for different industrial application. The main characteristic of the SMAs is Shape Memory Effect (SME), i.e., to remember their original shapes when returning to the pre-deformed shape upon heating related to the solid-state transformation of martensite to austenite, and vice versa [9, 10]. Furthermore, super elasticity, high damping capacity, and double shape memory effect are also discovered thermo-mechanical properties of SMA [11, 12]. Since their discovery in 1932, numerous researches have been carried out to date to find better technical and technological characteristics of alloys and their application in various branches of industry. The specific thermo-mechanical characteristics of SMA have enabled the application of these materials in various industries, such as the Automotive, Railway, Aircraft, and Maritime industries, as well as Medicine and Robotics [13, 14, 15, 16, 17].

The application of SMA in the marine environment has a special challenge for many researchers, both due to the application of new materials and due to the specific environmental conditions in which SMA is applied [18]. Numerous SMA applications occur in the maritime industry, both on vessels, fixed platforms, or in the deep seas [19].

From its inception until today, numerous SMA families based on Au, Al, Cu, Ni, Ti have appeared. These are mainly alloys with two or three elements, although there are also families of alloys with more than three elements. So far, Ni-Ti, Cu-Al-Ni, Cu-Zn-Al and Fe have been the most widely used. Ni-Ti-based alloys have proven to be the most

functional but expensive alloys. Cu-based alloys have shown excellent shape memory characteristics and low cost, while iron-based alloys have good machinability [20, 21].

Also, experimental research in real environmental conditions [22, 23] or in laboratory conditions [24] provided answers to numerous questions related to changes in the chemical composition of alloys, the appearance of different physical forms of corrosion, or the development of corrosion over time in SMA alloys.

From all the above, it becomes clear that the benefits of researching corrosive processes are immeasurable. However, researchers very often face the problem of a lack of sufficient empirical data. Measurements of corrosive damage on metal samples are very expensive and most researchers face the problem of an insufficient number of records in empirical databases. Therefore, it is necessary to upgrade empirical databases with artificially generated records to achieve sufficiently large database dimensions, which would allow reliable research, especially in cases where more demanding and sophisticated research techniques are applied, such as neural networks, machine learning, techniques of artificial intelligence, regression, etc.

The idea of this paper is based on the analysis of a technique of expanding the empirical database with artificially generated records, to achieve a greater number of adequate data, over which techniques could be applied that require more data than currently available.

The paper is divided into chapters as follows. The second chapter is devoted to the description of how the empirical database was formed. In addition, this chapter outlines the technique of adding new synthetic data to an existing empirical database, thus forming an expanded database. The third chapter describes the probabilistic methodology used for the analysis of empirical and extended databases. A comparative presentation of the results obtained with empirical databases and the same techniques applied to extended synthetically generated databases, provide insight into the effects produced by embedded artificially generated data. The fourth chapter gives conclusions based on the conducted research.

## 2. EMPIRICAL AND EXTENDED CORROSION DEPTH DATABASES

This paper uses empirical databases formed based on corrosion damage measurements on samples of two forms of nitinol alloys produced by two different processes and one CuAlNi alloy. Pure metals were used to produce NiTi alloys: Ni (99.99 wt.%) And Ti (99.99 wt.%) Supplied by Zlatarna Celje d.o.o. Slovenia. The NiTiAs Cast alloy is produced by classic casting and rolling in the form of a disc with a diameter of 42.3 mm and a thickness of 3.4 mm. The NiTiCC alloy was produced by a combination of vacuum melting and continuous casting methods. Its shape is a rod with a diameter of 11.9 mm and a length of 50 mm. A total of 18 NiTi alloy samples (9 NiTiAs Cast samples and 9 NiTiCC samples) were used during the experiment. CuAlNi SMA rods (9 samples) were produced by a continuous casting process using a laboratory device for vertical continuous casting, Technica Guss, which was connected to a medium frequency furnace (4 kHz) with a vacuum melting induction (VIM) of 60 kW, Leybold Hereaus. The traction parameters were programmable, so an almost arbitrary time velocity curve can be achieved (limits are set by engine performance and inertia of moving parts). The formed empirical databases were previously used in the works of the authors [23, 25].

