



Modeling residual biomass from mechanized wood harvesting with data measured by forest harvester

CARLA K. RODRIGUES¹, EDUARDO S. LOPES¹, AFONSO FIGUEIREDO FILHO¹,
ALLAN L. PELISSARI² and MATHEUS K.C. SILVA¹

¹Programa de Pós-Graduação em Ciências Florestais, Universidade Estadual do Centro-Oeste/UNICENTRO, PR 153,
Km 7, 84500-000 Irati, PR, Brazil

²Departamento de Ciências Florestais, Universidade Federal do Paraná/UFPR, Av. Prof. Lothário Meissner, 632,
80210-170 Curitiba, PR, Brazil

Manuscript received on February 21, 2019; accepted for publication on May 13, 2019

How to cite: RODRIGUES CK, LOPES ES, FIGUEIREDO FILHO A, PELISSARI AL AND SILVA MKC. 2019. Modeling residual biomass from mechanized wood harvesting with data measured by forest harvester. *An Acad Bras Cienc* 91: e20190194. DOI 10.1590/0001-3765201920190194.

Abstract: The lack of accurate models for estimating residual biomass in wood harvesting operations results in underutilization of this co-product by forestry companies. Due to the lack of this information, forestry operations planning, such as chipping and transport logistics, are influenced, with a consequent increase in costs. Thereby, the aim of this study was to propose and evaluate statistical models to estimate residual biomass of *Eucalyptus* sp. in wood harvesting operations by means of tree variables measured from harvester processing head. Generalized linear models were composed through stepwise procedure for estimating residual biomass by tree covariates: diameter at breast height, commercial height, commercial limit diameter, and stem commercial volume, considering also their transformations and combinations. Residual biomass distributions with positive skew support the application of generalized linear model and Gamma distribution in random component, since normality assumption in traditional linear regression was a requirement not satisfied in this study. By stepwise procedure, tree variables associated to forest biomass were selected, whose linear combinations resulted in models with high statistical efficiency and accuracy. Thus, models developed in this study are innovative tools to estimate residual biomass in mechanized wood harvesting, in which can be inserted into harvester's hardware to provide real-time information.

Key words: bioenergy, full-tree harvesting method, forest measurement, forest planning.

INTRODUCTION

Planted forest stands represent relevant importance for economy, with 6.2% participation of the Brazilian Gross Domestic Product (GDP) and contributing to the environmental preservation and 3.7 million jobs (IBÁ 2017). Forestry sector is responsible for the supply of various wood products, such as sawlog, veneer

Correspondence to: Carla Krulikowski Rodrigues
E-mail: carlacr@gmail.com
ORCID: <https://orcid.org/0000-0001-5551-1914>

log, pulpwood, and firewood. In addition, forest biomass for energy purposes has been highlighted in recent years due to the need to reduce carbon dioxide emissions and other gaseous pollutants (Lauri et al. 2014).

Forest biomass can be obtained in the wood harvesting operations of stands with multiple uses, called residual biomass (Yemshanov et al. 2014). In the full-tree harvesting method, residual biomass can be composed by tops, branches, leaves, and barks of harvested trees arranged at the edge of the fields for cutting (Esteban et al. 2011) and posterior using as electricity cogeneration by the thermoelectric plants in forest industries (IBÁ 2017).

The lack of accurate models for estimating residual biomass produced by harvesting operations results in underutilization of this co-product by forest companies. Due to the lack of this information, forest planning and industrial operations are influenced, as chipping and transport logistics, with a consequent increase in production costs.

According to Palander et al. (2009) and Vesa and Palander (2010), it is possible to estimate forest biomass in wood harvesting operations by means of data measured through harvester processing head and stored in the on-board computer. Thus, these data have potential to be inserted into machine's hardware to produce real-time information on the residual biomass produced by forest harvesting.

In the wood harvesting by harvester machine, lengths, diameters, and volumes of the logs are recorded (Nieuwenhuis and Dooley 2006, Mederski et al. 2018). With these data, reports of harvested trees containing individual-level variables can be used as covariates for composing regression models. This information can be useful for estimating forest biomass in real-time by harvester's on-board computer.

