UDC 575.633
https://doi.org/10.2298/GENSR1901313G
Original scientific paper

AN ALTERNATIVE STATISTICAL MODEL FOR THE ASSESSMENT OF DRY MATTER ACCUMULATION IN COOL SEASON CEREALS: COX REGRESSION

Zeynep GÖKKUŞ[1*] and Mevlüt AKÇURA[2]

[1]Kastamonu University, İnebolu Vocational School, Department of Computer Technologies, Kastamonu, Turkey
[2]Canakkale Onsekiz Mart University, Faculty of Agriculture, Department of Field Crops, Canakkale, Turkey

Gökkuş Z. and M. Akçura (2019): *An alternative statistical model for the assessment of dry matter accumulation in cool season cereals: Cox regression*.- Genetika, Vol 51, No.1, 313-322.

In this study, the growing behaviors of some cool season cereals (bread wheat, rye, durum wheat and barley) cereals were modeled simultaneously during the two growing seasons. For this purpose, Cox Regression was proposed as an alternative to the preferred regression methods in previous studies. In the study, based on the seasonal data of two different season growing seasons (2012-2013, 2013-2014 and both), each of which has 5 replicates 27 samples, growth rates of these cereals via dry matter accumulation quantities were explained in three different models. For this purpose, the dry matter accumulation amounts were fitted to the survival data and Cox Regression method, which uses the hazard function, the rate of occurrence of a particular event, was preferred. As a result, each model was found to be very important (p <0.000). It was determined that *i)* the fastest growing species was barley, *ii)* dry matter accumulation decreased as temperature increased, and *iii)* dry matter accumulation in crops changed during each growth season.

Keywords: Cox regression model, cool season cereals, growing model

INTRODUCTION

Cool season cereals in the world and Turkey are the most cultivated and used crops. Especially, bread wheat and barley among these cereals have highest usage and cultivation area. Due to representing less variable cereal yields, they have a higher yield per unit area according to

Corresponding author: Zeynep Gökkuş, Kastamonu University, İnebolu Vocational School, Department of Computer Sciences, Kastamonu, Turkey, email:z.gokkus@gmail.com

the others. For this reason, they are preferred mostly by farmers. As in many plants, cereal yields, however, depend directly on climate factors that cannot be changed by farmers. Meanwhile, heredity level of cereal yield is low in the cool season cereals (KAYA *et. al*., 2006).

Generally, cool season cereals are cultivated in rain-dependent conditions. A detailed study of the extent to which climatic factors play a role in the cultivation of cool climate cereal in a region is very important in terms of determining ideal cultivation techniques (HOCAOGLU and COSKUN, 2017). Cereal yield is the result of the accumulation of dry matter in a growing season. Generally, the green area season of the genotypes is high in the accumulation of dry matter and thus the yield of cereal is also high (SADRAS *et al*., 2019).

Different statistical methods are used to explain the accumulation of dry matter in plants (KARADAVUT *et al*., 2005). However, different regression models were found to be important in different plants (DAMGAARD *et al*., 2002; KARADAVUT, 2009). These are the parametric regression methods which are selected in accordance with the sampling and distribution of data. However, certain assumptions must be provided for the application of parametric methods (BOLIN, 2004; DAVIDIAN, 2017). Non-parametric regression methods should be used when these assumptions cannot be achieved (BREIMAN, 2017). If assumptions are provided, parametric methods must be used for reliability (FARAWAY, 2016). In addition, there is a semi-parametric regression method called Cox regression that requires few assumptions and is as reliable as parametric methods (COX, 2018).

Cox Regression has been developed to measure the effects of covariates on survival time, which generally have not normal distribution. This method is used when the assumption is made that the dependent variable does not need to show a certain parametric distribution and that the explanatory variables are proportional (COX, 1972). Today, it is widely used in order to reveal the relationship between dependent and independent variables in a wide range from health to industry, from science to social sciences.

In this study, it is aimed to model both directly the cereal yields of the crops and indirectly growth rates by using dry matter accumulation amounts. The dry matter accumulation in the cereals representing the growth rates of the crops during the sampling was the survival variable. Based on the assumption that the variables of temperature, precipitation and growing season have a proportional effect on cereal species regardless of the distribution of the dependent variable, Cox regression method was found suitable in this study. According to this, the function of survival, dependent variable, was the dry matter amount and the covariates were the cereal species, temperature, precipitation, and growing season. Three models (2012-2013; 2013-2014; both seasons) were developed to determine whether dry matter varies according to cereal species, temperature, precipitation and growing seasons and all seasons were investigated in terms of dry matter amounts according to the species.

