DESIGN OF PREDICTIVE MODELS BY NONLINEAR AND GAUSIAN KERNEL SMOOTHING

Jozef RÉDL[1], Grigory BOYKO[2], Patrik KÓSA[1], Robert SZABO[1]

[1]*Institute of Design and Engineering Technologies, Faculty of Engineering, Slovak University of Agriculture in Nitra, jozef.redl@uniag.sk, patrik.kosa@uniag.sk, robert.szabo.rps@seznam.cz.*
[2]*Department of Technical Operation and Design of Cars, Faculty of Automobile Transport, Volgograd State Technical University, Russian Federation, grigory.v.boyko@gmail.com.*

Abstract
This paper presents the design and application of analysis of variance for predicting the dislocation of the center of gravity in tractors. Various polynomial, rational, and nonlinear regression models were tested and validated. The analysis is based on previously published Monte Carlo simulation results, utilizing a dataset derived from 53 OECD tractor test reports in which center of gravity (COG) dislocation was measured. Model performance was evaluated using covariance analysis and Pearson's correlation coefficient (R), leading to the derivation of an optimal approximation function. All mathematical computations and modeling procedures were performed using Mathcad Prime software.

Key words: *simulation; curve fitting; parameters estimation; random process.*

INTRODUCTION
In technical practice is often occurring a situation when providing the experimental measurement is too difficult or impossible. Especially the determination of inertial parameters of heavy vehicles must be estimated. Usually, the estimation is based on Monte Carlo simulation *(Rédl, Boyko & Kalantari, 2024)*. The results of simulation are often the random dataset with defined random probability. The further processing of simulation results is required to derive an approximation equation. The approximation of nonlinear logarithmic function was conducted and where the three methods were analyzed with optimal initial parameters estimations *(Chen & Couch, 2015)*. The categorized problems of algebraic regression were analyzed, and probabilistic regression problems were defined. The first problem of probabilistic regression is a bias problem. The bias problem in probabilistic regression is simultaneously determined by first moments as well as second central moments by inhomogeneous multilinear, namely bilinear, variance-covariance component estimation. The solution space is divided into those classes:

- strong nonlinear,
- weak nonlinear,
- linear.

Weak nonlinearity is defined by a nonlinear system which allows a Taylor series approximation *(Grafarend, Zwanzig & Awange, 2022)*. To design a predictive model, the design process must be supported by the data analysis process It is necessary to have a good understanding of a phenomenon, and it is possible to make predictions about it. Data analysis helps us to make this possible through exploring the past and creating predictive models. The data analysis process is composed of following steps:

- the statement of problem,
- collecting data,
- cleaning the data,
- normalizing the data,
- transforming the data,
- exploration statistics,
- exploratory visualization,
- predictive modeling,
- validating the model,
- visualizing and interpreting the results,
- deploying the solution. *(Cuesta & Kumar, 2016).*

The Kernel smoothing is the most widely used image smoothing technique in brain image analysis. To increase the signal-to-noise ratio (SNR) and smooth- ness of data required for the subsequent random field theory based statistical inference, some type of smoothing is necessary. Among many images smoothing methods, Gaussian kernel smoothing (GKS) has emerged as a de facto smoothing technique among brain imaging researchers due to its simplicity in numerical implementation. GKS also increases statistical sensitivity and statistical power as well as Gaussian. GKS can be viewed as weighted averaging of voxel values. Then from the central limit theorem, the weighted average should be more Gaussian *(Chung, 2014)*. Some of the applied methods are based on the least square method (LSM) or use it itself. The linear least-squares technique (LLS) is characterized by its mathematical simplicity, in the sense that the estimates are obtained applying matrix algebra operations in a one-shot computational procedure. The regression techniques can be applied to nonlinear models as well because the basic principle is to minimize the sum of squares of the errors between the measurements and model response. As for the cost function, it is immaterial whether the errors result from a linear or nonlinear model. Only the minimization procedure is model type dependent. Nonlinear least-squares (NLS) problems can be solved only iteratively *(Ravindra,2015*). The aim of this article is to derive the most accurate approximation function for estimating the height dimension of a tractor's center of gravity.

