

A Generalized Profile Function Model Based on Artificial Intelligence

PERO RADONJA¹, SRDJAN STANKOVIC², and DRAGANA DRAZIC¹

Institute of Forestry, 11030 Kneza Visislava 3, Belgrade

²Faculty of Electrical Engineering, University of Belgrade, Belgrade
SERBIA

radonjap@eunet.yu stankovic@etf.bg.ac.yu drazicd@yubc.net

Abstract - A generalized profile function model (*GPFM*) provides an approximation of individual profile functions of the objects (trees) in a region. It is shown in this paper that this generalized model can be successfully derived using artificial computational intelligence, that is, neural networks. The generalized model (*GPFM*) is obtained as a mean value of all the available normalized individual profile functions. Generation of *GPFM* is performed by using the basic dataset, and verification is done by using a validation data set. As an example of the application of the proposed *GSPM* in volume computing, 42 objects from the same region are considered. Statistical properties of the original, measured data and estimated data based on the generalized model are presented and compared. Testing of the obtained *GPFM* is performed also by regression analysis. The obtained correlation coefficients between the real data and the estimated data are very high, 0.9946 for the basic data set and 0.9933 for the validation dataset.

Key Words: - Profile function, Generalized model, Neural networks, Histogram, Scatterplot

1 Introduction

In this paper a derivation of the generalized profile function model (*GPFM*) based on neural networks is described. It is known that neural networks (NN) can be used very successfully in the process of modeling of many biological processes [1, 2, 3, 4]. The generalized model provides an approximation of the individual profile function of any object in region [5, 6, 7]. A region can contain many thousands of individual objects (spruce trees) with their own individual profile functions. Measurements on all objects is practically impossible and because of that we shall try to find a generalized model that enables getting any individual profile functions without detailed measurements on the objects themselves. In other words, we are trying to get a generalized model, in order to be able to get an approximation of any individual profile function, only by using the basic measured values D and H that is, the sets of values, data pairs, $(1.3, D)$ and $(H, 0)$, where D is the diameter of a tree at breast height, 1.3m, H is the total height of a tree, and, generally, the set of values or data pairs (h, r) which denote the tree radius r at the height h .

*Research is supported by the Ministry of Science and Environmental Protection of Serbia, as a part of the project EE - 273015.B.

Using the individual profile functions we can get very accurate volumes of all the objects [7]. On the other hand, using the obtained approximations of the individual profile functions we can compute only approximate values of these volumes. Obviously, it is important that the errors in volume computing are unbiased. In this case the sum of the volumes of all objects from a region will be computed with minimal error. The overall volume of all objects from whole region is important because of sustainable management of the considered ecological system [8, 9, 10, 11].

2 Input Data Specification

The total data set contains 260 data which are divided into two datasets [12]. The basic data set will be used for obtaining the generalized profile function model and the validation dataset will be used for verification of the obtained generalized model. In our case we have used 13 data sets (h, r) for every object. In fact, in the considered case 14 objects will be used to get the generalized profile function model and verification will be performed on 6 new objects.

The basic statistics of the measured data, that is, of the input data set, are summarized in Table 1.

Table 1. Summary statistics of the input datasets

	Numb. of data	Aver.	Med.	SD	Min.	Max.
<i>Total Dataset</i>	260	8.1008	7.2000	5.9418	0.0	29.50
<i>Basic Dataset</i>	182	7.4975	6.3250	5.9019	0.0	29.50
<i>Validation Dataset</i>	78	9.5083	10.4250	5.8315	0.0	23.75

Histograms of the measured values of the total, basic and validation dataset are shown in Figs.1, 2 and 3.

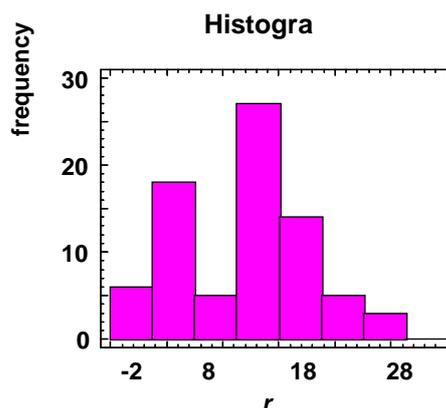


Fig.3. Histogram of the measured values of the validation dataset

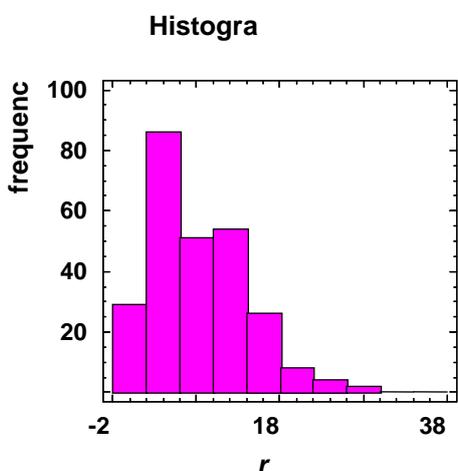


Fig.1. Histogram of the measured values of the total dataset

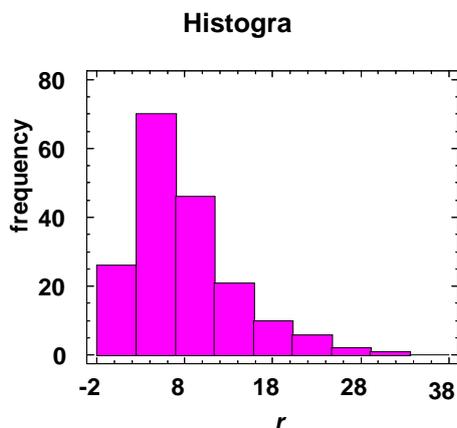


Fig.2. Histogram of the measured values of the basic dataset

The scatterplots and time sequence plots of the measured values of the total, base and validation datasets are presented in the figures in Appendices A, B and C, respectively.

