# An Integrated Approach to Improve Effectiveness of Industrial Multi-factor Statistical Investigations

Victoria Miroshnichenko and Alexander Simkin



March 19, 2020

# An Integrated Approach to Improve Effectiveness of Industrial Multi-factor Statistical Investigations

Victoria Miroshnichenko[1[0000-0002-5956-7867]], Alexander Simkin[2[0000-0002-9939-7866]]

[1]Priazovsky State Technical University, University str., 7, 87555, Mariupol, Ukraine
miroviktoria@gmail.com
[2]Priazovsky State Technical University, University str., 7, 87555, Mariupol, Ukraine
simkin@ukr.net

**Abstract.** An approach was developed providing fully objective, mathematically comprehensive, scientifically grounded and physically interpretable description of the manufacturing factor effects on the performance of an industrial product, based on computer statistical analysis of the big, multi-dimensional arrays of industrial technological parameters. The approach integrates a basic Data Mining exploratory technique, multiple regression models construction and Monte-Carlo simulations. The approach was applied to industrial statistical data investigations for the ASTM A514 steel. The results obtained are in a good accordance with the known Material Science data and were confirmed in industry

**Keywords:** multi-dimensional data; exploratory technique; multiple regression models; Monte-Carlo simulations.

## 1 Introduction

One of the current trends of the modern stage of the Industry 4.0 development is to improve the big manufacturing technology data analysis techniques for increasing effectiveness of the Industry 4.0 platform components [1, 2]. The basic finishing goals of the components are, as it known, to increase manufacturing productivity, improve quality and reliability of a product by eliminating employed technologies lacks with minimal expenses. Cardinal role in the situation is played by the industrial computer statistical investigations because of: their high potentials in treating multifactor industrial phenomena; principal low effectiveness of laboratory researches, not enabling to simulate exactly the real industrial manufacturing conditions; practical impossibility to conduct the real, in depth industrial experiments in an operating plant. Nevertheless, the statistical techniques currently applied in the industry are not enough effective in meeting actual practical and theoretical challenges.

## 2 Brief literature overview

Typically as-recorded, raw industrial data require the preliminary treatment. The most widely currently used relevant tool in the case is the Data Mining technology, which is the collection of several computer aided statistical techniques [3-5]. Among the techniques the most effective ones today are the artificial neural networks (ANN) [6-8] and classification and regression trees (C&RT) [9, 10]. Both they have been effectively applied to solving a number of technology improvement problems [8, 11], particularly in the fields of Material Engineering, in contrast [12-14] to the direct use of the traditional statistical analysis techniques: multiple regression models, MANOVA, ANOVA etc. Nevertheless, both ANN and C&RT have some lacks, considerably decreasing their application effectiveness. Particularly, ANN is not capable of to express a regression dependence revealed in the conventional visual, mathematic and physically interpretable forms, while C&RT does not provide the discovered visual relations in a quantitative form of a regression equation. Such features are typical for the most of the statistical techniques that restricts the ability of the modern computer modeling and simulation technologies to be applied effectively in industrial practice and fundamental multifactor phenomena researches.

Aim of the paper, is to outline the main features and some obtained results of an industrial statistical investigation approach [11] integrating a data exploratory technique, quantitative multiple regression models building together with conducting the Monte-Carlo (MC) simulations.

Such an approach is aimed to provide under the real industrial conditions:

- revealing the statistically valuable industrial technology factors effecting each performance index of a product or process considered;
- on-line, semi-quantitative characterizing the factors separate and collective effects to solve possible actual technology problems;
- corresponding adequate regression models specifying;
- comprehensive on-line computer control of an industrial technology process;
- off-line computer investigations of the collective and separate effects for the revealed industrial factors with possible novel synergetic phenomena discovering;
- specifying the fields of possible further industrial technology improvements, deep specific laboratory applied or fundamental researches.