All three alloys were observed under the influence of three marine environments, where the first is the constant influence of the atmosphere, the second is the constant influence of the sea, while the third influence represents the changing influences of the sea and the atmosphere. All samples were placed at clearly defined locations in the boundary zone of the sea and atmosphere. Samples exposed to the atmosphere were placed near the sea, three meters above sea level, samples exposed to the sea were immersed in the sea near the shore at a depth of three meters, while samples exposed to changing tides were located on the sea surface. Due to corrosive processes caused by the marine environment, the damage is created on the samples, the depth of which is measured by a focused ion beam (FIB) on a scanning electron microscope. After 12 and 18 months of exposure, the corrosion depth expressed in nm was detected on the surface of the samples by the FIB method, thus completing the process of forming empirical databases. When measuring the depth of corrosion, no additional factors influencing corrosive processes (temperature, pressure, salinity, particle flow, conductivity, etc.) were considered.

The formed empirical databases served as a basis for the creation of artificially generated extended databases. Namely, we have systematically supplemented the empirical databases with additional data, in the following way. Based on photometric representations and FIB measurements, we were able to detect areas on samples where corrosive processes occurred, but which were not considered in the FIB analysis process. By comparing the measured values of the corrosion depth for the points for which we knew the values expressed in nm, we were able to read the adequate values of the corrosion depth on sample areas that were not part of the empirical database. For precision, we used image analysis software tools in this process. In this way, approximate but realistic values of corrosion depth were obtained for each of the three considered alloys, in all three marine environments, and extended databases with a fixed, predefined number of records were formed. In the continuation of the research, both types of databases were comparatively analyzed: empirical and artificially generated databases. A total of 18 databases were formed and analyzed:

- 3 empirical bases for the influence of air on the depth of corrosion of CuAlNi, NiTiAs Cast, and NiTiCC alloys;
- 3 empirical bases for tide influence on corrosion depth of CuAlNi, NiTiAs Cast, and NiTiCC alloys;
- 3 empirical bases for the influence of the sea on the depth of corrosion of CuAlNi, NiTiAs Cast, and NiTiCC alloys;
- 3 artificially generated (extended empirical) bases for the influence of air on the depth of corrosion of CuAlNi, NiTiAs Cast, and NiTiCC alloys;
- 3 artificially generated (extended empirical) bases for the effect of tide on the corrosion depth of CuAlNi,

NiTiAs Cast, and NiTiCC alloys;

- 3 artificially generated (extended empirical) bases for the influence of the sea on the corrosion depth of CuAlNi, NiTiAs Cast, and NiTiCC alloys.

The basic descriptive characteristics of all listed databases from the point of view of the depth of formed corrosion on samples of CuAlNi, NiTiAs Cast, and NiTiCC alloys measured after 12 and 18 months of exposure to the influence of the marine environment are shown in Tables 1.-3.

**Table 1:** Descriptive statistics of empirical and extended databases related to the three different alloys and their corrosion depth formed under the influence of air

AIR	CuAlNi				NiTiAs Cast				NiTi CC			
	Original		Extended		Original		Extended		Original		Extended	
Stat./months	12	18	12	18	12	18	12	18	12	18	12	18
Sample Size	47	68	200	300	25	55	200	300	82	82	200	300
Range	90.28	90.3	109.9	109.9	5.6	5.6	5.6	5.6	5.4	5.3	5.3	5.3
Mean	38.8	39.8	39.3	40.4	5.0	3.8	4.6	4.0	3.9	3.9	4.6	4.1
Std. Dev.	22.4	19.2	21.4	18.3	2.0	1.7	2.0	1.8	1.4	1.4	1.5	1.4
Min	9.7	9.7	8.8	8.8	2.1	2.1	2.1	2.1	2.3	2.3	2.3	2.3
Q1	16.0	27.9	18.2	29.2	2.7	2.6	2.7	2.7	2.9	2.9	3.1	3.1
Median	37.5	41.4	38.5	41.6	5.8	3.0	4.4	3.2	3.2	3.2	4.2	3.5
Q3	51.4	49.8	51.7	50.4	6.5	5.6	6.3	5.8	5.2	5.2	5.8	5.6
Max	100.0	100.0	118.6	118.6	7.6	7.6	7.6	7.6	7.6	7.6	7.6	7.6

**Table 2:** Descriptive statistics of empirical and extended databases related to the three different alloys and their corrosion depth formed under the influence of tide