MATERIALS AND METHODS

The adaptive smoothing functions are described in Mathcad support web page *(MathCad – Data Analysis Functions, 2025)*. The *supsmooth* function performs super smoothing, a fast algorithm that uses an adjustable window to calculate a localized linear fit to the data. The form of function is as follows:
supsmooth(vx, vy). Function returns a vector created by the piecewise use of a symmetric nearest neighbor linear least-squares fitting on each element in *vy*, in which the number of nearest neighbors is adaptively chosen. The *supsmooth* function is most useful when your data lies along a band of relatively constant width. The *supsmooth* algorithm utilizes a local smoother that performs a localized linear fit. As is the case for median smoothing, the algorithm moves through the data, focusing on a window of values. The *x* and *y* values within the window are used to determine a local linear least-squares fit. The window length is calculated for each *x* value using cross-validation estimation. The localized window-adjustment makes *supsmooth* particularly useful in cases where data display different degrees of noise in different portions of the measurement. Some types of data are smoothed with one type of algorithm over another. You may wish to compare this method with median smoothing or Gaussian Kernel smoothing. The *loess* polynomial regression technique is also an effective smoother. The Gaussian Kernel Smoothing function has a next form:

$$f(y_i) = \frac{\sum_{j=1}^{n}\left[K\cdot\left(\frac{f(x_i)-f(x_j)}{b}\right)\cdot f(y_j)\right]}{\sum_{j=1}^{n}\left[K\cdot\left(\frac{f(x_i)-f(x_j)}{b}\right)\right]},\ \text{where } K \text{ is} \quad (1)$$

$$K(t) = \left[\frac{1}{\sqrt{2\cdot\pi}\cdot(0{,}37)}\right]\cdot e^{\left(-\frac{t^2}{0{,}2738}\right)}. \quad (2)$$

To perform curve fitting, to fit a single function, in a least-squares sense, through all data points. This method contrasts with interpolation, where piece-wise functions are fitted through adjacent data points. To further analyze the data, or determine the suitability of the chosen regression, it is necessary to apply the statistical functions for data analysis like:

1. *Linear and Median-Median Regression*
   With functions *line, slope, intercept, stderr*—Least-squares linear regression for data, and the standard error associated with linear regression, *medfit*—Median-median line regression for data.
2. *Polynomial and Rational Function Regression*
   With functions *loess*—Localized polynomial regression, *rationalfit, rationalfitnp*—Least-squares rational function regression*, polyfit, polyfitc, polyfitstat*—Multivariate polynomial regression.
3. *Nonlinear Regression*
   With functions *genfit* —Least-squares nonlinear regression for arbitrary fit functions, *expfit*—Least-squares exponential regression, *lnfit, logfit*—Least-squares logarithmic regression, *lgsfit*—Least-squares logistic curve regression, *pwrfit*—Least-squares power curve regression, *sinfit*—Least-squares sinusoidal regression, the *LeastSquaresFit* - include additional information

about the data or the parameters for any of the above fits, such as standard deviation in the data, bounds on the parameters, or constraint functions, it is allows to use function to do the calculation in a more detailed way.

4. Other Functions: *linfit*—Least-squares regression for an arbitrary linear combination of functions, *LeastSquaresFit*, *confidence*, *multidfit*—General multivariate fit.

The Regression Analysis Functions allows to use the following functions to carry out regression analysis: *polyfit*—Creates multivariate polynomial regression surfaces, *polyfitc*—Finds the coefficients of a multivariate polynomial regression, *polyfitstat*—Views the statistical analysis of a multivariate polynomial regression, *multidfit*—Finds the parameters of a multivariate fit function.