## 3 Methodologies

As the components of the approach the following techniques were consistently used:

- C&RT procedure resulting in the dendrogram building for each product performance index depicting the responsible technology factors and their effects with the 98 % confidence probability;
- based on the revealed variables construction of the multiple regression model for every control (dependent) characteristic;

- computer experiment workability verification for each built regression model with the possible model coefficients correction to achieve the highest adequacy;
- MC simulations of the traditional pair regression scatter plots obtained under the simultaneous change conditions for all responsible factors, which are typically built in the course of the conventional industrial quality analysis using unsorted, raw experimental data, as an additional workability verification tool for each regression model;
- MC simulations of the separate effects for each regressor under the constant values of the rest ones in a regression model, for the research purposes of the manufacturing technologies improvements.
- Multi-purpose optimization of the revealed technology parameters with applying a MC extremum searching technique.

## 4 Results and Discussion

Some results obtained by the above approach industrial application are considered below. The final goal of the conducted computer statistical research was to specify chemical composition and heat treating technology parameters values providing for ASTM A514 steel the standard mechanical properties combination which exceeds the technical requirements with 98% confidence probability. Actuality of the researches is caused by the extreme performance instabilities for the thick sheets made of the Boron containing steel due to its complex alloying and heat treating technologies interactions.

According to the methodology proposed, the first step of the investigations was C&RT analysis, resulting in the dendrograms building for each steel performance index. As it follows from the dendrogram for the steel yield stress shown on Fig. 1, the following statistically valuable technology factors effect the static steel strength: cooling duration at the steel quenching $t_{cool}^{Q}$ and tempering $t_{cool}^{temp}$, holding temperatures at austenitizing $T_A$ and tempering $T_{temp}$ together with the following chemical elements concentrations: V and B. Semi-quantitative characterizing the factors separate and collective effects on the static steel strength may be also visually obtained from the dendrogram.

Further step of the approach was to elaborate the multiple regression models, based on the corresponding dendrogram related data. The model developed for the steel yield stress based on the dendrogram shown on Fig. 1 is as follows:

$$YS = 400 + 63(B-4) + 15(V-35) + 0.03T_A\left\{1 + 25\cdot10^{-5}\cdot T_{temp}\cdot t_{cool}^{Q}\cdot\left[14-(0.12V+B)\right]\right\} \quad (1)$$

where B = % B·$10^5$ and V = % V·$10^3$; % B (V) – chemical element concentration, wt.%