MATERIAL AND METHOD

Field trials were conducted in Çanakkale Onsekiz Mart University Agricultural Experiment Station in Dardanos, Çanakkale in two consecutive growing seasons (2012-2013 and 2013-2014) in the same plot. Four cool season cereal cultivars [Bereket (bread wheat), Aslım-95 (rye), Fırat-95 (durum wheat), Akhisar-98 (six rowed barley)] were used as plant material. Field trials were sowed with plot seeder in 7 November 2012 and 30 October 2013 for the first and second years of experiment, respectively (Hocaoglu and Çoskun, 2018). Soil analysis showed that experimental area had loamy texture with low salinity and were slightly alkaline pH (7.9) with low organic matter (around 1%). Potassium content was found very high (41.3 ppm) and phosphorus and iron concentrations were lower than usual, 2.4 and 3.12 ppm respectively. Study area also contained adequate amounts of Copper (1 ppm), Manganese (2.36 ppm) and Zinc (4.08 ppm).

All cultivar plots were conducted with 500 plants m^{-2} density in 6 m^2 plots with 8 rows each year. 0.77 kg ha^{-1} pure N fertilization is applied with two splits, one with sowing (0.27 kg ha^{-1}) and another with the beginning of tillering stage (0.43 kg ha^{-1}) in ammonium nitrate form. Phosphor fertilization is applied in Di ammonium phosphate form along with sowing, at 0.69 kg ha^{-1} P_2O_5. Weed control is maintained by hand.

Plants within plots were continuously sampled with one week intervals beginning from seedling emergence for 27 weeks in each years of experiment. Each sampling consisted of 5 randomly chosen triticale plants to be collected and segregated into roots, stems, leaves, (spikes also are categorized after ear emergence) each week. Fresh plant materials were put in drying oven at 60°C for 48 hours (JONES, 1981), providing enough time for plant samples to reach a constant value, then dry weights are measured in laboratory. Dry weight averages of five samples for one sampling date were used as actual growth data. Plant were sampling from seedling to the Zadoks 90 scale in both growing season.

Average temperature and precipitation data of Çanakkale suggested a warm and temperate climate. It is classified as Mediterranean hot climate (Csa) in Köppen Gauger climate classification as mentioned by Peel et al. (2007). Some meteorolojical data, for both growing seasons of Çanakkale, are available in Hocaoglu and Coskun (2018).

Cox Regression Model

The dependent (or described) variable examined in a follow-up study is positive-defined and cumulative survival data (y). The survival function (Survival function, S (y)) is the probability that the data accumulated in an experimental unit until the occurrence of a particular event is greater than the observed value of the expected value. This is also called the Survival curve (Cantor, 2003). The independent variable/s (or covariate/s) are numerical or categorical variables that influence this variable. If there is no assumption about the distribution of the dependent variable and there is a proportional effect between the independent variables, the regression model used to reveal the cause-effect relation between the dependent variable and the independent variables is called Cox regression (or proportional hazard regression) (COX and OAKES, 1984).

Hazard function (rate) is used in this model. The Hazard function is the probability (risk) of a particular event occurring at a given time. In this method, proportional variable regression model was developed based on the assumption that the covariate values of hazard ratios are log-linear functions. In this model there is a proportional effect between covariates. That is, the ratio of hazard functions of two experimental units with different sets of covariates is not time dependent. Therefore, the method can be applied to other study data as well as the survival data (HOSMER *et al*., 2008).

When z_i is the covariate vector of $\boldsymbol{z}$ and the dependent variable is y, the hazard function can be taken as $h(y;\boldsymbol{z})$ according to the covariates of z_i. Accordingly, proportional hazard regression model according to the covariates of $\boldsymbol{z}$ is written as;

$$h(y;\boldsymbol{z}) = h_0(y)exp(\beta'\boldsymbol{z}) \quad (1)$$

In this model β is the regression coefficient vector and when **z**=0, $h_0(y)$ represents the baseline hazard function. When $X_1, X_2, \ldots, X_p$ are covariates, Cox regression model is written as follows;

$$h(t) = [h_0(y)]e^{(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p)} \quad (2)$$