3 Specific Profile Function Models Based on Neural Networks

Artificial neural networks, in short, neural networks (NN), represent a very efficient and powerful tool of computational artificial intelligence. In modeling many different biological process NN ensure smaller modeling error than classical methods based on polynomials or exponential approximation functions [1, 2, 3].

A typical structure of a NN is the three-layer feed forward NN. The commonly used activation functions of neural networks include linear functions for output neurons, logistic sigmoid functions for hidden neurons, and identity functions for input neurons.

It is known that the best model for the observed data in the sense of generalization is obtained in the case of a minimal number of neural network layers and individual neurons (for a general discussion, see [13, 14]). In our case, network complexity is determined simply by the number of free model parameters, that is by the number of *tansig* neurons in the second (hidden) layer. The *tansig* neurons have logistic sigmoid tangent hyperbolic transfer function. The method of *structural risk minimization* (SRM) provides a general systematic procedure for achieving the best performances by controlling the so-called VC dimension [13, 14]. In our case, we found the best network stucture by decreasing, step by step, the number of *tansig* neurons in the hidden layer [15].

In this study, a three layered feed forward NN with the back propagation algorithm and with 2 *tansing* neurons in the hidden layer was used.

As an example, we shall consider two very different individual profile functions. The process of NN learning (or training) for the first profile function is shown in Fig.4.

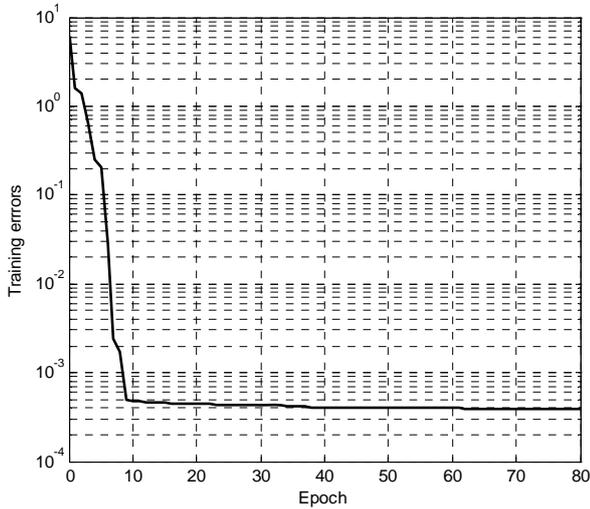


Fig.4. The process of learning for the first profile function

Because of good convergence properties, the Levenberg-Marquardt's algorithm is used in the process of learning [15]. Certainly, it is possible to use the EEM algorithm with tuning of the smoothing parameters [16].

In Fig.5, the first obtained (calculated) individual profile function is presented.

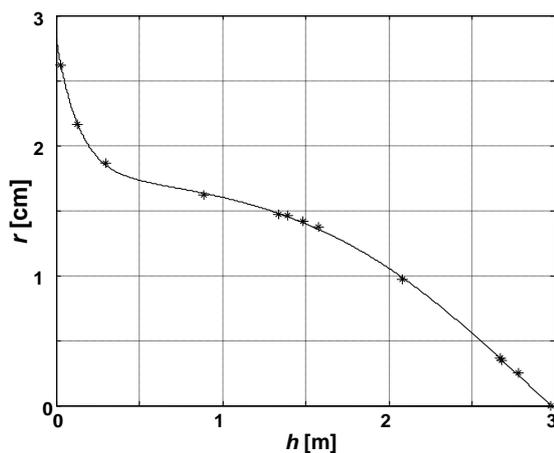


Fig.5. The first obtained individual profile function

Errors of fitting or modeling for the first profile function are shown in Fig.6.

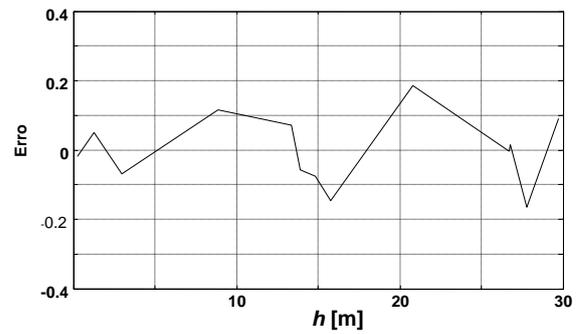


Fig.6. Errors of modeling for the first profile function

The process of learning for the second profile function is shown in Fig.7.

