

Application of statistical modeling techniques in forest research

AJIT SHARMA and SHILPA

Department of Basic Sciences, Dr YS Parmar University of Horticulture and Forestry
Nauni, Solan 173230 Himachal Pradesh, India

Email for correspondence: ajit.aziz.sharma@gmail.com

© Society for Advancement of Human and Nature 2017

Received: 2.11.2015/Accepted: 13.3.2016

ABSTRACT

Applied statistics in forest research has a significant role in collection, compilation, analysis and interpretation of data. In view of day to day rapid changes in research spectrum the scenario is becoming interesting for researchers. A model is defined as abstraction of real situations which aim to give the empirical content to relationships of variables and their interpretation. Modeling techniques are very common in basic as well as multidisciplinary research. This paper discusses various modeling techniques and prediction models viz straight line, second degree parabola, exponential, modified exponential, gompertz and logistic etc. Various thumb rules for the best fitted models are also discussed and array applications of these models are discussed in the field of forestry.

Keywords: Statistical model; prediction; forecasting; time series

INTRODUCTION

The model is a simplified design to describe the complex process using mathematical and statistical techniques. In 18th century simple mathematical models based on probability theory were used to understand the economics associated with insurance theory (Smith 1904). This theory was associated with estimation and chance theory which played an important role in development of probability theory and vital statistics (Edward and Martin 1990). These problems are described (Hald 2003) in the 3rd edition of Doctrine of Chances. Many mathematicians during the last decade have contributed in this field. A study was made on probability models based on linear assumption of the demand for attributes (Heckman and Synder 1977). These process-based models are typically developed and tested using experimental trials and thus offer the distinct advantage of leveraging decades of research on forestry, crop physiology, reproduction, agronomy and soil science among other disciplines. Yet these models also require extensive input data on cultivar, management and soil conditions that are unavailable in many parts of the world. The statistical models are also used for long term risk management.

Modeling provides powerful tools for investigating the dependence and nature of relationship

among the variables of interest. The relationship among variables must be determined for the purpose of predicting the values of one or more variables on the basis of observation on other variables. Model building is currently applied in many fields viz horticulture, forestry, agriculture, biometrics, econometrics, education, meteorology, industry etc (Prasad 2010).

It enables to study the past behaviour of the phenomenon under consideration ie to determine the type and nature of variations in the data. It helps to compare the changes in the values of different phenomena at different times. Statistical tests are used to carefully examine prior activities and then use these analyses to make informed predictions about future activities. Regardless of the statistical tests data are examined in a systematic manner so that decisions can be made with great degree of certainty. The opportunity of using existing data to predict future outcomes is viewed as model building. That is to say existing data are used to build a model of the future with a predetermined degree of error built into the model (Thomas 1998).

Central Statistical Organization (CSO) in 1994 assessed agricultural production and the variety of agricultural activities in the country. The objective was to develop a simple technique for farmers and planners

for predicting crop production in different districts of the country. This objective was in concert with CSO objectives of fostering data user-producer interaction and dialogue, collecting census data and the development of an agricultural data bank suited to the needs of the internal and external users and creation of analytical ability at the institution. Therefore this work partly assisted CSO in analyzing census data and in assessing data needs for various purposes (Henry 2011).

Forestry now a days has become highly input and cost intensive because the use of insecticides, pesticides and fungicides is increasing day by day to control the timber and other valuable species from insect pests and diseases attack at every stage. Under the changed scenario forecasting of various aspects relating to forestry are becoming very essential. But in spite of strong need for reliable and timely forecasts the current status is far from satisfaction. For most of the sectors there is no organized system of forecasting. The official forecasts (advance estimates) of major species are issued by Directorate of Economics and Statistics. The various statistical prediction models are classified as linear and non-linear models.

Linear models

Linear growth function was fitted to the data using OLS method on series data on area, yield and production of wheat for twenty years from 1970-71 to 1989-90. The function was specified as $Y = a + bX$ where Y = Index number of area, production and productivity, X = Time variable and a and b are estimated parameters (Ali and Singh 1995).

Wang et al (2000) developed tree growth prediction models based on grey system theory by utilizing data from 202 temporary plots and stem analyses of 206 trees of Dahurian larch (*Larix gmelinii*) in 10 forestry bureaux of Yakeshi Forestry Administrative Bureau in the Daxing'an mountains of the inner Mongolia autonomous region (Nei Menggu), China. The precision of the models was evaluated by residual and posterior residual tests for dbh and volume. They concluded that the grey system models performed better than statistical models.