Thereby, the aim of this study was to propose and evaluate statistical models to estimate residual biomass of *Eucalyptus* in wood harvesting operations by means of tree variables measured from the harvester processing head. The proposed regression models aiming to assist the planning of chipping and logistics in *Eucalyptus* stands, considering the hypothesis that the variables obtained from the processing head enable accurate estimates of residual biomass.

MATERIALS AND METHODS

STUDY AREA

This study was carried out in wood harvesting *Eucalyptus* stands located at Paraná State, Brazil, between the coordinates 24° 26' S and 50° 45' W. The predominant region's climate was subtropical Cfa and Cfb, according to the Köppen classification, with annual average temperature between 18 °C and 20 °C and average annual rainfall of 1,400 to 1,600 mm, in relief with average slope of 8% at 940 m.a.s.l. (Alvares et al. 2013).



The forest stands were composed of clones of *Eucalyptus saligna* Smith and *Eucalyptus grandis* W. Hill ex Maiden × *Eucalyptus urophylla* S. T. Blake interspecific hybrids, both with 7 years-old and initial density of 1,111 trees per hectare at a spacing of 3.75 m × 2.40 m. The stands characteristics are described in Table I.

The full-tree harvesting method used by the company was composed of feller buncher, skidder and harvester. In this method, feller buncher performed the felling and stacking trees in bundles inside the field, in which the skidder dragged the bundles to the edge of the field. The harvester performed the final wood process with a processing head (Table II), cutting logs in 7.2 m for pulp with commercial limit diameters

TABLE I
Mean characteristics of *E. saligna* and *E. grandis* × *E. urophylla* stands.

Variables	<i>E. saligna</i>	<i>E. grandis</i> × <i>E. urophylla</i>
Age (years)	7	7
Diameter at breast height 1.3 m (cm)	18.6	19.0
Total height (m)	31.3	32.0
Dominant height (m)	34.0	34.2

TABLE II
Harvester machine and processing head specifications for wood harvesting in *E. saligna* and *E. grandis* × *E. urophylla* stands.

Equipment	Brand	Model	Technical specifications
	Doosan	DX300	Power rating: 147 kW or 202 hp Operating weight: 29,300 kg Maximum crane reach at ground level: 10 m Continuous track system
	Waratah	622B	Operating weight: 2,280 kg Maximum hydraulic pressure: 32 MPa Maximum delimiting diameter: 650 mm Maximum roller opening: 730 mm Maximum cutting diameter: 750 mm

equal to 8, 10, 12, and 14 cm, aiming for attending to different energy parameters of residual biomass of the tree canopy.

SAMPLING AND DATA COLLECT

Before the wood harvesting operations, a forest inventory was carried out to characterize the diameter distributions of *Eucalyptus* stands. Thus, thirty trees were randomly selected by stand and proportional to five diameter classes determined by Sturges’ rule (Table III), resulting in a sample of sixty trees to measure individual tree variables.

These trees were scaling in logs with a length of 7.2 m, from a height of 0.1 m to the commercial limit diameters (d_i) of 8, 10, 12, and 14 cm, aiming for simulating harvester processing head, in which commercial volumes (v) were determined by Smalian’s formula. Additionally, diameter at breast height at 1.3 m above ground (d) and stem commercial height (h) were measured to determine the tree variables at individual-level that can be recorded by the harvester’s on-board computer.

Subsequently, biomass compartments of tops with bark, branches, and leaves were quantified by destructive method, in which samples were randomly selected for drying in oven with air circulation at

TABLE III
Diameter tree distribution and sampling in *E. saligna* and *E. grandis* × *E. urophylla* stands.

Diameter class	Lower limit (cm)	Midpoint (cm)	Upper limit (cm)	<i>f</i>	<i>n</i> <i>Esa</i>	<i>n</i> <i>Eug</i>
1	7.0	9.0	10.9	35	2	2
2	11.0	13.0	14.9	186	5	5
3	15.0	17.0	18.9	885	11	11
4	19.0	21.0	22.9	914	10	10
5	23.0	25.0	27.5	51	2	2
Total				2,071	30	30

f is absolute frequency in the forest inventory, *n* is number of trees sampled for *Eucalyptus saligna* (*Esa*) and *Eucalyptus grandis* × *Eucalyptus urophylla* (*Eug*).

