Chapter 1: Introduction

1.1 Background

Growth can be defined as variation and development in tissues and organs of an organism over time. 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 [89].

A mathematical model is a mathematical description or an equation of a real world phenomenon. When a system is represented as a model, the model becomes a tool for us to understand the system by helping us to sift through its complexity and to focus on the important, relevant aspects. Consequently, a mathematical model is one type of simplified representation of a real system which allows us to reach a mathematical conclusion about the behavior of the real system, as illustrated in **Figure 1.1**[26]. For a real-world problem, the first job is to formulate a mathematical model by identifying the independent and dependent variable and making assumptions that simplify the phenomenon enough to make it mathematically controllable. If there is no any

physical law associate with the problem, some sets of data may be considered and examine the data in the form of a table in order to discern a pattern. In the second stage, some mathematical conclusions will be derived by using mathematics on the mathematical model. The third step is to take those mathematical conclusions and take them as information about the original real world phenomenon by making predictions. The final step is to test the predictions by checking against new real data. If the prediction doesn't fit well then one need to refine the model or to formulate a new model and start the cycle again. There is a correspondence between the variables of the model and observable quantities. The purpose of the model is to understand the phenomenon and perhaps to make predictions about future behavior [78]. It helps us to understand, predict and control a system in a more organized or methodological manner. A mathematical model not only tells us what are the essential components of a real system but more importantly, they also tell us how and to what degree do each of these components interact with each other.

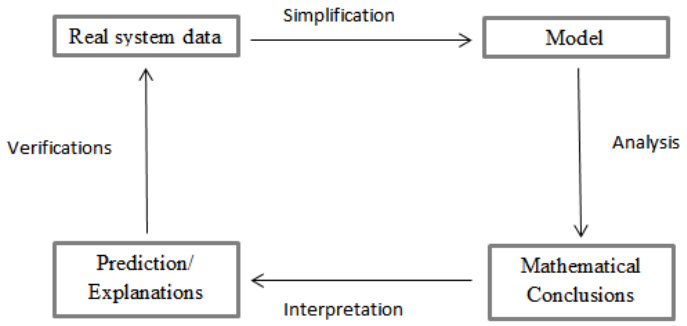

Figure 1.1: A flow of the modelling process begins with an examination of real data.

Growth patterns are traditionally classed in two groups: determinate and indeterminate ones. Determinate growth is usually defined as growth that stops when an organism reaches a certain size. Usually, growth stops during the reproductive

stage. Indeterminate growth is defined as growth that continues past maturation and may continue to the end of life [42].

Determinate growth is observed in bacteria and other unicellular organisms, all birds, some plants, fish, insects, and most mammals. Whereas Indeterminate growth is characteristic of a large number of invertebrate taxa – some kinds of algae, clams, cladocerans and crayfish, mollusks, many insects, echinoderms, many modular animals such as corals and sponges, and "lower" vertebrate taxa, such as most fish, amphibians, and reptiles (lizards and snakes). It is also found in perennial plants and most trees.

The growth models may also be classified as a linear and nonlinear model based on the order of their parameters. Linear models are linear in the parameters which have to be estimated but not necessarily in the independent variable. Any model that nonlinear in the parameters is known as a nonlinear growth model. Linear models though rigid, have been used frequently and may fit a set of data with a limited range of the dependent variable. Even though linear models can represent a wide variety of relationships there are many situations in which a model of this form is not appropriate. Nonlinear growth models play a very important role in science and engineering and they usually arise from the knowledge of the underlying mechanism that describes the phenomena. Indeed, nonlinear models are sometimes called mechanistic models, distinguishing them from linear models, which are typically thought of as empirical models [55]. Also, a theoretical nonlinear mathematical model can provide the basis for an objective method of estimating growth potential and the sustained yield of a crop. Vanclay [89] has also pointed out that polynomial models do little to further understanding of the process involved in the system being modelled and are notorious for their poor extrapolation. Though polynomial models

describe growth data satisfactorily the associated parameters provide no forestry meaning or interpretation.

Development of simple and efficient models that allow forest managers to determine the height of the trees with respect to time is a prime objective in forest management. Knowledge of the relation between these variables permits to obtain, without investing large amounts of money in height measurement, dominant height of the stands and height diameter ration etc.