TIDE	CuAlNi				NiTiAs Cast				NiTi CC			
	Original		Extended		Original		Extended		Original		Extended	
Stat./months	12	18	12	18	12	18	12	18	12	18	12	18
Sample Size	41	62	200	300	40	66	200	300	47	79	200	300
Range	262.4	316.3	340.8	340.8	10.8	10.8	10.8	10.8	45.3	47.4	68.9	70.8
Mean	185.0	140.3	186.5	141.0	4.5	3.9	4.6	4.1	11.2	13.1	11.8	13.4
Std. Dev.	64.3	82.0	60.4	81.3	2.5	2.1	2.5	2.2	9.0	9.9	10.8	10.7
Min	91.0	37.0	12.5	12.5	1.7	1.7	1.7	1.7	4.4	2.2	4.4	2.4
Q1	138.2	58.5	145.0	54.2	2.5	2.6	2.5	2.7	6.9	6.9	6.3	6.9
Median	180.8	138.2	184.2	145.6	4.0	3.1	4.0	3.1	7.9	8.8	8.1	8.8
Q3	211.3	200.2	220.0	197.8	6.3	4.9	6.3	5.0	9.6	17.7	9.7	17.2
Max	353.3	353.3	353.3	353.3	12.5	12.5	12.5	12.5	49.7	49.7	73.2	73.2

**Table 3:** Descriptive statistics of empirical and extended databases related to the three different alloys and their corrosion depth formed under the influence of sea

SEA	CuAlNi				NiTiAs Cast				NiTi CC			
	Original		Extended		Original		Extended		Original		Extended	
Stat./months	12	18	12	18	12	18	12	18	12	18	12	18
Sample Size	40	61	200	300	52	77	200	300	46	75	200	300
Range	106.3	117.2	140.0	151.4	74.7	74.7	74.7	74.7	50.0	57.5	61.0	61.0
Mean	148.9	136.9	150.5	139.2	26.3	26.2	24.7	25.7	22.1	25.3	23.5	27.4
Std. Dev.	25.5	31.3	27.5	32.2	19.4	16.6	22.4	19.2	10.6	12.1	10.6	13.0
Min	80.4	69.4	75.0	63.6	6.6	6.6	6.6	6.6	7.5	7.5	7.5	7.5
Q1	132.3	110.2	135.0	114.4	8.4	14.5	7.4	8.3	14.7	15.3	15.1	16.2
Median	151.7	142.8	152.5	142.6	21.2	23.8	9.7	21.9	18.0	22.2	20.0	24.3
Q3	167.5	163.3	170.0	165.7	36.5	31.3	41.7	38.0	28.1	32.9	28.2	35.6
Max	186.7	186.7	215.0	215.0	81.3	81.3	81.3	81.3	57.5	65.0	68.5	68.5

As can be seen from Tables 1.-3., for all three considered alloys, extended databases are formed so that they have a constant number of data, regardless of the number of corresponding empirical data. For extended databases related to corrosion depth detected after 12 months of exposure to the environment, the number of records is 200, while extended databases related to corrosion depth values after 18 months of exposure to air, tide, and sea have 300 data each. Extended databases also contain original, empirical data, in addition to the inserted adequate realistic values of corrosion depth from the same samples on which the original empirical measurements were performed. The values of all descriptive statistics related to extended databases are the same or very close to the corresponding values of descriptive statistics of the corresponding empirical databases. This fact indicates the systematic insertion of artificial data into empirical databases. In this way, the basic statistical characteristics of empirical databases were preserved and the requirement that the inserted artificial data simulate real empirical data was met.

### 3. PROBABILISTIC CORROSION DEPTH MODELING

In many articles dealing with the study of corrosive processes [6, 7, 23, 25, 26, 27], it has been proved that the wear of the plate thickness,  $d(t)$ , can be successfully modeled as a function of time, in the following way

$$d(t) = c_1(t - T_{cl})^{c_2}. \quad (1)$$

Time ( $t$ ) is usually expressed in months or years and corrosion depth  $d(t)$  in mm or nm. In expression (1)  $T_{cl}$  is the service life of the coating while  $c_1$  and  $c_2$  are positive real coefficients. Parameter  $c_1$  can be considered as corrosion rate expressed in mm/year or nm/month, while the coefficient  $c_2$  is usually taken to be 1 or 1/3 [27]. In our previous works that analyze the corrosive behavior of SMA, we start from the basic assumption that  $T_{cl} = 0$  months, i.e., that the samples were not treated with anticorrosive coatings. In addition, in all our previous works studying the linear corrosion model [6, 7, 23, 25] the coefficient  $c_1$  is not considered as a linear constant, but we observe corrosive processes as stochastic quantities influenced by many different factors. More precisely, the coefficient  $c_1$  is considered as a continuous random variable.