Let the k observation value in which the specific event occurs be sorted in the form of $y_1 < y_2 < \cdots < y_k$. The general hazard rate to be determined by taking into account all variables that will affect this variable is determined in the form of hazard ratio. The multiplication of *k* different values gives the partial similarity function and the regression coefficients are estimated with this. When m_i is the observation number of y_i, the maximum likelihood estimation of β coefficients is calculated as follows;

$$L(\beta) = \{\prod_{i=1}^{k} exp(\beta' z_i)/[\sum_{j\in R} exp(\beta' z_i)]^{m_i}\} \quad (3)$$

The hypothesis required for testing β coefficients' significance are as follows;

$H_0: \beta = 0$

$$H_1: \beta \neq 0 \quad (4)$$

These hypotheses are generally tested with Wald and Likelihood Ratio (LRT) tests.

The Wald statistics are obtained by the ratio of the regression coefficient to the standard error and hence it shows the standard normal distribution. This statistic can also be calculated as

$$W = z^2 = (\beta/SE_\beta)^2 \quad (5)$$

and each can be expressed as a 1-degree of freedom chi-square distribution.

Likelihood Ratio test is a more general test than Wald test. When I_0 is the greatest likelihood statistics of the model and I_v represents the v different regression number included in the model, if $\beta = 0$, likelihood ratio (LR) is calculated as follows;

$LR = -2log(I_0/I_v)$

$$LR = -2(L_0/L_v) \quad (6)$$

In the Cox regression model, variables whose $\beta = 0$ have no effect and H_0 hypothesis cannot be rejected. Otherwise, it is rejected and the effect of the variable on the model is explained with Exp(β)value (COLLET, 2015).

The analyzes were performed with NCSS 11 and SPSS 20 programs.

RESULTS AND DISCUSSIONS

In all models, Group dummy variable was used to explain the effect of grain species on models in accordance with Cox regression method. According to this process, the effects of bread wheat on the models were kept constant and dummies were assigned to rye, durum wheat and barley in Group1, Group2 and Group3 respectively. Accordingly, the following models were created.

Cox Regression Model for the2012-2013 Growing Season

Model-1: Estimated Cox Regression Model

$$\boldsymbol{h}(\boldsymbol{y}) = \boldsymbol{h}_0(\boldsymbol{y})\boldsymbol{exp}(-0{,}314 * \text{Temperature} + 0{,}009 * \text{Precipitation} - 0{,}305 * \text{Group1} - 0.097 * \text{Group2} - 1.167 * \text{Group3})$$

Table 1. Model-1 Results

Variable	β	SE	Wald	df	p	Exp(β)	Decision
Temperature	-0.314	0.016	396.242**	1	0.000	0.731	Significant
Precipitation	0.009	0.002	19.601**	1	0.000	1.009	Significant
Group			85.640**	3	0.000		
Group1	-0.305	0.124	6.043*	1	0.014	0.737	Significant
Group2	-0.097	0.122	0.629	1	0.428	0.908	Not Significant
Group3	-1.167	0.136	73.103**	1	0.000	0.311	Significant
Full Model	-2LogLik Value=5002.725; Chi-Square Value=720.955; df=5; p-value=0.000						Model important

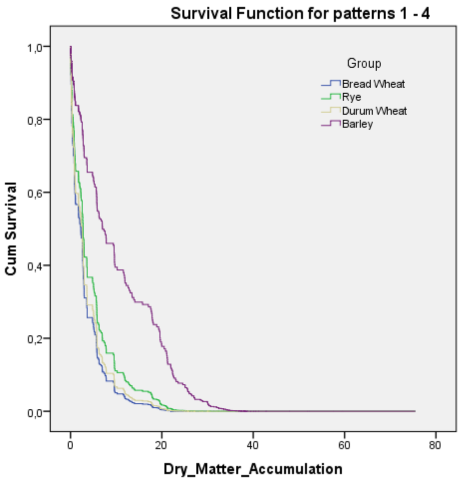

Figure 1. Dry Matter Accumulation curves of the 2012-2013 growing season of four cereal groups

According to Wald statistics (χ^2values) and p values of Table 1, variables other than Group2 are significant. In other words, these five variables together compose the model and the p value of the full model's chi-square (χ^2) test result is seen to be quite significant (<0.000). Accordingly, the precipitation is positive and the temperature has a negative β coefficient. So in this model, the amount of dry matter increased as the temperature increased while the precipitation increased. As a result of comparison of cereal varieties, it is concluded that the amount of dry matter accumulation in bread-wheat is significantly different from rye and barley but it is not different from durum wheat. When the survival curves in Figure 1 are examined, the amount of dry matter accumulation in the cereals from more to less can be sort as: barley, rye, durum wheat and bread wheat.