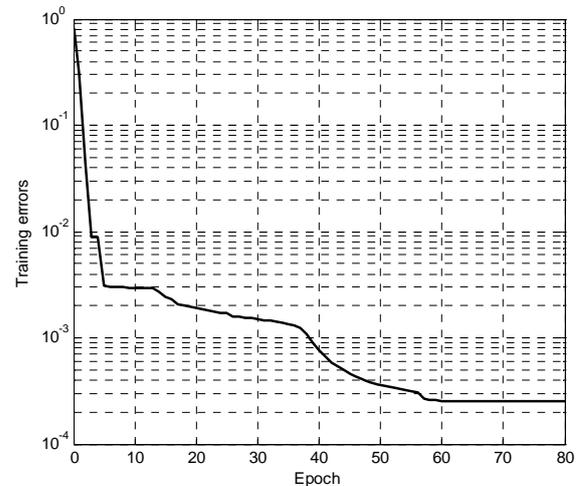


Fig.7. The process of learning for the second profile function

The second obtained individual profile function is presented in Fig. 8.

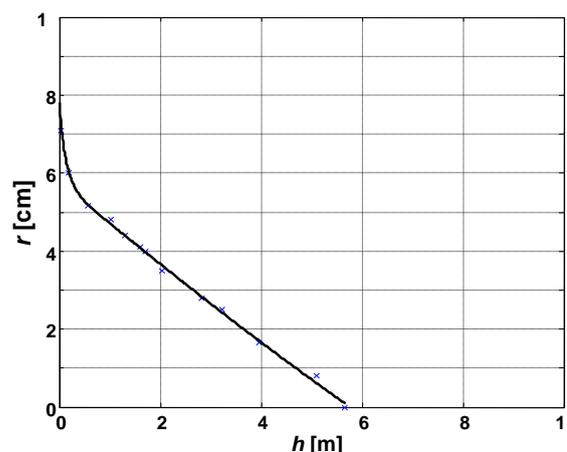


Fig.8. The second individual profile function

Errors of modeling for the second profile function are shown in Fig.9.

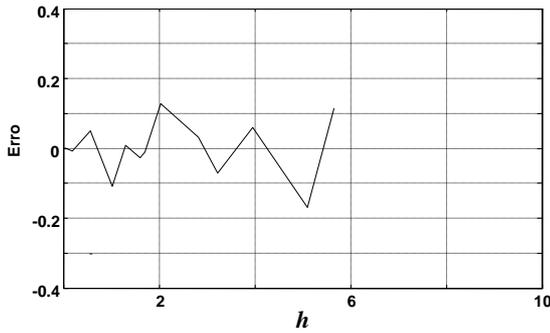


Fig.9. Errors of modeling for the second profile function

4 Generalized Profile Function Model

In order to get the generalized model, it is necessary, in the first step, to calculate the normalized individual profile functions of all available objects. Normalization is performed by using the largest values on the x and y axes. In fact, the normalization of x axis is performed by using H , and of y axis by using $r(0)$. Two normalized profile functions are presented in Figs. 10 and 11.

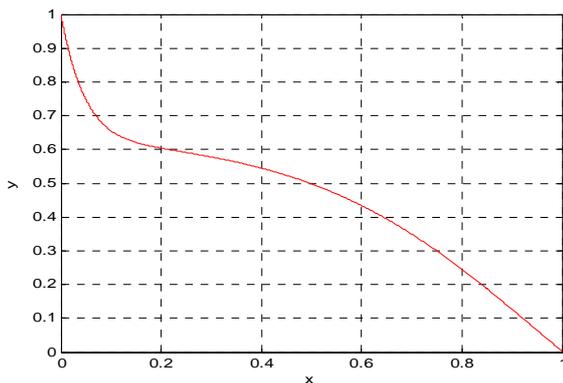


Fig. 10 The first normalized profile function

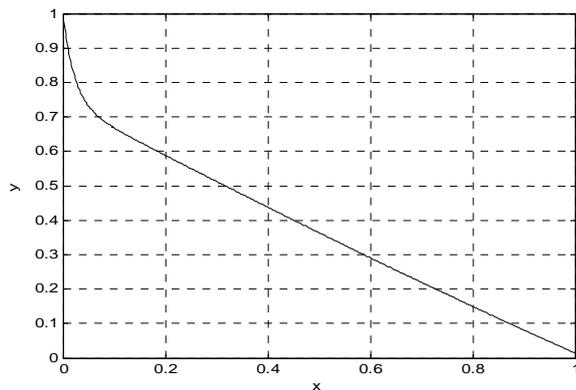


Fig. 11 The second normalized profile function

The generalized model (*GPFM*) is obtained as a mean value of all the available normalized individual profile functions. The obtained generalized model is presented in Fig.12.

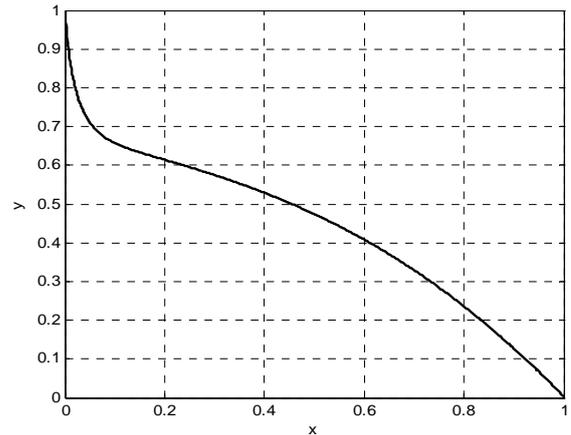


Fig.12. The generalized model

Now the generalized model presented in Fig.12 can be used for generating approximations of all the individual profile function. Renormalization is performed so that each individual (renormalized) profile function passes through the characteristic points (x_0, y_0) , (x_0 breast height, 1.3m, $y_0 = D/2$ radius at breast height) and the final point $(H, 0)$. In accordance with this, renormalization per x axis is performed using H , and per y axis using $y(0)$ where:

$$y(0) = [(D/2) / [y (1.3/H)]] \dots\dots\dots(1)$$

The value $y(1.3 / H)$ is obtained using the generalized model, Fig. 12.

