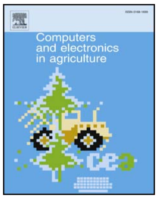
Computers and Electronics in Agriculture 144 (2018) 190–198

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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



Original papers

The gamma shape mixture model and influence of sample-unit size on estimation of tree diameter distributions: Forest modelling

Rafał Podlaski[a,*], Dariusz Wojdan[a], Monika Żelezik[b]

[a] Department of Nature Protection and Plant Physiology, Institute of Biology, Jan Kochanowski University, ul. Świętokrzyska 15, 25-406 Kielce, Poland
[b] Chair of Environment Shaping and Protection, Jan Kochanowski University, ul. Świętokrzyska 15, 25-406 Kielce, Poland

ARTICLE INFO

Keywords:
Gamma shape mixture (GSM) model
Gamma distribution
Survey data
Optimal sample unit size
Complex and stratified stands
Natural and near-natural forests

ABSTRACT

The gamma shape mixture (GSM) model is based on a mixture of gamma distributions. A general Bayesian approach for estimating the unknown parameters of the GSM model was employed. Tree diameter at breast height (DBH) sets of complex and stratified forests are ideal data sources for testing the usefulness of theoretical distributions applied to modelling and simulating strongly differentiated data sets. The GSM model was useful in fitting these DBH structures. During the identification of homogenous forest patches of the similar DBH structures one should select sample plots of at least 0.15 ha in area, with a minimum of 30 trees. The use of GSM model has the potential to facilitate the presentation of not only DBH structures, but also other empirical distributions, describing e.g. various biological stages and processes. The GSM model can be employed especially for modelling multimodal, asymmetrical and heavy-tailed survey data.

1. Introduction

Forest stands composed of trees of all ages and with multi-layered canopies (complex and stratified stands) have the most complicated structure that is significant for environmental interactions, regeneration, growth, and biotic habitat (e.g., Parker and Brown, 2000). Many structural and environmental variables have actual or potential connections with canopy functions and they can shape the vertical forest structure. The dynamics of the temperate complex and stratified forests usually is presented as birth, growth, and death of forest patches rather than of individual trees; it is described as a sequence of stages and phases forming a developmental cycle (e.g., Shugart and West, 1977; Weishampel and Urban, 1996; Nakashizuka, 2001). The area of patches is more or less equal to the area of gaps created by disturbances. The sum of all disturbances and their interactions is referred to as the tree community's "disturbance regime" (Runkle, 1985). Disturbances are caused in the first place by abiotic (e.g., wind and low temperatures in winter) and biotic (e.g., insect outbreaks and fungal infections) factors. The patches in different stages and phases were being distinguished mainly on the basis of vertical forest structure (e.g., Korpeľ, 1982, 1995; Leibundgut, 1993). In the uneven-aged forests tree height and tree age can usually be translated into tree diameter; that is why forest patches are often described using diameter at breast height (DBH) distribution (e.g., von Oheimb et al., 2005; Westphal et al., 2006). For shade-tolerant tree species, forming the uneven-aged forests, generally, there is a strong relationship between height, age and DBH of trees. The highest and the oldest trees usually have the largest DBHs.

In the near-natural, uneven-aged, complex and stratified forests, the shape of the DBH distribution can be negative–exponential, rotated sigmoid or multimodal, asymmetrical, they often show a positive skewness, that is, asymmetry towards the right (Goff and West, 1975; West et al., 1981; Parker et al., 1985; Goodburn and Lorimer, 1999; Piovesan et al., 2005; Zasada, 2013). Moreover, DBH distributions of this type are characterised by variable local irregularity with local maxima (Podlaski, 2008). These extremes may have a random character, or may be connected with the significant, from a biological point of view, tree generations, which in a given patch are represented only by several trees. In the first case one should apply a theoretical function that removes local irregularities, while in the second one it would be advisable to use a flexible model that approximates local maxima.