Cox Regression Model for the 2013-2014 Growing Season
Model-2: Estimated Cox Regression Model

$$h(y) = h_0(y)exp(-0{,}329 * \text{Temperature} - 0{,}025 * \text{Precipitation} + 0{,}483 * \text{Group1} + 0.265 * \text{Group2} - 0.612 * \text{Group3})$$

Table 2. Model-2 Results

Variable	β	SE	Wald	df	p	Exp(β)	Decision
Temperature	-0.329	0.016	447.979**	1	0.000	0.719	Significant
Precipitation	-0.025	0.003	52.809**	1	0.000	0.969	Significant
Group			60.218**	3	0.000		
Group1	0.483	0.128	14.199**	1	0.000	1.621	Significant
Group2	0.265	0.128	4.277*	1	0.039	1.014	Significant
Group3	-0.612	0.141	18.705**	1	0.000	0.411	Significant
Full Model	-2LogLik Value=5099.794; Chi-Square Value=623.231**; df=5; p-value=0.000						Model important

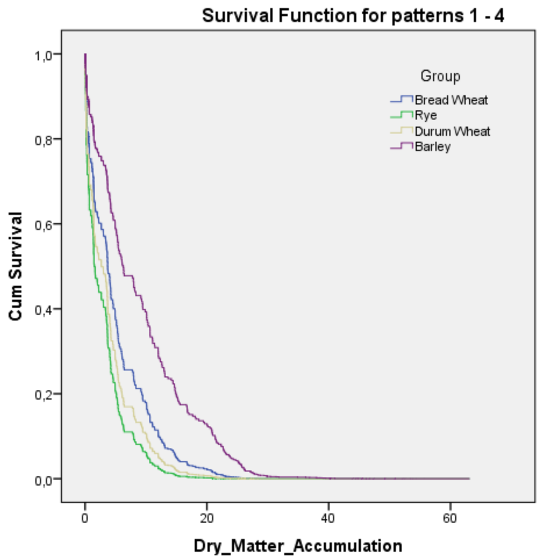

Figure 2. Dry Matter Accumulation curves of the four cool season cereals for the growing season 2013-2014

According to Wald statistics(χ^2) values and p values incident to this in Table 2, it is seen that all variables are significant. In other words, these five variables together compose the model and when the p-value for the chi-square (χ^2) test result of the full model is considered, it is seen that the model is significant (< 0.000). Accordingly, both precipitation and temperature have negative β coefficients. In this model, both the precipitation and temperature increased while the amount of dry matter decreased. As a result of comparison of cereal varieties, it is concluded that dry matter accumulation amounts of rye, durum wheat and barley are significantly different from bread wheat. When the survival curves in Figure 2 are examined, the amount of dry matter accumulation in the cereals, from more to less, can be sort as; barley, bread wheat, durum wheat and rye.

Cox Regression Model for Data in the2012-2013 and 2013-2014 Growing Seasons
Model-3: Estimated Cox Regression Model

$$h(y) = h_0(y)exp(-0{,}329 * \text{Temperature} - 0{,}002 * \text{Precipitation} + 0{,}091 * \text{Group1} - 0.953 * \text{Group2} + 0.295 * \text{Group3} + 0.295 * \text{Growing Season})$$

Table 3. Model-3 Results

Variable	β	SE	Wald	df	p	Exp(β)	Decision
Temperature	-0.329	0.011	891.905**	1	0.000	0.729	Significant
Precipitation	-0.002	0.002	1.101	1	0.294	0.998	Not Significant
Group			144.327**	3	0.000		
Group1	0.092	0.087	1.115	1	0.291	1.097	Not Significant
Group2	0.091	0.088	1.078	1	0.299	1.095	Not Significant
Group3	-0.953	0.099	92.838**	1	0.000	0.385	Significant
Growing Season	0.295	0.065	20.675**	1	0.000	1.183	Significant
Full Model	-2LogLik Value=11682.773; Chi-Square Value=1253.037**; df=6; p-value=0.000						Model important

