ISFOR 2017 International Symposium on New Horizons in Forestry
18-20 October 2017 | Isparta - Turkey

Oral presentation

Comparison of some autoregressive models for removing temporal autocorrelation effect in stem analysis data

Ferhat Bolat[1,*], İlker Ercanlı[1], Muammer Şenyurt[1]

[1]Çankırı Karatekin University, Faculty of Forestry, Çankırı, Turkey
[*] Corresponding author: fbolat@karatekin.edu.tr

Abstract: Some data characteristics (such as cause of heteroscedasticity or autocorrelation problems) can complicate fairly understanding the forest growth behavior. As the stand conditions can change over time, the measurements can be more similar each other in same time period than to measurements from other time periods. Therefore, the data obtained from the forest inventory are temporally dependent, which is called "temporal autocorrelation problem". Nonlinear least square (NLS) regression developed by using diameter increment values ,originating from stem analysis, leads to biased estimates of parameter confidence interval and prediction interval in forest growth models when NLS regression are applied without removing temporal autocorrelation effect. Therefore, this study focus on solving "temporal autocorrelation problem" caused by multiple measurements from individual trees. To overcome this problem, we compared the Negative Exponential, the Gompertz and the Logistic nonlinear models accounting for p-order autoregressive process. The results showed that the AR2 process based on the Gompertz model contributed considerably for increasing the adjusted determination coefficient (0.6545 to 0.9630) and removed the temporal autocorrelation effect (Durbin-Watson test value was 2.0201). AR2 process, also, based on the Gompertz model produced a considerable decreasing for the root mean squared error (69.7048 to 22.8368). The autoregressive model proved that the forest managers may confidently use the diameter increment values obtained from multiple measurements over time in forest ecosystems.
Keywords: Diameter increment, Nonlinear regression, Time series, Autocorrelation, Tree rings

1. Introduction

As developing forest growth models, the forest biometricians should pay more attention to modeling techniques, data characteristics, and statistical assumptions. These are crucial factors affecting model performance and limiting model applicability (Soares, et al. 1995; Thürig, et al. 2005)

In literature, the forest growth models have been developed using regression analysis. Before these models are fitted to forest inventory data, the some statistical assumptions should be ensured, such as the normal distribution of data, the homogeneity of variance, and the independence of observations (Zuur, et al. 2010). If this assumptions do not being considered cautiously, the forest growth model most likely will be inadequate for fitting to real data in a model. The data derived from stem analysis have a temporally hierarchal structure, meaning that the repeated measurements on the same individuals result in correlated data. This leads to violation of independence assumption among the observations. This data structure causes the biased parameter estimations, unexpected results, and consequently misunderstanding inferences (Zuur, et al. 2010; Paine, et al. 2012). Autoregressive modeling approach used for specifying the temporal dependence has been used widely in time-series analysis, and recently it has often been applied in the forest growth studies (Monserud 1986; Huang and Titus 1999; Fox, et al. 2001; Zhao, et al. 2013; Saud, et al. 2016; Kiaei, et al. 2017)

The diameter growth trend may changes over time depending on the plant inner characteristic, the resource availability, and the competition among species (Paine, et al. 2012). Thereby, the diameter growth shows a sigmoid curve, having an inflection point and reaching asymptote after a given time. Nonlinear models are asymptotic and concave-down, allowing to capture rapid increment at the earlier ages and slow increment at the older ages, and consequently to ensure the biological growth trend. Recently, therefore, the forest biometricians prefer to nonlinear models for fitting to diameter data as a function of age or height (Fekedulegn, et al. 1999; Bi, et al. 2012).

In this study, the first purpose is to develop some the nonlinear models as a function of age in order to predict diameter increments at different ages. The second purpose is to remove the temporal autocorrelation effects in stem analysis data by using the first-order and the second-order autoregressive models.