RESULTS AND DISCUSSION

The data analyzed was obtained from Monte Carlo simulation and are shown in Fig.1. The *supsmooth* smoothing function was used, and correlation factor was solved with value $R = 0{,}7093$ *(Rédl, Boyko & Kalantari, 2024)*.

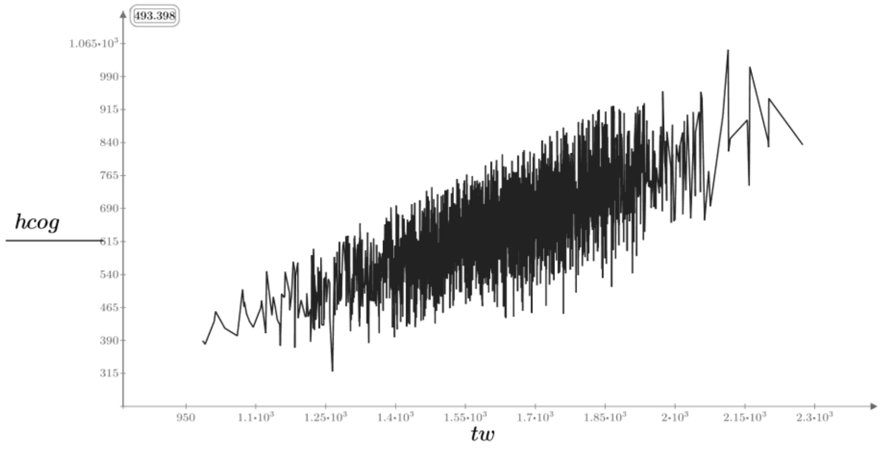

Fig. 1 Monte Carlo simulation data *(Rédl, Boyko & Kalantari, 2024)*.

The covariance was calculated by equation (3) as follows:

$$cvar = \frac{1}{m \cdot n} \cdot \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \left[\left(A_{i,j} - mean(A) \right) \cdot \left(B_{i,j} - mean(B) \right) \right], \quad (3)$$

and Pearson correlation factor was calculated by equation (4) as follows:

$$R = \frac{cvar(A,B)}{stdev(A) \cdot stdev(B)}, \text{ where} \quad (4)$$

A, B are the X, Y dataset, and $stdev$ is a standard deviation function. For Kernel smoothing we used the function $Yks := ksmooth(tw, hcog, b)$, where tw is a track width of vehicle, $hcog$ is a height of center of gravity of vehicle and b is a band width, the smoothed function is shown in Fig.2. The $R = 0{,}73989$. For the $b = 5$ was the R value as a best choice.

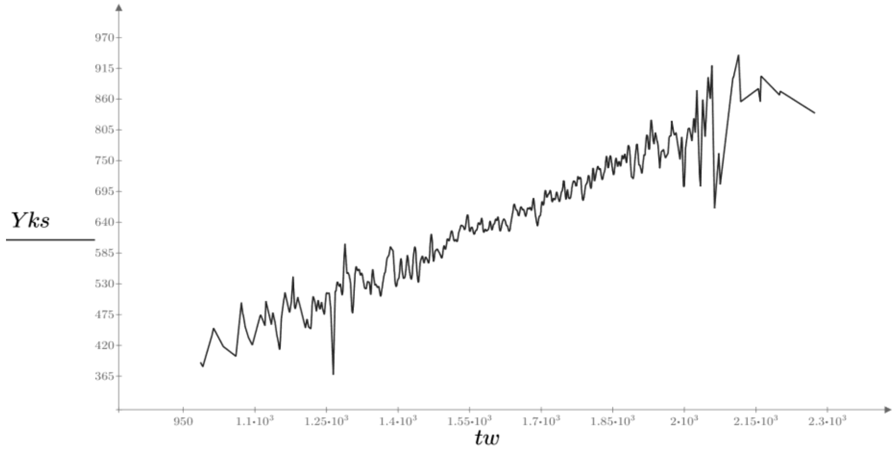

Fig. 2 Kernel smoothing of simulated dataset

In the next analysis we focused on the design of the best approximation function with the $linfit$, $expfit$ and $lnfit$ functions. Approximate functions forms are in Tab. 1. The simulation results are shown on Fig.3.