A mathematical growth equation never gives a complete accurate picture of a physical situation, it is an *idealization*. A good model simplifies real enough to permit mathematical calculations but is accurate enough to provide valuable conclusions. It is important to realize the limitations of a model [78].

1.2 Introduction to the growth models

The first attempts to establish a method of classifying forests through growth potential was based on subjectively defined relationships of height on age. Since then several mathematical models and empirical polynomial equations have been used to summarize and define growth trends.

A number of height growth equations for tree have been developed for various tree species. Among a variety of mathematical equations, nonlinear sigmoid growth models are widely used in the study of growth of trees. Some nonlinear growth may produce large extrapolation error when applied beyond the range of model development data. Therefore, the model's prediction capabilities should be carefully evaluated and validated before they are used [92].

The different growth models used in this work are listed in **Table 1.1**. For all models, w is the dependent growth variable, t is the independent variable, A, β, b, k and m are parameters to be estimated, and $\exp(e)$ is the base of the natural logarithms.

Table 1.1: The integral forms of the growth models along with the source.

S/N	Growth Models	Integral form of the models $(w(t))$	Source
1	Monomolecular	$A(1-\beta e^{-kt})$	[18]
2	Gompertz	$Ae^{-\beta e^{-kt}}$	[42]
3	Logistic	$\frac{A}{1+\beta e^{-kt}}$	[57]
4	Weibull 4 parameter	$A-\beta\ exp(-kt^m)$	[64]
5	Weibull 3 parameter	$A(1-exp(-kt^m))$	[30]
6	Weibull 2 parameter	$A(1-\ exp\ (-kt))$	[19]
7	Chapman Richards 4 parameter	$A\{1-\beta e^{-kt}\}^d$	[59]
8	Chapman Richards 3 parameter	$A\{1-e^{-kt}\}^d$	[84]
9	von Bertalanffy 4 parameters	$\{A^{1-m}-\beta e^{-kt}\}^{\frac{1}{1-m}}$	[15]
10	von Bertalanffy 3 parameters	$A-(A-b)e^{-kt}$	[54]
11	von Bertalanffy 2 parameters	$A(1-e^{-kt})$.	[38]

Despite the fact that there are a couples of theoretical models formulated particularly for forestry applications most developed in other disciplines are believed to have a considerable potential for modelling forest growth and yield parameters. Up to the time theoretical nonlinear mathematical models explicitly formulated to offer consistent growth and yield estimates have not been engaged in modelling forest growth parameters. This is mainly attributed to the fact that forestry conformance with the model's assumption, forestry meaning of the associated parameters and the

methodology used for fitting the models to forest growth data are all highly correlated with the mathematics of the models which is hardly understood in a forestry context.

1.3 Literature review

The first growth model was proposed by Pierre Francois Verhulst in 1838, which is a three-parameter model for the growth of single-species populations that shows a logistic sigmoid growth curve for time. The same model was also used independently by Pearl and Reed in 1920 [73]. Together with Verhulst model, the three parameter growth models reported by Gompertz [28] and Bertalanffy [4] have been used as fundamental growth models in various biological studies. Some attempts in which shape parameters were added to three parameters growth models aimed at accounting for the flexible form of growth curves [[5], [68], [74], [64]].

Many mathematical models were used to summarize and define forestry growth trends. The most frequently used nonlinear models include: Monomolecular, Logistic, Weibull, Gompertz, Chapman Richards and Von Bertalanffy growth models.

Klingaman and Oliver [44] applied the Monomolecular growth model to evaluate the interference potential of Palmer amaranth on growth and yield of soybean by using the nonlinear procedure of Statistical Analysis System (SAS).

Fekedulegn et al [22] used the Negative Exponential, Monomolecular, Mitcherlich, Gompertz, logistic, Chapman-Richards, von Bertalanffy, Weibull and Richard's nonlinear growth models to top height age of Norway spruce (Picea abies L.) from the Bowmont Norway Spruce Thinning Experiment. The parameters were estimated using the Marquardt iterative method of nonlinear regression. They also specified some formulas that provide good initial values of the parameters.

In 1999, Heinen [31] compared two ways of representation of exponential, monomolecular, Gompertz, logistic and Richards's growth models. This paper also addressed some properties of those growth models and applied for the data of lettuce grown on nutrient film.