2. Material and methods

The data used in this study were collected from even-aged Scotch pine (*Pinus sylvestris* L.) stands located in the Çankırı, Yapraklı and Yenice (Ilgaz) Planning Unit, Ankara Forest District Directorate, Turkey. In this study, dominant or co-dominat trees being as the 100 trees of greatest height per hectare were sampled for stem analysis in the sample plots. These 106 sampling trees for stem analysis were felled, and also the cross-sectional cuts were made at the first 0.3 m and every 2 m throughout the tree stem. On each tree the annual rings were counted at 0.3 (m) and the diameter (mm) at different ages were derived from stem analysis measurements. Summary statistics of stem analysis data were shown in Table 1.

Table 1. Summary statistics of the diameters (mm) at 0.3m

Min	Max	Arithmetic mean	Std. Deviation
2.0	1250.0	172.2	120.7

A number of statistical growth functions have been used to model the diameter increment - age relationship in forest literature. In this study, the different three nonlinear models were fitted to stem analysis data using SAS PROC Model procedure. The nonlinear models considered in this study were presented in Table 2.

Table 2. The nonlinear growth models used in this study

Model name	Functional form
Negative Exponential	$d = \beta_0(1 - \exp(-\beta_1 t)) + \varepsilon$
Gompertz	$d = \beta_0 \exp(-\beta_1 \exp(-\beta_2 t)) + \varepsilon$
Logistic	$d = \frac{\beta_0}{1 + \beta_1 \exp(-\beta_2 t)} + \varepsilon$

β_0, β_1 and β_3 are the model coefficients, t is age (year), d is diameter (mm)

To remove temporal autocorrelation effects in the diameter increment prediction, we used the first-order (AR1) and the second-order (AR2) autoregressive models using SAS software (Appendix A).

$$d(t) = c + \phi_1 d_{t-1} + \varepsilon_t \quad \text{(AR1)}$$

$$d(t) = c + \phi_1 d_{t-1} + \phi_2 d_{t-2} + \varepsilon_t \quad \text{(AR2)}$$

$$-1 < \phi < +1$$
$$\varepsilon_t \backsim N(0, \sigma^2)$$

Where: d (the diameters at different ages) and t (age) are observations, c and ϕ are determined by nonlinear least square, ε is random with mean zero and serially independent.

Model comparison was carried out based on the root mean squared error (*RMSE*) and the adjusted determination coefficient (R^2_{adj}) in order to identify the best model. Durbin-Watson (DW) test was used for quantifying temporal autocorrelation existing among the error terms. The effects of autoregressive process were evaluated using residual graphics of the best model selected based on model selection criteria.

3. Result and discussion

We developed nonlinear regression models including the Negative Exponential, the Gompertz and the Logistic models and predicted the parameters. Parameter prediction and its significance (at the 5% probability level), model performance criteria and Durbin-Watson statistic results were given in Table 3. For all the models used in this study, the parameter predictions were found to be significant at the 0.05 probability level (P < 0.05). The DW test results showed that there was positive autocorrelation (0 < DW < 2) among the error terms for all used nonlinear models. The best predictive model was the Negative Exponential model with the second autoregressive process (AR2) having adjusted determination coefficient of 0.9637 and the root mean squared error of 22.5845 mm in terms of the model selection criteria. However, Fekedulegn, et al. (1999) reported that the Negative exponential model was not adequate for describing the biological growth trend because it has not inflection points and is not sigmoid shaped. Thereby, we decided to perform the Gompertz model for describing the diameter increment, which its adjusted determination coefficient was 0.9630 and the root mean squared error value was 22.8368 mm. Tjørve and Tjørve (2010) corroborated the decision, the Richards-model family such as the Gompertz, and the logistic produced a satisfactory result for the organisms (plant or animal) which exhibit a sigmoid growth. Tewari and Kumar (2005) found that the Gompertz model was the proper for diameter prediction in Dalbergia Sissoo plantations.

