Proceeding Paper

A Proposal for a Hybrid Model Based on the Weibull Growth Equation in the Adjustment of Growth Curves applied to Pine Forest Species in Northern Mexico [†]

Marcos de Almeida Souza [1], Miguel Ricardo Hernández-Velazco [1], João Marcelo Brazão-Protázio [2,*], Jonathan Gabriel Escobar-Flores [3], Carlos Antonio López-Sánchez [4], Jose Ciro Hernández-Díaz [5] and Christian Wehenkel [5]

[1] Maestría Institucional en Ciencias Agropecuarias y Forestales, UJED, Durango 34120, Mexico, marcossouza2705@gmail.com (M.d.A.S.); mrhv2@outlook.com (M.R.H.-V.)
[2] Postgraduate Program in Mathematics and Statistics, UFPA, Belém 66810-010, Brazil
[3] Centro Interdisciplinario de Investigación para el Desarrollo Integral Regional, Unidad Durango, Instituto Politécnico Nacional, Durango 34220, Mexico; jonas286@hotmail.com
[4] GIS-Forest Group, Departamento of Organisms and Systems Biology, University of Oviedo, 33071 Oviedo, Spain, lopezscarlos@uniovi.es
[5] Instituto de Silvicultura e Industria de la Madera, UJED, Durango 34120, Mexico; jciroh@ujed.mx (J.C.H.-D.); wehenkel@ujed.mx (C.W.)
[*] Correspondence: mprotazio@gmail.com
[†] Presented at the 1st International Electronic Conference on Forests—Forests for a Better Future: Sustainability, Innovation, Interdisciplinarity, 15–30 November 2020; Available online: https://iecf2020.sciforum.net.

Abstract: In this work, we present a hybrid methodology based on ADA, applied to the simulation of pine species from northern Mexico. From algebraic transformations and certain premises, we were able to reduce the parameter space of the Weibull growth function from three parameters to just one, the maximum diameter of the individual at the end of growth. From the point of view of complexity, we have a decrease in computational cost, because now we have only one parameter to estimate, and from a practical point of view, it is an easy parameter to obtain, even in the field. To present the quality of the method, we used the mean absolute percentage error and the quantile regression to visually present the quality of the fit. Then we compare this proposed methodology with the GADA versions for the Chapman–Richards and Hossfeld models. The results presented for the study area, to avoid generalizations, show that the proposed hybrid method was more accurate in the estimates.

Keywords: GADA method; growth equations; forest models; nonlinear regression

1. Introduction

For many purposes in forestry, it is helpful to be able to make precise future predictions of the mean values of growth variables based on repeated measurements through time made on units that are grouped hierarchically. Many forest management decisions are based on yield projections that crucially depend on projections of plot level averages of tree diameter and height, basal area, and other dentrometric variables [9].

A stand growth model is an abstraction of the natural dynamics of a forest stand, and may encompass growth, mortality, and other changes in stand composition and structure. Common usage of the term "growth model" generally refers to a system of equations which can predict the growth and yield of a forest stand under a wide variety of conditions [1].

Growth and yield are mathematically related. From the biological point of view, growth is the increase in dimensions of one or more individuals in a forest stand over a given period of time. Yield corresponds to final dimensions at the end of a certain period. In even-aged stands, a growth equation might predict the growth of diameter, basal area, or volume in units per annum as a function of age and other stand characteristics, whereas

a yield equation would predict the diameter, stand basal area, or total volume production attained at a specified age. In an uneven-aged stand, yield is the total production over a given time period, and growth is the rate of production [1].

Whole stand models are often simple and robust, but may involve complexities not possible in other approaches. On the other hand, single-tree models is the most detailed approach which uses the individual tree as the basic unit of modeling. The minimum input required is a list specifying the size of every tree in the stand. Some models also need the spatial position of the tree, or tree height and crown classes. Single-tree models usually are very complex for modeling branches and internal stem characteristics, and may be linked to harvesting and conversion simulators [2]. Therefore, our study aim was to test a new and simple method to model tree growth.