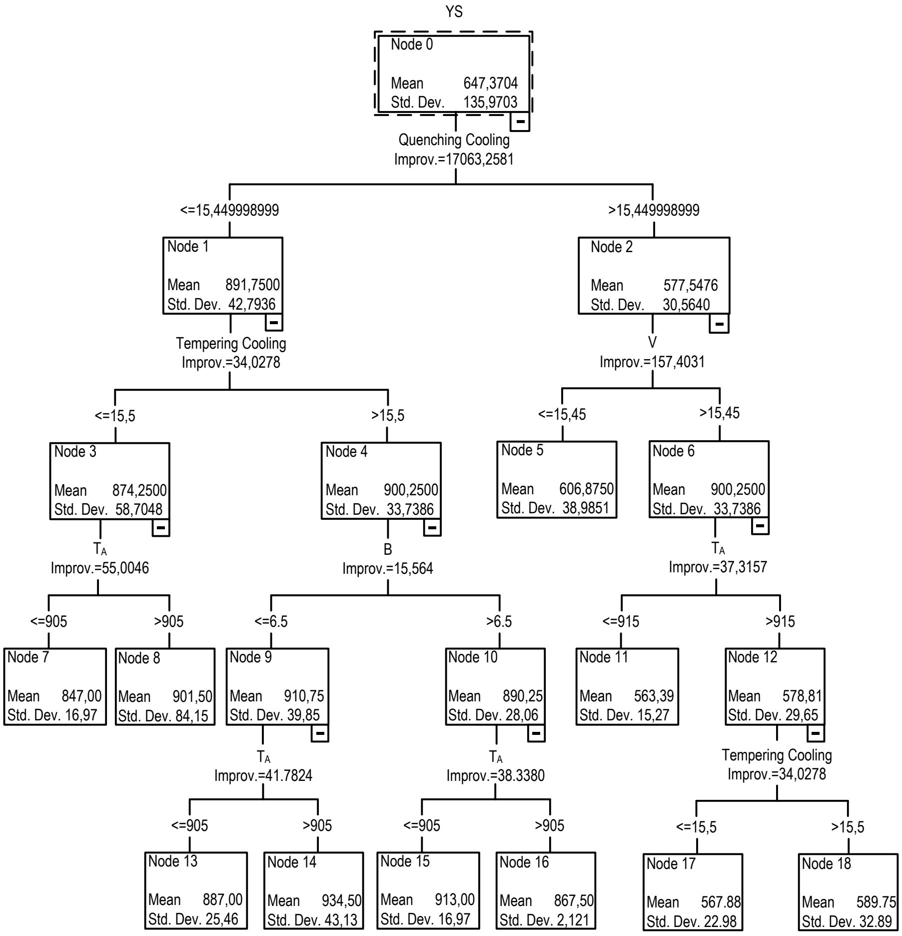


**Fig. 1.** C&RT dendrogram showing the statistically valuable industrial factors and their effects on yield stress of ASTM A514 steel in thermally improved state.

The necessary in such a case adequacy verification for the obtained regression models was conducted by the use of MC technique to build the frequency distributions for each performance index considered in the current investigation.

Taking into account the statistical comprehensiveness of such a mathematic description of a measured quantity, the procedure employed may be considered as a regression model workability verification. The MC distributions simulated based on the obtained regression models were compared with the real experimental ones. The simulation results for the considered above performance index obtained using the corresponding regression model are shown on Fig. 2 as the frequency distribution line. As it seen, good correspondence of the simulated (line) and real experiment (histogram) distributions had been reached.

An important formal advanced feature of the regression models like the shown above one should be outlined. Namely, the obtained models provide high workability in the performance descriptions by taking into account only real values of the technology parameters and their multiplications without using the terms of two or higher power. It allows to propose a real, physically grounded interpretation of such equations in terms of the industrial factors separate effects and their interactions. In turn, such conclusions may be further used for the corresponding phenomena mechanisms investigations. Workability of the models was also verified by the MC simulations of two dimensional scatter plots corresponding to pair regression relations of a performance index vs. an industrial factor, under the conditions of all the factors simultane-

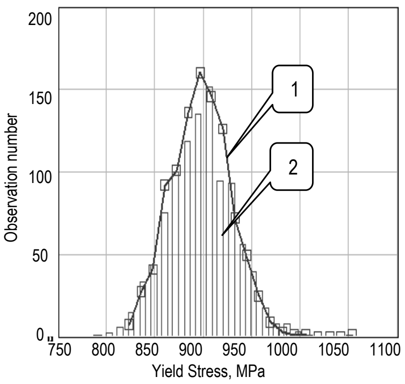


**Fig. 2.** Frequency distributions for yield stress of ASTM A514 steel in thermally improved state according to results of MC simulations (1) and real experiments (2)

ous variations. Such effect of the factor should evidently be considered as collective one caused by interactions of all the factors simultaneously changed.

Some examples of the MC simulated and real experimental scatter plots for the steel currently studied are shown on Fig. 3. As it seen, good agreement of the simulat-