Quality and performance of the obtained *GPFM* will be analyzed in the subsequent sections.

5 Basic Statistics of the Estimated Data

In applications of artificial intelligence (neural networks) and verifications of obtained results, it is typical to use two datasets, the basic (model) and the validation dataset. The validation dataset can have the same volume as the basic dataset, but more frequently it is smaller. In our case, the validation datasets was about 40% of the basic dataset.

Testing the model accuracy using the validation dataset we get the **real** model accuracy. Evidently, it is better to use more validation datasets. If we validate the model with the same data used during model derivation, we get the **initial** model accuracy. Typically, the initial model accuracy is higher than the real one.

Summary statistics of the estimated data for the validation and basic datasets are given in Table 2.

Table 2. Summary statistics of the estimated data of both the validation and basic datasets

r [cm]	Num. of data	Aver.	Med.	SD	Min.	Max.
Validation Dataset	78	9.2297	10.0040	5.6736	0.0671	22.6973
Basic Dataset	182	7.4744	6.3828	5.8483	0.0204	29.6145

As we expected, a better agreement between the measured and estimated data, from the point of view of the average value, exists for the basic dataset, 7.4975 and 7.4744, than for the validation dataset, 9.5083 and 9.2297, Tables 1 and 2.

The histogram of the estimated values of the validation dataset is shown in Fig. 13.

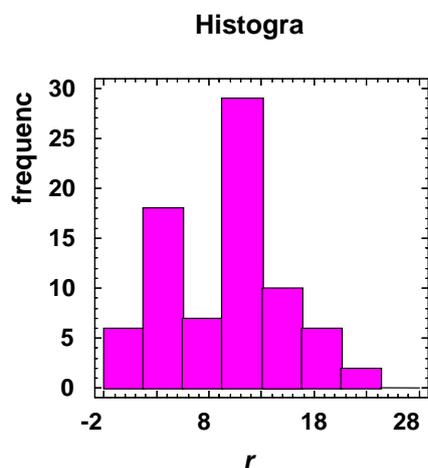


Fig. 13. Histogram of the estimated values of the validation dataset

Now we shall compare the histograms of the measured and estimated values of the validation dataset, Figs. 3 and 13. Based on Fig.3, we conclude that there are more data in the first range of the histogram than in the third. The conclusion is oposite in the case of Fig.13. However, the difference is very small, and the general conclusion is that a good agreement between the histograms for the measured and the estimated values is achieved.

The histograms of the measured and the estimated values of the basic dataset are given in Figs. 2 and 14. We can see that the histograms presented in Figs. 2 and 14 are very similar. Based on the presented figures, we can conclude that a better agreement between the histograms is achieved for the basic, than for the validation dataset.

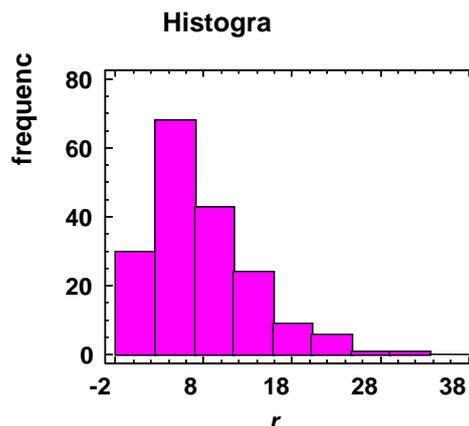


Fig. 14. Histogram of the estimated values of the basic dataset

6 Testing of the Obtained Generalized Models by Regression Analysis

Usability of the obtained generalized model will be tested by comparing the measured data and the corresponding estimated data. Testing of the accuracy of the generalized model is practically performed by applying the regression analysis technique. Undoubtedly, it is possible to use the forward regression algorithm based on M-estimators [17]. A comparison of the measured and the corresponding estimated data is performed for both available datasets. The result of testing for the validation dataset is presented on Fig.15.

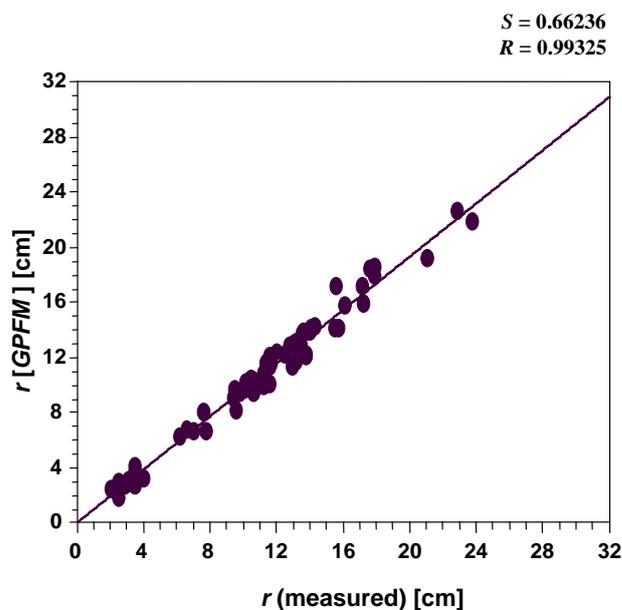


Fig. 15. Statistical performance in radius estimation (validation data set)