Table 3. The parameter estimations of the models and model selection criteria results

Model	Parameters	RMSE	R^2_{adj}	Durbin-Watson	Pr<DW	Pr>DW
Negative Exponential		69.2303	0.6592	0.1106	<0.0001	1.0000
	β_0=1819.3150					
	β_1=0.0016					
Negative Exponential (AR2)		22.5845	0.9637	2.0190	0.6755	0.3245
	β_0=1365.5020					
	β_1=0.0023					
	ϕ_1 =1.0345					
	ϕ_2=-0.0932					

β_0, β_1 and β_2 are the parameter coefficients
Pr<DW was less than 0.05, meaning that there is the autocorrelation among the error terms
DW= 2 means no autocorrelation
0 < DW < 2 means positive autocorrelation
2 < DW < 4 means is negative autocorrelation

Table 3 (Continue). The parameter estimations of the models and model selection criteria results

Model	Parameters	RMSE	R^2_{adj}	Durbin-Watson	Pr<DW	Pr>DW
Gompertz		69.7048	0.6545	0.1170	<0.0001	1.0000
	β_0=444.3968					
	β_1=2.9328					
	β_2=0.0185					
Gompertz (AR2)		22.8368	0.9630	2.0201	0.6797	0.3203
	β_0=511.4391					
	β_1=2.7159					
	β_2=0.0170					
	ϕ_1 =1.0481					
	ϕ_2=-0.1009					
Logistic		70.2359	0.6492	0.1242	<0.0001	1.0000
	β_0=376.5500					
	β_1=9.8121					
	β_2=0.0340					
Logistic (AR2)		23.0781	0.9621	2.0170	0.6527	0.3473
	β_0=480.1943					
	β_1=6.8316					
	β_2=0.0280					
	ϕ_1 =1.0597					
	ϕ_2=-0.1000					

β_0, β_1 and β_2 are the parameter coefficients
Pr<DW was less than 0.05, meaning that there is the autocorrelation among the error terms
DW= 2 means no autocorrelation
0 < DW < 2 means positive autocorrelation
2 < DW < 4 means is negative autocorrelation

The temporal autocorrelation impacts greatly the regression model results in terms of the adjusted coefficient of determination (R^2_{adj}) and the root mean squared error (RMSE). In our study, we found that R^2_{adj} was approximately 0.70 for all the considered models when the autocorrelation effects did not remove (table 3). For increasing the models performance, firstly, the first-order autoregressive process (AR1) based on the Gompertz model was applied to remove the positive autocorrelation effects, but this attempt was failed to remove the autocorrelation effects whereas the R^2_{adj} increased considerably (∽45%) and all the parameters were statistically significant (at 5% level, table 3). AR2 process based on the

Gompertz model provided both the solution to eliminate the autocorrelation effects and a significant increase for R^2_{adj} (∽45%). Also, the AR2 process contributed greatly to being reduced RMSE (∽65%, table 3).

