



TODOR KABLESHKOV UNIVERSITY OF TRANSPORT



向更美好而变-

- Change for the better

变化的时刻 -

- Time to improve



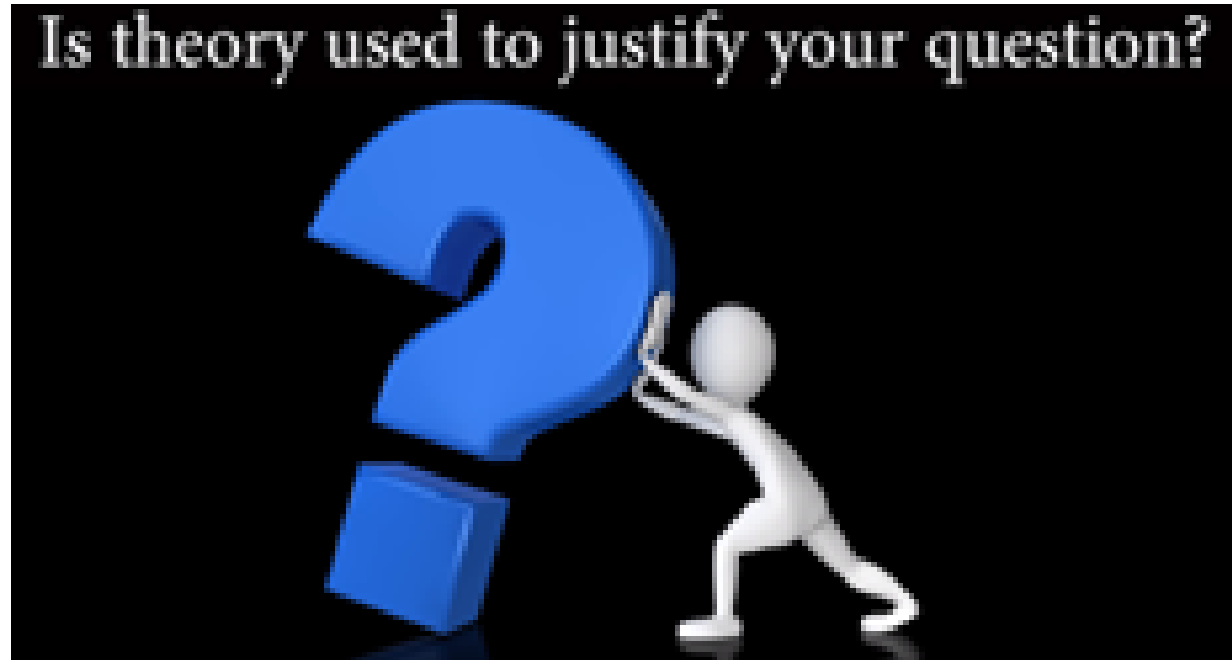
Real data science,
fast and simple.

Key Features for
Multi - Criteria
Support for
Decision - Making
by Shifting
Restrictions

Prof. Nikolay TONTCHEV,



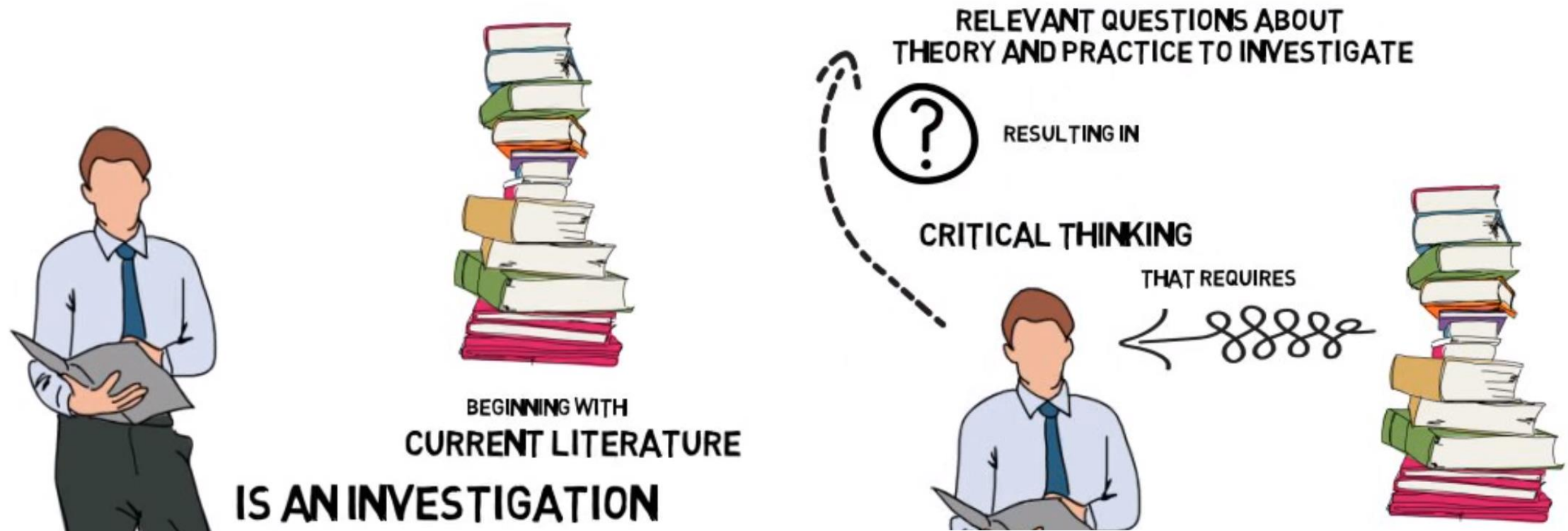
A complex, fully completed study is what is based on an existing theory, and the obtained results are experimentally confirmed. In the following below presentation, for the accredited method, the obtained adequate models need to confirm the predicted results that are not involved in the model's output. Everything described below is applied to just a single or a system of regression models. This approach can be applied to single-criterion and multi-criteria optimization, for which the approximation is performed with a neural model.





The timeliness of this activity is determined on the basis of the benefits achieved, as in any other optimization process. Unlike classical optimization, technical decision making takes place under more than one criteria, with a different number of control parameters. For this reason, our team has built several software applications to support the process of this application from engineering practice.

