Drewno 2016, Vol. 59, No. 197
DOI: 10.12841/wood.1644-3985.C16.06

Ewa Bakinowska, Piotr S. Mederski, Anna Szczepańska-Álvarez, Zbigniew Karaszewski, Mariusz Bembenek

THE PARALLEL APPLICATION OF TWO PROBABILITY MODELS, LOGIT AND PROBIT, FOR THE ACCURATE ANALYSIS OF SPRUCE TIMBER DAMAGE DUE TO THINNING OPERATIONS

Logit and probit models belong to the class of generalised linear models. A few applications of both models have been documented in the field of forestry. The objective of this paper was to test the parallel use of these models to discover the differences in damage to a spruce stand after thinning using the full tree system, the long wood system and the short wood system. In particular the aim was to ascertain the general damage probability caused by the harvesting systems (HS) and the particular damage class probability in each HS. When the general damage probability was calculated the logit model was used. When nine damage classes were taken into account, however, the probit model was found to fit the data better. In this case, the results obtained gave accurate information on the probability of the appearance of a particular damage class for each HS. It was concluded that the probit and logit models should be considered in parallel in order to obtain the best possible goodness of fit and to get accurate information on the distribution of damage classes.

Keywords: generalised linear models, categorical data, goodness of fit, Akaike criterion

Introduction

Thinning operations in Polish forests are performed once per decade. Zasady Hodowli Lasu (Principles of Silviculture) [2012] for spruce stands recommend early and late thinning, before and after 40 years of age, respectively. During this

Ewa Bakinowska✉ (*ewa.bakinowska@put.poznan.pl*), Institute of Mathematics, Poznan University of Technology, Poznan, Poland; Piotr S. Mederski (*piotr.mederski@up.poznan.pl*), Mariusz Bembenek (*mariusz.bembenek@up.poznan.pl*), Department of Forest Utilisation, Poznan University of Life Sciences, Poznan, Poland; Anna Szczepańska-Álvarez, (*anna.szczepanska-alvarez@up.poznan.pl*), Department of Mathematical and Statistical Methods, Poznan University of Life Sciences, Poznan, Poland; Zbigniew Karaszewski, (*z_karaszewski@itd.poznan.pl*), Wood Science and Application Department, Wood Technology Institute, Poznan, Poland

process, damage occurs to some of the remaining trees. Aside from the human factor, the level of damage can depend on, among other things, thinning intensity, the method of thinning, the operation and machine applied and the harvesting system (HS) [Bembenek et al. 2013a, b; Karaszewski et al. 2013]. The differences in the level of damage can be analysed with simple data comparison using the Kruskal-Wallis or U Mann-Whitney tests [Stańczykiewicz et al. 2015a]. There are also studies based on the χ^2 Pearson test [Lageson 1997], T-tests [Modig et al. 2012] and Student's *t* distribution [Glöde and Sikström 2001; Stańczykiewicz et al. 2015b]. Sirén et al. [2013] used mixed-model analysis for nested data structures, including both fixed and random effects.

Tree damage in lowland spruce stands caused by early and late thinning was also analysed by Bembenek et al. [2013a, b]. These authors used the χ^2 Pearson test and the Fisher test to study the relationship between the system of wood harvesting and percentage of trees with damage.

The above-mentioned experiment by Bembenek et al. [2013a, b] only gave three sets of information: 1) the percentage of trees with damage in each damage class, but without division into HSs, 2) the percentage of trees with damage in each HS, but without division into damage classes, and 3) the statistical differences in the frequency of trees with damage (without division into classes) in each HS. It was hypothesised, however, that the use of a model could give more exact results, such as the probability of the appearance of a particular damage class in each HS separately.

The general objective of this paper was to find out:

1) the probability of damage (without division into classes) for each HS: full tree system (FTS), long wood system (LWS) and short wood system (SWS)
2) the probability of the appearance of a particular damage class (from one to nine) separately in each HS
3) if there are (statistically significant) differences for 1) and 2)
4) if there are differences between the results obtained and previous work based on simple statistics.