105 °C until constant mass. By means of wet biomass and humidity for components, residual biomass for each tree (*w*) was determined.

MODELING RESIDUAL BIOMASS

Generalized linear models (GLM) were composed for estimating residual biomass (*w*) by the covariates: diameter at breast height (*d*), commercial height (*h*), commercial limit diameter (*d_i*), and commercial volume (*v*), considering their transformations in logarithm, inverse, and root-square, as well as the combinations *dh*, *d²h*, *d_ih*, and *d_i²h*. The covariates were selected based on the stepwise procedure through the MASS package (Venables and Ripley 2002) in R programing (R Core Team 2018).

The use of GLM was motivated by the generalization of ordinary linear regression for dependent variables with error non-normally distributed. Thus, for residual biomass with non-normal distribution by the Shapiro-Wilk’s test at 5% of significance level, as well as positive skew, we considered the Gamma distribution (1) for the random component (error), which is related to the systematic part (linear model) by means of the canonical link function (2) (Faraway 2016).

$$f(y; \mu, \theta) = \frac{\left(\frac{\theta}{\mu}\right)^\theta}{\Gamma(\theta)} y^{\theta-1} \exp\left(\frac{-\theta y}{\mu}\right), y > 0 \tag{1}$$

$$\eta = \frac{1}{\mu} \tag{2}$$

Where: μ and θ are Gamma distribution parameters, and η is link function.

In addition, we consider the identification of discrepant values or outliers for the residual biomass estimates, as well as the possibility to removing them to improve the fits. For this, the studentized residuals (3) were plotted per leverage values (*h_{ii}*), which correspond to the importance of each observation for fitting the models, through the car package in R (Fox and Weisberg 2019). Thus, studentized residuals (*t_i*) greater than -2 or 2 with leverage close to zero were identified as outliers of weak contribution to modeling and could be removed without loss of phenomenon characteristics.

$$t_i = \frac{\hat{\epsilon}_i}{\sqrt{\hat{\sigma}^2(1-h_{ii})}} \quad (3)$$

Where: $\hat{\epsilon}_i$ is ordinary residuals, $\hat{\sigma}^2$ is residual variance, and h_{ii} is leverage.

The models were evaluated based on significance of regression coefficients (b_i) at the 1% of significance level by t-test; Akaike's information criterion (AIC), aiming to select the model whose combination of covariates results in the lowest value (4); coefficient of determination (r_{yy}^2), given by the square of correlation between observed and estimated biomass values; normalized root-mean-square-error (NRMSE) as a measure of accuracy, where values close to zero indicate low residual variance (5); and graphical analysis of Pearson's residuals (r_i^p) to indicate the absence of systematic tendency in the estimates (6).

$$AIC = -2 \ln(L) + 2p \quad (4)$$

$$NRMSE = \frac{RMSE}{\bar{y}}, \text{ where: } RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (5)$$

$$r_i^p = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{V}_i}} \quad (6)$$

Where: $\ln(L)$ is log-likelihood function, p is number of model parameters, \bar{y} is mean value of residual biomass, y_i is observed value of residual biomass, \hat{y}_i is estimated residual biomass, n is number of observations, $\hat{\mu}_i$ is estimated mean of residual biomass, and \hat{V}_i is estimated variance of residual biomass.

Additionally, half-normal plots with confidence envelopes at the 95% of significance level were performed to indicate if the models were adequately specified. For this, studentized residuals were plotted by the theoretical quantiles and 99 sample simulations were carried out using the estimates of fitted models through the *hnp* package in R (Moral et al. 2017).