Modelling of DBH distributions in complex and stratified stands usually requires the use of functions enabling the detailed approximation of several extremes. Of the parametric methods the best one is to use a mixture of probability density functions (Zhang et al., 2001; Zasada and Cieszewski, 2005; Podlaski, 2011a,b; Zasada, 2013). A very promising model of that type is the gamma shape mixture (GSM) model (Venturini et al., 2008). A particularly important advantage of this model is a possibility to use a great number of mixture components. A general Bayesian approach, allows to create a flexible function characterised by the set of mixture weights and a single scale parameter for

---

[*] Corresponding author.
E-mail address: r_podlaski@pro.onet.pl (R. Podlaski).

https://doi.org/10.1016/j.compag.2017.12.009
Received 14 September 2017; Received in revised form 28 November 2017; Accepted 7 December 2017

all the gamma components (Jasra et al., 2005; Venturini et al., 2008). The use of a different number of mixture components enables modelling DBH distributions with an adequate precision to smooth local irregularities, needed in certain type of studies, or approximating local maxima.

An appropriate sample size is a fundamental requirement to distinguish correctly different stages and phases and next, characterise patch mosaic in the complex and stratified stands (e.g., Bobiec et al., 2000; Winter and Brambach, 2011; Král et al., 2014). The use of the no optimal area of sample plots is cost-effective and cost-efficient. The relationship between sample size and model-based inference has been partially explored in several recent studies; generally, the prediction error increases as the area of sample plots diminishes (Zeide, 1980; Williams et al., 2001; Gray, 2003; Lombardi et al., 2015). The optimal area of sample plots changes in dependence on forest structure; therefore, the selection of an optimal area requires an understanding of DBH data variability. The greatest problem is to establish a suitable size of sample plots for the complex and stratified stands because in this type of stands patches are composed of at least several tree generations and show the much differentiated shape of DBH distributions (Rubin et al., 2006; Alessandrini et al., 2011). The use of fixed-size plots is outlined as one of the most effective options to describe the DBH structures (e.g., Gregoire and Valentine, 2008; Lombardi et al., 2015).

The purposes of this study are to (1) examine the consequence of using a different DBH sample unit size for the approximation of patch empirical DBH data, and (2) determine the optimal area of sample plots needed to adequately fit and then simulate random DBH data from the GSM model.

2. Methods

2.1. Study sites

The investigations were carried out in the Świętokrzyski National Park (Central Poland; 50°50′–50°53′N, 20°48′–21°05′E). The main soil types are Distric Cambisol and Haplic Luvisol (subtypes according to IUSS Working Group WRB (2006)). Mean annual precipitation ranges from 700 to 850 mm (Olszewski et al., 2000). The study area is dominated by the following plant associations: *Dentario glandulosae-Fagetum* Klika 1927. em. W. Mat. 1964, and *Abietetum polonicum* (Dziub. 1928). Br.-Bl. & Vlieg 1939 (in the Święty Krzyż forest section), as well as *Tilio Carpinetum abietetosum* with *Larix decidua* Mill. subsp. *polonica* (Racib.) Domin (in the Chełmowa Góra forest section) (nomenclature after Matuszkiewicz (2008)). The main forest-forming species are: fir *Abies alba* Mill., beech *Fagus sylvatica* L., larch *Larix decidua* Mill. subsp. *polonica* (Racib.) Domin, as well as oaks *Quercus robur* L. and *Q. petraea* [Matt.] Liebl.. The stands are affected mainly by fine-scale disturbances (below 0.04 ha in size) and intermediate-scale disturbances (0.04–0.50 ha in extent), predominating in these areas (Podlaski, 2008). Disturbances and other natural processes form forest patches not exceeding 1.5 ha in area.