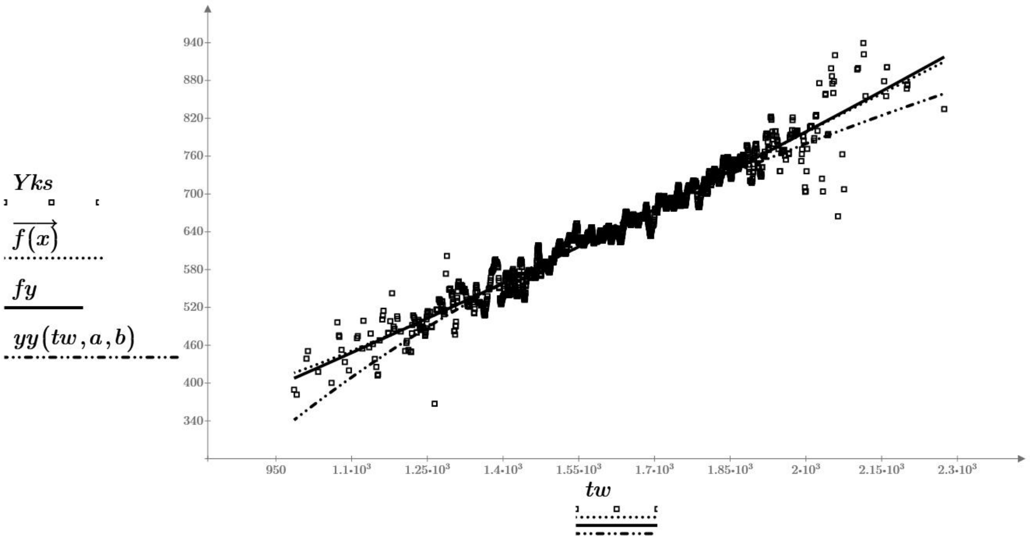

Fig. 3 Regression functions

Tab. 1 Proposed correlation functions

Mathcad function	Proposed correlation function	R
$linfit$	$f(x) = c_0 \cdot \ln(x) + c_1 \cdot \sqrt[3]{x} + c_2$	0,9718
$expfit$	$f(y) = \beta_0 \cdot e^{\beta_1 \cdot tw} + \beta_2$	0,9997
$lnfit$	$yy(tw, a, b) = a \cdot \ln(tw) + b$	0,9658

Finally, the resulting form of calculation of height of center of gravity could be finalized in the Tab.2.

Tab. 2 Resulting functions

Mathcad function	Resulting function
$linfit$	$hcog_1 = -1{,}733 \cdot 10^3 \cdot \ln(tw) + 607{,}252 \cdot \sqrt[3]{tw} + 6315$
$expfit$	$hcog_2 = 1{,}6113 \cdot 10^3 \cdot e^{\beta_1 \cdot tw} - 1{,}5205 \cdot 10^3, \beta_1 = 1{,}8211 \cdot 10^{-4}$
$lnfit$	$hcog_3 = 619{,}721 \cdot \ln(tw) - 3930{,}744$

To verify the derived results, we choose the three certain tractors COG known dislocation dimensions and tracks with and try to approximate $hcog$ with equations in Tab.2. The verification results are in Tab.3. The values of $tw - real$ and $hcog - real$ were obtained from the research work *(Rédl, Boyko & Kalantari, 2024)*.

