Mathematical Biosciences 249 (2014) 60–74

Contents lists available at ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

Modelling diameter distributions of two-cohort forest stands with various proportions of dominant species: A two-component mixture model approach

Rafał Podlaski [a,*], Francis A. Roesch [b]

[a] Department of Nature Protection, Institute of Biology, Jan Kochanowski University, ul. Świętokrzyska 15, 25-406 Kielce, Poland
[b] USDA Forest Service, Southern Research Station, 200 WT Weaver Blvd., Asheville, NC 28804-3454, USA

A R T I C L E I N F O

Article history:
Received 13 February 2013
Received in revised form 14 January 2014
Accepted 20 January 2014
Available online 31 January 2014

Keywords:
Two-component mixture model
Parameter estimation
Initial values
Weibull distribution
Gamma distribution
Kernel density estimator

A B S T R A C T

In recent years finite-mixture models have been employed to approximate and model empirical diameter at breast height (DBH) distributions. We used two-component mixtures of either the Weibull distribution or the gamma distribution for describing the DBH distributions of mixed-species, two-cohort forest stands, to analyse the relationships between the DBH components, age cohorts and dominant species, and to assess the significance of differences between the mixture distributions and the kernel density estimates. The data consisted of plots from the Świętokrzyski National Park (Central Poland) and areas close to and including the North Carolina section of the Great Smoky Mountains National Park (USA; southern Appalachians). The fit of the mixture Weibull model to empirical DBH distributions had a precision similar to that of the mixture gamma model, slightly less accurate estimate was obtained with the kernel density estimator. Generally, in the two-cohort, two-storied, multi-species stands in the southern Appalachians, the two-component DBH structure was associated with age cohort and dominant species. The 1st DBH component of the mixture model was associated with the 1st dominant species *sp*1 occurred in young age cohort (e.g., sweetgum, eastern hemlock); and to a lesser degree, the 2nd DBH component was associated with the 2nd dominant species *sp*2 occurred in old age cohort (e.g., loblolly pine, red maple). In two-cohort, partly multilayered, stands in the Świętokrzyski National Park, the DBH structure was usually associated with only age cohorts (two dominant species often occurred in both young and old age cohorts). When empirical DBH distributions representing stands of complex structure are approximated using mixture models, the convergence of the estimation process is often significantly dependent on the starting strategies. Depending on the number of DBHs measured, three methods for choosing the initial values are recommended: min.k/max.k, 0.5/1.5/mean, and multistart. For large samples (number of DBHs measured $\geqslant$80) the multistage method is proposed – for the two-component mixture Weibull or gamma model select initial values using the min.k/max.k (for $k = 1,5,10$) and 0.5/1.5/mean methods, run the numerical procedure for each method, and when no two solutions are the same, apply the multistart method also.

1. Introduction

Management and disturbances play an important role in the forest dynamics as well as in shaping the spatial and dimensional structure of forest stands (e.g., [62,18,37]). After cuttings and disturbances regeneration processes start in the gaps and under open stand canopy; and as a result, among others mixed-species, two-cohort, two-storied and partly multilayered stands are created. These resulting stands are usually characterised by strongly skewed and irregularly descending diameter at breast height (DBH) distributions. There are two general approaches to fitting empirical DBH distributions. The first approach is nonparametric and therefore does not require the estimation of parameters. The second, and usually preferred, approach is to identify an appropriate parametric distribution, such as e.g., the Weibull or the gamma distribution, and then estimate the unknown parameters. There are several reasons to prefer the latter approach, for instance, nonparametrically binning the data does not provide information beyond the range of the sample data, whereas some extrapolation is possible when a parametric model is applied [7].

* Corresponding author. Tel.: +48 41 3496322; fax: +48 41 3496292.
E-mail address: r_podlaski@pro.onet.pl (R. Podlaski).

http://dx.doi.org/10.1016/j.mbs.2014.01.007
0025-5564/

When modelling forest dynamics, one must determine the DBH distributions of tree species representing particular cohorts and stand layers (e.g., [5,58,11]), among other parameters. Procedures that allow determination of these parameters from DBH measurements alone, without associated assessments of tree age and height, are particularly valuable. If the overall DBH distribution of a stand is treated as a compound of the distributions of trees belonging to different groups (e.g., cohorts or stand layers) one may adopt a finite-mixture distribution approach (e.g., [73,69,71,46,47]). Mixture distributions are an appropriate tool for modelling heterogeneous populations (e.g., [12,60,35]). Because of their usefulness as an extremely flexible method of modelling, finite mixture models are continuing to receive increasing attention in forestry, from both practical and theoretical points of view. Indeed, in the past decade, the extent and the potential of the applications of finite mixture models have widened considerably [55].

The single and mixture Weibull and gamma models have often been used to approximate empirical DBH distributions because of their flexibility in shape (e.g., [20,32,73,29,72,46,47]). These distributions can conform to a wide variety of DBH data. Because the overall shape of the empirical DBH distribution is often composed of multiple basic shapes, a natural alternative is to utilise a mixture distribution for DBH modelling.