Based on the formed empirical and extended databases, and applying model (1), our research in this paper is focused on modeling the depth of corrosion of all three observed alloys in three different marine environments. The probabilistic approach to modeling the  $c_1$  coefficient is based on fitting the Cumulative Distribution Functions (CDF) of known continuous distributions into data describing the corrosion depth of each alloy in each considered seawater environment [23, 25]. In this procedure, a total of 65 multiparameter (MP) distributions were considered (number of parameters vary between 1 and 6), of which 27 were three-parameter (3P) and 27 two-parameter (2P). We used the well-known continuous distributions, such as Log-Logistic (Log-Log), Generalized Extreme Value (GEV), Phased Bi-Weibull (PBW), Generalized Pareto (GenPareto), Inverse Gaussian (InvGaussian), etc.

From each set of continuous distributions, based on the value of the Kolmogorov-Smirnov test, all distributions were ranked according to the quality of fitting and thus the three best distributions were determined that can adequately describe the corrosion depth behavior under the influence of the marine environment. Standard Anderson-Darling (AD) and Chi-Squared ( $\chi^2$ ) tests were used as additional tests to determine the goodness of fit. Based on these three tests, the null hypothesis was tested, which claims that the data follow the chosen theoretical distribution, where the values of 0.2, 0.1, 0.05, 0.02, and 0.01 for significance level were considered. The hypothesis regarding distribution form was rejected for the selected significance level if the test statistically exceeded a predefined critical value. The p-value was calculated based on the test statistics and marked the significance level threshold. The null hypothesis was rejected if the p-value was lower than the selected critical value, i.e., it could be concluded that the theoretical distribution does not describe the observed data for the selected level of significance. The results of the described procedure of fitting continuous distributions into data distributed in 18 different databases are shown in Tables 4-9. In these tables, cells are color-coded as follows:

- the green color indicates the fact that the presented theoretical distribution has passed all statistical tests for all the stated significance levels, that is, that the null hypothesis is not rejected in any of the cases;
- the yellow color of the cell indicates that certain problems occur that can be overcome either by changing the statistical test or by reducing the levels for significance level;
- the red color of the cell indicates that the null hypothesis is rejected for each of the proposed statistical tests and each proposed significance level.

Table 4. shows the three best-fitted multiparameter, two-parameter, and three-parameter distributions, for all three considered seawater environments (air, tide, and sea) for the empirical depth of corrosion caused by CuAlNi alloy, after 12 and 18 months of exposure to the environment. The ranking of the three best distributions is defined by the values of KS test statistics. The goodness of fit was additionally tested with AD and  $\chi^2$  tests, for all the stated significance levels. The green color in all cells of the table indicates that all tests for all considered significance levels showed that the null hypothesis cannot be rejected, that is, that the stated theoretical distributions describe well the empirical data of the corrosion depth of CuAlNi alloy.

**Table 4** Three best fitted multiparameter, two-parameter, and three-parameter distributions related to the CuAlNi alloy corrosion depth empirical database

CuAlNi		12			18		
		Air	Tide	Sea	Air	Tide	Sea
MP	1	GenPareto	GenLogistic	JohnsonSB	Wakeby	Wakeby	Wakeby
	2	Weibull	Burr	Dagum	Normal	GenPareto	JohnsonSB
	3	Normal	Wakeby	Wakeby	Chi-Squared	JohnsonSB	GenGamma
2P	1	Weibull	FatigueLife	GumbelMin	Normal	Weibull	Uniform
	2	Normal	Lognormal	Weibull	Chi-Squared	Reciprocal	GumbelMin
	3	Nakagami	Log-Log	InvGaussian	Logistic	Normal	Weibull
3P	1	GenPareto	GenLogistic	Dagum	Error	GenPareto	GEV
	2	Error	Burr	GEV	GEV	Error	PowerFun
	3	GEV	Log-Pearson3	Log-Pearson3	Dagum	PowerFun	GenPareto

NiTi alloys show significantly more complex corrosive behavior in all marine environments, compared to CuAlNi alloy.

These differences are especially notable if the air and tide environments are observed [25] because NiTi alloys show similar corrosive behavior in these two environments. Tables 5. and 6. show the goodness of fit results for the empirical databases of the two NiTi alloys. The green marked cells indicate that there were no restrictions for any of the considered statistical tests or for any of the selected significance levels, that is, that the presented theoretical distributions adequately follow the empirical data. The yellow background of the cell indicates that individual constraints must be considered to accept the null hypothesis. It is evident that more than two parameters are needed to adequately describe the complex corrosive processes of these alloys, that is, two-parameter distributions are not an adequate choice for NiTi alloys (two red cells appeared, indicating that all statistical tests rejected the null hypothesis). In addition, as the exposure time to the environment increases, the corrosive behavior of these alloys changes and becomes more complex, thus after 18 months of exposure to the environment, some deviations may appear in the data, which may affect the quality of distribution fitting. Therefore, outliers' removal techniques should be additionally introduced into the analyzes [25], which is especially indicated by a large number of yellow cells in Tables 5. and 6.