In order to achieve this, two generalised linear models were used: logit and probit. Two models were used mainly to find out if they would be characterised by different goodness of fit, and if so, the one (model) with a better goodness of fit would be used for the final analysis.

The use of both generalised linear models in this work aims to extend the results published by Bembenek et al. [2013a, b]. The parallel use of two models has also been applied in medicine [Zhou and Wong 2008], animal breeding [Röhe and Kalm 2000] and in economic studies [Layton and Katsuura 2001].

In certain publications, for example, there are applications of, the logit model, to compare varieties of seed pea in respect to lodging [Bakinowska and Kala 2007] and to analyse downy mildew infection of field pea varieties [Bakinowska et al. 2015]. In forestry, the logit model has been used to analyse tree damage [Sirén 2001]. Taking into account the above-mentioned studies, however, in which only the logit model was used, it was assumed that there was a risk that it would not give the best (possible) goodness of fit to describe a particular phenomenon.

Material and methods

Experiment

In the experiment, the number of damaged trees and the damage class formed during early and late thinning were studied. Trees with damage were divided according to nine classes (tab. 1).

Table 1. Tree damage classes

Damage class	Damaged part of tree		Damage type[1]
1	Bark	trunk or root collar	bark damaged
2	Bark	roots	bark damaged
3	Cambium	trunk or root collar	bark damaged, opened or damaged cambium
4	Cambium	roots	bark damaged, opened or damaged cambium
5	Wood	trunk or root collar	wood fibres wounded
6	Wood	roots	wood fibres wounded
7	Tree	bent	slightly bent tree with opportunity for further growth
8	Tree	bent significantly	tree bent significantly with little chance of good development
9	Tree	damaged	broken or pulled-off

The damage data was analysed from sample plots selected in North and North-West Poland. Early thinning was carried out on lowland spruce stands of second age class (AC2, 21-40 years old) in the forest districts of Zaporowo and Różańsko, on eight sample plots, each measuring 0.25 ha. On all the sample plots, the total number of observations amounted to 2381 trees, 2163 and 218, respectively, with and without damage (tab. 2).

Table 2. Number of observations (trees) evaluated statistically

Trees	With damage		Without damage		Total	
	Et	Lt	Et	Lt	Et	Lt
FTS	84	60	454	147	538	207
LWS	109	61	577	381	686	442
SWS	25	12	1132	574	1157	586
Total	218	133	2163	1102	2381	1235

Et – early thinning, Lt – late thinning, FTS – full tree system, LWS – long wood system, SWS – short wood system.

Late thinning was provided on lowland spruce stands in AC4 (61-80 y.o.) in the forest districts of Dretyń, Wejherowo, Mieszkowice and Różańsko, on ten sample plots, each with an area 0.25 ha. On all the sample plots, the total number of observations amounted to 1235 trees, 1102 and 133, respectively, with and without damage (tab. 2).

During the thinning process, three different systems of wood harvesting were applied: FTS, on two plots in both early and late thinning, LWS, on two plots in early thinning and on four plots in late thinning, as well as SWS, on four plots in both early and late thinning.

Generalised linear model

For analysis of the experiment, both the logit and probit models were used. These models belong to a wider class of models called generalised linear models [Agresti 1984; McCullagh and Nelder 1989], which in matrix form can be written as:

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} \tag{1}$$

where the vector $\boldsymbol{\eta}$ is called the link function, $\mathbf{X}$ is the matrix of covariates, and $\boldsymbol{\beta}$ is the vector of unknown parameters.