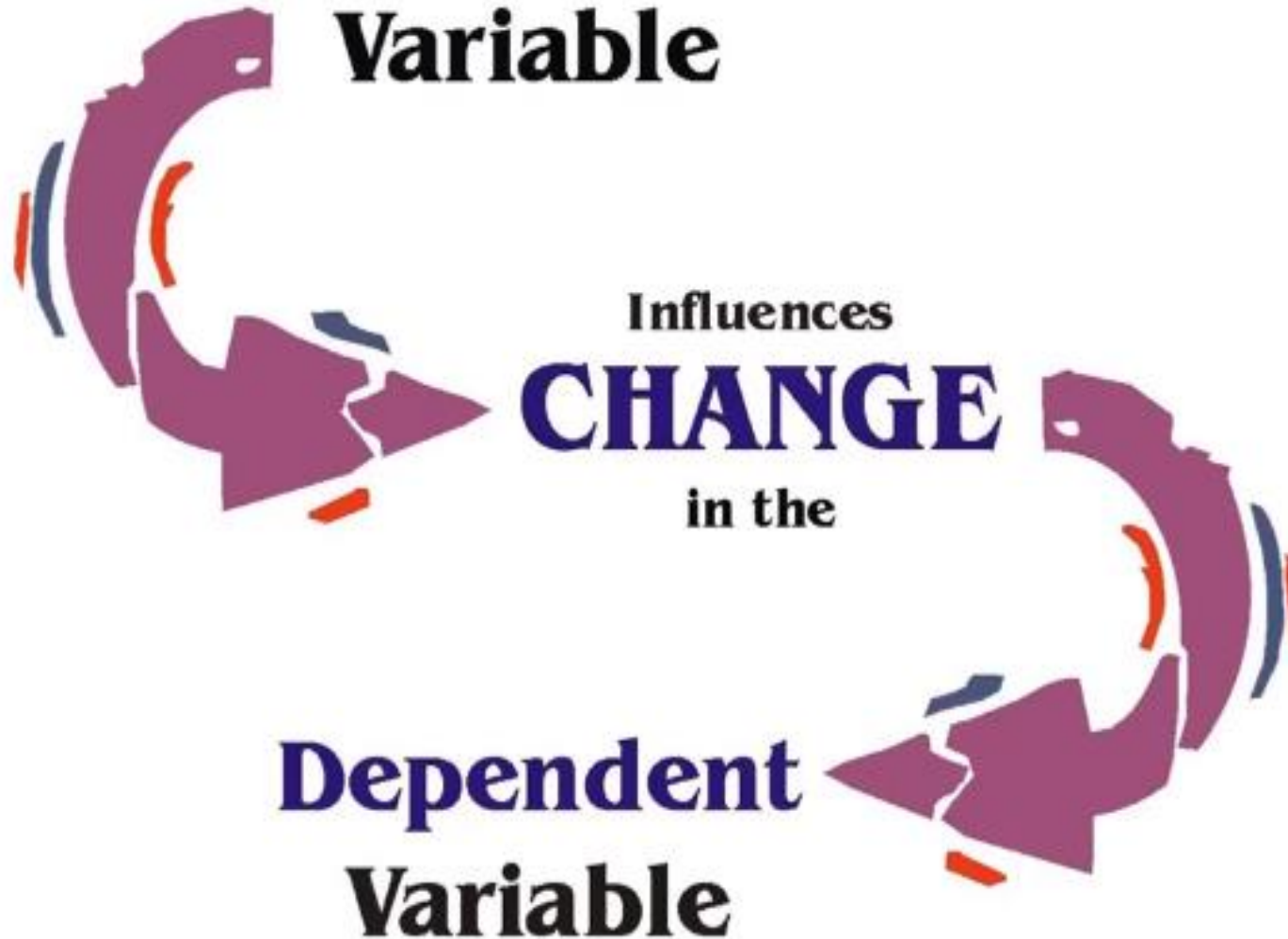
The problem definition begins exploring the available literature. If a suitable database from an experimental or simulation research is found in the available bibliography, it is possible to clarify the boundaries and objectives of the problem to plan the new experimental program by means of a preliminary research. In this process, critical thinking is most important .



Approved software can be defined as a multifunctional tool to reach the desired solution. In management theory independent variables are defined as input factors /varying within a defined range/ and dependent variables are the output parameters. The method embedded in the software is extremely suitable for systems where these two sets of parameters are not optimized. The system / object / process optimization can be performed for several outgoing parameters at predefined requirements. The software defines all effective combinations of input factors that fulfill the desired output parameter preferences.