**Table 5** Three best fitted multiparameter, two-parameter, and three-parameter distributions related to the NiTiAs Cast alloy corrosion depth empirical database

NiTiAs Cast		12			18		
		Air	Tide	Sea	Air	Tide	Sea
MP	1	Beta	GenPareto	Wakeby	Pearson5	Burr	Wakeby
	2	GenPareto	Wakeby	JohnsonSB	Pearson6	GenPareto	Gamma
	3	Wakeby	Pert	GenPareto	Frechet	Wakeby	Weibull
2P	1	Uniform	Gamma	Weibull	Frechet2P	Frechet	Gamma
	2	Normal	FatigueLife	Gamma	Exponential	Log-Gamma	Weibull
	3	Weibull	InvGaussian	Log-Log	Pareto	InvGaussian	GumbelMax
3P	1	GenPareto	GenPareto	GenPareto	Pearson5	Burr	GenGamma
	2	Error	Pert	Gamma3P	Frechet	GenPareto	GenLogistic
	3	GEV	Log-Pearson3	GEV	Log-Log	Log-Log	GEV

**Table 6** Three best fitted multiparameter, two-parameter, and three-parameter distributions related to the NiTiCC alloy corrosion depth empirical database

NiTi CC		12			18		
		Air	Tide	Sea	Air	Tide	Sea
MP	1	Burr	Burr	Burr	Burr	Log-Log 3P	JohnsonSB
	2	JohnsonSB	GenPareto	GEV	JohnsonSB	Log-Log	GenPareto
	3	Dagum	Wakeby	Log-Log	Dagum4P	Log-Gamma	Burr4P
2P	1	Frechet	Chi-Squared	Pearson5	Frechet2P	Log-Log	Pearson5
	2	Exponential	Cauchy	Log-Gamma	Exponential	Log-Gamma	Log-Gamma
	3	Erlang	Frechet	Log-Log	Erlang	FatigueLife	Frechet
3P	1	Burr	GenPareto	Burr	Burr	Log-Log3P	GenPareto
	2	Log-Log	GEV	GEV	Log-Log	Dagum	Dagum
	3	Frechet3P	GenLogistic	Log-Log	Frechet	Frechet3P	InvGaussian

Tables 7., 8., and 9. show the results of fitting the best continuous distributions into an artificially formed extended database. In the process of testing the null hypothesis, the same technique was applied as in the case of testing empirical databases, with the same statistical tests, ranking, and significance level values.

Even though the expanded databases were obtained by inserting records in empirical databases in a systematic way, and not by random, completely artificially generated values, as can be seen in Tables 7.-9. significant problems arise in selecting adequate theoretical distributions. The number of yellow cells has increased significantly, but an even more worrying fact is the drastic increase in the number of red cells.

**Table 7** Three best fitted multiparameter, two-parameter, and three-parameter distributions related to the CuAlNi alloy corrosion depth extended database

CuAlNi		12			18		
		Air	Tide	Sea	Air	Tide	Sea
MP	1	GenPareto	Log-Log	Dagum	Wakeby	Beta	Wakeby
	2	Nakagami	Burr	GEV	Normal	Ph. Bi-Weibull	GEV
	3	Weibull	Dagum	LogPearson3	Dagum	Wakeby	JohnsonSB

2P	1	Nakagami	Nakagami	Weibull	Normal	Uniform	Weibull
	2	Weibull	Logistic	InvGaussian	Logistic	Weibull	Uniform
	3	Chi-Squared	Chi-Squared	GumbelMin	Hypersecant	Lognormal	Normal
3P	1	GenPareto	Log-Log	GEV	Dagum	GenPareto	GEV
	2	Dagum	Dagum	LogPearson3	GEV	Error	Burr
	3	Burr	GenLogistic	Burr	Error	Gamma	Weibull3P

**Table 8** Three best fitted multiparameter, two-parameter, and three-parameter distributions related to the NiTiAs Cast alloy corrosion depth extended database

NiTiAs Cast		12			18		
		Air	Tide	Sea	Air	Tide	Sea
MP	1	Beta	GenPareto	FatigueLife	GenGamma	Burr	Kumaraswamy
	2	PBW	Wakeby	GenGamma	FatigueLife	JohnsonSB	Beta
	3	Rayleigh	JohnsonSB	Log-Log	Log-Log	Dagum	JohnsonSB
2P	1	Uniform	Exponential	Levy	Frechet	Frechet	Nakagami
	2	Weibull	Pearson5	Frechet	Exponential	Log-Gamma	GumbelMax
	3	Rice	Lognormal	Pareto	Pareto	Exponential	Gamma
3P	1	Error	GenPareto	FatigueLife	FatigueLife	Burr	GenPareto
	2	GenPareto	Pert	Log-Log	Log-Log	Frechet3P	PowerFunction
	3	Pearson6	Weibull	Lognormal	Lognormal	Log-Log3P	Gamma3P