Tsoularis and Wallace [88] reviewed thirteen growth models including Monomolecular, Von Bertalanffy, Richards and Gompertz etc. They identified some limitations and restrictions by analyzing the properties of these models. They also introduced a generalized form of the logistic growth curve which incorporates these models as special cases.

In 2004, Colbert et al [15] fitted four different sigmoidal growth models (Richards, Weibull, Chapman- Richards and Von Bertalanffy) to the basal area data of oak trees. The Marquardt method in the NLIN procedure of SAS was used to estimate the parameters.

The paper by Lei and Zhang in 2004 [48] explained the different features of the Bertalanffy-Richards growth model based on the different conditions of the allometric parameter. They introduced an assessment software to easily get the partial derivatives with respect to each parameter for estimating the parameters of any nonlinear models.

The paper by Chandrasekhar [9] fitted the three growth models namely Richards, Gompertz and the Monomolecular to identify the most parsimonious and biological reasonable model for describing the girth growth of young rubber trees.

In 2012, Lekwadi et al [49] demonstrated top height-age site classification for Sitka spruce plantations in Ireland with Chapman-Richards models (Four parameters

Chapman- Richards and three parameters Chapman- Richards) using Gauss- Newton NLS procedure in R 2.13.2.

Paine et al. [61] stated that the relative growth rate almost universally decreases with increasing size. They discussed the three non-asymptotic growth models; Linear, Exponential, Power and four asymptotic growth models; Monomolecular, Logistic (three parameters and four parameters) and Gompertz growth models. They observed that the exponential model was the only model in which the relative growth rate was constant with respect to time and biomass.

The paper by Goncalves et al. [29] characterized the behavior of European pear cultivars under quince rootstocks to Entomosporium leaf spot (ELS) using three empirical models Logistic, Monomolecular and Gompertz using SAS. They found that the Logistic and Gompertz models were the most suitable to describe the ELS progress in the edafoclimatic condition of Southern Brazil.

Araneda et al in 2013 [3] modified the Gompertz, von Bertalanffy and putter growth model in order to include the effect of initial density and used to the growth of a population in aquaculture, including the phenomenon of size heterogeneity. The estimation of the growth models were performed using nonlinear regression methods. They concluded that the modified von-Bertalanffy model was the most effective of the three models in predicting growth.

1.4 The importance of the methods of estimation

The mathematical models are theoretical analysis of dynamic systems and these models describe the input and output properties of the system. The parameters of a model through which a system can be described may have a lot of unknown values. So in order to cope with these uncertainties, the parameters of a model are to be estimated. With the help of several statistical methods the parameters are to be

estimated to get a better prediction of the study. The parameters of the models have been estimated as they can give the future predictions, which are very important. Prediction has a great impact on decision making and it is also a major component of reverse estimation. Accurate predictions will help to have knowledge of how much of product will be needed in a particular year, and how to get that much production and how much to invest and the way to manage all the things. In that it will available in the market and it will help us in getting economic growth.

1.5 Objectives of the Study

The objectives of this thesis are:

- To discuss the mathematical properties of the models mentioned in **Table 1.1** for giving biological implications of their parameters.
- To estimate the parameters of the growth models using the various method of estimation and to develop some new method of estimation in order to provide better estimates.
- To specify the initial guess value selection criteria for the use of certain iterative method of estimation.
- To select the best fit model based on a specific selection procedure for some of the well-known forestry data sets using suitable methods of estimation.

1.6 Methodology

The nonlinear growth models can be written as,

$$w_i = f(t_i, \boldsymbol{B}) + \varepsilon_i \tag{1.1}$$

$i = 1,2,\cdots,n$, where n be the number of observations, w_is are the response variable, $t_i s$ are the independent variable, **B** is the vector of the parameters (say the parameters be $b_1, b_2, \cdots, b_p$) of those models. ε_i is a random error in the model with mean zero and constant variance [66].

1.6.1 Importance of good starting value

The iteration methods require an initial value for each parameter being estimated. The better these initial estimates are the faster will be the convergence to the fitted value. Initial value specification is one of the most difficult problems encountered in estimating parameters of the nonlinear model [18]. If the initial values are too far away from the actual value then the iteration may not be convergent or one needs larger iteration [62]. There is no any general method for obtaining initial estimates. One uses whatever information is available. All available prior information can be used to make these starting values as reliable and realistic as they possibly can be.