2. Materials and Methods

2.1. Study Area

The study was conducted in the mixed and irregular forests of the Sierra Madre Occidental of Durango (Figure 1), more precisely in the longitudinal gradient between the states of Durango and Sinaloa, from September to November 2017. Six plots, 50 m × 50 m (0.25 ha), were established, where the diameter at breast height (DBH) was measured, growth rings were analyzed to describe the age and diameter increments of each tree based on growth nuclei, the pines were located within the plot, and the species were recorded. For the analysis presented here, we applied the methodology only to the species found in plot 5 (Figure 2), which had the largest number of individuals.
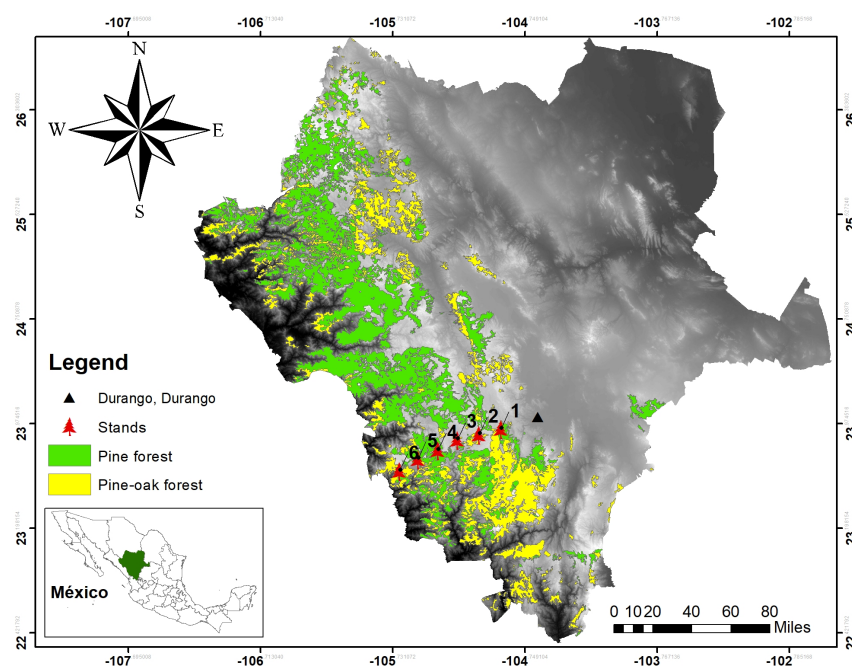

Figure 1. Study area and the locations of the six data collection plots.
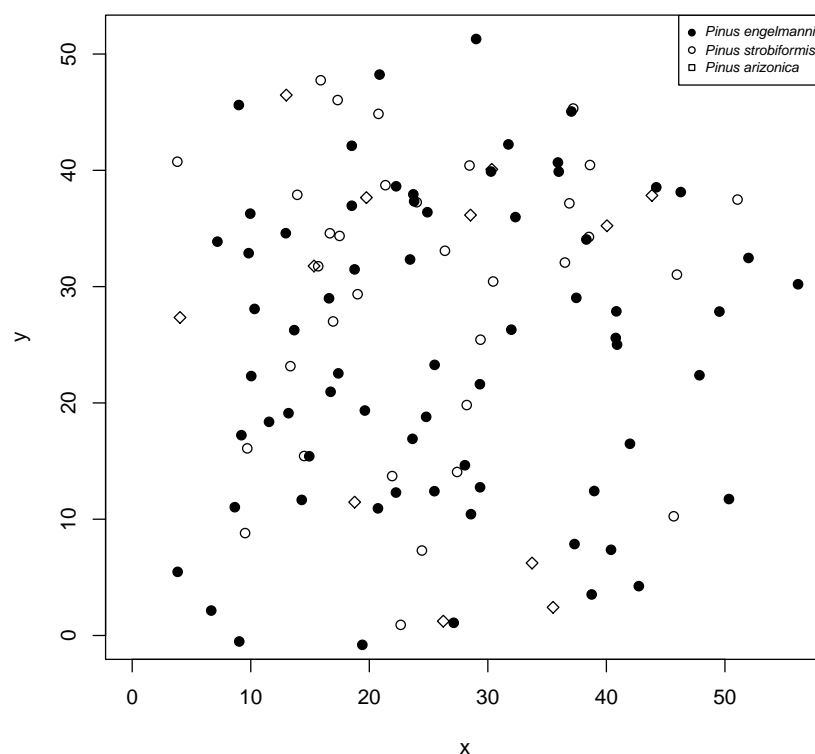
Figure 2. Plot 5 of the study area and location of each tree and species within the plot.