**Table 9** Three best fitted multiparameter, two-parameter, and three-parameter distributions related to the NiTiCC alloy corrosion depth extended database

NiTiCC		12			18		
		Air	Tide	Sea	Air	Tide	Sea
MP	1	JohnsonSB	Burr	Burr	JohnsonSB	Frechet	JohnsonSB
	2	PowerFun	Log-Log	JohnsonSB	Burr	Burr	Wakeby
	3	Reciprocal	Dagum	GenPareto	Chi-Squared2P	JohnsonSB	GenPareto
2P	1	Reciprocal	Cauchy	Frechet	Chi-Squared2P	Frechet	Gamma
	2	Uniform	Frechet	Log-Gamma	Exponential	Pearson5	Log-Gamma
	3	Exponential2P	Pearson5	InvGaussian	Frechet2P	Log-Gamma	InvGaussian
3P	1	PowerFun	Log-Log	Burr	Burr	Burr	GenPareto
	2	GenPareto	GenPareto	GenPareto	Frechet3P	GenPareto	FatigueLife
	3	Error	Burr	Log-Log	Pearson5	Dagum	Weibull3P

Yellow cells indicate problems with the application of statistical tests, but they can be easily overcome by changing the statistical test, or possibly by lowering the chosen significance level. Depending on the type of research being conducted, it is possible, for example, to choose the AD test as the leading test, instead of the KS test. Decreasing the value of significance level allows the null hypothesis to be rejected. However, lowering the significance level reduces the probability of the occurrence of a Type I Error, i.e., it is more difficult to reject the null hypothesis in case it is not true.

The empirical database related to CuAlNi alloy and its corrosive behavior in all three marine environments shows the most stable behavior when all three considered alloys are compared, and thus, inserting artificially generated data into the existing empirical database for CuAlNi alloy does not produce significant problems concerning the process of fitting theoretical distributions into the resulting extended databases. With small interventions over the formed extended databases, very reliable results of fitting continuous distributions can be obtained. A characteristic result was obtained for 18 months of exposure to the tide, where all statistical tests indicated that the null hypothesis should be rejected. The influence of changing sea level with the additional influence of air proves to be the most dominant factor in the occurrence of corrosive processes. This is the characteristic behavior in the case of CuAlNi alloy as well, which results in a complex data structure of the formed extended database. Even small interventions on such measurements significantly change the image of the corresponding histograms of frequencies of corrosion depth values. In particular, in CuAlNi alloys, several peaks appear in histograms due to redundant values of corrosion depth. By simply deleting these multiple values, the problem of fitting theoretical distributions is eliminated, and statistical tests show that the selected best distributions fit well into the observed values of corrosion depth.

A large number of red fields occur primarily in extended databases related to the NiTi alloys. As it is already mentioned, the reason for this is the very complex corrosive behavior of these alloys. The increase in the number of

considered data further complicates the analysis of the formed extended databases. In the case of NiTi alloys, the proposed statistical tests can hardly be applied in all cases. The reason for this is the large number of outliers detected in the empirical data. The insertion of new data further increases the frequency of these non-standard corrosion depth values, resulting in the problematic fitting of theoretical distributions into extended databases. In addition, many redundant data were detected, which significantly increases the frequencies of individual corrosion depth values. This increase in frequencies affects the structure of the histogram that describes the observed data. The associated histograms show a tendency of very high bars for individual values of corrosion depth, which makes it almost impossible to fit standard shapes of theoretical distributions into the formed histograms. To overcome this problem, it is necessary to implement redundant data reduction techniques and eliminate outliers before inserting synthetic data into empirical databases. In addition, for alloys that show such complicated corrosive properties as NiTi alloys, it is necessary to conduct specific research based on extreme value theory methodologies.