Tab. 3 Regression functions verification

Function	$tw - real$ mm	$hcog - real$ mm	$hcog - simulated$ mm
$hcog_1$	945, 1245, 1518	691, 728, 619	401, 496, 600
$hcog_2$	945, 1245, 1518	691, 728, 619	393, 501, 604
$hcog_3$	945, 1245, 1518	691, 728, 619	315, 486, 609

CONCLUSIONS

Kernel smoothing, nonlinear regression, and least squares regression models were employed to estimate the unknown height dimension of the center of gravity (COG) of a tractor. These three methods were introduced and evaluated based on their goodness of fit using the Pearson correlation coefficient. The best-fitting function, denoted as $hcog_2$, was an exponential approximation with a correlation coefficient of $R = 0{,}997$. The other two functions had correlation coefficient values of approximately $0{,}97$ and $0{,}96$, respectively. The verification dataset included boundary and mid-range values within the interval. Despite some evident inaccuracies, the proposed method proves useful in scenarios where only the track width of the vehicle is available and experimental measurements are not feasible.

ACKNOWLEDGMENT

This research work has been funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V05-00018 and it has been also supported by the Scientific Grant Agency of the Ministry of Education, Research, Development and Youth of the Slovak Republic under grant number VEGA 1/0691/23.

REFERENCES

1. Cuesta, H., Kumar, S. (2016). Practical Data Analysis (2nd Edition) - A Practical Guide to Obtaining, Transforming, Exploring, and Analysing Data Using Python, MongoDB, and Apache Spark. *Packt Publishing*. Retrieved from https://app.knovel.com/hotlink/toc/id:kpPDAPMDBV/practical-data-analysis/practical-data-analysis.
2. Grafarend,E.W., Zwanzig,S. & Awange, J.L. (2022). Applications of Linear and Nonlinear Models: Fixed Effects, Random Effects, and Total Least Squares. *Springer Nature Switzerland AG*. https://doi.org/10.1007/978-3-030-94598-5.
3. Chen,T., Couch,K.A. (2015). An approximation of logarithmic functions in the regression setting. *Statistical Methodology*, 23, 50-58.
4. Chung, M.K. (2014). Statistical and Computational Methods in Brain Image Analysis. *Taylor & Francis Group*.
5. Jakaitiene, A. (2025). Nonlinear Regression Models. Encyclopedia of Bioinformatics and Computational Biology (Second Edition), *Elsevier*, 309-316.
6. MathCad – Data Analysis Functions.(2025). https://support.ptc.com/help/mathcad/r9.0/en/index.html#page/PTC_Mathcad_Help/about_data_analysis_functions.html#.
7. Montgomery, D. C. (2013). Design and Analysis of Experiments (8th Edition). *John Wiley & Sons*. Retrieved from https://app.knovel.com/hotlink/toc/id:kpDAEE0013/design-analysis-experiments/design-analysis-experiments.
8. Ravindra V. J. (2015). Flight Vehicle System Identification - A Time-Domain Methodology (2nd Edition) - Progress in Astronautics and Aeronautics, 245. *American Institute of Aeronautics and Astronautics/Aerospace Press (AIAA)*. Retrieved from https://app.knovel.com/hotlink/toc/id:kpFVSIAT04/flight-vehicle-system/flight-vehicle-system.
9. Rédl, J., Boyko, G. & Kalantari, D. (2024). Prediction of Machine's Centre of Gravity Dislocation with Monte Carlo Simulation. *Acta Technologica Agriculturae*, 27(3), 133-141.
10. Schittkowski, K. (2002). Numerical Data Fitting in Dynamical Systems: A Practical Introduction with Applications and Software. *Springer-Science+Business media, b.v.*

Corresponding author:

Ing. Patrik Kósa, PhD., Institute of Design and Engineering Technologies, Faculty of Engineering, Slovak University of Agriculture in Nitra, Tr.A. Hlinku 2 Nitra, Nitra 949 76, Slovak Republic, phone: +421 37 641 4104, e-mail: patrik.kosa@uniag.sk