2.2. Base-Aged-Specific Model (BAS Model)

The BAS models implemented in order to estimate and evaluate the respective diametric growth of the pine trees in the study area were taken from the literature [3–5]. In this case, we only used versions that depend on three parameters (Table 1).

Table 1. List of some growth models to be implemented at this work, where θ_1, θ_2, and θ_3 are the model parameters; and d and t are, respectively, the diameter (*dbh*) and the age of the individual.

BAS Model	Function	Reference
Chapman–Richards	$d_1(t) = \theta_1(1 - e^{-\theta_2 t})^{\theta_3}$	[3]
Hossfeld	$d_3(t) = \theta_1(1 + \theta_2 t^{-\theta_3})^{-1}$	[4]
Weibull	$d_4(t) = \theta_1 - \theta_2 e^{-\theta_3 t}$	[5]

2.3. Algebraic Difference Approach (ADA) and Generalized Algebraic Difference Approach (GADA)

The Algebraic Difference Approach (hereinafter ADA) was introduced by Bailey and Clutter [8] and it consists replacing a parameter of the base model with its initial condition solution. The ADA technique allows the derivation of dynamic or non-static functions that are capable of producing anamorphic or polymorphic curves, from a BAS model chosen a priori. The GADA methodology [6] arises when we assume that two or more parameters of the BAS model are simultaneously site-specific.

2.3.1. Chapman–Richards (Hereinafter C-R) GADA Model

The C–R GADA model [6] implemented here is defined as:

$$d_4(t) = d_0 \left(\frac{1 - e^{b_1 t}}{1 - e^{b_1 t_0}} \right)^{b_2 + b_3 / \chi_0} \tag{1}$$

where $\chi_0 = 0.5(\ln d_0 - b_2 l_0 \pm ((\ln d_0 - b_2 l_0)^2 - 4 b_3 l_0)^{0.5}$ and $l_0 = \ln(1 - e^{-b_1 t_0})$.

2.3.2. Hossfeld GADA Model

The Hossfeld GADA Model [6] implemented here is defined as:

$$d_5(t) = \frac{b_1 + \chi_0}{1 + (b_2/\chi_0)t^{-b_3}} \tag{2}$$

where $\chi_0 = 0.5(d_0 - b_1 + ((d_0 - b_1)^2 + 4 b_2 d_0 t_0^{-b_3}))^{0.5}$.

2.4. Proposed Methodology Based on the Algebraic Difference Approach (ADA)

The procedure proposed here consists of first reducing parameters to immediately apply the ADA methodology, assuming in this case that there is any intrinsic relation between the parameters θ_1 and θ_3. In such case, we can rewrite the Weibull BAS model as:

$$d(t) = \theta_1 - (\theta_1 - d_r)e^{-(\theta_3 = f(\theta_1))t}. \tag{3}$$

where θ_1 is the the maximum diameter reached by the individual; $d_r = d(t = 0)$ is the recruitment diameter, which has been considered the same for all individuals. This restructuring of the equation restricts the model to the unique parameter θ_1, of course, only if the relationship $\theta_3 = f(\theta_1)$ exists.

For the population in study, in order to avoid generalization, we observed a strong inverse relationship between the variable θ_1 and θ_3, such that:

$$\theta_3 = \frac{a}{\theta_1}, \tag{4}$$

hence, it is possible rewrite Equation (3) as:

$$d(t) = \theta_1 - (\theta_1 - d_r)e^{-\frac{a}{\theta_1}t}, \tag{5}$$

where a is a scale factor.