If the link function $\boldsymbol{\eta}$ in model (1) is the logit function, then the model is called the generalised linear model with logit link function, or, in short, the logit model or the logistic regression model [Rao and Toutenburg 1999]. This model is often used in the analysis of experiments in which the observed units are assigned to separate categories. The scalar form of the logit model [McCullagh and Nelder 1989; Miller et al. 1993; Rao and Toutenburg 1999; Bakinowska and Kala 2007] is as follows:

$$\eta_{ji} = \log \frac{\pi_{1i} + \pi_{2i} + \ldots + \pi_{ji}}{1 - (\pi_{1i} + \pi_{2i} + \ldots + \pi_{ji})} = \theta_j + \tau_i, \quad j = 1,2,\ldots,k-1, \quad i = 1,2,\ldots,s \tag{2}$$

where θ_j is the border (cut point) of the jth category, τ_i is the effect of the ith treatment (so $\theta_j + \tau_i$ represents the cut point of the jth category for the ith treatment), π_{ji} are the probability of success of the jth category, $j = 1, 2, ..., k$ for the fixed ith treatment, which equal to one: $\sum_{j=1}^{k} \pi_{ji} = 1$. If the link function $\boldsymbol{\eta}$ in model (1) is the vector of inverses Φ^{-1} of the standard normal cumulative distribution functions, then the model is called the generalised linear model with a probit link function or, in short, the probit model [Rao and Toutenburg 1999]. For a fixed ith treatment, this model can be written in the following scalar form:

$$\eta_{ji} = \Phi^{-1}(\pi_{1i} + \pi_{2i} + \ldots + \pi_{ji}) = \theta_j + \tau_i, \quad j = 1,2,\ldots,k-1, \quad i = 1,2,\ldots,s. \tag{3}$$

For this research, it was assumed that the vector of unknown parameters $\boldsymbol{\beta}$ in model (1) is of the form $\boldsymbol{\beta}^T = (\boldsymbol{\theta}^T, \boldsymbol{\tau}^T)$, where $\boldsymbol{\theta}$ denotes the vector of cut points and $\boldsymbol{\tau}^T = (\tau_1, \tau_2, \ldots, \tau_s)$ is the vector of effects. The system of hypotheses was tested:

$$H_0\text{: } \boldsymbol{\beta}=\boldsymbol{\beta}_* \quad \text{against} \quad H_1\text{: } \boldsymbol{\beta}\neq\boldsymbol{\beta}_* \tag{4}$$

using the Wald statistic, which has the form:

$$(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_*)^T \hat{\mathbf{F}}(\boldsymbol{\beta}_*)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_*)$$

and has approximately χ^2_p distribution (p being the order of $\boldsymbol{\beta}_*$), where $\hat{\boldsymbol{\beta}}_*$ is an estimate of $\boldsymbol{\beta}$, and $\mathbf{F}(\boldsymbol{\beta})$ is the Fisher information matrix of $\hat{\boldsymbol{\beta}}$ [Agresti 1984; McCulloch and Searle 2001].

The aim of the analysis was to estimate the unknown probabilities in model (2) which can be expressed as:

$$\pi_{1i}+\pi_{2i}+\ldots+\pi_{ji}=\frac{\exp(\theta_j+\tau_i)}{1+\exp(\theta_j+\tau_i)}, \quad j=1,2,\ldots,k-1, \quad i=1,2,\ldots,s, \tag{5}$$

or in model (3) which can be written as:

$$\pi_{1i}+\pi_{2i}+\ldots+\pi_{ji}=\Phi(\theta_j+\tau_i), \quad j=1,2,\ldots,k-1, \quad i=1,2,\ldots,s, \tag{6}$$

based on the experimental data. In order to select the model which fitted the experimental data better, one of the goodness of fit statistics, namely the Akaike information criterion was used:

$$AIC=-2l(\hat{\pi})+2p,$$

where $l(\hat{\pi})$ is the log-likelihood function obtained in the current model, and p is the number of parameters in the model. The smaller the Akaike information criterion, the better the model fits the observed data [Lindsey 1997].

The Akaike criterion was considered the most suitable for the number of observations in the analysis. Other statistics were also tested in order to measure the goodness of fit, e.g. the Schwarz Information Criterion (SIC), -2logL and Deviance, however they produced the same results as the Akaike criterion.