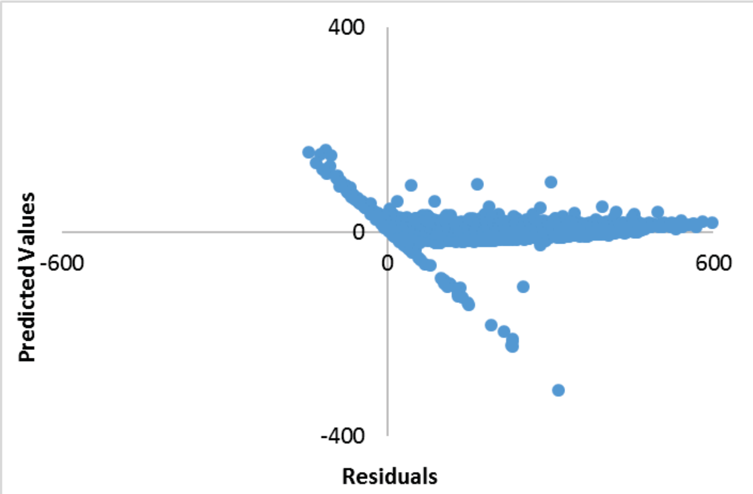

Figure 1. The residual distributions of Gompertz model

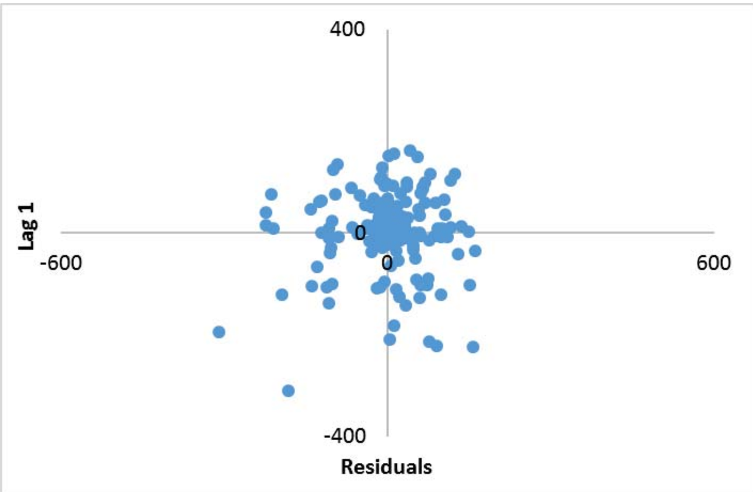

Figure 2. The residual distributions of Gompertz model based on AR1

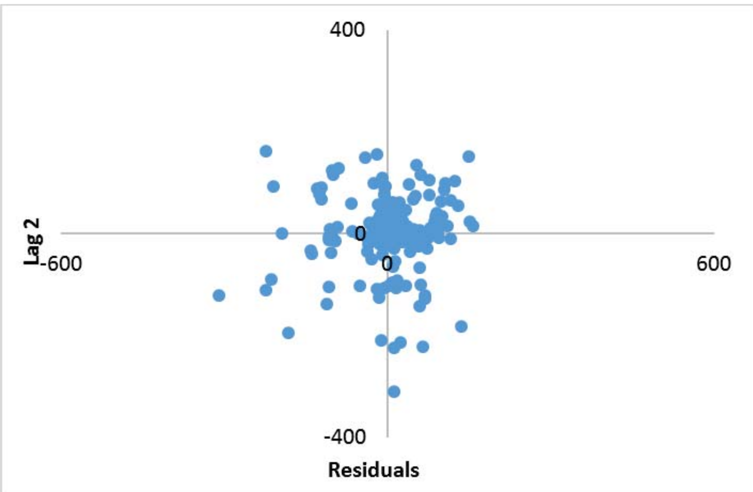

Figure 3. The residual distributions of Gompertz model based on AR2

Zuur, et al. (2010) reported that the parametric statistical analysis assume to be the independent of residuals, but this assumption is often violated in forest populations. Theoretically, if the observations are independent, the model residuals are randomly scattered around zero. In our study, the Gompertz model residuals exhibited a dispersed distribution because the observations were highly correlated (figure 1, table 3). The AR1 process produced a normally residual distribution, which is close to zero, but it could not eliminate the temporal autocorrelation in diameter increments (table 3, figure 2). AR2 process showed a scattered residual distribution about zero (figure 3). The Durbin-Watson (DW) test for AR2 process exhibited to be non-autocorrelation in the residuals (table 3). Monserud (1986) reported that AR1 process was adequately not account for the error variance as it was expected and the higher-order autoregressive process (e.g. AR2, AR3) was better choice in order to removing temporal autocorrelation effects in tree-ring chronologies. In contrast Biging and Gill (1997) reported that AR1 process provided an adequate solution to removal of temporal autocorrelation effects and also to ensured assumption relevant to homogeneity of error variance for describing the tree crown profile. Huang and Titus (1999) found that whereas the nonlinear models including AR1 process contributed slightly a decrease of RMSE, the autoregressive process provided the normal distribution of error terms for tree height predictions.