**Independent
Variable**

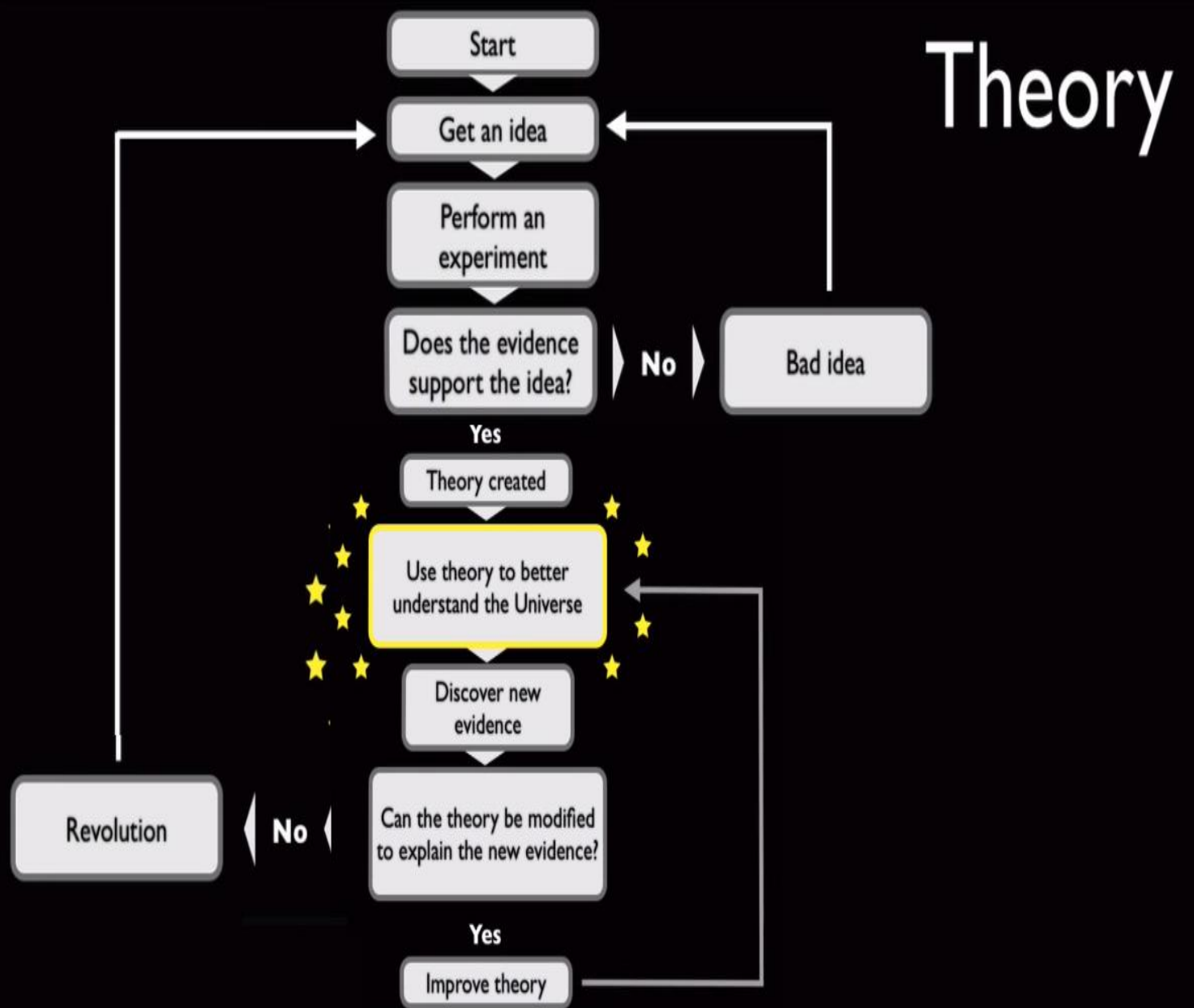


The software is extremely useful in exploring a set of quality indicators, as in the material science is the complex of properties after applied processing. Processing parameters are process input control parameters, and quality indicators are output controlled reactions. Multicriteria optimization defines these process modes of the research process, for which the user has explicitly defined certain preferences of the quality indicators.

The formulation of the problem relates the ideas and objectives of the research to its subject. If there is enough evidence supporting the idea after a numerical or real experiment, then the research is justified and it can be conducted. The approach, through the software, discovers new system / object / process / laws and evidences that explain better the theory by visualizing mutual influences. Emphasis is placed on visualization because the suggested friendly approach graphically visualizes processes and objects up to 5D – dimensions for graphical analysis or 11D for numerical calculations. The basis of this visualization is the principle of multi-criteria support for decision-making on saving energy or materials.

Indeed, how is this approach applied to energy and materials savings. Among the many technologies investigated with the software, a chemico-thermal process was developed to enhance the working properties of heat-resistant steels. Several solutions with a desired set of properties have been identified to solve the task. This is characteristic of any multi-criterion task. Each particular solution corresponds to different modes that vary considerably over the duration of the process or the pressure of the gas used. Thus, among these solutions, a mode with a shorter chemico-thermal treatment time is chosen which is more energy efficient and less gas consuming. Thus, on the one hand, a compromise solution is established that satisfies all the quality indicators and, on the other hand, provides less energy or material consumption.

The number of previously developed four parameter controls can be considered as optimal. For tasks with more established influences, they can be transformed and solved in steps / parts.



The experimental method embedded in the generalized algorithm can identify and optimize optimization output outcomes that are the result of different combinations of input, technological factors when a process is the objective. The methodology can be equally applied for both real-world experiments imitating a given manufacturing process and also for simulation tests from the CAE procedure applied in the design process.

THE EXPERIMENTAL METHOD

...IS A WAY TO ISOLATE AND
IDENTIFY THE CAUSE OF
SOMETHING

The Experimental Method

- The only research method that can identify cause-effect relationships

Dependent and independent variables define the problem. When executing a certain plan (purposefully modifying combinations, through input-factor levels), their output parameters are determined.

Independent Variable



The variable that is changed
on purpose by the experimenter



Also known as cause, stimulus, reason,
or manipulated variable

Dependent Variable



The variable that responds



Also known as effect, result, or
responding variable

Regression analysis explores the relationship between a quantitative response variable and one or more explanatory variables.

In the generalized algorithm of the theory through regression analysis the validity of the idea is checked.

The regression analysis gives the structure and the coefficients of the model (s). They help define multi-criteria problems. Models can be polynomials up to third degree with up to four input factors and with no restrictions for the output parameters.

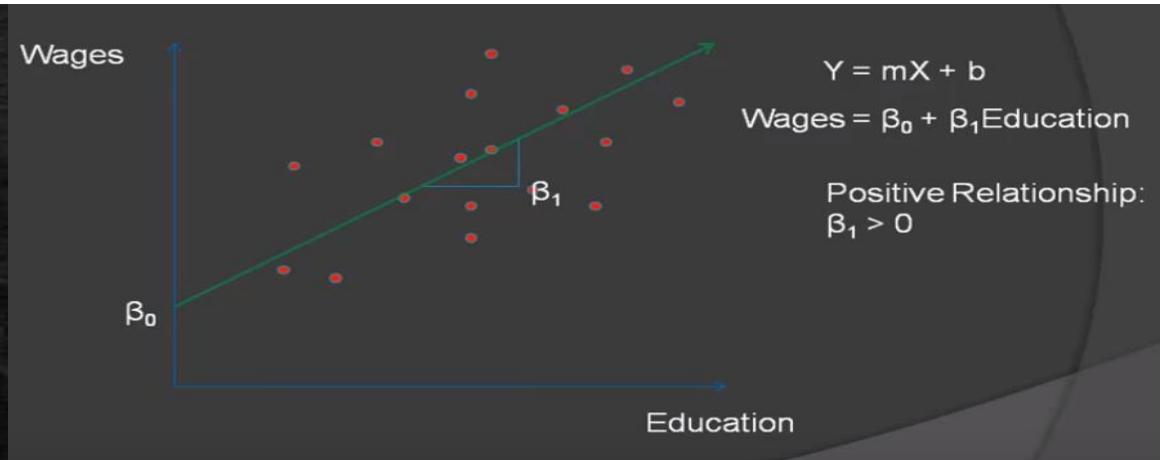


In statistics, **regression analysis** includes any techniques for modeling and analyzing trends between a dependent variable and an independent variable.

Regression analysis helps us make predictions outside of the given data. If the prediction is within the range of the given data it is called **interpolation**. If the prediction is outside the range of given data values it is called **extrapolation**.

$$y = \beta_0 + \beta_1 x + \varepsilon$$

- y is the dependent variable
- x is the independent variable
- β_0 is the constant or intercept
- β_1 is x 's slope or coefficient
- ε is the error term



The regression analysis gives the structure and the coefficients of the model (s). They help define multi-criteria problems. Models can be polynomials up to third degree with up to four input factors and with no restrictions for the output parameters.

The algorithm analyzes and optimizes parameters after an engineering experiment. Since actual experimental data is used, if the models obtained prove the necessary checks, it means that the models are adequate and the forecasts obtained are reliable and can be used in engineering practice.

The algorithm analyzes and optimizes parameters after an engineering experiment. Since actual experimental data is used, if the models obtained prove the necessary checks, it means that the models are adequate and the forecasts obtained are reliable and can be used in engineering practice.

