IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

Classification analysis of tensile strength of alloyed steels

To cite this article: G L Dobrev and I P Iliev 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **878** 012066

Classification analysis of tensile strength of alloyed steels

G L Dobrev[1] and I P Iliev[1]

[1] Technical University Sofia – Branch Plovdiv, 25 Tzanko Diusstabanov Street, 4000 Plovdiv, Bulgaria

E-mail: dobrev.1975@abv.bg

Abstract. Based on experiment data about 63 steels, a statistical models and a methodology are developed to assess the influence of chemical composition on the tensile strength of steels. Using the methods of factor and cluster analysis, the independent quantities are grouped according to correlation and similarity criteria into factors and clusters. The degree of influence of the chemical composition on the tensile strength of some alloyed steels is determined.

1. Introduction

The construction of machine components poses special physical, chemical and technical requirements. Ductility, tensile strength, and resistance to deformation depend to a large degree on the right combination of chemical elements, the presence of phases, which cause strengthening, the emergence of compounds, that are a factor for the shift in the fatigue value.

Modern statistical and mathematical methods are actively used to study the status and to predict the behavior of various metals and alloys. This makes it possible to find new dependencies when studying the behavior of the examined quantity, which cannot be discovered using any other experimental or theoretical method.

Statistical prediction is performed in [1] for the average tensile strength of 316L stainless steel using the Weibull distribution method. Paper [2] presents the results of mechanical testing at room temperatures and elevated temperatures of 17-7PH stainless steel with modified chemical composition. Regression analysis is performed, revealing the statistically significant variables for predicting tensile strength and yield strength of 17-7PH steel with modified chemical composition. The regression analysis describes the dependency between the content of chromium, nickel, aluminum and their interactions, and the variables R_m and $R_{p0.2}$ (the tensile strength and yield strength of the steel 17-7PH). Second order equations are obtained. [3] describes the results of tensile strength testing of AISI type 304, deformed at room temperature and elevated temperatures. It has been found that the parameters of the Weibull distribution show a trend towards reduction with temperature elevation and the coefficient of variance tends to increase as the temperature is elevated. In [4], regression models are used as a tool to reduce the time and costs related to the development and selection of new metal alloys. A multiple regression model is developed, which can predict accurately the tensile strength of high-strength low alloy steel production based on its chemical composition and processing parameters. A linear regression model is developed to include the chemical elements C, Mn, Cr, Ni, Mo, Cu, N, V, plate thickness, solution processing and aging temperatures.

Classification analysis is used to qualify the variables according into specific target groups. When these target groups are known in advance, the methods of logistic regression and discriminant analysis are used, and if the groups are unknown – factor and cluster analysis. Classification analysis is also used to study the state and to predict the behavior of complex systems, which are dependent on a large number

IOP Conf. Series: Materials Science and Engineering **878** (2020) 012066 doi:10.1088/1757-899X/878/1/012066

of independent variables [5]. The basis for classification analysis is the availability of a lot of experiment data.

As it is known, the tensile strength of steels depends on multiple independent variables. The objective of this paper is to study the influence of chemical composition on this property of steels. The use of classification analysis to study the tensile strength of steels enables the solution of the following problems: 1) classification of independent variable quantities (chemical composition of steels) into macro categories and their grouping into several significant groups; 2) determining the degree of influence each variable has over the dependent variable – tensile strength; 3) determining the position of each independent variable in a general hierarchy; 4) determining the distance of each independent variable to the dependent variable – tensile strength. The resulting classifications and classification groups of independent variables can serve as the basis for subsequent regression models. These models will enable the forecasting of the behavior of the studied dependent variable in order to plan the experiment and to create new steels with enhanced tensile characteristics. Classification analysis allows the expression of new properties and dependencies, which cannot be found using known theoretical and experimental methods.

This paper uses the results for 63 steels with regard to the influence of chemical composition on tensile strength [6, 7].

The statistical studies are performed using the SPSS software package [8].

2. Results of multivariate factor analysis

The main objective of factor analysis is: 1) Reduction in the number of independent variables (reduction of data dimensionality); 2) Classification of variables, determining the relationships between independent variables, as well as between independent variables and the dependent variable. The most commonly used method for extracting the significant factors and factor structure is the method of principal component analysis (PCA).

The classification through factor analysis is based on the presence of multicollinearity between the variables. The objective is to predetermine the known number of macrovariables (factors), which group within themselves the input variables according to their mutual correlation. The resulting factors are mutually independent, which allows them to be used subsequently to build regression models and to predict the behavior of systems.