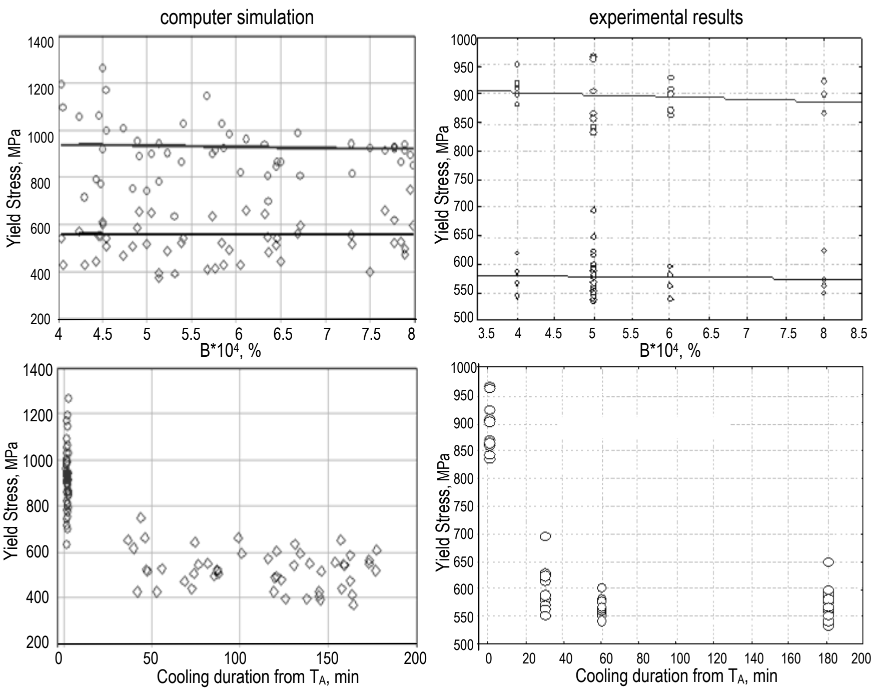


**Fig. 3.** Collective effects of some industrial factors on yield stress of ASTM A514 steel.

ed and experimental data is provided that is an additional confirmation of the regression model high workability.

An important role in control of multi-factor systems of complex physical-social-economical nature such as industrial technological processes, finished products etc. is played by specifying the sole effects of each valuable manufacturing factor on the performance indexes. As a rule, such information is unavailable under real industrial conditions due to simultaneous variations of numerous effecting manufacturing factors. Such research methodology restrictions can be avoided by analyzing currently available industrial data using the considered approach. An example of the revealed separate effects of the valuable manufacturing factors on the yield stress of ASTM A514 steel is shown on Fig. 4. The corresponding shown regression dependences were simulated for each valuable manufacturing factor which varies under some different constant values of the rest variables. These constant values for each accompanying variable were chosen randomly within the intervals of its possible variation in the steel. As it seen, varying the values of accompanying variables considerably influencing the steel yield stress level or even the general character of its dependence from a considered factor. Particularly, the boron dependence of the steel yield stress changes from decreasing to increasing type under simultaneous meeting the conditions: $> \sim 0.5$ % V; cooling duration from austenite temperature $< \sim 45$ min; cooling duration from subcritical temperatures $\neq \sim 10$ min; austenitizing temperature $< 900$ °C or $> 930$ °C. It should be additionally noted that the above results are in a good accordance with the known specific Material Science data concerned with the considered factors effects on structure and properties of corresponding steels and allow to explain the discrepancies often observed for boron containing steels in the literature.

Based on the results obtained using the applied integrated approach, some predictions were made aimed to improve of the industrial manufacturing technologies for the steel, particularly, its chemical composition and heat treatment technology. The technology parameters such predicted provide guaranteed exceeding the technical requirements to the steel performance indexes with 98% confidence probability. The adequacy of the corresponding technology recommendations was verified in real industrial conditions: the following combination of the performance indexes for thick sheets made of the researched steel was provided: $YS = 950 \pm 40$ MPa, $\psi = 56.5 \pm 4\%$, $KV = 44 \pm 2$ $J$. It should be noted the considerably low standard deviations for the just given standard mechanical properties characteristics, that is in violent contrast with the previously obtained industrial data, particularly shown in Fig. 1 and Fig. 2. So, the results of the developed integrated statistical investigation approach employment show considerable improvement of the finished industrial product reliability comparing with the same product made in traditional industrial conditions.