RESULTS

The residual biomass of *Eucalyptus* stands with non-normal distributions, at the 5% of significance by Shapiro-Wilk's test, and positive skew (Figure 1) corroborate the use of Gamma distribution for the random component of generalized linear models in this study. Thus, it was shown that the application of traditional statistical techniques, whose normality hypothesis is required, are not entirely adequate for the analysis of residual biomass variable in wood harvesting operations.

By means of the studentized residuals by the leverages (Figure 2), discrepant values were observed in the modeling of *E. saligna* (Figure 2a) and *E. grandis* × *E. urophylla* stands (Figure 2b), which resulted in studentized residuals higher than -2 and 2 in the estimations. However, in these observations, five points in Figure 2a and seven in Figure 2b presented leverage values closer to zero, indicating that they have a low contribution for modeling and can be removed.

The stepwise procedure resulted on the selection of the covariates: commercial volume (v), diameter at breast height (d), commercial height (h), commercial limit diameter (d_i), log d transformation, and the combinations of dh , d^2h , d_ih , and d_i^2h . Thus, the following regression models were composed for *E. saligna* (7) and *E. grandis* × *E. urophylla* (8) stands for estimating residual biomass (\hat{w}):

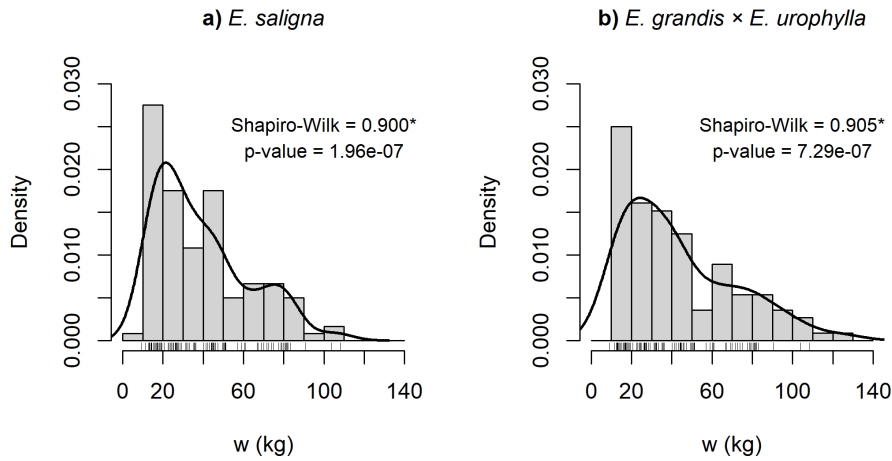


Figure 1 - Residual biomass distribution (*w*) of wood harvesting in *E. saligna* (a) and *E. grandis* × *E. urophylla* (b) stands.

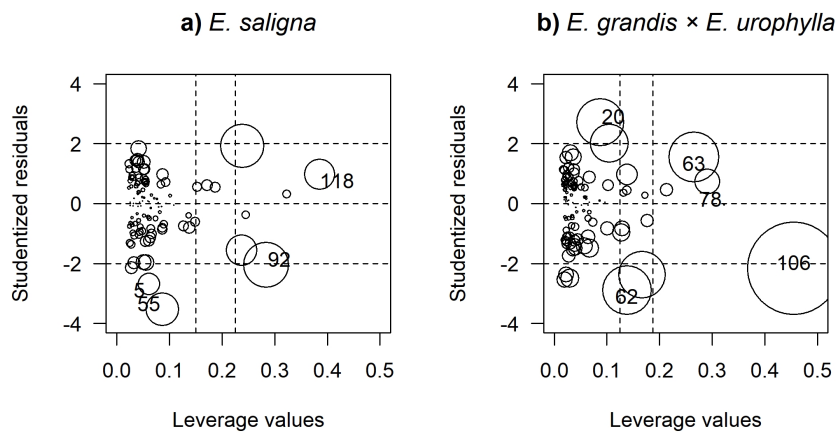


Figure 2 - Studentized residuals by leverage values in the residual biomass estimation of wood harvesting in *E. saligna* (a) and *E. grandis* × *E. urophylla* (b) stands.