Ares et al (2003) conducted a study to know trends in tree growth and understorey yield in silvipastoral practices with southern pines and examined data from published/existing field trials to determine mid and long term trends in tree growth and understory

yields in silvipastoral practices with southern pines in United States. Understory affected dbh more than height and therefore dbh-height relationships differed among practices. Sigmoidal models predicted that the tree height would peak at different age depending on tree spacing and understory type. These changes may affect the accuracy of site indices and wood yield predictions. Livestock gains decrease linearly with increasing stand basal area and stand age although for age yields sometimes decay exponentially. ANOVA, regression analysis and test of differences between regression slopes were performed. ANOVA was used to test the effect of grazing, fertilization and grass types on mean dbh at different ages. Linear regression analysis was used to examine mean dbh growth versus age, mean total height versus mean dbh and livestock gains versus age and stand basal area relationships.

Raizada et al (2007) developed biomass prediction models for 17 year old *Acacia nilotica* trees raised on salt-affected vertisols of the semi and tropics in Karnataka, India. *A. nilotica* was raised at 8×8 m spacing with an understory of three grass species- *Cenchrus ciliaris*, *Dicanthium annulatum* and *Chloris bournellii* for the production of fodder. Wide variations occurred in the trees sampled by random selection in the plantation with respect to diameter (3.1 to 16 cm) and tree height (3.5 to 5.1 m). Leaf biomass varied from 0.5 to 3.1 kg/tree and contribution by big branches (>2 cm diameter) varied from 3.81 to 24.13 kg/tree. Total above ground biomass ranged from 26.5 to 100.74 kg/tree. Prediction models with the best fit were in the linear form with R^2 values of 0.826, 0.916 and 0.866 for predicting bole, utilizable and total above ground biomass respectively.

Swetha (2009) worked out impact of rainwater harvesting on farming economy by using nine models (linear, logarithm, power, compound, S-curve, logistic, growth, exponential and inverse models). It was observed that linear model was showing significant results and was the best model. The R^2 was used to compare and choose the best fit model. The variation in groundwater level was found to increase over the periods as the temperature increased.

The trends in extraction of medicinal plants in Himachal Pradesh were estimated by making use of linear growth rates. The extracted species were classified on the basis of habitat, habit, economic part used and their present status. Secondary data including quantity of medicinal plants collected/produced/traded

were obtained from the records of the state forest department for a period of eighteen years (1994-95 to 2011-12). The study was divided into Period-I from 1994-95 to 2002-03 and Period-II from 2002-03 to 2011-12. On the basis of habitat only medicinal plants under cultivation showed a positive and significant growth (29.01%/annum) during P-I while during P-II upper hill subtropical (28.74%/annum) and cold desert (24.71%/annum) showed significant negative growth. The entire plant showed a negative and significant growth of 21.4 per cent per annum in P-I and seeds in P-II showed negative and significant growth (45.00%/annum). But in overall bark entire plants and rhizomes showed negative significant growth rates (13.43%/annum), (19.44%/annum) and (10.50%/annum) respectively. On the other hand Barah flowers showed positive and significant growth of 18.34 per cent per annum. The roots yielded maximum average volume and maximum revenue. The linear growth of medicinal plants based on their present status indicated that only commonly available medicinal plants showed a significant declining growth rate of 10.67 per cent per annum and 9.26 per cent per annum in P-II and overall respectively (Shilpa et al 2015).

Non-linear models

Chengappa (1981) conducted a study on growth rates of area, production and yield of coffee in India. Linear and exponential models were used and corresponding growth rates were worked out. The exponential model provided a good fit for Arabica and Robusta coffee production.

Umakapila (1982) computed the growth rate of groundnut for all India and state-wise for the period 1951-52 to 1974-75 by using exponential growth function and reported that growth rate was negative (-0.56%) and growth rate of output was 0.55 per cent. The major source of growth in output in the states of Andhra Pradesh and Karnataka was found to be the area under the crop.

Fialar (1985) analysed the production pattern in marketing of cocoa in Ghana by using exponential model. The study indicated that the rate of growth under cocoa for the world as a whole has decreased. However the total production per annum has increased. The negative growth rate for exports for the world as a whole reflected that the international trade on cocoa was declining.