#### 4. CONCLUSION

Some techniques of data analysis and modeling of physical processes require a large amount of empirical data, which are often inaccessible to researchers. This group of important physical processes includes the corrosive processes of various metals and alloys. Focused ion beam techniques for measuring the depth of corrosion due to the high-cost, limit the number of available empirical data to interested researchers conducting experiments on metal samples. On the other hand, researchers often need large databases that contain significantly more realistic data than the number of empirical measurements available. As a result, there is often a need to insert additional, artificially generated data into existing empirical databases, which achieves a sufficient number of records and expands the possibility of using more techniques for modeling corrosive processes. These modeling methods typically include regression analysis, neural networks, machine learning, and similar artificial intelligence techniques. However, the insertion of synthetic data can significantly disrupt the structure of the empirical database and thus affect the quality of the results and the reliability of the conclusions reached during the research.

In this paper, the effects of inserting synthetic sequences of corrosion depth values into existing empirical databases related to the corrosive behavior of two NiTiInol and one CuAlNi alloy were studied. The depth of corrosion on the samples of all three alloys, which was caused by the influence of air, tide, and sea, was detected by the FIB method. Synthetic data were not generated completely artificially, at random, but a systematic method was applied, comparing micro views of existing real measurements of corrosion depth, with micro views of samples on which measurements were not performed. Using the software, and comparing the depths of corrosion, approximate values were obtained with which the empirical databases were expanded, up to the desired number of records.

Statistical analysis was performed on all databases and showed consistency and systematic formation of new databases. Based on the linear corrosion model, which has been previously tested and successfully applied to modeling the empirical corrosion depth of these alloys, the best continuous distributions have been determined that can adequately describe the data in extended databases. Namely, in empirical and extended databases the depth of corrosion was observed as a continuous random variable, and then in the process of fitting theoretical distributions, based on statistical tests, a set of three best multiparameter, two-parameter, and three-parameter distributions was determined, which can well describe the corrosive processes of observed alloys. A comparative presentation of the goodness of fit results showed that synthetically inserted data have a significant impact on the quality of results and conclusions reached based on the distribution fitting procedure.

In most cases, the problem with artificially generated databases related to the depth of corrosion arises due to data redundancy or due to the insertion of a large number of extreme values that affect the shape, location, and skewness of formed histograms. These problems can be overcome by removing outliers as well as eliminating data that is frequently repeated, whereby the corresponding data histograms can be adjusted to follow the structure of empirical histograms. Also, if researchers have sufficient knowledge of mathematical statistics, it is possible to obtain relevant conclusions by changing the statistical tests, or possibly by reducing the significance level. In the case of very complicated corrosive behaviors of alloys, it is possible that the solution to the problem of fitting continuous distributions lies exclusively in changing the research methodology, and it is logical to explore other directions, such as extreme value theory.

In general, it can be concluded that it is very important that researchers pay attention to techniques for inserting artificially generated data into empirical databases that tend to depict complex corrosive behaviors of alloys. In addition to the systematic way of inserting synthetic data into empirical databases, it is possible to consider other known techniques for artificial data generation, such as probability sampling (Samples from Probability Distributions, Simple Random Sampling, Cluster Sampling, Multi-stage Sampling, etc.) and non-probability sampling (Quota Sampling, Convenience Sampling, Purposive Sampling, etc.).

#### REFERENCES

- [1] Uhlig HH. Corrosion handbook. Hoboken, NJ: John Wiley & Sons, 2011: 598–599.
- [2] G. Koch, Trends in Oil and Gas Corrosion Research and Technologies: Production and Transmission Woodhead Publishing, (2017), Cambridge, England.
- [3] <https://www.g2mtlabs.com/corrosion/cost-of-corrosion/>
- [4] Schumacher, M.: Seawater corrosion handbook, NJ: Noyes Data Corporation, 1979