1.6.2 Selection of a growth equation

After fitting the growth models using different methods of estimation, the best-fit model is selected based on the following selection criteria. The selection criteria consist of six distinct steps.

Step I: Logical and Biological consistency: In this step the logical consistency and biologically realistic of the estimated parameters are checked. The growth models with non-consistent and non-natural consistency and poor statistical properties are excluded.

Step II: Chi-Square Goodness-of-Fit Test ($\boldsymbol{\chi^2}$): This test enables us to see how well does the growth model fit to the observed data. The Chi-Square is defined as

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i},$$

where y_i is the observed value and $\hat{y}_i$ is the predicted value for $i = 1, 2, \cdots, n$. If the calculated value of χ^2 is greater than the tabulated value of χ^2 with $n - 1 - p$ degrees of freedom (where p is the number of parameters of the growth model and n is the number of observations) then the null hypothesis is rejected otherwise accepted.

Step III: The Root Mean Square Error (RMSE): The RMSE measures to aggregate the residuals into one measure of predictive power. The RMSE of a model prediction with reference to the calculable variable is defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(w_i - \widehat{w}_i)^2}{n}},$$

where w_i is the observed values and $\widehat{w}_i$ is the predicted values for $i = 1, 2, \cdots, n$.

Step IV: Coefficient of determination $(\boldsymbol{R^2})$ and adjusted coefficient of determination $(\boldsymbol{R_a^2})$: The R^2 value indicates how well the data point fits a growth model. Generally the value of (R^2) lies between 0 and 1$(0 \leq R^2 \leq 1)$. But it is not possible for R^2 to actually attain 1, if pure error exist. In practice, sometime negative value of R^2 may occur. Theoretically the value, 1 indicates a perfect fit, 0 reveals that the model is not a better than the simple average and the negative value indicates a poor model [79]. If the value of R^2 is above 0.9, it is accepted as efficient [59]. The mathematical formulation of the coefficient of determination is,

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2},$$

where $\bar{y}$ is the mean of the response variables.

The R_a^2 value is an endeavor to redress the propensity for over fitting of R^2 by adjusting both the numerator and the exterminator by their respective degree of freedom and defined is as

$$R_a^2 = 1 - (1 - R^2)\left(\frac{n-1}{n-p}\right).$$

The R_a^2 is used to compare growth models not only to a specific set of data but also two or more entirely different sets of data. The equation with the least standard error of the estimate have the maximum R_a^2.

Step V: Confidence interval: In this step, the confidence intervals of the estimated parameters are found. Suppose B is the vector of the parameters (say the parameters are $\beta_1, \beta_2, \cdots, \beta_p$) of the growth models. Confidence limits for the true value of the parameters B are evaluated on the basis of the linearized approximation, evaluated at the predicted value of the parameters $\hat{B}$. The $100(1-\alpha)\%$ confidence intervals for the parameters B may be written as

$$\hat{B}_i \pm t_{\frac{\alpha}{2}, n-p}\, se\left(\hat{B}_i\right),$$

where $t_{\frac{\alpha}{2}, n-p}$ is the $t-value$ at $n-p$ degrees of freedom and $se\left(\hat{B}_i\right) =$

$$\left\{appropriate\ diagonal\ element\ of\ \left(\hat{Z}'\hat{Z}\right)^{-1} S^2\right\}^{\frac{1}{2}}.$$

Here $S^2 = S(\hat{B})/(n-p)$; $S\left(\hat{B}\right) = \sum_{i=1}^{n}\left\{y_i - f(t_i, \hat{B})\right\}^2$ and

$$\hat{Z} = \begin{bmatrix} \frac{\partial f(t_1,B)}{\partial \beta_1} & \frac{\partial f(t_1,B)}{\partial \beta_2} & \cdots & \frac{\partial f(t_1,B)}{\partial \beta_p} \\ \frac{\partial f(t_2,B)}{\partial \beta_1} & \frac{\partial f(t_2,B)}{\partial \beta_2} & \cdots & \frac{\partial f(t_2,B)}{\partial \beta_p} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f(t_n,B)}{\partial \beta_1} & \frac{\partial f(t_n,B)}{\partial \beta_2} & \cdots & \frac{\partial f(t_n,B)}{\partial \beta_p} \end{bmatrix}.$$

The final estimate of the parameters with ~95% confidence band excluding zero, indicating that there are only non-zero values of the parameters and then they are always significant. In this step, those results with negative confidence interval are eliminated.