Now assuming $a = \chi$ as site-specific at Equation (5) and subject to the initial conditions $d(t_0) = d_0$, we have:

$$\chi = \frac{\theta_1}{t_0} log \left(\frac{d_r - \theta_1}{d_0 - \theta_1} \right) \tag{6}$$

and by replacing this result in Equation (5), we finally have:

$$d_8(t) = \theta_1 - (\theta_1 - d_r) \left(\frac{d_r - \theta_1}{d_0 - \theta_1} \right)^{-t/t_0}, \tag{7}$$

which now depends only on the parameter θ_1.

2.5. Mean Absolute Percentage Error (MAPE)

In order to quantify the error associated with each of the models implemented in our analysis, we use, for its robustness and simplicity, the mean absolute relative error [7], here defined by the equation:

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{o_i - e_i}{o_i} \right|, \tag{8}$$

where e_i and o_i are respectively the estimated and the observed values, and n is the number of observations.

3. Results and Discussion

Here we provide a concise and precise description of our results, their interpretation and the conclusions concerned.

We can see (Figure 3) that it is possible to write θ_3 as a function of θ_1 of the form $\theta_3 = f(\theta_1) = a/\theta_1$. Then, using nonlinear regression methods with the nlsLM function on R [10], we estimated the parameters for each species and model analysed in this study and presented in summary in Table 2.
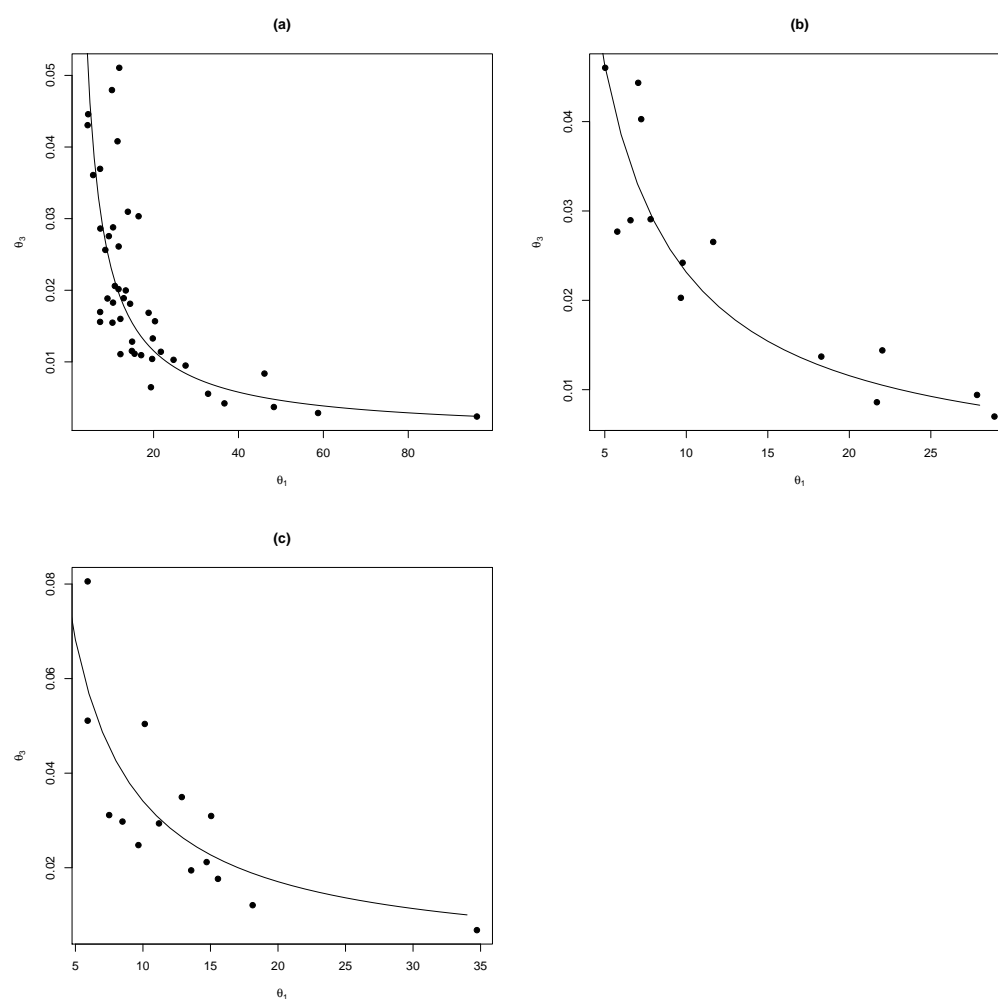