The classification analysis in this paper includes 10 independent parameters: 9 chemical elements: Mn, Cr, Ni, Si, S, P, Al, Mo, C (% content) and the diameter of the specimen for testing tensile strength: d (mm). The tensile strength is considered as the dependent variable, *MPa*.

The factor analysis of the 10 considered independent parameters needs to be analyzed for admissibility. Following the factor analysis process in SPSS, the results are given in Table 1.

The results show that the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) is equal to 0.657 (this indicator needs to be between 0.5 and 1) and Bartlett's Test of sphericity is significant (Sig.= 0.000). These results demonstrate the applicability of factor analysis in this particular case.

Table 1. Results for testing of admissibility of factor analysis.

KMO and Bartlett's Test		Value
Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.657
Bartlett's Test of Sphericity	Approx. Chi-Square	275.249
	Df	45
	Sig.	0.000

The next step is to determine the number of components or factors, into which the 10 independent parameters are grouped. Using the PCA method, all components (factors) are obtained with aggregation of experiment data as a percentage, Table 2.

From this table, 10 independent parameters are selected to be grouped in 5 factors. In this case, 85.3 % of experiment data are described. We consider this process to be acceptable for subsequent studies.

Table 2. Total variance explained using PCA.

Component	Total	% of Variance	Cumulative %
1	3.506	35.060	35.060
2	1.887	18.867	53.927
3	1.520	15.204	69.131
4	1.011	10.109	79.240
5	0.606	6.060	85.300
6	0.505	5.050	90.350
7	0.442	4.419	94.769
8	0.225	2.246	97.015
9	0.179	1.790	98.805
10	0.119	1.195	100.000

Next, the 10 independent variables are grouped into 5 factors, in keeping with the principle that the variables in each factor are strongly correlated and the variables in the different factors are not correlated or correlate weakly. The results are obtained using the PCA and Oblimin rotation method with Kaiser Normalization as shown Table 3.

Table 3. Grouping of predictors into factors[a].

Variable	Factors				
	F1	F2	F3	F4	F5
Mn	0.902				
Cr	0.735				
Ni	0.718				
Si	0.665				
S		-0.945			
P		-0.871			
Al			0.916		
Mo			0.884		
d (mm)				0.998	
C%					0.930

[a] Loadings smaller than 0.5 are omitted.

Table 3 shows that the first factor groups the quantities Mn, Cr, Ni and Si. This means that they are strongly correlated and describe 35% of the sample variance in the experiment data, Table 2. The second factor groups together with a negative sign the quantities P and S, which influence 18.9% of experiment data, Table 2. Because they are in one group (factor) they are also correlated. The third factor includes quantities Ai and Mo, which influence 15.2 % of experiment data. They are also correlated. The last two factors consist of a single quantity, d and C, which do not correlate with any of the other 9 quantities and influence experiment data by 10.1% and 6% respectively. The clear differentiation of factors confirms the adequacy of this factor analysis.

3. Cluster analysis

Cluster analysis (CA) is the name given to a group of various calculation procedures used to classify objects. Usually in CA, the number of groups is known in advance and is determined by the researcher during the study. Formed groups (clusters) need to be mutually homogeneous within and mutually heterogeneous among each other according to specified characteristics. The quantitative assessment of the term "similarity" is related to the term "metric".

In engineering and technical sciences, and physics, clusters are usually defined using hierarchy agglomeration methods. The application of these methods produces a dendrogram (tree diagram).

Table 4 shows the distribution of 10 independent quantities into 4 to 8 clusters using the between-groups linkage method.

Table 4. Grouping independent variables into clusters.

Variable	8 Clusters	7 Clusters	6 Clusters	5 Clusters	4 Clusters
P	1	1	1	1	1
Mo	2	2	2	2	2
Al	3	2	2	2	2
d (mm)	4	3	3	3	3
C%	5	4	4	4	3
Mn	6	5	5	5	4
Si	7	6	6	5	4
Cr	7	6	6	5	4
Ni	8	7	6	5	4
S	1	1	1	1	1

The table shows that with 4 clusters, the grouping of chemical variables is the same as that in factors, Table 3. Clusters are clearly differentiated. The first cluster includes the variables S and P, the second cluster – Mo and Al, the third – d, the fourth – C, the fifth –Mn, Si, Cr and Ni. Even though the grouping is done according to different criteria – correlation in the factor analysis and proximity in the cluster analysis, the results are identical, which is an indication of their reliability.

The following Fig. 1 shows a dendrogram built using the between-groups linkage method. The horizontal axis visualizes the distance between the variables in relative units, expressed as the squared Euclidean distance.