- [5] Guo, J., Wang, G., Ivanov, L. and Perakis, A.N.: "Time-varying ultimate strength of aging tanker deck plate considering corrosion effect", *Journal of Marine Structures*, Vol. 21 No. 4, October, 2008, pp. 402-419
- [6] Ivošević, Š.; Meštrović, R.; Kovač, N. Probabilistic estimates of corrosion rate of fuel tank structures of aging bulk carriers. *Int. J. Nav. Archit. Ocean Eng.* 2019, 11, 165–177
- [7] Ivošević, Š.; Meštrović, R.; Kovač, N. A Probabilistic Method for Estimating the Percentage of Corrosion Depth on the Inner Bottom Plates of Aging Bulk Carriers. *J. Mar. Sci. Eng.* 2020, 8, 442
- [8] Vukelić, G., Vizentin, G., Brnic, J., Brcic, M., Sedmak, F., L, Long-Term Marine Environment Exposure Effect on Butt-Welded Shipbuilding Steel, *Journal of Maritime Science and Engineering*, 2021, 9(5), 491; <https://doi.org/10.3390/jmse9050491>
- [9] T. N. Raju, Sampath, *Transactions of the Indian Institute of Metals*, 64 1(2011) 165.
- [10] S. Sathish, U. S. Mallik, T. N. Raju, *Journal of Minerals and Materials Characterization and Engineering*, 2 (2014) 71.
- [11] A. Ölander, "An electrochemical investigation of solid cadmium-gold alloys," *J. Am. Chem. Soc.*, vol. 54, no. 10, pp. 3819–3833, Oct. 1932, doi: 10.1021/ja01349a004.
- [12] Ivošević, Š., Majerić, P., Vukićević, M., Rudolf, R., "A Study of the Possible Use of Materials With Shape Memory Effect in Shipbuilding", *Journal of Maritime & Transportation Sciences*, Vol. Special edition No. 3, 2020, pp. 265-278
- [13] Dye, T. E., (1990), "An Experimental Investigation of the Behavior of Nitinol", Master's Thesis, Department of Mechanical Engineering, Virginia Polytechnic Institute and State University
- [14] Weimin Huang, (1998), *Shape Memory Alloys and their Application Disertation*, University of Cambridge Department of Engineering, to Actuators for Deployable Structures
- [15] J. Mohd Jani, M. Leary, A. Subic, and M. A. Gibson, "A review of shape memory alloy research, applications and opportunities," *Mater. Des.*, vol. 56, pp. 1078–1113, 2014, doi: 10.1016/j.matdes.2013.11.084.
- [16] S. Nedovic, Š. Ivošević, *Corrosion of SMA materials*, 10th International Conference of Social and Technological development, 3-6.06.2021., Trebinje, Book of Abstracts, 104
- [17] Jose San Juan, *Applications of Shape Memory Alloys to the Transport Industry*, International Congress on Innovative Solutions for the Advancement of the Transport Industry, At San Sebastian, Spain, October 2006
- [18] Ivošević, Š., Rudolf, R., Kovač, D., "The overview of the varied influences of the seawater and atmosphere to corrosive processes", 1st International Conference of Maritime Science & Technology, NAŠE MORE, Dubrovnik, 17 - 18 October, 2019, Croatia. pp.182-193
- [19] Špiro Ivošević: "An Approach to the Corrosion Analysis of the Bulk Carrier Hull Structure", *Scientific Journal of Polish Naval Academy*, (ISSN:2159-5275), Vol. 202 (3), p. 23-32, 2015. DOI: 10.5604/0860889X.1178568
- [20] Mishra, M., Ravindra, A.A., (2014), "Acomparision of conventional and iron based shape memory alloys and their potential in structural applications", *International Journal of Structural and Civil Engineering Research*, 3 (4)
- [21] Amini, R., Mousavizad, S.M., Abdollahpour, H., Ghaffari, M., Alizadeh, M. Okyay, A.K., (2013), "Structural and microstructural phase evolution during mechanosynthesis of nanocrystalline/amorphous CuAlMn alloy powders", *Advanced Powder Technology*, 24,1048–1053.
- [22] Ivošević Š., Vastag G., Majerić P., Kovač D., Rudolf R. (2020) Analysis of the Corrosion Resistance of Different Metal Materials Exposed to Varied Conditions of the Environment in the Bay of Kotor. In: . *The Handbook of Environmental Chemistry*. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/698\\_2020\\_644](https://doi.org/10.1007/698_2020_644)
- [23] Ivošević, Š., Kovač, N., Vastag, G., Majerić, P., Rudolf, R.: A Probabilistic Method for Estimating the Influence of Corrosion on the CuAlNi Shape Memory Alloy in Different Marine Environments, *Crystals* 2021, 11, 274. <https://doi.org/10.3390/cryst11030274>
- [24] Volkov-Husovic T., Ivanic i., Kožuh S., Stevanovic S., Vlahovic M., Martinovic M., Stopic S., Gojic M, Microstructural and Cavitation Erosion Behavior of the CuAlNi Shape Memory Alloy, *Metals* 2021, 11, 997. <https://doi.org/10.3390/met11070997>
- [25] Kovač, N., Ivošević, Š., Vastag, G., Vukelić, G., Rudolf, R.: Statistical approach to the analysis of the corrosive behavior of NiTi alloys under the influence of different seawater environments, submitted.
- [26] Qin, S.; Cui, W. Effect of corrosion models on the time-dependent reliability of steel plated elements. *Mar. Struct.* 2003, 16, 15–34
- [27] Paik, J.K.; Thayamballi, A.K. Ultimate strength of ageing ships. *J. Eng. Marit. Environ.* 2002, 1, 57–77