Analysis

In the experiment, observations were made of the trees (experimental units), which were assigned to fixed disjoint categories. As the data were categorised, the analysis was carried out using the generalised linear models: logit and probit. All the calculations were performed in SAS [SAS Institute 1997] using the *logistic glm* procedure (with *logit* or *probit* link function). Based on the results of the analysis, estimates of the unknown parameters in models (2) and (3) were obtained. A comparison was then made using Akaike's goodness of fit for both models – the lower the value, the better the goodness of fit.

Finally, the model with the better goodness of fit was selected. Based on the estimates of the parameters of the selected model, the unknown probability of success for each category was evaluated.

It transpired that when the values of the *AIC* statistics were identical for both the models, the calculated probability also had the same values. In this case, the results of the logit model were selected (as they proved easier to interpret) for presentation in tables 2 and 3. In these tables, the number in brackets '(n)' denotes the standard error of the estimate, asterisk '*' denotes significance at level 0.05, and two asterisks '**' denote significance at level 0.01. In all the cases considered, the unknown probabilities were estimated under the restriction that the effect $\tau_s = 0$, and the system of hypotheses $H_0 : \tau_i = 0$, against $H_1 : \tau_i \neq 0$, for $i = 1,2,\ldots,s-1$, was tested.

Results and discussion

Early thinning

Initially, a comparison of the various timber HSs was made to estimate the probability of tree damage during early thinning. The trees remaining after thinning were divided into two categories: damaged trees (category one) and undamaged trees (category two). In this case, it was observed that both models fit the data with the same goodness of fit statistics $AIC_{\text{logit}} = AIC_{\text{probit}} = 1314.018$. The results showed that the FTS and LWS differed significantly (α=0.01) from the SWS with respect to the level of trees with damage (the estimates of the parameters and values of the Wald statistic for the logit model (2) (tab. 3)).

Additionally, using the logit model to compare the FTS with LWS and based on the value of the Wald statistics, it was found that these systems did not differ significantly. These results were not different to those presented by Bembenek et al. [2013a], where simple statistical analysis (χ^2 Pearson test and the Fisher test) also gave the same differences.

Based on the estimates of the parameters (tab. 3), the probabilities of tree damage for each timber HS were calculated using formula (5). The probabilities of tree damage were similar for the FTS and LWS, and were close to 0.16, while the SWS was the method causing the lowest level of damage (fig. 1).

Further analysis was carried out for the damage class to obtain the probabilities of different classes of damage caused to the trees due to the various timber HSs. In this case, based on the *AIC* fit statistics, the probit model fitted the experimental data better (AIC_{logit}=617.818, AIC_{probit}=617.561), therefore the analysis was based on model (3) with the use of estimate(s) for early thinning

Table 3. Significance of estimates of parameters and Wald statistics for trees with and without damage (two categories) of three analysed HSs

Thinning	Parameter		Estimate	Wald statistic
Early	Intercept	θ	-3.8128 (0.20)	355.61**
	FTS	τ_1	2.1255 (0.23)	82.16**
	LWS	τ_2	2.1463 (0.23)	88.95**
	SWS	τ_3	0	–
Late	Intercept	θ	-3.8673 (0.29)	175.87**
	FTS	τ_1	2.9712 (0.33)	81.36**
	LWS	τ_2	2.0354 (0.32)	39.82**
	SWS	τ_3	0	–

Significance level: (*) $\alpha = 0.05$, (**) $\alpha = 0.01$.

