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

Souza, M. A.[1], Hernández-Velazco, M. R.[1], Protázio, J. M. B.[2], Escobar-Flores, J. G.[3], Lopez-Sanchez, C. A.[4], Hernández-Diaz, J. C.[5], and Wehenkel, C. A.[5]

[1]Mestría Institucional en Ciencias Agropecuárias y Forestales, UJED, Durango, Durango, México.
[2]Postgraduated Program in Mathematics and Statistics, UFPA, Belém, Pará, Brasil.
[3]Centro Interdisciplinario de Investigación para el Desarrollo Integral Regional, Unidad Durango, Instituto Politécnico Nacional, Durango, Durango, México.
[4]GIS-Forest Group, Departamento of Organisms and Systems Biology, University of Viedo, Mieres, Spain.
[5]Instituto de Silvicultura e Industria de la Madera, UJED, Durango, Durango, México.

October 31, 2020

Abstract

In this work, we present a hybrid methodology based on the ADA methodology, 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 3 parameters to just 1, 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.

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 (Hall and Robert 2001).

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 (Vanclay 1994).

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, while growth is the rate of production (Vanclay 1994).

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 is which use 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 class. Single-tree models may be very complex, modelling branches and internal stem characteristics, and may be linked to harvesting and conversion simulators (Mitchell 1988).

2 Materials and Methods

2.1 Study Area

The study was conducted in the mixed and irregular forests of the Sierra Madre Occidental de Durango (Figure 1), more precisely in the longitudinal gradient between Durango and Sinaloa, from September to November 2017. Six sites, 50 m x 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 increment of each tree based on growth nuclei, the pines were located within the plot and the species were recorded. For the experiment presented here, we applied the methodology only to the species presented in stand 5 (Figure 2), which has the largest number of individuals.

2.2 Base-Aged-Specific Model (BAS Model)

The BAS models implement in order to estimate and evaluate the respective diametric growth of the pine trees at the study area at were: Chapman-Richard (1959), Hossfeld (1822) and Weibull (1951). At this case, we are only using versions that depend on three parameters (vide 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	$d_1(t) = \theta_1(1 - e^{-\theta_2 t})^{\theta_3}$	Richards (1959)
Hossfeld	$d_3(t) = \theta_1(1 + \theta_2 t^{-\theta_3})^{-1}$	Hossfeld (1822)
Weibull	$d_4(t) = \theta_1 - \theta_2 e^{-\theta_3 t}$	Weibull (1951)

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

Bailey and Clutter (1974) introduced the concept using a technique now known as the Algebraic Difference Method (hereinafter ADA) approach. The approach consists of