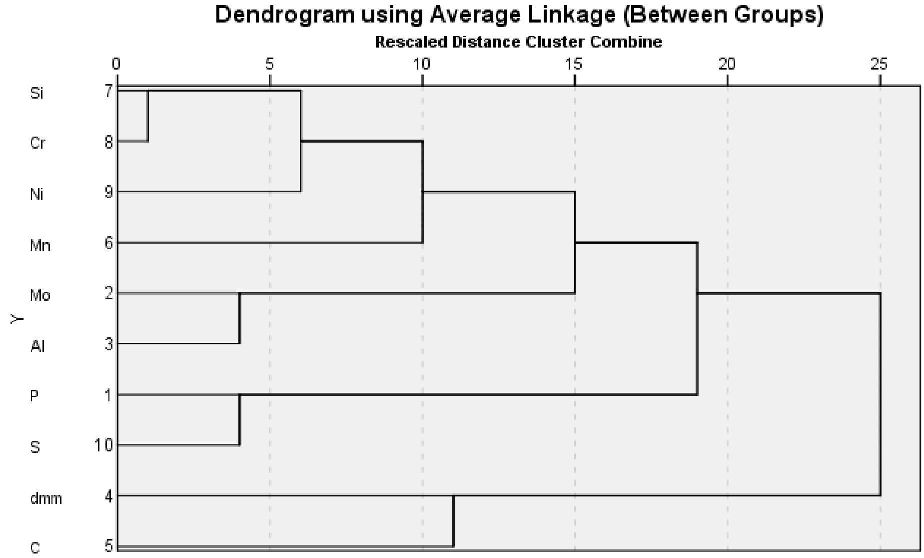

Figure 1. Dendrogram of the independent variables.

The next step is to indicate the place of the dependent variable – tensile strength – among the 10 independent variables. This is given in Table 5.

The grouping of the 10 variables into 5 clusters is maintained, with the dependent variable σ (Mpa) grouped in the second cluster together with quantities S and P. In both cases, tensile strength σ (Mpa) is significantly affected by the variables S and P.

The next figure 2 shows a dendrogram based on Table 5. It clearly shows that tensile strength σ (Mpa) is grouped in a cluster together with variables S and P.

Table 5. Distribution of independent variables and dependent into clusters.

Variable	8 Clusters	7 Clusters	6 Clusters	5 Clusters	4 Clusters
C (%)	1	1	1	1	1
σ (Mpa)	2	2	2	2	2
Mn	3	3	3	3	2
Si	4	4	3	3	2
Cr	4	4	3	3	2
Ni	5	4	3	3	2
S	6	5	4	2	2
P	6	5	4	2	2
Mo	7	6	5	4	3
Al	7	6	5	4	3
d (mm)	8	7	6	5	4

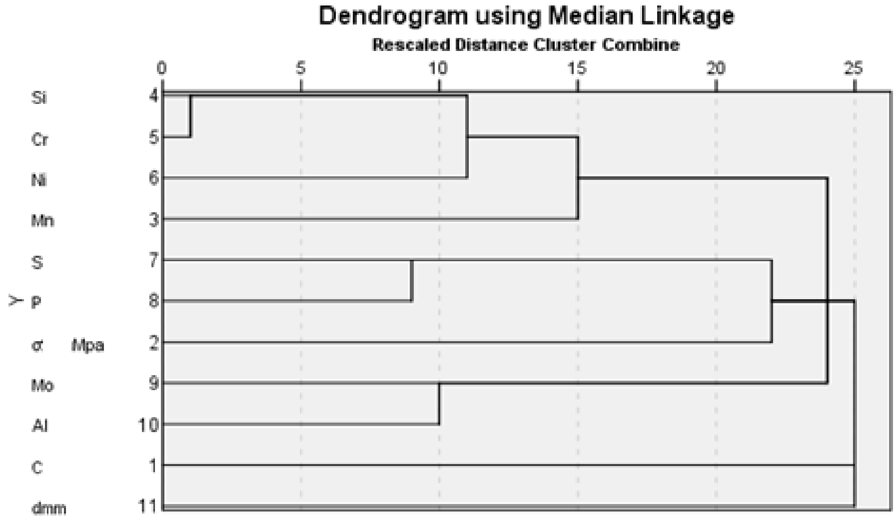

Figure 2. Dendrogram of the independent variables and dependent.