The first pair of results of data comparison represents statistical performance in data estimation. This pair contains the regression parameters: standard error of the estimate S and the correlation coefficient R . The second pair of the results contains the slope of the regression line and the intercept on the y axis. Evidently, the slope of the regression line obtained by comparing the estimated and the measured data must have in the ideal case the angle of 45° and has to start from the origin. In other words, parameter b has to have its value near 1.0 and parameter a , that is the intercept on the y axis, has to have its value near 0.

Residuals which determine the standard estimation error for the validation data set are presented in Fig. 16.

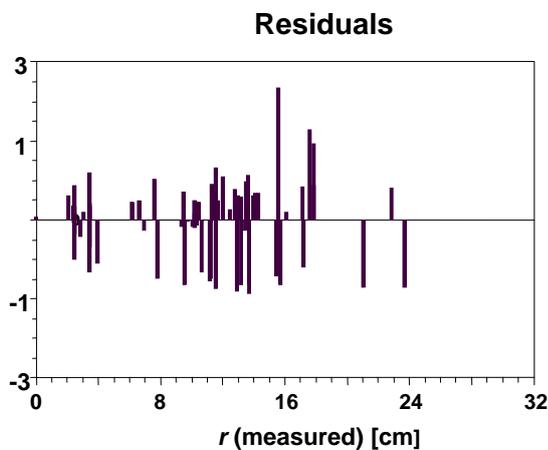


Fig. 16. Residuals (validation data set)

Data comparison between the measured and the estimated data for basic data set is done, as well. The result of testing is presented in Fig. 17.

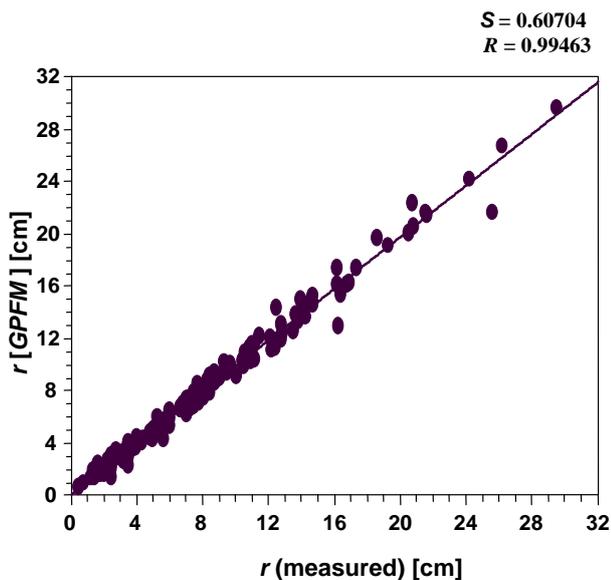


Fig. 17. Statistical performances in radius estimation (basic data set)

Residuals which determine the standard estimation error in the case of the basic data set are presented in Fig. 18.

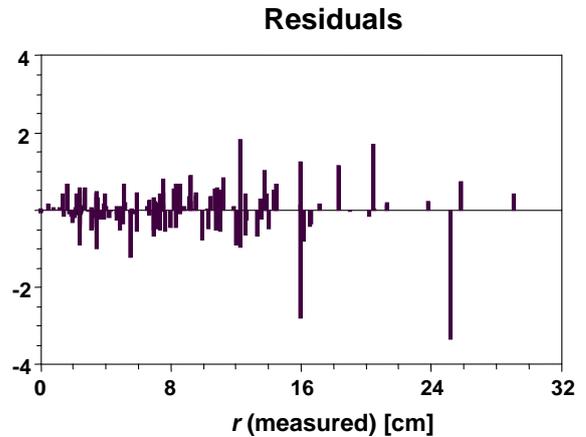


Fig. 18. Residuals (basic data set)

Values of the statistics of data comparison by regression analysis are presented in Table 3.

Table 3. Values of the statistics of data comparison by regression analysis

	Statistics of the r comparison				
	Interc. a	Slope b	Std.Err. S	Corr. coeff. R	R^2 [%]
Validation Dataset	0.0412	0.9664	0.6624	0.9933	98.65
Basic Dataset	0.0849	0.9856	0.6070	0.9946	98.93

As we expected, the parameter b , the slope of the regression line, is closer to 1.0 in the case of data from the basic data set, 0.9856, than in the case of the validation dataset, 0.9664. Also, the obtained standard error 0.6070 is lower in comparison with 0.6624, and the correlation coefficient is higher (0.9946 compared to 0.9933). However, the intercept on y axis is lower for the validation dataset, 0.0412, in comparison with 0.0849 for the basic dataset. Using the same dataset for both modeling and validation, Korol and Gadov [6] have obtained a lower standard error, 0.526, in comparison with 0.607 in the our case, but a worse correlation coefficient, 0.9820, in comparison with 0.9946 in our case.