In the modeling area, when the regression model is displayed, it is necessary to specify the connections between the control and the managed parameters as input data. They can be pre-planned or consecutively executed, unassigned in the so-called passive experiment. For a larger number of data processing observations, a different pattern structure can be applied. Each structure is evaluated with two estimates

Example

$$\hat{y} = 6.211 + 0.014x_1 + 0.383x_2 - 0.607x_3$$

variables

intercept

coefficients

Estimated Multiple Regression Equation

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3$$

$b_0, b_1, b_2, \dots, b_p$ are the estimates of $\beta_0, \beta_1, \beta_2, \dots, \beta_p$

\hat{y} = predicted value of the dependent variable

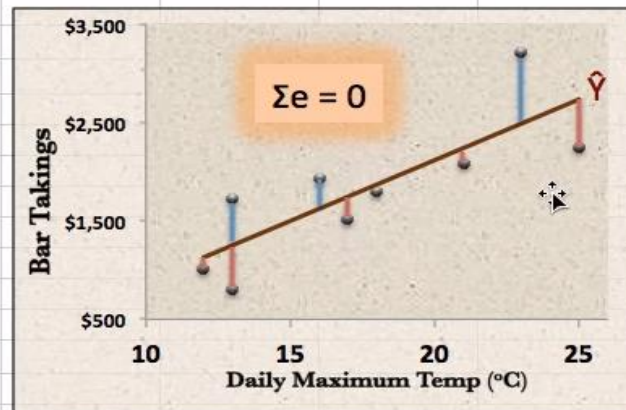
The decision-maker chooses the best structure for these ratings. The structure determines the respective coefficients of the regression model. The determined coefficients define the magnitude examined. Several dimensions investigate define the criteria in the multi-criterion task, with preferences for them.

The basis of model evaluation lies with the residuals between the experimental / numerical values and those obtained through the model.

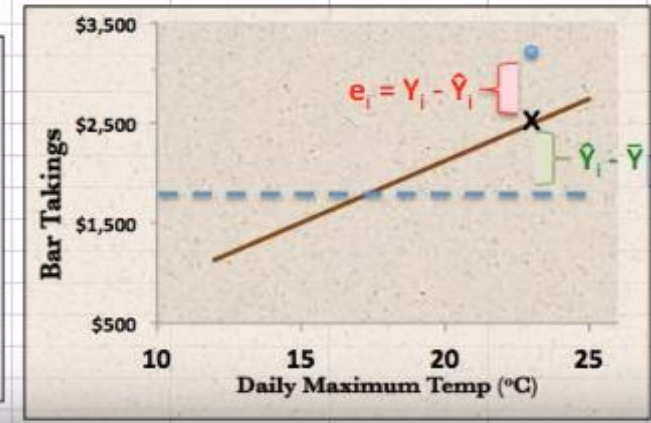
Day	Takings	Temp (°C)
3-Jun	\$3,213	23
10-Jun	\$2,089	21
17-Jun	\$2,253	25
24-Jun	\$1,801	18
1-Jul	\$801	13
8-Jul	\$1,934	16
15-Jul	\$1,720	13
22-Jul	\$1,514	17
29-Jul	\$1,017	12

SAMPLE REGRESSION LINE

$$\hat{Y} = -353.11 + 123.54X$$



$$\begin{aligned} SSR &= \Sigma (\hat{Y}_i - \bar{Y})^2 \\ SSE &= \Sigma (Y_i - \hat{Y}_i)^2 \\ SST &= SSR + SSE \\ SST &= \Sigma (Y_i - \bar{Y})^2 \\ R^2 &= SSR/SST \end{aligned}$$



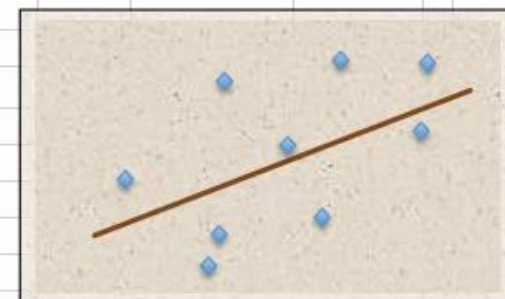
$$\{ (X_1, y_1), (X_2, y_2), \dots, (X_n, y_n) \}$$

Residual Error of sample i :

$$e_i = y_i - \hat{y}_i$$

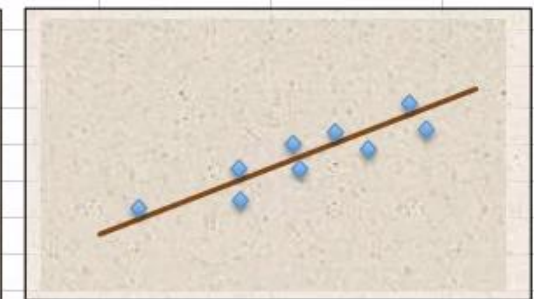
Sum of Squared Residuals (RSS):

$$RSS = \sum_{i=1}^n e_i^2$$



High SSE

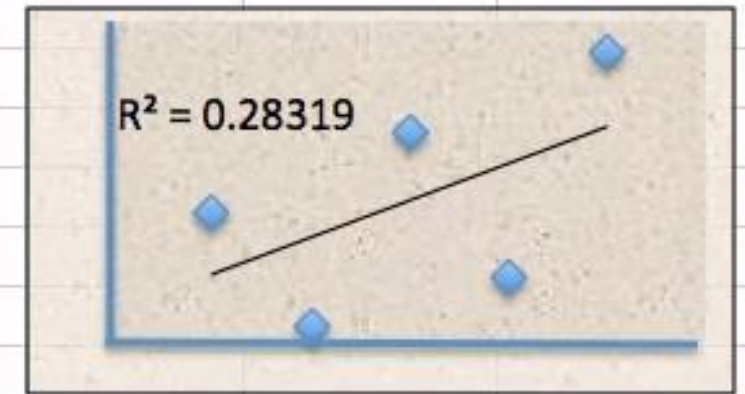
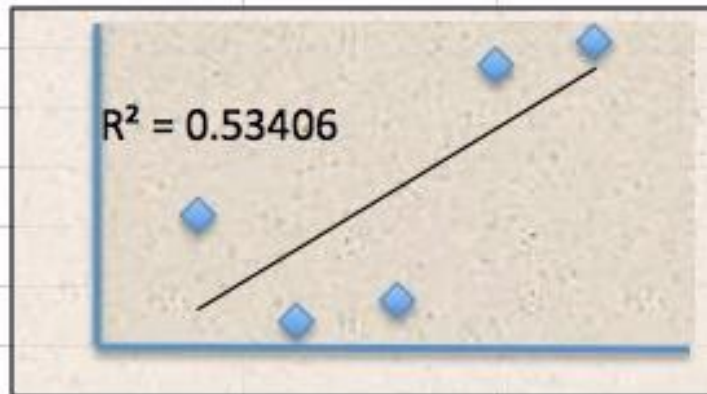
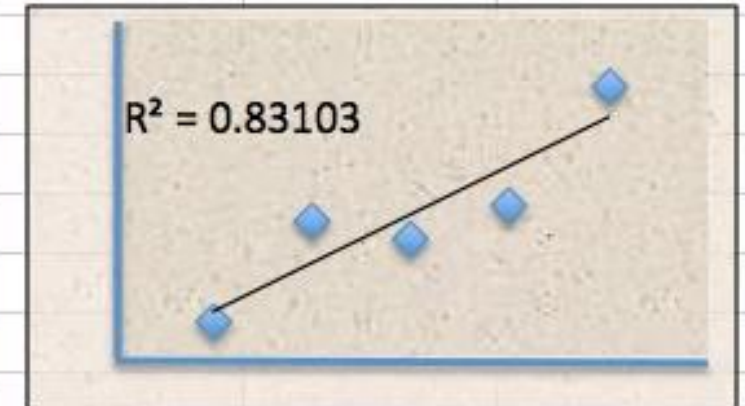
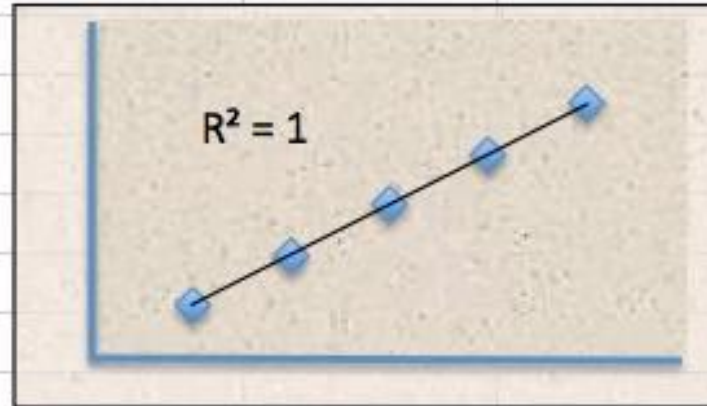
Low R^2