Sadasivan (1989) analyzed the pattern of pulse production in India from 1955-56 to 1984-85. The analysis was done at the disaggregate levels of seasons, periods, states and crops. He used exponential model to estimate growth rates. It was concluded that stagnancy in pulse production at all India level was the net result of changing situation at disaggregate level rather than on overall stagnancy.

Devi et al (1990) used semi-log, exponential and quadratic models to analyze the trends in area, production and productivity of banana in Kerala. The annual average rate was computed from these functions using the relationship $r = (\text{antilog } b \cdot t) \times 100$, where Y= Area, production and yield, t= Time period, a and b are parameters estimated.

Kaushik (1993) studied the growth of oilseeds production in India for the period of 1968-69 to 1991-92. The production was divided into two periods, period 1 (1980-81 to 1991-92) and period 2 (1968-69 to 1979-80) to clearly bring out the trend in the more recent period. The exponential growth model was used. The study revealed that during period 1 most of the growth in oilseeds output was due to growth in area whereas in period 2 it was mainly due to improvement in productivity.

Tesfaye (1994) conducted a study on growth performance of some multipurpose trees and shrubs in the semi-arid areas of southern Ethiopia. They studied the growth performance of some multipurpose trees and shrubs for five years at two locations in semiarid areas of southern Ethiopia and found that there existed strong correlation between diameter and volume than between height and volume. Growth curves were fitted for all the species and they reported that periods of growth coincided with rainy season.

Pandey et al (1998) conducted a study on felled trees of *Populus deltoids* G-3 from a sample population in a plantation at the experimental farm in Himachal Pradesh to investigate the suitability of six volume prediction models based on tree height (h) and diameter at breast height (d). The measurements made on the trees included h, d and actual volume. The models were assessed using both the criteria of $R^2 > 0.9$ and a cross-validation technique. Models with $R^2 > 0.9$ did not all pass the validation test. Two volume estimation models were recommended on the basis of

the analysis one based on a function of (dh) and the other on a function of dh, both provided values of volume.

Rahaman and Ahmad (2000) utilized the data from 171 plots in Chittagong hill tracts, Bangladesh and studied the current growth estimation and future yield prediction models of gamar (*Gmelina arborea*). All plots were laid in well-stocked Gamar plantation of every age class from 1 to 17 years. Twenty of the 171 sample plots were selected and kept separately to validate the growth and yield models. Data from the remaining 151 plots were used to formulate the models which included stand diameter and height function, number of trees per hectare prediction model, basal area and stand volume equation. Results indicated that the derived models would help to determine the optimal harvest age of Gamar plantation and prescribe the best financial investment in forestry compared with other competing uses of land.

Kumar et al (2005) studied the price forecasting of different classes of teak by the application of exponential smoothing model. A single-parameter exponential smoothing model was used to forecast prices of different classes of teak in the Dandeli timber depot in Karnataka, India. Price data for the period May 1987 to May 2001 were used and both ex post and ex ante forecasts were made. The results of the ex post forecasts revealed that the predicted prices were close to the actual prices. Mohanty et al (2006) worked out the impact of puddling, tillage and residue management on wheat (*Triticum aestivum* L) seedling emergence and growth in a rice-wheat system using non-linear regression models. The non-linear regression model study indicated that the logistic model predicted the shoot growth of wheat under different tillage and residue management practices better than the Gompertz model. On the other hand for root growth the monomolecular model fitted well with the experimental data.

Rizvi and Khare (2006) developed prediction models for timber weight of *Populus deltoids* planted on farmland in Haryana. They estimated the fresh green timber of poplar tree and evaluated growth process based non-linear models for fresh timber weight. The models viz $W = 1.398 D_1 (0.608)$ and $W = 9.975 \exp(1 + 7.768) \exp(-1.299) (D_2 H) - (0.217)$ where W- fresh timber weight, D- diameter at breast height and H- height of the tree were found to be good fit. The mean

errors in prediction of timber weight by these models were 10.4 and 7.0 kg.