4. Analysis of results from factor and cluster analysis

The grouping of 9 chemical elements into 4 factors allows for the reduction of dimensionality of the problem when performing experiment studies in order to develop new steels with enhanced strength properties. The nine independent parameters, chemical elements Si, Cr, Ni, Mn, P,S, Mo, Al and C are replaced by 4 new latent independent variables (factors) F_1, F_2, F_3, F_4, F_5. The reason for this is that in each factor the independent variables are grouped together due to high mutual correlation. This means that the variables in each factor should change **at the same time**. In this manner, they lose their identity and start to represent a single independent variable, that of the factor.

The performed cluster analysis is confirmatory. It validates the results of the factor analysis.

In accordance with the results in Table 3, the coefficients for impurities S and P have a negative sign, which means that the values of S and P need to be reduced. As it is well-known, sulfur in steel produces FeS or MnS compounds and causes hot brittleness; therefore, its content should not exceed 0.003% to 0.005%. Where sulfur forms FeS, there is hot brittleness and cracks which form in the temperature interval (1000–1250) 0C. Phosphorus causes cold brittleness, which makes shaping more difficult and for this reason its content should not exceed 0.03%-0.04%. In accordance with the results in Table 3, alloy elements Mn, Cr,Ni and Si are found in one factor. This means that they are strongly correlated and should be modified together. They increase deformation resistance and hardness more strongly with higher carbon content. Si content from 0.17% to 0.35% and of Mn between 0.5% and 0.8% do not

IOP Conf. Series: Materials Science and Engineering **878** (2020) 012066 doi:10.1088/1757-899X/878/1/012066

influence ductility, however over these values ductility deteriorates. This further reduces the dimensionality of the problem of predicting new experiments.

The results obtained by factor and cluster analysis are significant when performing regression analysis. One of the crucial requirements for regression models is for predictor quantities to be mutually independent, i.e. they should not be strongly correlated. In this case, the factor analysis indicates strong correlation between the initial quantities and the use of linear regression analysis is statistically reliable. The development of regression models [4] in our opinion should always be preceded by factor analysis in order to establish the collinear dependence of the predictors. The development of regression models of second or higher order excludes the presence of multicollinearity between predictors and better reflects the nonlinear properties of processes in nature. For this reason, the second order regression model developed in [2] provides very good fit between experiment data and predicted results.

5. Conclusion

Factor and cluster analysis are performed on 9 independent quantities and the dependent variable – tensile strength. Independent quantities are grouped into 5 factors according to their degree of correlation. The results of the cluster analysis are confirmatory. It is found that the linear regression analysis with independent quantities cannot be statistically reliable due to the lack of strong correlation between some of these. It is recommended to perform regression analysis of second or higher order to improve fit with experiment results.

Acknowledgement
This work was supported by the Research and Development Sector at the Technical University of Sofia, Bulgaria.

References

[1] An Z L, Chen T, Cheng D L, Chen T H and Wang1 Z Y 2017 Statistical analysis and prediction on tensile strength of 316L-SS joints at high temperature based on Weibull distribution 1st International Workshop on Materials Science and Mechanical Engineering IOP Publishing. IOP Conf. Series: Materials Science and Engineering **281** 012062 doi:10.1088/1757-899X/281/1/012062

[2] Fakić B, Ćubela D, Burić A and Horoz E 2019 Regression analysis of tensile strength testing results steel 17-7 PH with modified chemical composition 14th International conference on Accomplishment in Mechanical and Industrial Engineering-DEMI 2019, Banja Luka, 24-25 May 2019, 691–697

[3] SungHo P, NoSeok P and JaeHoon K 2010 A statistical study on tensile characteristics of stainless steel at elevated temperatures 15th International Conference on the Strength of Materials (ICSMA-15) IOP Publishing *Journal of Physics: Conference Series* **240** 012083 1–4 doi:10.1088/1742-6596/240/1/012083,

[4] Golodnikov A, Macheret Y, TrindadeA A, Uryasev S and Zrazhevsky G 2005 Statistical modelling of composition and processing parameters for alloy development IOP Publishing *Modelling Simul. Mater. Sci. Eng*. **13** 633–644 doi:10.1088/0965-0393/13/4/013

[5] *Atlas Specialty Metals Technical Handbook of Bar Products* 2005 www.atlasmetals.com.au

[6] Bringas J E (ed) 2004 *Handbook of Comparative World Steel Standards* ASTM Intern. DS 67B.

[7] Izenman A J 2008 *Modern Multivariate Statistical Techniques. Regression, Classification, and Manifold Learning* (New York: Springer)

[8] SPSS IBM Statistics https://www.ibm.com/analytics/spss-statistics-software