Low SSE

High R^2

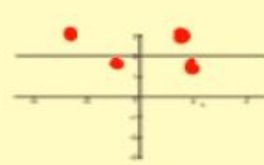
From the scattering of the debris to the model curve, plane or hyper-surface, the determination coefficient is determined. In the attached example, graphical data representations are presented for linear regression, for which the determination coefficients were defined. The software performs one more verification before giving a conclusion on adequacy.



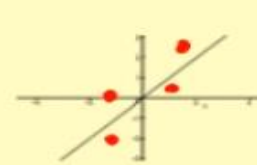
The chosen model depends on the data distribution to the input factor boundaries. Selected approximations have the most optimal distribution with respect to the residuals.

Graphs of Basic Functions

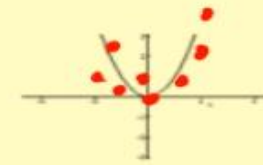
What
function
best fits the
pattern of
the data?



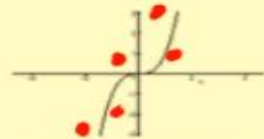
constant function



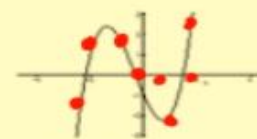
linear function



quadratic function



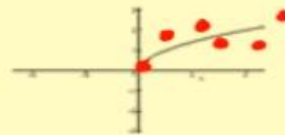
cubic function



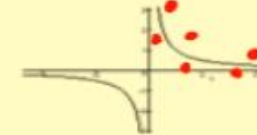
cubic function



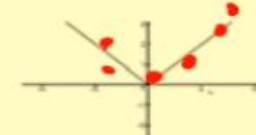
polynomial function



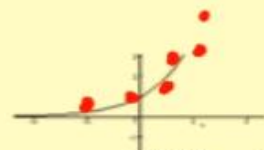
square root function



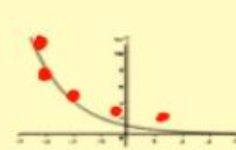
rational function



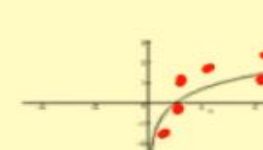
absolute value function



exponential function



exponential function



logarithmic function

In the multiparametric non-linear approximation, the software performs similar calculations, that are improved on the research. Through them, the regression coefficients are determined for a chosen structure of the model.