7 Applications: Volumes computing

As an example of applications of the obtained *GSPM* in volume computing, consider 42 objects from the same region. The basic statistical properties of the input dataset are presented in Table 4.

Table 4. A review of the basic statistical properties of the input dataset

	<i>D</i> [cm] <i>H</i> [m]	Aver.	Med.	SD	Min	Max.
Total Data set	<i>D</i>	27.57	28.25	10.58	6.00	48.40
	<i>H</i>	24.28	24.77	7.86	5.65	36.15

The all available data, *D* and *H* are presented in Fig.19.

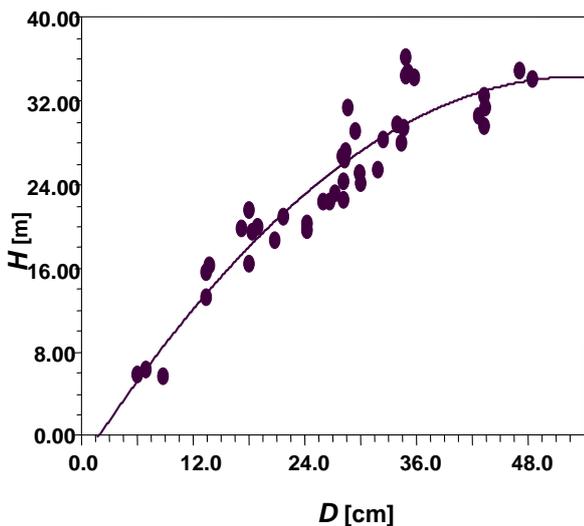


Fig. 19. The all available data, *D* and *H*

When NNs are used for obtaining the individual profile function, stem volumes can simply be calculated by using the *inner vector product* instead of numerical integration. With the assumption that all the values of *r(h)* define the *row vector R*, the volume can simply be calculated by the equation:

$$V(NN) = \pi \mathbf{R} \mathbf{R}^T \Delta h \dots\dots\dots(2)$$

where Δh is the height increment 5.65 - 36.15 mm. In the considered case, the row vector **R** contains 1000 values.

The basic statistical performance of the calculated volumes *V(NN)* based on Equation (2) are presented in Table 5.

Table 5. The basic values of the statistics of the calculated volumes

	<i>V</i> [m ³]	Aver.	Med.	SD	Min.	Max.
Total Data set	NN	0.9473	0.7953	0.7755	0.0100	2.8905

The histogram of the calculated volumes is presented in Fig. 20

Histogram

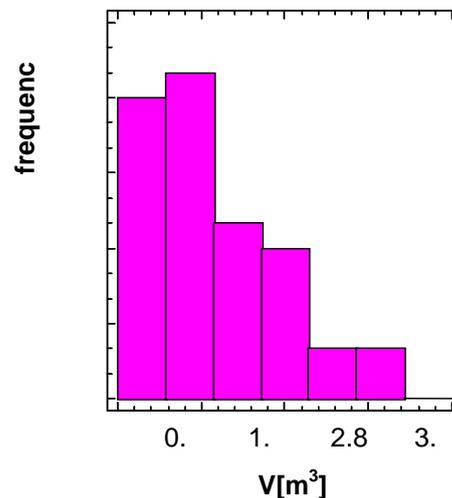


Fig. 20. The histogram of the calculated (referent) volumes

Testing of the accuracy of the obtained *GSPM* is performed by comparing the obtained estimated volumes with the referent volumes. **The estimated volumes** represent the volumes based on *GSPM*, *V(GSPM)*, and the **referent volumes** are in fact *V(NN)*.

In Table 6. values of the statistics of the estimated volumes based on *GSPM*, are given.

Table 6. Values of the statistics of the estimated volumes

	V [m ³]	Aver.	Med.	SD	Min.	Max.
Total Data set	GSPM	0.8938	0.7764	0.7124	0.0105	2.7375

The histogram of the estimated volumes is presented on Fig. 21.

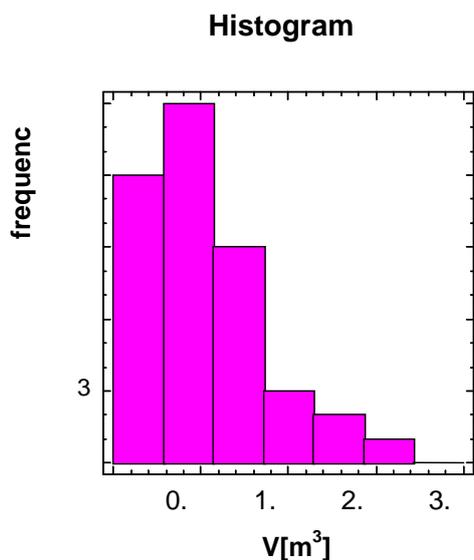


Fig. 21. The histogram of the estimated volumes

Based on Figs. 20 and 21 we can conclude that most of the data are in the second range of both histograms. Also, we can conclude, based on both histograms that more data are in the first than in the third range. Evidently, the largest difference between the histogram exists for the fourth range.

8 Testing of the Obtained Volumes by Regression Analysis

Testing of the efficiency of volume estimation can be performed by comparing the obtained estimated volumes with the referent volumes by the regression analysis. As in the case of radii comparison, in the ideal case it is necessary that four important regression parameters have already mentioned values.