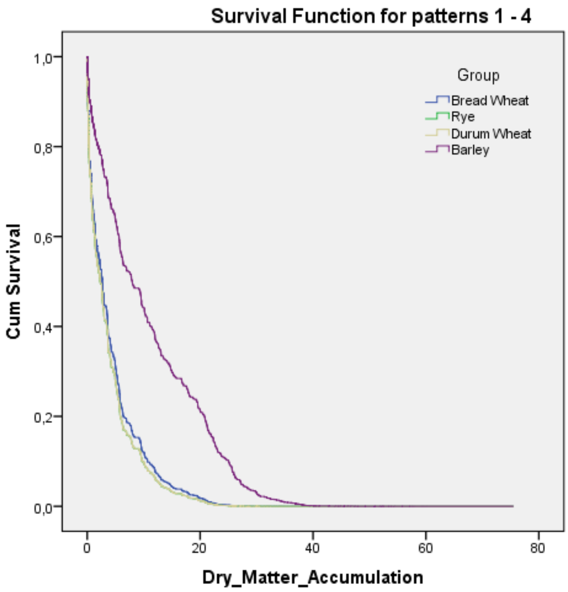

Figure 3. Dry Matter Accumulation curves of the four cereal groups season across growing seasons

With the addition of the growing season factor, Model-3 was created by using data of both seasons. Wald statistics ((χ^2) values) and p values in Table 2, all other variables except precipitation, Group1 and Group2 seem to be significant. Although the p-value of the precipitation variable was not significant (p = 0.294), it was observed that the precipitation factor was as effective as the previous models and the dry matter accumulation was increased by 1 (0.998) times when hazard rate Exp(β) was taken into consideration. Therefore, it should be included in the model. In other words, these five variables together compose the model and the p value of the full model's chi-square (χ^2) test result is seen to be quite significant(< 0.000). Temperature and precipitation variables have negative β coefficients, so that the amount of dry matter decreased as temperature and precipitation increased. As a result of comparison of cereal varieties, it is concluded that the amount of dry matter accumulation in bread wheat is significantly different from barley while not different from rye and durum wheat. When we look at the last line of Table 3 it is seen that the effect of growing season factor on the model is quite significant. That is, the accumulation of dry matter in cereals varies significantly from season to season. When the survival curves given in Figure 3 are examined, it is seen that the highest dry matter accumulation among cereals is in barley and bread wheat, respectively. Rye and durum wheat are not different from each other and have the least dry matter accumulation.

CONCLUSIONS

It was seen that all three models produced a significant result. The reason for this can be shown as the growing season factor in the third model. The hazard ratio of this factor is Exp(β)=1.183. In other words, this factor affected dry matter accumulation more than 1 time. In this case, it is concluded that dry matter accumulation in cereals should be modeled separately for each growing season.

In all three models, the temperature factor was found to be significant and had a negative β coefficient. Accordingly, it was concluded that dry matter accumulation increased as the temperature decreased.

The precipitation factor has positive β coefficients in the first model and negative in other models. Therefore, the dry matter accumulation increasing or decreasing effect varies according to the growing seasons. In the third model, this factor was not significant. However, because the $Exp(\beta)$ value is high (0.998), it is not appropriate to remove.

The effect of the Group factor on the models is significant. The amount of dry matter accumulation in cereal types is also important for all of the models. In all models, by keeping bread wheat dry matter constant, it was compared to the rye (Group1), durum wheat (Grup2) and barley (Grup3). In all models, the highest dry matter accumulation was found in barley. The order between the other cereals varies from model to model.

Based on these results, it can be said that the Cox Regression method is an alternative for modeling similar studies.

This study can be continued in two ways: 1. The effect of precipitation factor on dry matter accumulation can be generalized by the creation of models of different seasons, the dry matter accumulation quantities (pace) of rye, durum wheat and barley can be sorted. 2. With the addition of different cereals, soil and fertilizer factors, the growth performance of cereals can be studied extensively.

Received, November 12th, 2018
Accepted February 18th, 2019

REFERENCES

BOLIN, J.H. (2014): Hayes, Andrew F. 2013. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression Based Approach. New York, NY: The Guilford Press. J. Educational Measurement, *51*(3), 335-337.

BREIMAN, L. (2017): Classification and regression trees. Routledge.