2.2. Field work

In the years 2002–2013 four plots (SKRZ1, SKRZ2, CHMG1, and CHMG2), each 0.5 ha in area (83.33 m × 60.00 m), representing near-natural, uneven-aged, complex and stratified forests, were randomly selected. Plots SKRZ1 and SKRZ2 were situated in the Święty Krzyż forest section, in the area of the nature reserve established in 1924, plots CHMG1 and CHMG2 were situated in the Chełmowa Góra forest section in the nature reserve established in 1921. At the present time these two reserves are included in the Świętokrzyski National Park. In this area, there occur the very well-preserved patches of natural forests and natural plant communities. Since the establishment of the reserves these sites have been strictly protected, including no felling or removal of deadwood.

The plots were situated within boundaries of the homogenous forest patches of the similar vertical stand structure. The longer side of each plot has been traced more or less parallel to contour lines. While the plots were created, the stages and phases of forest development were determined. The main criteria used to determine the stages and phases were (Korpeľ, 1982): (1) vertical stand structure, (2) tree age distribution, and (3) tendency in volume increment (increasing or decreasing). A fraction of trees at variant canopy position is a useful information for vertical stand structures; the number of trees was analysed in three main stories: (1) upper ($> 2/3\ h_{max}$), (2) middle (between $1/3\ h_{max}$ and $2/3\ h_{max}$), and (3) lower ($< 1/3\ h_{max}$). Tree age was measured in individual layers in increment cores from several trees in each main story. The tendency in volume increment was derived from age distribution and vertical structure. When estimating tree age two main generations were distinguished: (1) in the SKRZ1 and the SKRZ2 an older generation was formed by fir and beech trees between 100 and 300 years and a younger generation containing fir and beech trees to 100 years, (2) in the CHMG1 an older generation was composed of larch trees between 130 and 330 years and a younger generation of fir and beech trees to 100 years, and (3) in the CHMG2 an older generation was consisted of oaks, fir, and beech trees between 120 and 300 years and a younger generation of beech trees to 80 years.

In each plot all living trees > 6.9 cm in diameter at 1.3 m above ground were marked. Collection of DBH sets above this minimum diameter threshold is common practice in forest ecosystem research. Measurement of trees below this threshold adds field time and provides relatively little ecological data significant at the level of the stand. In addition, these trees often have high turnover rates making them difficult to track over time on remeasured plots, resulting in a significant increase in effort searching for missing trees and tagging new cohorts at each periodic measurement (e.g., Mcgarrigle et al., 2011). Then, the DBHs were measured and tree locations were determined using a total station (Spectra Precision Focus 4).

2.3. Data analysis

To approximate DBH data the GSM model was used. This model has a probability density function (PDF) given by (Lehmann and Casella, 1998; Venturini et al., 2008):

$$f(x|\pi_1,...,\pi_J,\theta) = \sum_{j=1}^{J} \pi_j f_j(x|\theta) \tag{1}$$

with the gamma distribution:

$$f_j(x|\theta) = \frac{\theta^j}{\Gamma(j)} x^{j-1} e^{-\theta x} \tag{2}$$

where x is DBH in cm, J is a number of mixture components (known and fixed); $\pi_1,...,\pi_J$ is a vector of mixture weights (unknown), $1/\theta$ is a scale parameter for the whole GSM model (unknown), each gamma distribution in the GSM model is indexed by a component-specific shape parameter (j), $\Gamma(\cdot)$ is the gamma function.

A general Bayesian approach for estimating the unknown parameters of the GSM model is used. Fitting with the GSM model requires three hyperparameters: the J, and the α and β from the conjugate prior on θ (Venturini et al., 2008):

$$\pi_1,...,\pi_J \sim D_J\left(\frac{1}{J},...,\frac{1}{J}\right) \tag{3}$$

and

$$\theta \sim G(\alpha,\beta) \tag{4}$$

where $D_J(\cdot)$ is a Dirichlet distribution and $G(\cdot)$ is a gamma distribution.

During the approximation of the empirical DBH data using the GSM model, the value of $J = 250$ and the weight of the prior information $\omega = 0.35$ were assumed. With these assumptions the α and β values were