Comparison of the estimated and referent volumes done by regression analyses for the total data set is illustrated by Fig. 22.

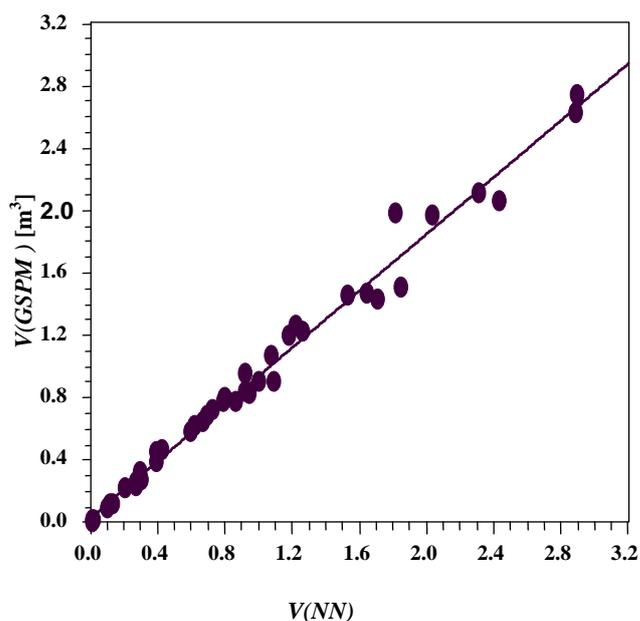


Fig. 22. Testing of volume estimation V(GSPM) versus V(NN)

We can see in Fig. 22 that the volumes based on GSPM do not deviate much (except four points) from the referent volumes, $S_{VE} = 0.0853$ and $R_{VE} = 0.9930$, Table 7. It can be seen that the regression line starts approximately from the origin, as the translation along y axis is only 0.0326.

Table 7. Values of the statistics of volume comparison by regression analyses

	GSPM
S_{VE}	0.0853
R_{VE}	0.9930
a	0.0326
b	0.9100

9 Conclusion

The estimated data based on the proposed GPFM are compared with real, measured data, using the basic statistics and the obtained histograms, as well as by the regression analyses. The obtained histograms for the measured and the corresponding estimated data are very similar. This conclusion is valid for the validation and the basic dataset. Also, correlation coefficients between the measured and the estimated data are very high, over 0.99, and the standard error of estimation very low, less than 0.7, for both datasets. It can be seen that the obtained

generalized profile function model can be very successfully used in the process of estimating the individual profile functions.

References:

- [1] Radonja, Pero, and Stankovic Srdjan, "Modeling of a highly nonlinear stochastic process by neural Networks", **In: Recent Advances in Computers, Computing and Communications** Editors: N. Mastorakis and V. Mladenov, Published by WSEAS Press., 2002, pp.259-264.
- [2] Radonja Pero, Srdjan Stankovic i Zoran Popovic: "Specific Process Models Obtained From General Ones", *WSEAS TRANSACTIONS on SYSTEMS*, Issue 7, vol. 3, Pub. by WSEAS Press, Athens, 2004, pp. 2490-2495.
- [3] Zhang, Q.-B., Hebda, R.J., Zhang, Q.-J. and Alfaro, R.I. "Modelling Tree-Ring Growth Responses to Climatic Variables Using Artificial Neural Networks", *Forest Science*, vol. 46, no. 2, Society of American Foresters, Grosvenor Lane, Bethesda, 2000, pp. 229-239.
- [4] Hanewinkel, M.: "Neural networks for assessing the risk of windthrow on the forest division level: a case study in southwest Germany", *Eur. J. Forest Res.* 124, 2005, pp. 243-249.
- [5] Hui, G.Y., von Gadow, K., "Entwicklung und Erprobung eines Einheitsschaftmodells fuer die Baumart *Cunninghamia lanceolata*". (in German) *Forstw Cbl* 116, 1997, pp. 315-321.
- [6] Korol M., von Gadow, K. "Ein Einheitsschaftmodell fuer die Baumart Fichte" (in German) *Forstw Cbl* 122, 2003, pp. 175-182.
- [7] Radonja, P., S. Stankovic, D. Drazic and B. Matovic.: "Development and Construction of Generalized Process Models by using Neural Networks and Multiple Linear Regression." *WSEAS TRANSACTIONS on SYSTEMS*, Issue 2, vol. 6, Pub. by WSEAS Press, Athens, 2007, pp. 287-293.
- [8] Lek, S., A. Belaud, P. Baran, I. Dimopoulos and M. Delacoste, "Application of neural networks to modelling nonlinear relationships in ecology", *Ecol. Model.*, vol. 90, 1996, pp.39-52.
- [9] Roadknight, C. M., G.R. Balls, G. E. Mills, and D. Palmer-Brown, "Modeling complex environmental data", *IEEE Trans. Neural Net.*, vol.8, 1997, pp. 852-862.
- [10] Vladimir Olej "Design of the Models of Neural Networks and the Takagi-Sugeno Fuzzy Inference System for Prediction of the Gross Domestic Product Development", *WSEAS TRANSACTIONS on SYSTEMS*, Issue 4, vol. 4, Pub. by WSEAS Press, Athens, 2005, pp. 314-319 .
- [11] Mehmet Kürümlüoğlu and Rita Nøstdal, "Management and Organizational Topics in Virtual Enterprises – Business Drivers, Characteristics and Issues in Collaborative Environments", *WSEAS TRANSACTIONS on SYSTEMS*, Issue 6, vol. 3, Pub. by WSEAS Press, Athens, 2004, pp. 2448-2453.
- [12] Maunaga Z. "Productivity and structural characteristics of same-age stands of spruce in Republic Serbian", (in Serbian). Ph. thesis, Faculty of forestry, Belgrade, 1995.
- [13] Haykin, S. *Neural Networks: A Comprehensive Foundation*, Macmillan College Publishing Company, New York, 1994.
- [14] Vapnik, Vladimir N., "An Overview of Statistical Learning Theory", *IEEE Transaction on Neural Networks*, vol.10, no.5, Sept. 1999, pp. 988-999.
- [15] Beale, Mark, "Neural Network Toolbox" NN Toolbox, Version 6.0.0.88, Release12, September 22, 2000, MATLAB 6 R12.
- [16] Jorge Santos, Joaquim Marques de Sá and Luís Alexandre: "Neural Networks Trained with the EEM Algorithm: Tuning the Smoothing Parameter", *WSEAS TRANSACTIONS on SYSTEMS*, Issue 4, vol. 4, Pub. by WSEAS Press, Athens, 2005, pp. 295-300 .
- [17] Xia Hong, Sheng Chen: "A Forward Regression Algorithm based on M-estimators", *WSEAS TRANSACTIONS on SYSTEMS*, Issue 1, vol. 5, Pub. by WSEAS Press, Athens, 2006, pp. 92-97 .