$$\hat{w} = \hat{\beta}_0 + \hat{\beta}_1 v + \hat{\beta}_2 d + \hat{\beta}_3 h + \hat{\beta}_4 \log(d) + \hat{\beta}_5 (dh) + \hat{\beta}_6 (d^2h) \tag{7}$$

$$\hat{w} = \hat{\beta}_0 + \hat{\beta}_1 v + \hat{\beta}_2 d + \hat{\beta}_3 h + \hat{\beta}_4 \log(d) + \hat{\beta}_5 (d_i h) + \hat{\beta}_6 (d_i^2 h) \tag{8}$$

In Table IV, these models resulted in lowest values of the Akaike information criterion (*AIC*), whose regression coefficients ($\hat{\beta}_i$) were statistically significant at the 1% of significance level, as well as efficient to explain residual biomass variance by the coefficients of determination (r^2_{yy}) greater than 0.9. In addition, the normalized root-mean square-error (*NRMSE*) of the models corroborated with their high statistical accuracy, resulting in values close to zero and indicating low residual variances.

By means of Pearson’s residuals (Figures 3a and 3b), we observed constant distributions of the residuals that showed the absence of heteroscedasticity and systematic errors for estimating biomass. In addition,

TABLE IV
Goodness of fits to estimate residual biomass in wood harvesting of *E. saligna* and *E. grandis* × *E. urophylla* stands.

$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	AIC	r_{yy}^2	NRMSE
<i>E. saligna</i>									
0.4021	-0.2582	0.0129	8.58×10^{-3}	-0.2162	-4.15×10^{-4}	1.57×10^{-5}	706.96	0.924	0.166
**	**	**	**	**	**	**			
<i>E. grandis</i> × <i>E. urophylla</i>									
0.3207	-0.0581	7.72×10^{-3}	1.02×10^{-3}	-0.1591	-1.51×10^{-4}	4.22×10^{-5}	613.08	0.962	0.127
**	**	**	**	**	**	**			

** significance at the 1% of significance level.

half-normal plots (Figures 3c and 3d) indicated that the fitted models were satisfactory, in which the studentized residuals were inside in the simulated envelopes at a 95% of significance level.

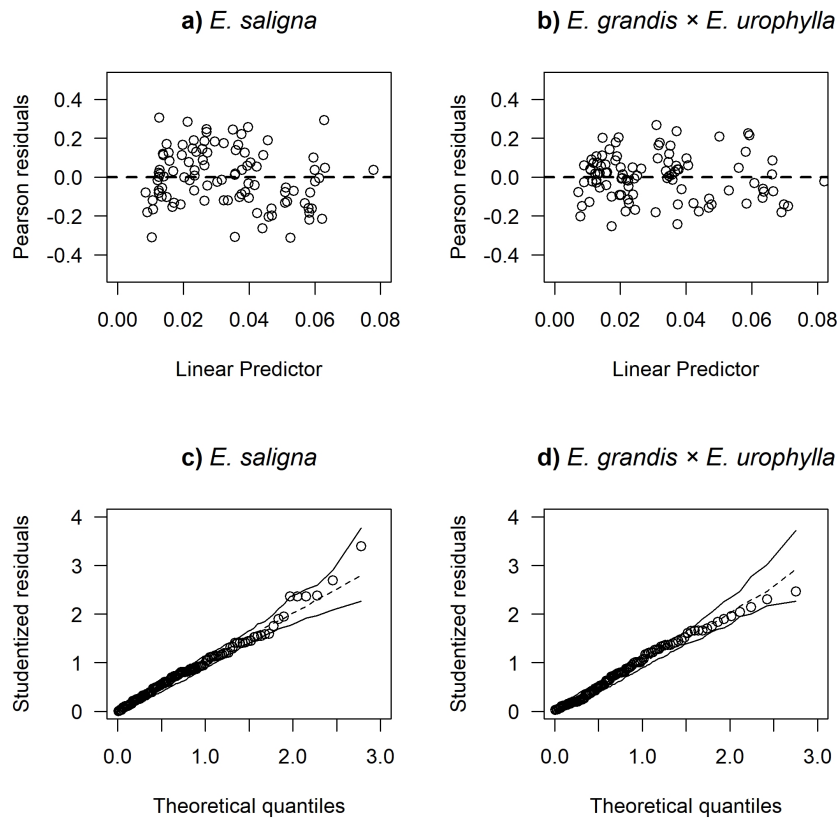


Figure 3 - Pearson's residuals in *E. saligna* (a) and *E. grandis* × *E. urophylla* (b) stands and half-normal plots in the residual biomass estimation of wood harvesting in *E. saligna* (c) and *E. grandis* × *E. urophylla* (d) stands.