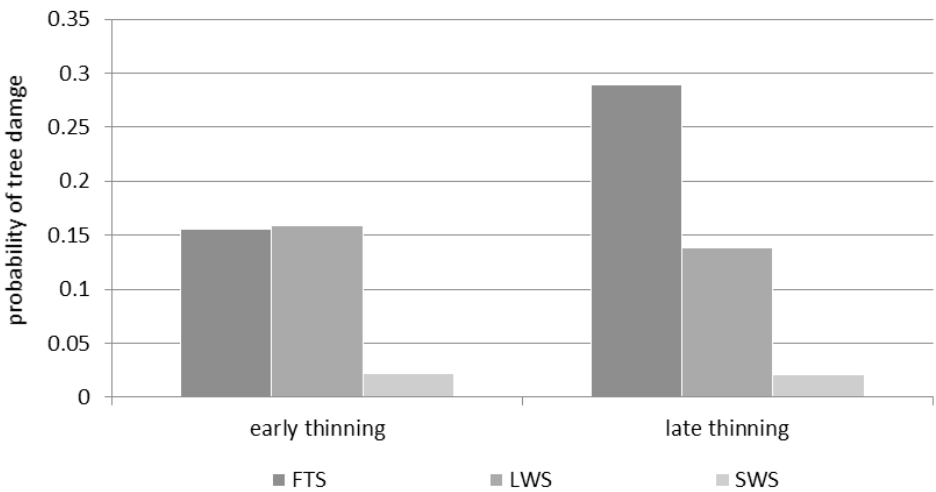

Fig. 1. Probability of tree damage

(tab. 4). It transpired that the differences between the FTS and LWS methods were not statistically significant from the SWS in respect of each damage class considered separately (tab. 4).

An analysis comparing the FTS and LWS was also conducted and the differences between these systems were also not significant. Based on the formula (6), the probability of the appearance of a particular tree damage class was calculated. The damage class most likely to appear in all the HSs was type 1 (damage class 1, fig. 2). The values of the probabilities were between 0.45 and 0.58 (fig. 2). The next most probable was damage class 3 at about 0.25. The trend for the distribution of the probabilities of tree damage type, however, was similar for all the systems: the FTS, LWS and SWS, and there were no statistically significant differences (fig. 2).

Table 4. Significance of estimates of parameters and Wald statistics for all considered damage classes (nine categories) of three analysed HSs

Parameter		Early thinning		Late thinning	
		estimate	Wald statistic	estimate	Wald statistic
Intercept	θ_1	0.2073 (0.23)	0.7874	-0.5973 (0.31)	3.6386
Intercept	θ_2	0.3003 (0.23)	1.6486	0.0296 (0.31)	0.0091
Intercept	θ_3	1.0514 (0.24)	19.1129**	0.5996 (0.31)	3.6835
Intercept	θ_4	1.2683 (0.25)	26.8885**	0.7471 (0.31)	5.6700*
Intercept	θ_5	1.6969 (0.26)	43.0485**	1.5820 (0.33)	22.6578**
Intercept	θ_6	1.9905 (0.28)	51.8391**	–	–
Intercept	θ_7	2.0581 (0.28)	53.2261**	–	–
Intercept	θ_8	2.7488 (0.41)	45.0548**	–	–
FTS	τ_1	-0.0421 (0.26)	0.0256	-0.3219 (0.33)	0.9331
LWS	τ_2	-0.2238 (0.25)	0.7719	-0.1881 (0.33)	0.3198
SWS	τ_3	0	–	0	–

Significance level: (*) $\alpha = 0.05$, (**) $\alpha = 0.01$.

The results based on formula (6) were different to those in the work of Bembenek et al. [2013a]. The application of the model(s) makes it possible to find out the particular probability of the appearance of a damage class in each analysed HS.

Late thinning

As in the case of early thinning, the trees remaining in the forest after late thinning were divided into two categories: damaged trees (category one) and undamaged trees (category two), and the analysis was performed based on models (2) and (3). In this case, the goodness of fit statistics were equal, $AIC_{\text{logit}}=AIC_{\text{probit}}=727.090$ and the logit model was used (due to ease of interpretation). The estimates obtained by the logit model (2) were selected for the analysis, and showed that the FTS and LWS differed significantly from the SWS with respect to tree damage (tab. 3, late thinning). In late thinning, further analysis also showed that differences between the FTS and LWS were significant. In contrast to early thinning, in all the studied systems, the differences in damage were statistically significant at $\alpha = 0.01$. These results were similar to earlier work done by Bembenek et al. [2013b]. In late thinning, the heaviest damage was caused by the FTS, with the probability of tree damage close to 0.3, while the least harmful was the SWS, with a probability of tree damage close to 0 (fig. 1). Not all of the tree damage classes that appeared in early thinning were present in late thinning: only the first six damage classes were observed (fig. 2).