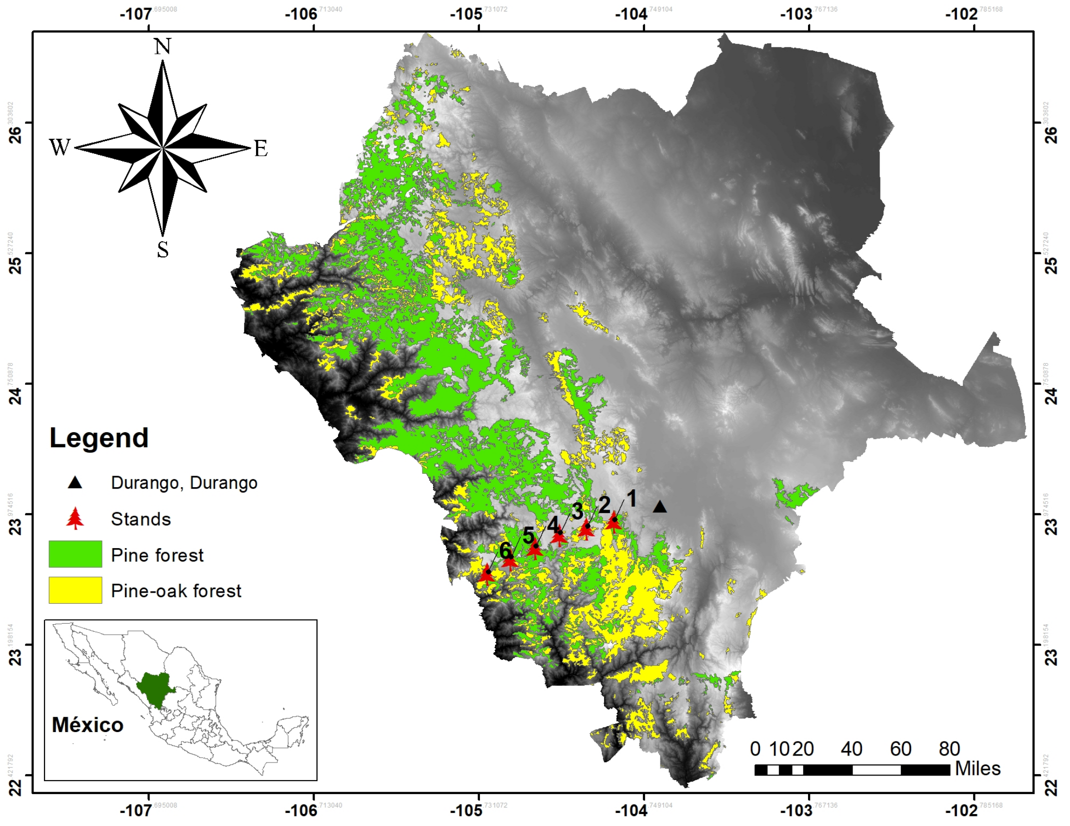

Figure 1: Study Area and the locations of the six data collection stands.

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 (Cieszewski, 2002) 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 (Cieszewski, 2002) implemented here is defines 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 (Cieszewski, 2002) implemented here is defines as:

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

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

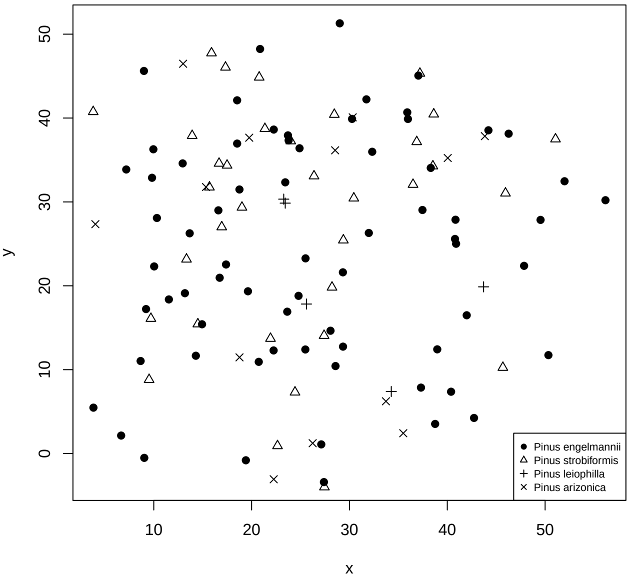

Figure 2: Study Area and the locations for the species presented at the experiment.

2.4 Proposed Methodology based on Algebraic Difference Approach (ADA)

The methodology proposed here consists on to apply first the the method of reducing parameters to immediately apply the ADA Methodology, supposing at this case, that there is any intrinsic relation between the parameters θ_1 and θ_3. At this 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 for one parameter θ_1, of course, only if the relationship $\theta_3 = f(\theta_1)$ exists.

For the population in question, in order to avoid generalization, we observed a strong inverse relationship between the variable $\theta 1$ and θ_3, such that:

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

than, it is possible rewrite the equation (3) as

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

where a is a scale factor. Now assuming $a = \chi$ as site-specific at the 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 replacing this result at the 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 to quantify the error associated with each of the models implemented in our experiment, we will use, for its robustness and simplicity, the mean absolute relative error (Tofallis 2015), 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 respectivelly the estimated and the observed values and n the number of observations.

3 Results and Discussion

Here we provide a concise and precise description of the experimental results, their interpretation as well as the experimental conclusions concerned to it.

We can see experimentally (vide Figure 3) that it is possible to write θ_3 as a function of θ_1 of the form $\theta_3 = f(\theta_1) = \frac{a}{\theta_1}$. Than using Nonlinear Regression Methods (nlsLM function on R), we estimated the parameters for each species and model presented in the experiment and presented in summary in Table 2.

The tests show that the results presented for the MAPE function calculated for the C-R GADA, Hossfeld GADA and Proposed Methodology were equivalent (vide Figure 4), considering each of the species presented in the experiment. The results presented 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). With a small advantage 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 (vide Table 3).