Multiple Regression

$$\{ (X_1, y_1), (X_2, y_2), \dots, (X_n, y_n) \}$$



Least Squares Criteria

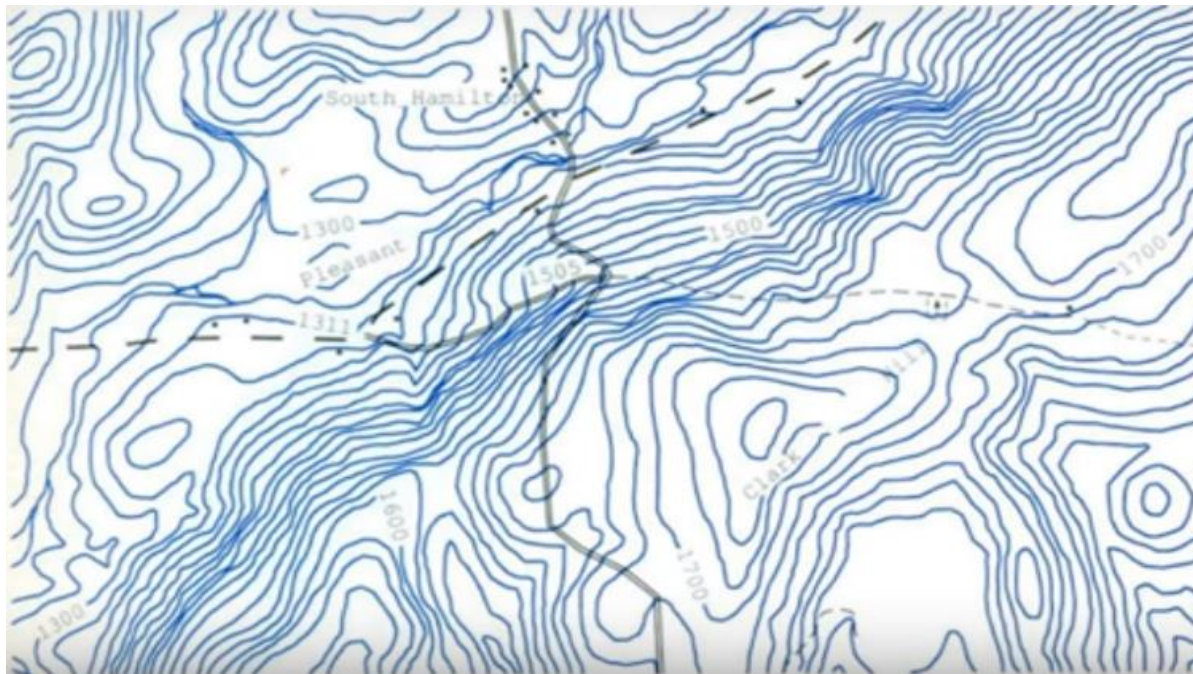
$$\hat{\beta} = (X^T X)^{-1} X^T y$$



$$\hat{y} = \widehat{f(X)} = X\hat{\beta}$$

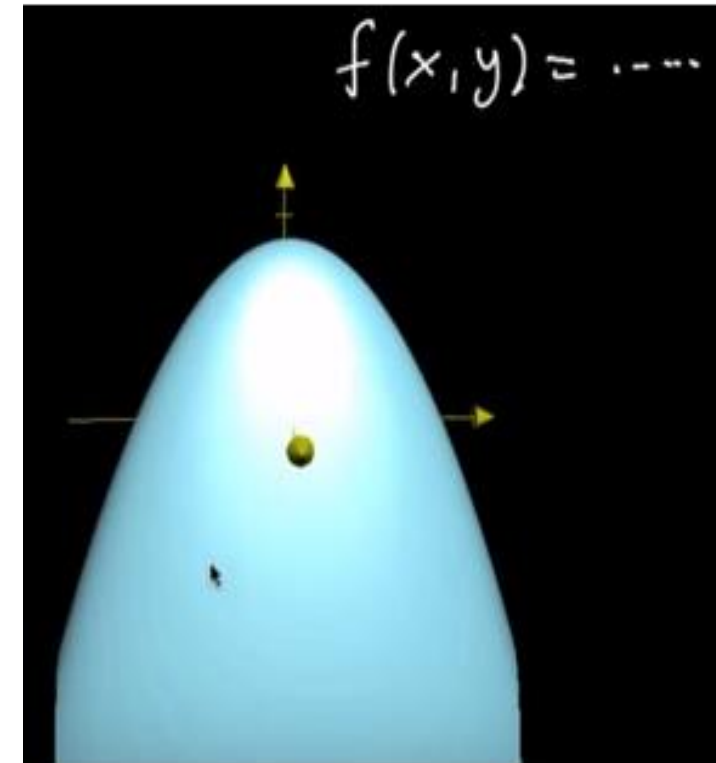
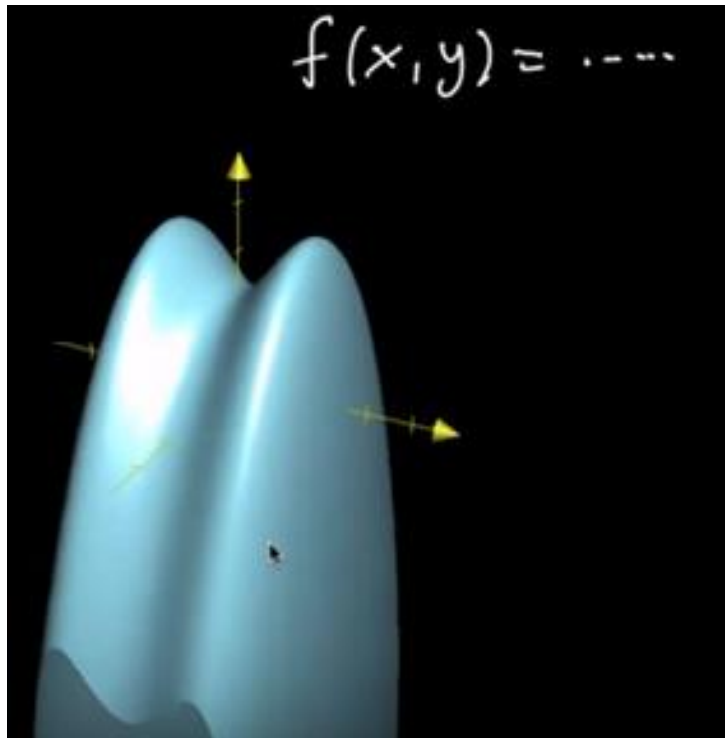
The analysis that is applied is user-friendly. This analysis is valuable because it provides solutions for multifactor processes. Various alternatives can be evaluated. In the future, under the proposed algorithm, there is an idea to develop further with control parameters for which the total number is up to ten parameters describing the technological regime.

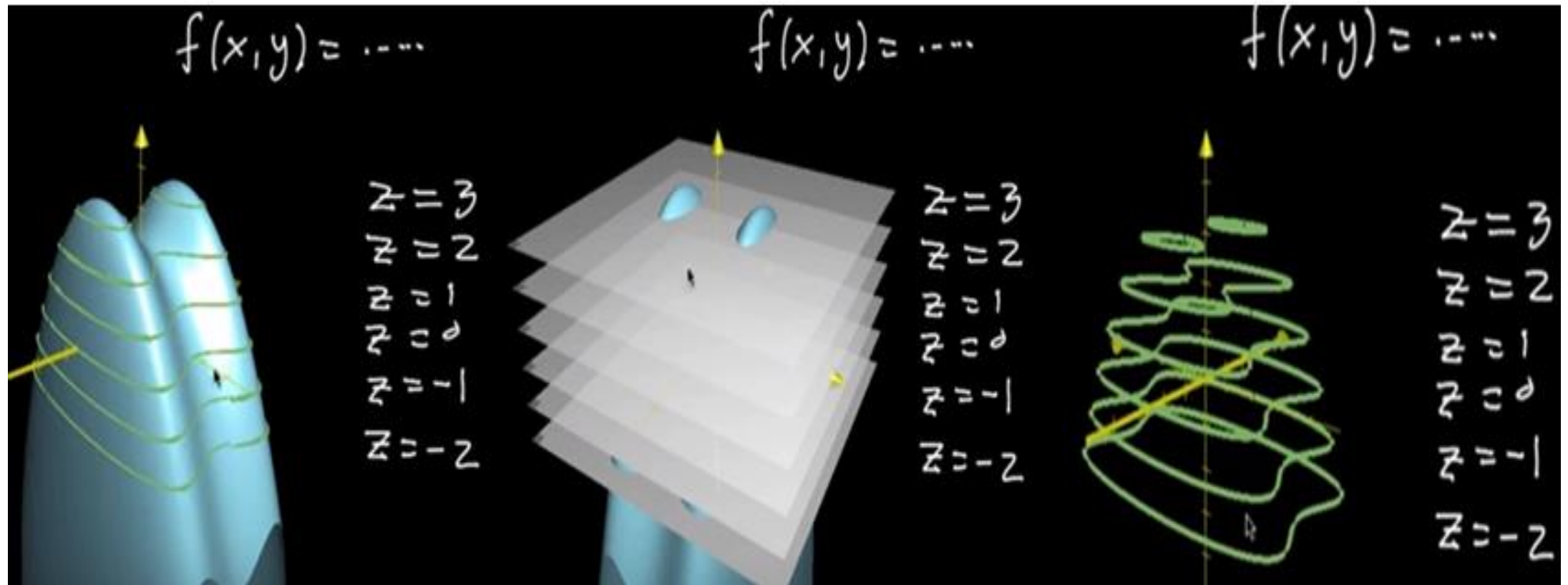
What is the idea of visualization and what's new in it?