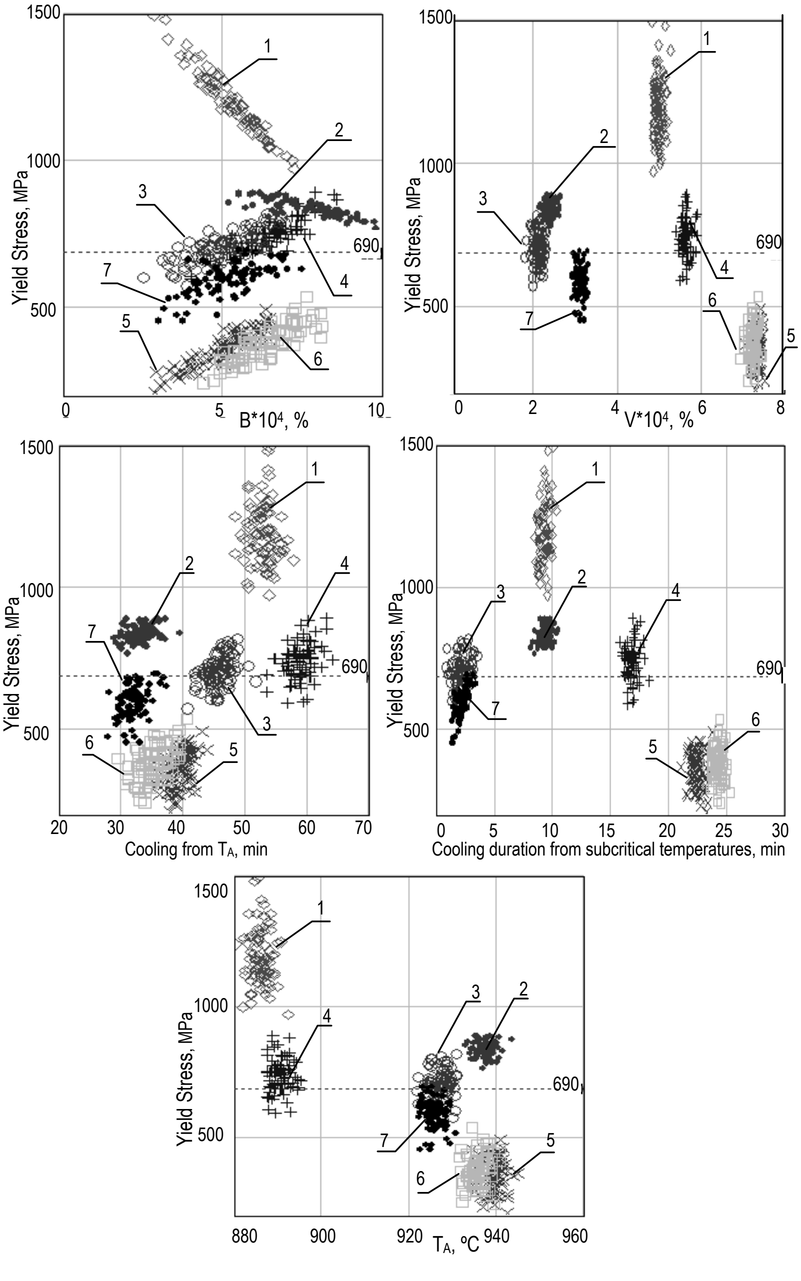


**Fig. 4.** Computer simulated scatter plots showing separate effects of the revealed valuable manufacturing factors on yield stress of ASTM A514 steel. Numerals on the plots correspond consecutive numbers of randomly chosen combinations of the variables values used in Eq(1)

## 5 Conclusions

1. In view of the Industry 4.0 needs, an approach to the industrial statistical investigations was developed aimed to improve effectiveness of big data, multi-dimensional array analysis and its results practical applications.
2. The developed approach integrates: C&RT technique, as a Data Mining procedure allowing the obtained results further physical interpretation and mathematical treatment; multiple regression models building to express the revealed draft regularities in a rigorous mathematical form; Monte-Carlo simulations to verify the regression models, to conduct computer investigations and outgoing product quality index predictions, to specify further research areas.
3. The approach developed was applied to solving some industrial quality and reliability problems for thick sheets made of boron-containing ASTM A514 steel concerned with its typical low and unstable yield stress and impact resistance on the levels: $YS = 900 \pm 100$ *MPa*, $KV = 35 \pm 12$ *J*.
4. The main performance indexes values obtained in industry for the steel with the confidence probability 98%, as a result of the approach application, are as follows: $YS = 950 \pm 40$ *MPa*, $KV = 44 \pm 2$ *J*.
5. As a result of the approach application the following technology advantages have been reached providing the guaranteed finished industrial product performance improvements:

- specification of industrial technology factors having valuable effects on the product quality and reliability;
- development of regression models providing the statistically comprehensive description of the revealed effects;
- determination of separate effects for each of the revealed factors and conditions of the effects realization;
- determination of the multipurposely optimized industrial technology parameters providing increase and stabilizing a combination of the finished product quality indexes.

## References


1. Erboz, G.: How To Define Industry 4.0: Main Pillars Of Industry 4.0. Proceedings of 7th International Conference on Management (ICoM'2017), Nitra, Slovakia, July 2017, pp. 245 – 249 (2017)
2. Hermann, M., Pentek, T., Otto, B.: Design Principles for Industries 4.0 Scenarios. Proceedings of 49th Hawaii International Conference on System Sciences (HICSS), Koloa, HI, pp. 3928-3937(2016)
3. Lee, J., Bagheri, B., Kao, H.: Recent Advances and Trends of Cyber-Physical Systems and Big Data Analytics in Industrial Informatics. 10.13140/2.1.1464.1920. (2014)

4. Lee, J., Lapira, E., Bagheri, B., Kao, H.: Recent advances and trends in predictive manufacturing systems in big data environment. Manufacturing Letters, vol. 1, pp. 38-41, doi.org/10.1016/j.mfglet.2013.09.005 (2013)
5. Witten, I., Frank,E., Hall, M.: Data Mining: Practical Machine Learning Tools and Techniques. 3rd Edition, Morgan Kaufmann (2011)
6. Bethge, M., Ecker, A., Gatys, L.: A Neural Algorithm of Artistic Style, arXiv: 1508.06576 (2015)
7. Ojha, V., Abraham, A., Snášel, V.: Metaheuristic design of feedforward neural networks: A review of two decades of research. Engineering Applications of Artificial Intelligence. vol.60, pp. 97–116, arXiv: 1705.05584 (2017)
8. Engineering Applications of Neural Networks. Proceedings12th International Conference, (EANN 2011) and 7th IFIP International Conference (AIAI 2011), Corfu, Greece, September 15-18, 2011, vol. 1 (2011)
9. Wang, F., Rudin, C.: Falling Rule Lists. Journal of Machine Learning Research. vol.38, pp. 21-25 (2015)
10. Gareth, J., Witten, D., Hastie, T., Tibshirani, R.: An Introduction to Statistical Learning. New York, Springer. p. 315 (2015)
11. Tkachenko, I., Tkachenko, K., Miroshnichenko V.: A Combined Approach to Provide a Guaranteed Combination of Steel Quality Characteristics. Contributed Papers from Materials Science and Technology 2018 (MS&T18). October 14-18, 2018, Greater Columbus Convention Center, Columbus, Ohio, USA. DOI 10.7449/2018mst/2018/mst_2018_1264_1270 (2018)
12. Tkachenko, I., Bolshakov, V., Nosenko, O.: An analysis of collective effects of chemical elements in multi-component alloy steels on their performance by means of Data Mining techniques, Metaloznavstvo ta termichna obrobka metalliv. Dnipro, Pridneprovskaya State Academy of Civil Building and Architecture. №4 (23), pp.36-44 (2003)
13. Tkachenko, I. Machine parts service efficiency forecasting based on the multipurpose optimization of material performance. Metal and casting of Ukraine №7-8, pp. 72-77 (2005)
14. Tkachenko, I.: Multi-purpose optimization of heat strengthening technology for thick sheets made of high strength weldable steels using the Data Mining computer technology. Visnik of the Priazovsky State Technical University: Scientific collected works. Mariupol, №14, pp. 111-117 (2004)