Table 2: Parameters ajusted for each species and models presented above

Model	C-R GADA			Hossfeld GADA			Hybrid Weibull
Specie	b_1	b_2	b_3	b_1	b_2	b_3	θ_1 with $d_r = 0.05$
Pinus *engelmannii*	0.003	0.843	0.236	513.109	0.032	0.882	27.185
Pinus *strobiformis*	0.009	1.019	-0.124	38.051	-0.325	0.983	19.899
Pinus *arizonica*	0.004	0.011	2.681	68.413	1.639	0.794	13.655

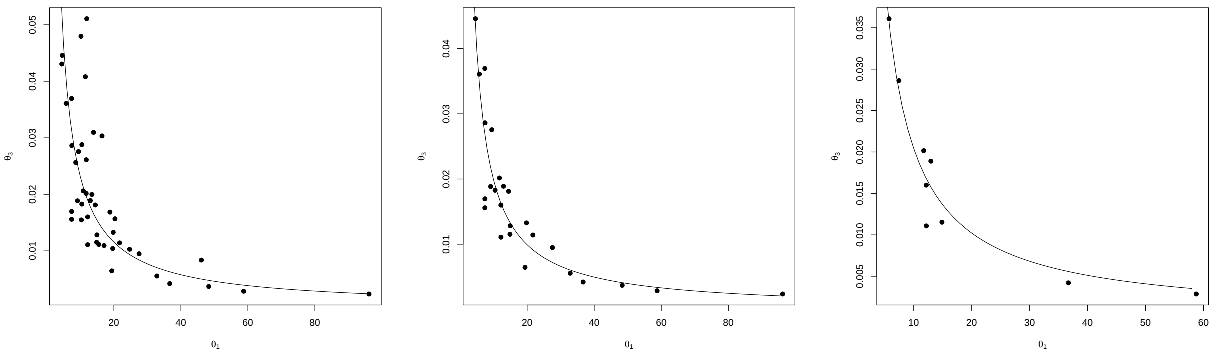

Figure 3: Relationship between the parameters θ_1 and θ_3 adjusted for the species Pinus *engelmannii* (left), Pinus *strobiformis* (middle) and Pinus *arizonica* (right).

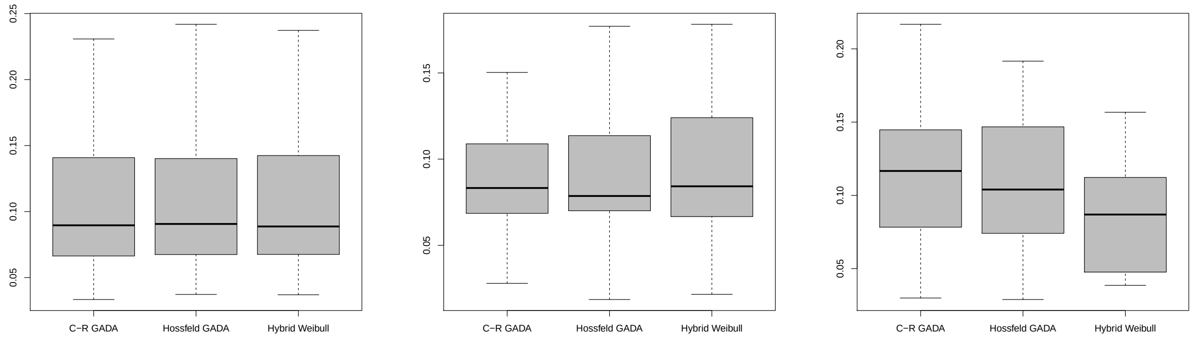

Figure 4: Boxplot of the MAPE function calculated for the species Pinus engelmannii (left), Pinus strobiformis (middle) and Pinus arizonica (right).

Table 3: Basic statistics calculated for the MAPE function for each species and model of the experiment

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 Conclusion

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

However, if this behavior presented by the parameters θ_1 and θ_3 reproduce 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 work to test the same hypothesis with other growth models and also see if this behavior repeats and still further try to understand from the biological point of view because this occurs.

5 Bibliography

Cieszewski, C. J. 2002. Comparing Fixed- and Variable-Base-Age Site Equations Having Single Versus Multiple Asymptotes, Forest Science, Volume 48, Issue 1, Pages 7–23,

Hossfeld, J.W. 1822. Mathematik für Forstmänner, Ökonomen und Cameralisten. Gotha, T. 4. Bd., S. 310 [in German].

Mitchell, K.J., 1988. Sylver: Modelling the impact of silviculture on yield,lumber value, and economic return. For. Chron. 64:127–131

Bailey, R. L. and Clutter, J. L. 1974. Base-Age Invariant Polymorphic Site Curves, Forest Science, Volume 20, Issue 2, Pages 155–159

Richards, F.J., 1959. A flexible growth function for empirical use J. J. Exp. Botany, 10(2): 290–300.

Tofallis, C. 2015. A Better Measure of Relative Prediction Accuracy for Model Selection and Model Estimation, Journal of the Operational Research Society, 66(8),1352-1362.

Vanclay, J. K. 1994. Modelling forest growth and yield: applications to mixed tropical forests.Schoolof Environmental Science and Management Papers, page 537

Weibull, W. 1951. A Statistical Distribution Function of Wide Applicability. Journal of Applied Mechanics, 18, 293-297.