Yusuf and Salau (2007) estimated the prediction of future production of citrus and mango in the medium term up to 2010. The prediction was based on the assumptions that past trends (area planted and yield) and existence of normal weather pattern will hold. Time trend model with specific emphasis on growth model was employed. The analysis delineated three different eras (period between 1961 and 2003, 1986 to 2003 and 1991-2003). These eras were used to simulate the different policy regimes of regulation, structural adjustment era and liberalization era. In general output of citrus and mango maintained upward trend over the years. However the growth rate was highest for the era including structural adjustment. Following from this output predictions over the medium term were highest for the analysis with structural adjustment era. Gaddour and Nazari (2008) worked on adjustment of the kid's growth curve in pure goat breeds and crosses under southern Tunisian conditions. Five non-linear statistical models viz Gompertz, Brody, Richards, logistic and exponential were tested to fit the kid's growth curve parameters and shape of indigenous goat, Alpine, Damascus and their crosses. Data from 16 years periodical weighing were used to adjust growth curve before 5 months age of 1687 suckled kids. Gompertz model was the best to adjust kid's growth evolution.

Sreekant (2008) developed an empirical regional cashew prediction model (ERCPM) based on secondary database for area and production from 1965-66 to 2002-2003 for different states viz Karnataka, Kerala, Maharashtra, Goa, Andhra Pradesh, Tamil Nadu, Orissa and West Bengal from which potential explanatory variables were derived. The empirical model developed for cashew production was a function of five derived variables viz estimated cultivated area, incremental area, initial yield, moving average of productivity and moving average for production growth rate. The results have proved that it is possible to predict cashew production in different states with reasonable precision using this model. The accuracy of these models in predicting cashew production for different states ranged from 85 to 99 per cent. Chethana (2009) examined prediction models in teak-based agroforestry systems in northern transitional zone of Karnataka. Different prediction models namely linear, quadratic, cubic, exponential,

growth, sigmoid, vapor pressure, hassel, MMF compound, logarithmic, logistic, Weibull, Gompertz, power etc have been tried to predict the diameter and height growth of teak tree. MMF model was found better followed by Gompertz and Weibull for diameter prediction whereas for height growth prediction MMF model followed by logistic and Richards was found to be better. Rana et al (2008) examined the impact of climate change in recent years on apple shift to higher altitude in Himachal Pradesh based on climate information and farmers' perceptions. It is evident that temperature in apple growing regions of Himachal Pradesh showed increasing trends whereas precipitation showed decreasing trends in the region. The chilling units showed decreasing trends up to 2400 m elevation from Bajaura, Kullu (1221 m) to Sharbo, Kinnaur (2400 m). The Dhundi station situated at 2700 m elevation showed increasing trend of chilling units at the rate of 25.0 CUs per year. The increasing trend of chill units at 2700 m suggested that area was becoming suitable for apple cultivation in higher altitudes. These findings have also been supported by the farmers' perceptions which clearly reflected that apple cultivation was expanding to higher altitude in Lahaul and Spitti. The climate change has demonstrated its impact of decreasing productivity of apple crop in recent years.

CONCLUSION

Research investigation and analysis are the part of a wider development of any nation with regard to finance, education, public health, agriculture etc that are indicators of better life of human beings. Any social phenomenon and especially those that can be characterized by numerical facts are the results of one or more causes of action. Finally it is concluded and suggested that policy makers, research workers and foresters should use the best-fitted models for the prediction of area, production, yield, basal area and volume of forest trees.