DISCUSSION

Some studies have been carried out to evaluate regression models to estimate stump biomass (Palander et al. 2009, Vesa and Palander 2010), productivity (Gallis and Spyroglou 2012, Palander et al. 2017, Brewer et al. 2018), and production and costs (Silayo and Migunga 2014) in wood harvesting methods. However, none of them proposed and evaluated statistical models to estimate residual biomass of *Eucalyptus* stands or other forest species by means of tree variables measured from the harvester processing head. This evidences that the results of this study are important for the development of science and for application in forest sector.

Residual biomass distributions with positive skew (Figure 1) support the application of generalized linear model and Gamma distribution in the random component, since the normality assumption in traditional linear regression was a requirement not satisfied in this study. In addition, the use of Gamma model is plausible for the analysis of non-negative continuous data (Manning et al. 2005, Faraway 2016), as for the residual biomass of wood harvesting.

Generalized linear models also has the advantage for eliminating the dependent variable transformation practice, in order to satisfy the classical linear regression assumptions (Cordeiro and Andrade 2009, Lo and Andrews 2015). Thus, statistical models composed in this study lead to a most explanation of the phenomenon under investigation. These results are especially feasible for data from forest harvesting researches.

The stepwise procedure selected variables associated with forest biomass (Table IV), such as the commercial volume, which has a relationship with wood density and, therefore, with woody components of residual biomass (Somogyi et al. 2007). Also, the variables diameter, height, and their combinations can be associated with the tree component mass that constitutes the tree canopy in residual biomass.

The linear combinations of these variables resulted in models with efficiency and statistical accuracy for estimating residual biomass from tree measurements recorded on the harvester's on-board computer. This was evidenced by the highest $r_{\hat{y}y}^2$ and lowest *NRMSE* values (Table IV), as well as by the uniformly distributed residuals (Figure 3), although the residuals with symmetrical distribution are not an assumption required for generalized linear models (Faraway 2016).

Thus, models developed in this study are innovative tools to estimate the residual biomass in mechanized wood harvesting of *Eucalyptus* stands, in which can be inserted into the harvester's hardware to provide real-time information for operational planning and logistics. In addition, we suggest that other factors that also influence on the biomass production, such as age, species or genetic material, and forest site productivity (Palander et al. 2009, Eloy et al. 2018), are investigated to be incorporated in the models.

AUTHORS CONTRIBUTIONS

RODRIGUES CK, LOPES ES, FIGUEIREDO FILHO A and PELISSARI AL conceived and designed the study; RODRIGUES CK and SILVA MKC performed the study; RODRIGUES CK and PELISSARI AL analyzed the data; LOPES ES and FIGUEIREDO FILHO A contributed to materials/analysis tools; RODRIGUES CK and PELISSARI AL wrote and revised the paper.

ACKNOWLEDGMENTS

This study was carried out with the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação Araucária, and Universidade Estadual do Centro-Oeste (UNICENTRO).