Appendix A

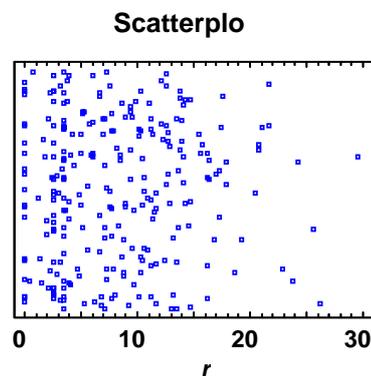


Fig.A1. Scatterplot of the measured values of the total dataset

Appendix C

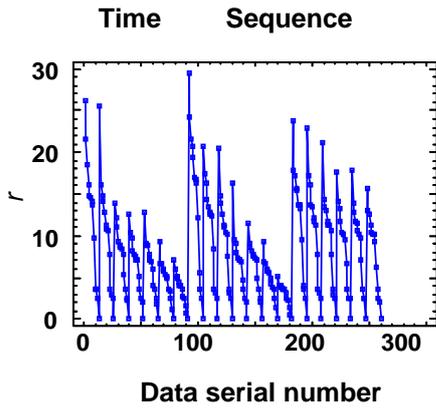


Fig.A2 Time sequence plot of the measured values of the total dataset

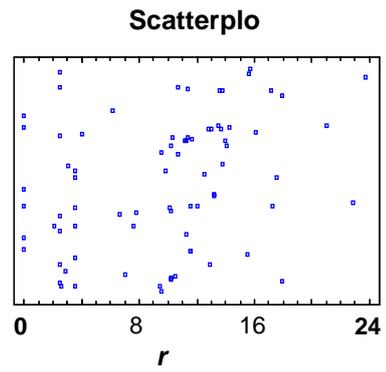


Fig. C1 Scatterplot of the measured values of the validation dataset

Appendix B

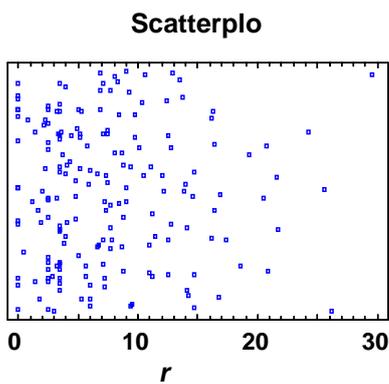


Fig. B1 Scatterplot of the measured values of the basic dataset

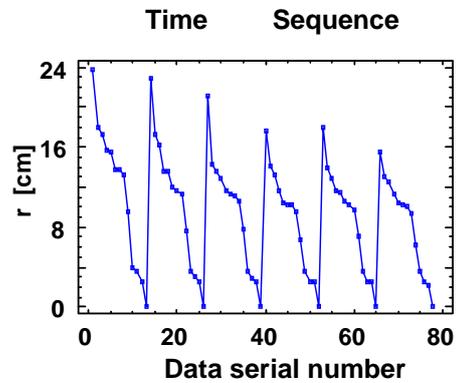


Fig. C2 Time Sequence Plot of the measured values of the validation dataset

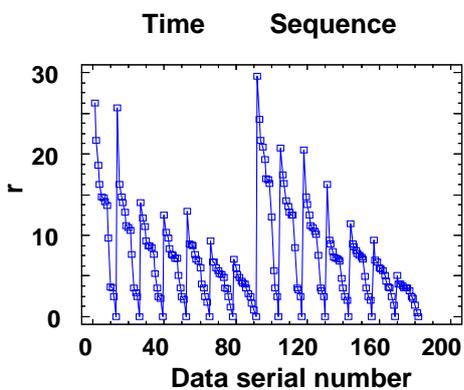


Fig. B2 Time Sequence Plot of the measured values of the basic dataset