Figure 3. The relationship between the parameters θ_1 and θ_3 adjusted for the model $\theta_3 = a/\theta_1$, for the species (**a**) *Pinus engelmannii* ($a = 0.231$, $p < 0.001$), (**b**) *Pinus strobiformis* ($a = 0.198$, $p < 0.001$), and (**c**) *Pinus arizonica* ($a = 0.204$, $p < 0.001$).

The tests show that the results for the MAPE function calculated for the C-R GADA, Hossfeld GADA and the proposed methodology were equivalent (Figure 4), considering each of the species analysed in this study. The results were: *Pinus engelmannii* (Wilcoxon–Mann–Whitney test, *p*-value = 0.983), *Pinus strobiformis* (Wilcoxon–Mann–Whitney test, *p*-value = 0.983), and *Pinus arizonica* (Wilcoxon–Mann–Whitney test, *p*-value = 0.198). There are small advantages for the proposed methodology for the species *Pinus engelmannii* ($\mu = 0.107$) and *Pinus arizonica* ($\mu = 0.084$), if we are considering only the mean MAPE (Table 3).

Table 2. Parameters ajusted for each species and model analysed in this study.

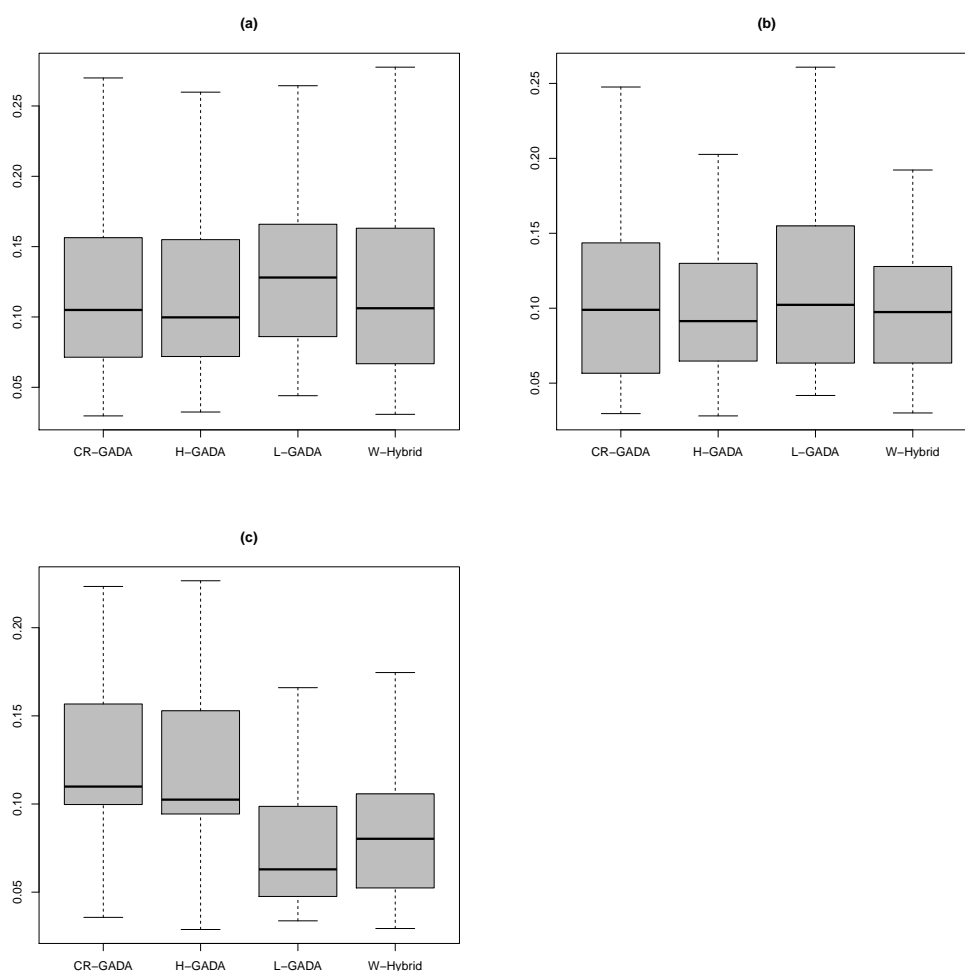| Model	C-R GADA			Hossfeld GADA		
Species	b_1	b_2	b_3	b_1	b_2	b_3
Pinus engelmannii	0.003	0.843	0.236	513.109	0.032	0.882
Pinus strobiformis	0.009	1.019	−0.124	38.051	−0.325	0.983
Pinus arizonica	0.004	0.011	2.681	68.413	1.639	0.794