Computation of the parameters for a mixture model can be carried out by various numerical algorithms, such as the expectation–maximisation (EM) algorithm and the Newton-type methods (e.g., [9,36]). The Newton-type methods include quasi-Newton methods, modified Newton methods, etc. These numerical procedures can be decomposed into three main parts: initialisation (in which initial values for all parameters and a criterion to stop the algorithm should be chosen), iteration, and completion when the criterion is met. In general, starting from suitable initial parameter values, the iterations are repeated until convergence is achieved. If the likelihood function is regular, these methods usually find the most likely estimates for mixture parameters. However, if the likelihood function is irregular and has finitely or infinitely many local maxima and minima, the algorithms become extremely unstable. Unfortunately, this concern is a serious obstacle to interpreting the results when applied to separating finite mixtures. Therefore, when the likelihood function is not regular, a combination of the EM algorithm and the Newton-type method is often employed (the EM algorithm improves the initial values, and the Newton-type method is used then to estimate the parameters). In many cases, suitable initial values are difficult to ascertain, especially for empirical DBH distributions representing uneven-aged stands of complex structure. An evaluation of the usefulness of the various methods for choosing the initial values is very important. The appropriate strategies allow one to estimate the parameters of the mixture models and to construct accurate DBH models, especially in difficult situations, such as when the DBH components of mixture models overlap, in which case the global maximum may not be found or the estimation process may fail to converge.

The purposes of this study are (1) to verify the two hypotheses that (a) in mixed-species, two-cohort, two-storied and partly multilayered stands, two-component mixtures of either the Weibull distribution or the gamma distribution would be appropriate models for the DBH distributions; (b) in these models, the DBH components, representing age cohorts (and usually stand layers), can be associated with dominant species; (2) to compare four methods for choosing initial values for the numerical procedure for estimating the parameters of mixture models; (3) to propose a new strategy for maximising the likelihood during parameter estimation for mixture models; and (4) to assess the significance of differences between the parametric (two-component mixture distributions) and the nonparametric (kernel density estimation) methods.

2. Material and methods

2.1. Study area

The plots investigated here were randomly sampled in mixed-species, two-cohort, two-storied and partly multilayered stands in which DBH distributions of two main age cohorts are partially overlapping.

The plots were located in the Święta Katarzyna and Święty Krzyż forest sections of the Świętokrzyski National Park (Poland; Świętokrzyskie Mountains; geographical coordinates: 50°50′–50°53′N, 20°48′–21°05′E); and in areas close to and including the North Carolina section of the Great Smoky Mountains National Park (USA; southern Appalachians; geographical coordinates: 34°59′–36°32′N, 78°43′–84°13′W).

In the Świętokrzyskie Mountains naturally regenerated near-natural forests chosen for this study are composed of native tree species. Soils are Distric Cambisols and Haplic Luvisols (subtypes according to Food and Agriculture Organization, International Soil Reference and Information Centre, and International Soil Science Society, [15]). Long-term mean annual temperature was 6 °C, mean January and July temperatures were −5 °C and 16 °C; the growing season was ca. 182 days (data from the Święty Krzyż meteorological station at 575 m a.s.l.). The highest temperatures and the highest precipitation usually occur in summer, in the middle of the growing season. Three associations occur: *Dentario glandulosae-Fagetum*, *Abietetum polonicum* and *Querco roboris-Pinetum* (nomenclature after [33]).

In the southern Appalachians, variations in elevation, rainfall, temperature, and geology provide habitat for nearly 1600 species of flowering plants, including 100 native tree species and over 100 native shrub species [38]. Great Smoky Mountains National Park contains some of the largest tracts of wilderness in the Eastern United States, including 66 species of mammals, over 200 varieties of birds, 50 native fish species, and more than 80 types of reptiles and amphibians. The study area is part of the Unaka Range a sub-range in the Appalachian chain, ranging in elevation from about 300–2040 m a.s.l. The climate and precipitation vary greatly in relation to elevation and landscape position. The precipitation averages from 1200 mm annually to approximately 2500 mm at the highest elevations. High precipitation and cool temperatures at the higher elevations produce brown, medium textured soils that have a high content of organic matter in the surface layer. The warmer temperatures at the lower elevations produce soils that are redder and that contain more clay in the subsoil. Mean January temperatures range from 2 °C to 10 °C and mean July temperatures range from 18 °C to 31 °C [61].

2.2. Field measurements

In the Świętokrzyskie Mountains, eleven 0.25 ha plots were measured in 2008 and 2009. In the southern Appalachians, nineteen 0.067 ha plots were measured from 2003 to 2008; data were selected from the USDA Forest Service's Forest Inventory and Analysis database (documented in [68]). The DBH of all live trees greater than 6.9 cm in the Świętokrzyskie Mountains and 12.9 cm in the southern Appalachians in diameter was measured.

2.3. Data analysis

The stands investigated were categorised in three groups. Group 1 (two species stands) consisted of 11 stands that were strongly dominated by two species, from the Świętokrzyski National Park. Group 2 (multi-species stands with two main species) consisted of 10 stands that were medium dominated by two