An established practice in the analysis of regression models is the use of bi-dimensional graphs with lines at a constant level . These diagrams are an established tool in cartography.

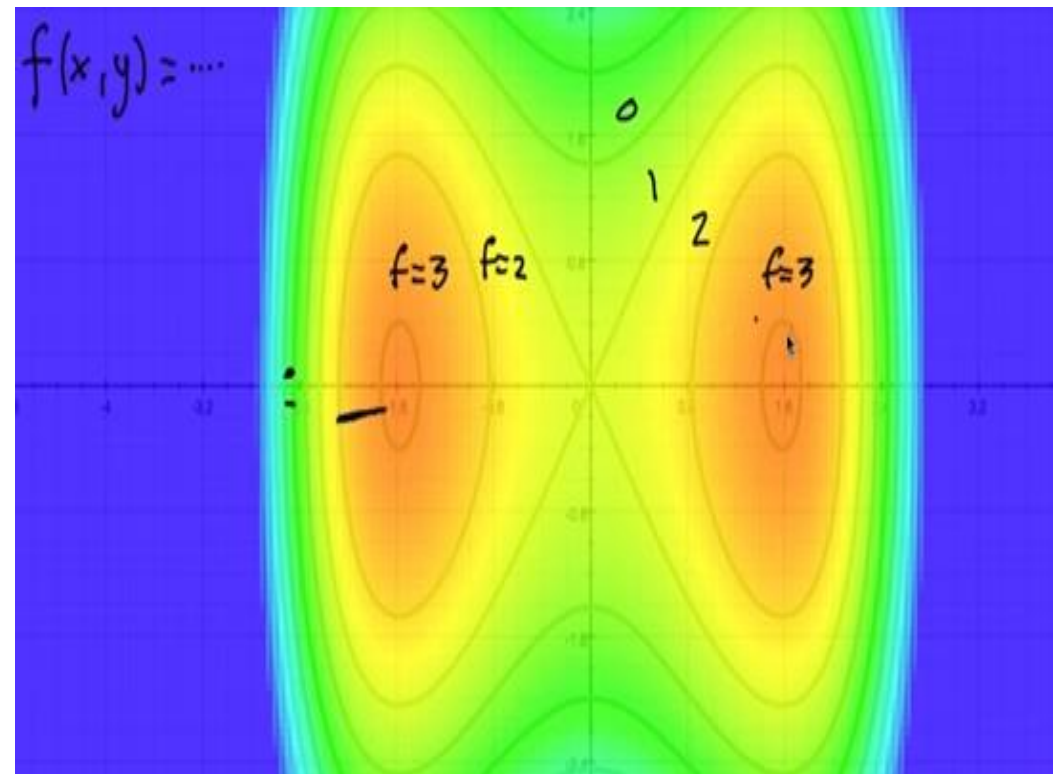
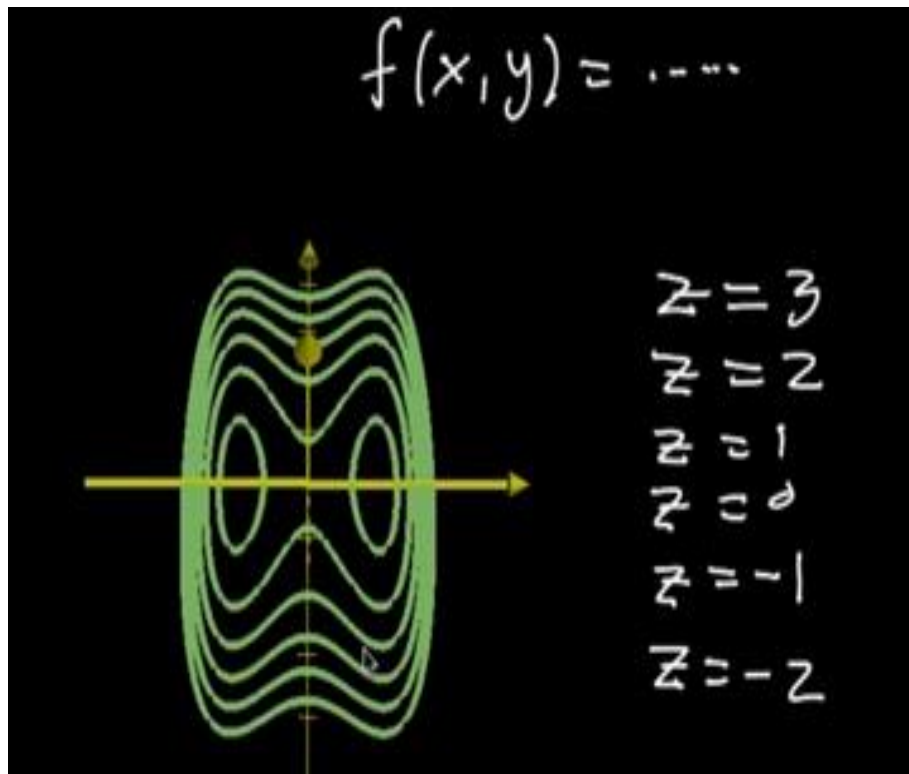
The visualization suggested in the approach uses elements of this analysis. In order to reveal the idea, the peculiarities and the differences between this method and the new one, the following example is considered. Let's look at the model represented by the images.

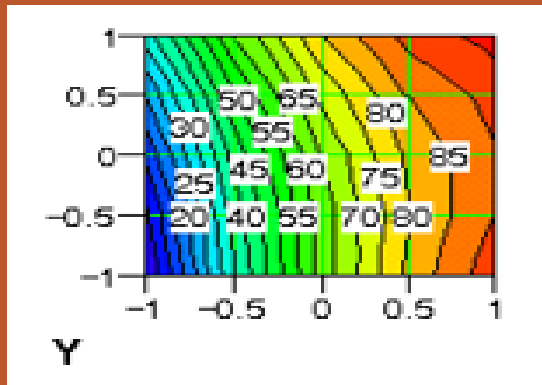




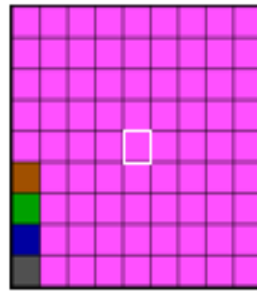
Boreholes cut the three-dimensional image of the model in height. The cut section is projected on each plane. This produces the corresponding contour line. At the last stage, the lines are gathered in a general image.

Valuable analysis of two and many parametric processes can be applied because the chosen approach takes place in the space of the variables.