Figure 4. Boxplot of the MAPE function calculated for the species (**a**) *P. engelmannii*, (**b**) *P. strobiformis*, and (**c**) *P. arizonica*.

Table 3. Basic statistics calculated for the MAPE function for each species and model of the study.

Model	C-R GADA		Hossfeld GADA		Hybrid Weibull	
Specie	μ_1	σ_1	μ_2	σ_2	μ_3	σ_3
Pinus engelmannii	0.110	0.062	0.112	0.064	0.107	0.055
Pinus strobiformis	0.093	0.046	0.094	0.047	0.098	0.051
Pinus arizonica	0.113	0.050	0.107	0.046	0.084	0.038

4. Conclusions

The results unambiguously show, at least for the population analyzed, the accuracy and feasibility of the proposed methodology presented in this study. Although this methodology cannot be classified as completely dynamic, such as the ADA or GADA methodologies, we consider it as semi-dynamic, since it will always need to estimate the parameter θ_1 for each of individual and assumes that one of the parameters is site-specific.

However, if this behavior presented by the parameters θ_1 and θ_3 can be reproduced in other data sets, this will greatly reduce the computational effort in obtaining the parameters of the models, because the parameter space is smaller.

It would be interesting in a future study to test the same hypothesis with other growth models and also see whether this behavior repeats, and still further try to understand from the biological point of view the reasons of such behavior.

References

1. Vanclay, J.K. *Modelling Forest Growth and Yield: Applications to Mixed Tropical Forests*; CAB International: Wallingford, UK, 1994; p. 537.
2. Mitchell, K.J. Sylver: Modelling the impact of silviculture on yield, lumber value, and economic return. *For. Chron.* **1988**, *64*, 127–131.
3. Richards, F.J. A flexible growth function for empirical use. *J. Exp. Botany* **1959**, *10*, 290–300.
4. Hossfeld, J.W. Mathematik für Forstmänner, Ökonomen und Cameralisten, Hennings Gotha, 4. Bd., S. 310; 1822.
5. Weibull, W. A Statistical Distribution Function of Wide Applicability. *J. Appl. Mech.* **1951**, *18*, 293–297.
6. Cieszewski, C.J. Comparing Fixed- and Variable-Base-Age Site Equations Having Single Versus Multiple Asymptotes. *For. Sci.* **2002**. *48*, 7–23.
7. Tofallis, C. A Better Measure of Relative Prediction Accuracy for Model Selection and Model Estimation. *J. Oper. Res. Soc.* **2015**, *66*, 1352–1362.
8. Bailey, R.L.; Clutter, J.L. Base-Age Invariant Polymorphic Site Curves. *For. Sci.* **1974**, *20*, 155–159.
9. Bailey, R.L.; Clutter, J.L. Modeling and Prediction of Forest Growth Variables based on Multilevel nonlinear Mixed Models. *For. Sci.* **2001**, *47*, 311–321.
10. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2017.