REFERENCES

- Ali MA and Singh AK 1995. Growth and fluctuation in area, production and productivity of wheat in Chhattisgarh region of Madhya Pradesh. *Agricultural Situation in India* **59(9)**: 609-614.
- Ares A, St Louis D and Brauer D 2003. Trends in tree growth and understory yield in silvipastoral practices with southern pines. *Agroforestry Systems* **59(1)**: 27-33.
- Chengappa PG 1981. Growth rates of area, production and productivity of coffee in India. *Indian Journal of Coffee Research* **11(2)**: 19-26.
- Chethana HM 2009. Prediction models in teak-based agroforestry systems in northern transitional zone of Karnataka. MSc thesis, University of Agricultural Sciences, Dharwad, Karnataka, India.
- Devi PI, Thomas EK and Thomas JK 1990. Growth and supply response of banana in Kerala. *Agricultural Situation in India* **45(4)**: 239-242.
- Edward FW and Martin CL 1990. Stochastic life contingencies with solvency considerations. *Transactions of Society of Actuaries* **42**: 91-148.
- Fialar SC 1985. Analysis production pattern and marketing of cocoa in Ghana. MSc thesis, University of Agricultural Sciences, Bangalore, Karnataka, India.
- Gaddour A and Najari S 2008. Adjustment of the kid's growth curve in pure goat breeds and crosses under southern Tunisian conditions. *Journal of Applied Animal Research* **34(2)**: 117-120.
- Hald A 2003. A history of probability and statistics and their applications before 1750. *Wiley Series in Probability and Statistics*, Wiley-Interscience, John Wiley and Sons, Inc, Hoboken, New Jersey.
- Heckman JJ and Synder JM Jr 1977. Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators. *Rand Journal of Economics* **28**: S142-S189.
- Henry M 2011. Prediction of crop production in Zambia using census data. *Ossrea Journal* **11(2)**: 3-9.
- Kaushik KK 1993. Growth and instability of oilseeds production. *Indian Journal of Agricultural Economics* **48(3)**: 334-338.
- Kumar TV, Bhat ARS and Patil SJ 2005. Price forecasting of different classes of teak by the application of exponential smoothing model. *Karnataka Journal of Agricultural Sciences* **18(3)**: 55-59.
- Mohanty M, Painuli DK, Misra AK, Bandyopadhyaya KK and Ghosh PK 2006. Estimating impact of puddling, tillage and residue management on wheat (*Triticum aestivum* L) seedling emergence and growth in a rice-wheat system using non-linear regression models. *Soil and Tillage Research* **87(1)**: 119-130.
- Pandey R, Dhall SP, Kanwar BS and Bharadwaj SD 1998. Some models for predicting volume of *Populus deltoides*. *Indian Forester* **124(8)**: 629-632.
- Prasad H 2010. Statistical modeling technique on export of fruit crops in India. *Indian Journal of Agricultural Sciences* **74(5)**: 619-625.

- Rahaman MM and Ahmad LV 2000. Growth and yield prediction model of Jamoon (*Gmelina arborea*) in Chittagong hill tracks Bangladesh. *Journal of Tropical Forest Science* **12(2)**: 276-285.
- Raizada A, Rao MSRM, Nambiar KTN and Padmaiah M 2007. Biomass production and prediction models for *Acacia nilotica* in salt affected vertisols in Karnataka. *Indian Forester* **133(2)**: 239-246.
- Rana RS, Bhagat RM, Kalia V and Lal H 2008. Impact of climate change on shift of apple belt in Himachal Pradesh. ISPRS Archies XXXVIII-8/W3 Workshop Proceedings, Impact of Climate Change on Agriculture, pp 131–137.
- Rizvi RH and Khare D 2006. Prediction model for timber weight of *Populus deltoids* planted on farmlands in Haryana. *Indian Journal of Agroforestry* **8(1)**: 77-85.
- Sadasivan S 1989. Pattern of pulse production: an analysis of growth trends. *Economic and Political Weekly* **24(52)**: A167-A179.
- Shilpa, Sharma R, Sharma S, Sharma A and Nisha 2015. Trends and variability study of medicinal plants in Himachal Pradesh. *International Journal of Farm Sciences* **5(1)**: 149-162.
- Smith A 1904. An inquiry into the nature and causes of the wealth of nations. 5th edn, Methuen and Co Ltd, London.
- Sreekant RD 2008. Empirical prediction model for regional cashew production in India. *Journal of Plantation Crops* **36(3)**: 508-511.
- Swetha KS 2009. Statistical study on the impact of rain water harvesting on farming economy. MSc thesis, University of Agricultural Sciences, Dharwad, Karnataka, India.
- Tesfaye A 1994. Growth performance of some multipurpose trees and shrubs in the semi-arid areas of southern Ethiopia. *Agroforestry Systems* **26**: 237-249.
- Thomas WM 1998. Regression, prediction and model building. *Journal of Agricultural Sciences* **74(5)**: 418-423.
- Umapila S 1982. Oilseed economy of India: a case study of groundnut. Agricole Publication Co, NDP, pp 28-34.
- Wang J, Hou YS, Weilin LI and Cheng W 2000. Application of grey system theory to tree growth prediction. *Journal of Forestry Research* **11(1)**: 34-36.
- Yusuf SA and Salau AS 2007. Forecasting mango and citrus production in Nigeria: trend analysis. *Nigerian Agricultural Development Studies* **1(2)**: 1-19.