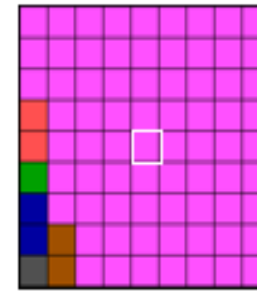




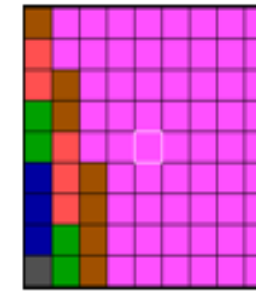
The chosen approach in our software selects the plane of the variables and normalizes the value of the research value in percentages from 0-100%. These features lead to the novelty of the proposed solution, which is the ability to vary with the number of moving planes and the variation in distance and color between them. The demonstration of this effect is indicated by movement(s) to the maximum and minimum of a two-parametric model, shown below.



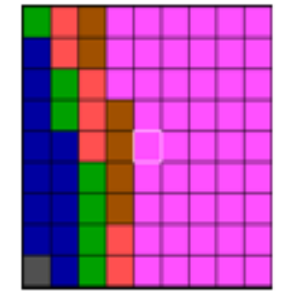
Y[1]	0	..	2	%
Y[2]	2	..	4	%
Y[3]	4	..	6	%
Y[4]	6	..	8	%
Y[5]	8	..	100	%



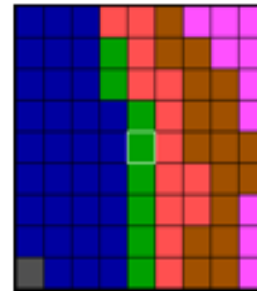
Y[1]	0	..	5	%
Y[2]	5	..	10	%
Y[3]	10	..	15	%
Y[4]	15	..	20	%
Y[5]	20	..	100	%



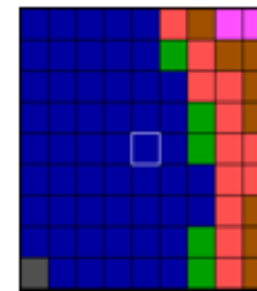
Y[1]	0	..	10	%
Y[2]	10	..	20	%
Y[3]	20	..	30	%
Y[4]	30	..	40	%
Y[5]	40	..	100	%



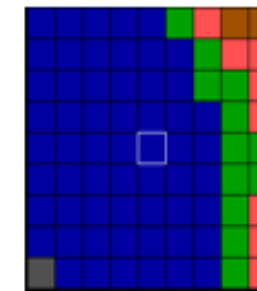
Y[1]	0	..	30	%
Y[2]	30	..	40	%
Y[3]	40	..	50	%
Y[4]	50	..	60	%
Y[5]	60	..	100	%



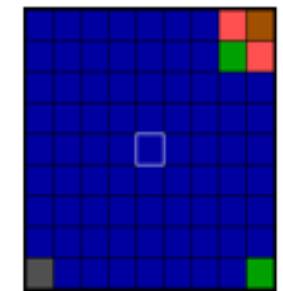
Y[1]	0	..	60	%
Y[2]	60	..	70	%
Y[3]	70	..	80	%
Y[4]	80	..	90	%
Y[5]	90	..	100	%



Y[1]	0	..	80	%
Y[2]	80	..	85	%
Y[3]	85	..	90	%
Y[4]	90	..	95	%
Y[5]	95	..	100	%

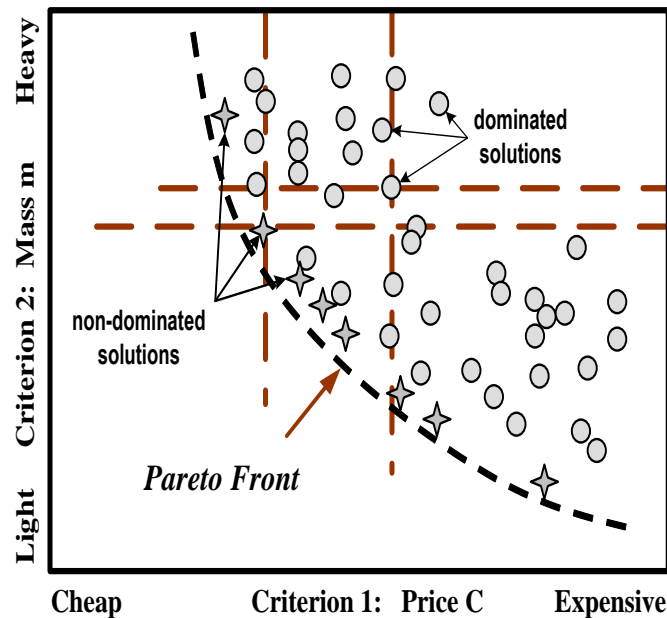


Y[1]	0	..	80	%
Y[2]	85	..	90	%
Y[3]	90	..	95	%
Y[4]	95	..	99	%
Y[5]	99	..	100	%



Y[1]	0	..	92	%
Y[2]	92	..	94	%
Y[3]	94	..	96	%
Y[4]	96	..	98	%
Y[5]	98	..	100	%

This visualization is the basis of a friendly multi-criteria optimization procedure.



Any effective solution, by its very nature, can also be innovative. Effective solutions are Pareto's solutions. These solutions are not improving optimal solutions. Strategies for determining effective solutions can be varied: average, geometric, and so on. Our approach uses the strategy of the pessimistic option. This strategy maximizes solutions in a matrix containing the smallest value of the criteria analyzed. The performance of this analysis is possible because of the visualization of decisions in the space of the control parameters. The investigated managed parameters are projected into the space of the technological modes of a certain color, which is determined by the corresponding value of the survey parameter. The value of this parameter is normalized in the range [0-100%]. In this scale are set the intervals of the corresponding coloring. In this color variation, both the number of the respective colors can be chosen, so the interval of the respective colorings. This tool is also used in the regression model analysis phase, and in the multi-criteria optimization stage.

<https://www.researchgate.net/publication/316889741> Principle of operation and applications of the MADML software

<https://www.researchgate.net/publication/309823792> GENERALIZED ALGORITHM FOR NUMERICAL ANALYSIS AND MULTICRITERIA OPTIMIZATION OF MULTIPARAMETRIC REGRESSION MODELS

<https://www.researchgate.net/publication/265292354> BOOK with methods to metallurgical design alloys - <https://www.morebooks.de/store/gb/book/materials-science-effective-solutions-and-technological-variants/isbn978-3-659-22604-5>

<https://www.researchgate.net/publication/321492511> APPLICATION OF CAD DESIGN OF TECHNOLOGICAL PROCESSES IN THE FIELD OF MATERIAL SCIENCE

The achieved results give grounds for future active international cooperation with scientific and production institutions for which a specific scientific product has been created and tested. The elaborated scientific product is a set of numerical methods beyond the scope of the originally intended application only for metallurgical and metalworking production. The developed methods are evaluated with the benefits of the relevant area. In the references, results have been achieved for saving materials and energy while maintaining the same level of quality indicator. This also indirectly reflects on the environmental protection. The importance of my research has been underpinned by the universality of the approach, and it has been developed for models of regression analysis and the artificial neural networks. Thus, a valuable and up-to-date methodology, thoroughly endorsed in a scientific style, has been apportioned and implemented.