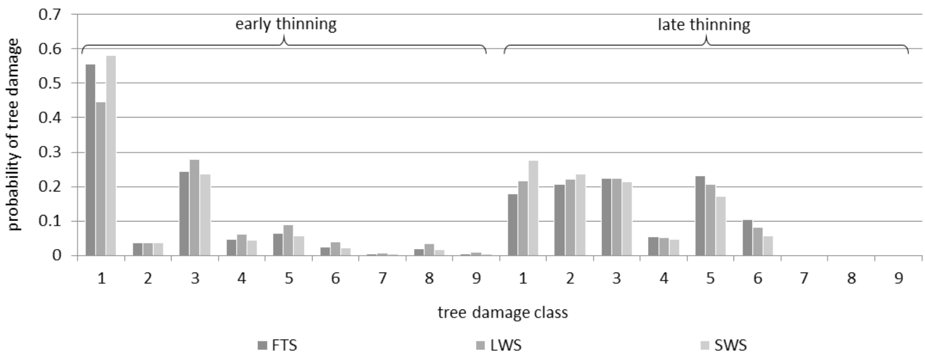

Fig. 2. Probability of tree damage with respect to all damage classes

Based on the goodness of fit statistics, AIC_{logit}=463.684 and AIC_{probit}=462.871, the probit model (3) was selected for the analysis. It was observed that, taking each damage class into account separately, all three HSs did not differ significantly (tab. 4). These results are different to those of Bembenek et al. [2013b], where statistically significant differences were observed between the damage class groups (joined), such as bark (class 1 and 2), cambium (class 3 and 4) and wood (class 5 and 6). The cited work, however, does not explain precisely what the probability was of the appearance of a particular damage class in each HS. The results obtained with the use of the model(s) are more accurate and, ultimately, more useful for the forester, who can precisely select less invasive HS with regards to a particular damage class. It is important when decisions are made concerning particular species. Spruce is particularly vulnerable to fungi infections when the bark is removed, therefore, focus on the probability of the appearance of this particular damage is important and possible when the model is used. Furthermore, the application of both models, logit and probit, gives an even higher level of accuracy in the prediction of the appearance of damage.

Conclusion

The parallel use of the logit and probit models made it possible to accurately distinguish the probability of damage in early and late thinning. When only the damaged trees were analysed, the differences between the SWS, and both the FTS and LWS, were statistically significant in early thinning. In contrast, in late thinning, those differences were statistically significant between all the HSs. The use of logit and probit models in this case did not give different results from earlier findings based on the χ^2 Pearson test and the Fisher test.

Further analysis of the probability of the appearance of a particular damage class in each HS was possible only when the models were used, but there were no differences observed which were statistically significant. In previous work, based on the χ^2 Pearson test and the Fisher test, only damage classes in groups (bark, cambium and wood) were analysed, although with statistically significant differences in some cases.

The new approach using the application of models gives useful results indicating an accurate level of a particular damage class probability in each HS. This can be useful when decisions are made concerning which HS to apply in spruce stands vulnerable to fungi disease.

Additionally the parallel use of the two models, logit and probit, made it possible to choose the one with the better goodness of fit when calculating the probability of the remaining stand damage after thinning. When two categories (damaged trees and undamaged trees) were considered, it was observed (based on the *AIC* goodness of fit statistics) that both models fit equally well, although the logit model was eventually used (as it was easier to interpret). When nine categories (damage classes) were taken into account, however, the probit model was found to fit the data better. Based on the above results, therefore, it can be concluded that to achieve the best and most accurate results, two models, probit and logit, should be considered in parallel to describe such phenomenon.