Step VI: Approximate R^2 for prediction: Finally calculate the approximate R^2 for prediction and it is given by

$$R^2_{prediction} = 1 - \frac{PRESS}{\sum(y_i - \bar{y})^2},$$

where $PRESS = \sum_{i=1}^{n}\left(\frac{e_i}{1-h_{ii}}\right)^2$ is known as the PRESS statistic.

Here $e_i = y_i - \hat{y}_i$ and h_{ii} are the diagonal elements of hat matrix $H = T(T'T)^{-1}T'$ and T is a $n \times 1$ matrix of the independent variables. This statistic gives some indication of the predictive capability of the model.

If the value of $R^2_{prediction}$ and R^2 are r and m respectively, then one can expect from the model to explain about $r\%$ of the variability in predicting new observations, as compared to the approximately $m\%$ of the variability in the original data explained by the fitting [55].

1.7 Outline and organization of the thesis

In this thesis, different methods of estimation have been developed for estimating the parameters of the growth models mentioned in **Table 1.1**. A software package has been developed in FORTRAN 77 and in MATLAB version 7.11 for the fitting of the models. This thesis comprises of six chapters including the introductory chapter.

In the introductory chapter, discussion has been made about the growth model and a brief account of the relevant works done by the earlier researchers.

In chapter 2, a suitable method of estimation has been determined to estimate the parameters of some non-linear growth models using various methods of estimation. Six different estimation methods are introduced and used for estimating the parameters of the Monomolecular, Gompertz and Logistic growth models. The maximum diameter data and top height growth of *babul (Acacia Nilotica) tree* are used for testing the validity of the methods. These sets of data are based on the analysis of sample plot data of Uttar Pradesh, Maharashtra and Madhya Pradesh [37]. The top height age, the cumulative basal area production and the mean diameter at breast height data, originated from the Bowmont Norway spruce thinning experiment, sample plot 3661 [[21] [22]] are also used.

In chapter 3, certain properties of three forms of Weibull growth models have been discussed in forestry viewpoint. The parameters of the Weibull growth models are estimated using Newton-Raphson iteration method for five well-known forestry data sets. The Generalised Newton-Raphson iteration method is applied by minimizing the sum of square residue under the assumption that random errors in the models have mean zero and constant variance. The initial value specifications of the parameters for use of any iterative methods of estimation are also provided.

In the chapter 4, the von Bertalanffy growth models are discussed in forestry viewpoint. Integral forms of the von Bertalanffy growth models are also discussed together with their varied re-parameterizations. The properties of the parameters are studied by observant its nature on biology. This chapter also introduces some economic methods to suit the models, which demands less computation and can be used for any growth data. The parameters of those models have been estimated using the methods for five well-known data sets.

In Chapter 5, the applicability of Chapman Richard growth models have been discussed in forestry. Integral forms and the limiting cases of the Chapman Richard growth model are also discussed along with their varied re-parameterizations. The biological properties of the parameters are studied briefly. This Chapter introduces two methods suitable for the models, which demand less computation and can be used for any growth data. Five well-known forestry data are used for testing the validity of the proposed methods based on the certain statistical test criteria. The results show that each method of estimation performs well.

In Chapter 6, a comparative study among the six growth models Monomolecular, Gompertz, Logistic, Weibull, Von Bertalanffy and Chapmen Richard growth models have been presented for describing the growth pattern of Teak (*Tectona grandis*) and

Babul (*Acacia Nilotica*) in India. The height and DBH growth data from Teak trees in Warangal state and Hoshangabad division of India [27] are considered in this study. This Chapter also presents a comparative study of six growth models for top height age, the mean diameter at breast height data and the cumulative basal area production originated from the Bowmont Norway spruce thinning experiment, sample plot 3661. The best fit model is selected based on the selection criterion.

The literatures cited in the thesis are listed at the bibliography section.

*****