REFERENCES

- ALVARES CA, STAPE JL, SENTELHAS PC, GONÇALVES JLM AND SPAROVEK G. 2013. Köppen's climate classification map for Brazil. *Meteorol Z* 22(6): 711-728.
- BREWER J, TALBOT B, BELBO H, ACKERMAN P AND ACKERMAN S. 2018. A comparison of two methods of data collection for modelling productivity of harvesters: manual time study and follow-up study using on-board-computer stem records. *Ann For Res* 61(1): 109-124.
- CORDEIRO GM AND ANDRADE MG. 2009. Transformed generalized linear models. *J Stat Plan Infer* 139(9): 2970-2987.
- ELOY E, SILVA DA, CARON BO, ELLI EF AND SCHWERZ F. 2018. Effect of age and spacing on biomass production in forest plantations. *Rev Árvore* 42(2): e420214.
- ESTEBAN TE, LAINA R, MARTÍNEZ-FERRARI R AND TORRIJOS YA. 2011. Recovering of forest biomass from Spanish hybrid poplar plantations. *Biomass Bioenerg* 35(7): 2570-2580.
- FARAWAY JJ. 2016. Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models, 2nd ed., CRC Press, Taylor & Francis Group, Boca Raton, 345 p.
- FOX J AND WEISBERG S. 2019. An R companion to applied regression, 3rd ed., Sage Publications, Thousand Oaks, California, 474 p.
- GALLIS C AND SPYROGLOU G. 2012. Productivity linear regression models of tree-length harvesting system in natural coastal Aleppo Pine (*Pinus halepensis* L.) Forests in the Chalkidiki area of Greece. *Croat J For Eng* 33(1): 115-123.
- IBÁ – INDÚSTRIA BRASILEIRA DE ÁRVORES (BRAZILIAN TREE INDUSTRY). 2017. Report 2017. Disponível at: http://iba.org/images/shared/Biblioteca/IBA_RelatorioAnual2017.pdf. Acessado 29/02/2019.
- LAURI P, HAVLÍK P, KINDERMANN G, FORSELL N, BÖTTCHER H AND OBERSTEINER M. 2014. Woody biomass energy potential in 2050. *Energ Policy* 66: 19-31.
- LO S AND ANDREWS S. 2015. To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Front Psychol* 6: 1171.
- MANNING WG, BASU A AND MULLAHY J. 2005. Generalized modeling approaches to risk adjustment of skewed outcomes data. *J Health Econ* 24(3): 465-488.
- MEDERSKI PS, BEMBENEK M, KARASZEWSKI Z, PILAREK Z AND ŁACKA A. 2018. Investigation of log length accuracy and harvester efficiency in processing of oak trees. *Croat J For Eng* 39(2): 173-181.
- MORAL RA, HINDE J AND DEMÉTRIO CGB. 2017. Half-normal plots and overdispersed models in R: the hnp package. *J Stat Softw* 81(10): 1-23.
- NIEUWENHUIS M AND DOOLEY T. 2006. The effect of calibration on the accuracy of harvester measurements. *IJFE* 17(2): 25-33.
- PALANDER T, KÄRHÄ K AND MEHTÄTALO L. 2017. Applying polynomial regression modeling to productivity analysis of sustainable stump harvesting. *Scand J Forest Res* 32(3): 204-212.
- PALANDER T, VESA L, TOKOLA T, PIHLAJA P AND OVASKAINEN H. 2009. Modelling the stump biomass of stands for energy production using a harvester data management system. *Biosyst Eng* 102(1): 69-74.
- R CORE TEAM. 2018. R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna.
- SILAYO DSA AND MIGUNGA GA. 2014. Productivity and costs modeling for tree harvesting operations using chainsaws in plantation forests, Tanzania. *IJFE* 3(4): 464-472.
- SOMOGYI Z, CIENCIALA E, MÄKIPÄÄ R, MUUKKONEN P, LEHTONEN A AND WEISS P. 2007. Indirect methods of large-scale forest biomass estimation. *Eur J Forest Res* 126(2): 197-207.
- VENABLES WN AND RIPLEY BD. 2002. Modern applied statistics with S. 4th ed., Springer, New York, 498 p.
- VESA L AND PALANDER T. 2010. Modeling stump biomass of stands using harvester measurements for adaptive energy wood procurement systems. *Energy* 35(9): 3717-3721.
- YEMSHANOV D, MCKENNEY DW, FRALEIGH S, MCCONKEY B, HUFFMAN T AND SMITH S. 2014. Cost estimates of post harvest forest biomass supply for Canada. *Biomass Bioenerg* 69: 80-94.