References

Agresti A. [1984]: Analysis of ordinal categorical data. Wiley, New York

Bakinowska E., Kala R. [2007]: An application of logistic models for comparison of varieties of seed pea with respect to lodging. Biometrical Letters [44]: 143-154

Bakinowska E., Pilarczyk W., Zawieja B. [2015]: Analysis of downy mildew data on field pea: an empirical comparison of two logistic models. Acta Agriculturae Scandinavica, Section B – Soil & Plant Science DOI:10.1080/09064710.2015.1072235

Bembenek M., Giefing D.F., Karaszewski Z., Mederski P.S., Szczepańska-Álvarez A. [2013a]: Uszkodzenia drzew w następstwie trzebieży wczesnych w nizinnych drzewostanach świerkowych. Sylwan 157 [10]: 747-753

Bembenek M., Giefing D.F., Karaszewski Z., Mederski P.S., Szczepańska-Álvarez A. [2013b]: Uszkodzenia drzew w nizinnych drzewostanach świerkowych podczas zabiegu trzebieży późnej. Sylwan 157 [12]: 892-898

Glöde D., Sikström U. [2001]: Two felling methods in final cutting of shelterwood, singlegrip harvester productivity and damage to the regeneration. Silva Fennica 35 [1]: 71-83

Karaszewski Z., Giefing D.F., Mederski P.S., Bembenek M., Dobek A., Stergiadou A. [2013]: Stand damage when harvesting timber using a tractor for extraction. Forest Research Papers 74 [1]: 27-34

Lageson H. [1997]: Effects of thinning type on the harvester productivity and on the residual stand. Journal of Forest Engineering 8 [2]: 7-14

Layton A., Katsuura M. [2001]: Comparison of regime switching, probit and logit models in dating and forecasting US business cycles. International Journal of Forecasting 17 [3]: 403-417

Lindsey J.K. [1997]: Applying Generalized Linear Models. Springer-Verlag, New York

McCullagh P., Nelder J.A. [1989]: Generalized linear models. 2nd. ed., Chapman and Hall, London

McCulloch C.E., Searle S.R. [2001]: Generalized, linear, and mixed models. Wiley, New York

Miller M.E., Davis C.S., Landis J.R. [1993]: The analysis of longitudinal polytomous data: generalized estimating equations and connections with weighted least squares. Biometrics 49: 1033-1044

Modig E., Magnusson B., Valinger E., Cedergren J., Lundqvist L. [2012]: Damage to residual stand caused by mechanized selection harvest in uneven-aged *Picea abies* dominated stands. Silva Fennica 46 [2]: 267-274

Rao C.R., Toutenburg H. [1999]: Linear Models. 2nd ed., Springer-Verlag, New York

Röhe R., Kalm E. [2000]: Estimation of genetic and environmental risk factors associated with pre-weaning mortality in piglets using generalized linear mixed models. Animal Science 70: 227-240

SAS Institute [1997]: SAS/STAT software: Changes and enhancements through release 6.12. SAS Inst., Cary, NC

Sirén M. [2001]: Tree damage in single-grip harvester thinning operations. Journal of Forest Engineering 12 [1]: 29-38

Sirén M., Ala-Ilomäki J., Mäkinen H., Lamminen S., Mikkola T. [2013]: Harvesting damage caused by thinning of Norway spruce in unfrozen soil. International Journal of Forest Engineering 24 [1]: 60-75

Stańczykiewicz A., Sowa J.M., Leszczyński K., Kulak D., Szewczyk G. [2015a]: Uszkodzenia drzew i odnowienia w wyniku pozyskania drewna z użyciem urządzeń agregowanych z ciągnikami rolniczymi w trzebieżowych drzewostanach świerkowych. Sylwan 159 [3]: 201-210

Stańczykiewicz A., Szewczyk G., Kulak D. [2015b]: Damage to advance growth resulting from timber harvesting during final cuttings. Baltic Forestry 21 [1]: 144-151

Zaou X., Wong S.T.C. [2008]: Computational systems bioinformatics – methods and biomedical applications. World Scientific Publishing Co. Pte. Ltd.

Zasady Hodowli Lasu (Principles of Silviculture) [2012]: Centrum Informacyjne Lasów Państwowych (The State Forests Information Centre)

Submission date: 15.04.2015

Online publication date: 26.07.2016