Consequently, we emphasized that the violation of statistical assumptions had a great effect on nonlinear model fitting to stem analysis data in terms of error variance and model performance. We demonstrated the AR2 process was capable of removing temporal autocorrelation effects in stem analysis data. In next studies relevant to the temporal autocorrelation, forest biometricians should consider the autoregressive modeling techniques.

References

Bi H, Fox JC, Li Y, Lei Y, Pang Y (2012). Evaluation of nonlinear equations for predicting diameter from tree height. Canadian Journal of Forest Research 42:789-806

Biging GS, Gill SJ (1997). Stochastic models for conifer tree crown profiles. Forest Science 43:25-34

Fekedulegn D, Mac Siúrtáin MP, Colbert JJ (1999). Parameter Estimation of Nonlinear Models in Forestry.

Fox JC, Ades PK, Bi H (2001). Stochastic structure and individual-tree growth models. Forest Ecology and Management 154:261-276

Huang S, Titus SJ (1999). An individual tree height increment model for mixed white spruce–aspen stands in Alberta, Canada. Forest ecology and management 123:41-53

Kiaei M, Mortazavi SJ, Veylaki M (2017). The Semi-parametric Prediction in Nonlinear-autoregressive Model for Annual Ring Width of Pinus eldarica. CHIANG MAI JOURNAL OF SCIENCE 44:715-720

Monserud RA (1986). Time-series analyses of tree-ring chronologies. Forest Science 32:349-372

Paine C, Marthews TR, Vogt DR, Purves D, Rees M, Hector A, Turnbull LA (2012). How to fit nonlinear plant growth models and calculate growth rates: an update for ecologists. Methods Ecol Evol 3:245-256

Saud P, Lynch TB, KC A, Guldin JM (2016). Using quadratic mean diameter and relative spacing index to enhance height–diameter and crown ratio models fitted to longitudinal data. Forestry 89:215-229

Soares P, Tomé M, Skovsgaard J, Vanclay JK (1995). Evaluating a growth model for forest management using continuous forest inventory data. Forest Ecology and Management 71:251-265

Tewari V, Kumar V (2005). Growth and yield functions for Dalbergia sissoo plantations in the hot desert of India grown under irrigated conditions. J Trop for Sci 87-103

Thürig E, Kaufmann E, Frisullo R, Bugmannc H (2005). Evaluation of the growth function of an empirical forest scenario model. Forest Ecology and Management 204:53-68

Tjørve E, Tjørve KM (2010). A unified approach to the Richards-model family for use in growth analyses: why we need only two model forms. J Theor Biol 267:417-425

Zhao L, Li C, Tang S (2013). Individual-tree diameter growth model for fir plantations based on multi-level linear mixed effects models across southeast China. Journal of forest research 18:305-315

Zuur AF, Ieno EN, Elphick CS (2010). A protocol for data exploration to avoid common statistical problems. Methods Ecol Evol 1:3-14

Appendix A.

The SAS software interface for the Gompertz model used in this study is given below.

```
data isparta;                    (Write your worksheet name)
input t d;                       (Write your dependent and independent name)
cards;

5    7
15   20
20   30                          (Write your own dependent and independent values)
.
.
.
;
run;

PROC Model data=isparta;

 parameters  b0=380.2443 b1=2.966557 b2=0.022082;     (You should determine proper starting values)

d=b0*exp(-b1*exp(-b2*t));            (Write equation)

%ar (d, 2;                           (This means the second-order autoregressive process)

fit d/dw dwprob;

run;
```

NOTE: If you want to print the all predictions and the all residuals, you must add codes expressed as italic;

fit artim/dw dwprob *out=resid outall* ;

run;

proc print data=resid;

run;