CANTOR, A. (2003): SAS Survival Analysis Techniques for Medical Research (Second edition). SAS Institute, 248, North Carolina.

COLLETT, D. (2015): Modelling Survival Data in Medical Research. Chapman and Hall/CRC, 548, Florida.

DAVIDIAN, M. (2017): Nonlinear models for repeated measurement data. Routledge.

COX, D.R. (1992): Regression models and life-tables. In: Breakthroughs in statistics (pp. 527-541). Springer, New York, NY.

COX, D.R. (2018): Analysis of survival data. Routledge.

COX, D.R., D., OAKES (1984): Analysis of Survival Data. Chapman and Hall/CRC, 114, New York.

Damgaard, C., Weiner, J.,Nagashima, H. 2002. Modeling individual growth and competition in plant populations: growth curves of Chenopodium album at two densities. J. Ecology, *90*:666-671.

DOUGLAS, M.B., W.G., DONALD (1988): Non linear regression and its applications. John Wiley & Sons Inc. Canada.

FARAWAY, J.J. (2016): Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models (Vol. 124). CRC press.

HOCAOGLU, O., Y., COSKUN (2018): Evaluation of Dry Matter Accumulation in Triticale by Different Sigmodial Growth Models in West Anatolia of Turkey, Genetika, *50*(2):561-574.

HOSMER, D.W., S., LEMESHOW, S., MAY (2008): Applied Survival Analysis: Regression Modeling of Time-to-Event Data (Second Edition). John Wiley & Sons, 67-91, United States.

JONES, D.I.H. (1981): Chemical composition and nutritive value. In: Handson, J., Baker, R.D., Davies, A., Laidlows, A.S., Leawer, J.D. (eds.), Sward Measurement handbook. The British grassland society, Berkshire, UK, pp: 243–265.

KARADAVUT, U., A., GENÇ, A., SINAN, A., KARAKOCA, Ç., PALTA, Ş., AKSOYAK (2005): Determination of growth curves in Dagdas wheat plants. Statistical Researches Symposium 9-13 May, 2006. Belek-Antalya.

KARADAVUT, U. (2009): Path analysis for yield and yield components in lentil (Lens culinaris Medik.). Turk. J. Field Crops, *14*(2): 97-104.

KAYA, Y., M., AKÇURA, S., TANER (2006): GGE-biplot analysis of multi-environment yield trials in bread wheat. Turk. J. Agric. Forestry, *30*(5): 325-337.

PEEL, M.C., B.L., FINLAYSON, T.A., MCMAHON (2007): Updated world map of the koppen-geiger climate classification. Hydrology and earth system sciences discussions, European geosciences union, *4* (2): 439-473.

ALTERNATIVNI STATISTIČKI MODEL ZA PROCENU AKUMULACIJE SUVE MATERIJE U HLADNIM SEZONIMA: COX REGRESIJA

Zeynep GÖKKUŞ[1*], Mevlüt AKÇURA[2]

[1]Kastamonu Univerzitet, İnebolu Vocational Škola, Departman za kompjuterske tehnologije, Kastamonu, Turska
[2]Canakkale Onsekiz Mart Univerzitet, Poljoprivredni fakultet, Departman za ratarstvo, Canakkale, Turska

Izvod

U ovom radu, rađen je model na osnovu porasta nekih žitarica (hlebne pšenice, raži, durum pšenice i ječma) u toku dve hladne sezone. U tu svrhu predložena je Cox regresija kao alternativa preferiranim metodama regresije u prethodnim istraživanjima. U studiji, na osnovu sezonskih podataka za dve vegetacione sezone (2012-2013, 2013-2014 i obe), od kojih svaka ima 5 ponavljanja 27 uzoraka, stope rasta ovih žitarica kroz količine akumulacije suve materije su objašnjene u tri različitih modela. U tu svrhu, količine akumulirane suve materije su prilagođene podacima o preživljavanju i Cox regresionoj metodi, koja je bila je poželjna, jer koristi funkciju hazarda i brzinu pojavljivanja određenog događaja. Rezultat je bio da je svaki model bio veoma značajan ($p < 0.000$). Utvrđeno je da je i) najbrže rastuća vrsta bio ječam, ii) da se akumulacija suve materije smanjuje kako se temperatura povećava, i iii) akumulacija suve materije u usevima se menja tokom svake sezone rasta.

Primljeno 28.XI.2018.
Odobreno 18. II. 2019.