

RAFAŁ PODLASKI, FRANCIS A. ROESCH

Aproksymacja rozkładów pierśnic w dwugeneracyjnych drzewostanach za pomocą rozkładów mieszanych.

I. Estymacja parametrów

Approximation of the breast height diameter distribution of two-cohort stands by mixture models. I. Parameter estimation

ABSTRACT

Podlaski R., Roesch F. A. 2013. Aproksymacja rozkładów pierśnic w dwugeneracyjnych drzewostanach za pomocą rozkładów mieszanych. I. Estymacja parametrów. Sylwan 157 (8): 587-596.

Study assessed the usefulness of various methods for choosing the initial values for the numerical procedures for estimating the parameters of mixture distributions and analysed variety of mixture models to approximate empirical diameter at breast height (dbh) distributions. Two-component mixtures of either the Weibull distribution or the gamma distribution were employed. The study plots, representing two-cohort stands, were located in the Świętokrzyski National Park (central Poland) and in the Southern Appalachian Mountains (eastern USA). A new strategy using three methods for choosing initial values (min. k /max. k for $k=1, 5, 10$; 0,5/1,5/mean; wp) for maximizing the likelihood during parameter estimation for mixture models for small and large plots is proposed.

KEY WORDS

two-component mixtures, tree diameter distribution, Weibull distribution, gamma distribution, starting strategy

ADDRESSES

Rafał Podlaski ⁽¹⁾ – e-mail: r_podlaski@pro.onet.pl

Francis A. Roesch ⁽²⁾ – e-mail: froesch@fs.fed.us

⁽¹⁾ Zakład Ochrony Przyrody; Uniwersytet Jana Kochanowskiego; ul. Świętokrzyska 15; 25-406 Kielce

⁽²⁾ Southern Research Station, USDA Forest Service; 200 WT Weaver Blvd., Asheville, NC 28804-3454 USA

Wstęp

Drzewostany dwugeneracyjne powstają zarówno w naturalnym cyklu rozwoju lasów o charakterze pierwotnym, jak i w wyniku stosowania różnych rębni [Bernadzki 1994; Korpel 1995; Jaworski 1997]. Mogą to być lasy o zbliżonej postaci na stosunkowo rozległych obszarach lub lasy charakteryzujące się większym zróżnicowaniem przestrzennym, w których obok siebie występują płaty reprezentujące różne gatunki drzew [Korpel 1995; Olano, Palmer 2003; Jaworski, Podlaski 2007]. Jednym z ważniejszych czynników wpływających na dynamikę tych lasów są zaburzenia o różnej skali, generujące luki o różnej wielkości [Runkle 1985].

Rozkłady pierśnic charakterystyczne dla dwugeneracyjnych drzewostanów są zazwyczaj dwumodalne, z dodatkowymi, niewielkimi ekstremami powstałymi głównie w wyniku wydzielania się kilku drzew o zbliżonej grubości lub wskutek intensywnego przyrostu grupy drzew w luce. Często rozkłady dla młodszej i starszej generacji stopniowo zachodzą na siebie, co powoduje przekształcenie rozkładów dwumodalnych w rozkłady obrócone sigmoidalne lub jednomodalne [Westphal i in. 2006; Podlaski 2010]. Do modelowania tego typu rozkładów stosowano

m.in. pojedyncze rozkłady teoretyczne. Najlepsze wyniki osiągnięto, stosując elastyczne rozkłady, np. Weibulla, gamma czy rozkłady z rodziny Burra (typ III i XII) [Zasada 1995; Rymer-Dudzińska, Dudzińska 1999, 2001; Gove i in. 2008]. Rozkłady pojedyncze nie oddają jednak dwumodalności rozkładów rzeczywistych. W celu wyeliminowania tych ograniczeń wykorzystano rozkłady mieszane [Zhang, Liu 2006; Podlaski 2011a, b; Jaworski, Podlaski 2012]. W Polsce prekursorami stosowania rozkładów mieszanych w naukach leśnych byli Siekierski [1991] i Zasada [2003], a w Stanach Zjednoczonych m.in. Zhang i in. [2001] czy Liu i in. [2002]. Analiza rozkładów pierśnic ma szereg zastosowań teoretycznych i praktycznych. Jest wykorzystywana do modelowania struktury grubości drzewostanów oraz jest jednym z etapów budowy modeli wzrostu drzew [Bruchwald 2001; Bruchwald, Zasada 2010; Bruchwald i in. 2011].

Rozkłady mieszane mają skomplikowaną funkcję gęstości, określoną przez przynajmniej kilka parametrów. Standardową procedurą estymacji tych parametrów jest wykorzystanie metody największej wiarygodności (ang. maximum likelihood estimation; MLE). Funkcja wiarygodności ma złożoną, wielowymiarową postać. Prawidłowa estymacja parametrów rozkładu mieszanego wymaga znalezienia jej globalnego ekstremum. W tym celu stosowane są metody numeryczne, najczęściej algorytm EM (ang. expectation-maximization) i metody Newtona [Dempster i in. 1977; McLachlan, Krishnan 2008]. Procedury numeryczne rozpoczynają poszukiwanie ekstremum globalnego od zadanych z góry wartości startowych, a następnie powtarzają określone iteracje aż do spełnienia założonych kryteriów końcowych. Nie zawsze po zakończeniu iteracji znajdowane jest ekstremum globalne i nie zawsze uzyskuje się zbieżność procedury numerycznej. W przypadku nieregularnej postaci funkcji wiarygodności z wieloma ekstremami lokalnymi (także nieskończonymi) i punktami przegięcia algorytmy numeryczne są niestabilne. W tej sytuacji kluczowe znaczenie ma wybór wartości startowych. Stosowanych jest wiele metod określania wartości startowych, a różnice między nimi sprowadzają się m.in. do sposobu określenia przestrzeni początkowej oraz do stopnia skomplikowania algorytmu ich wyboru [Böhning i in. 1994; Seidel i in. 2000a, b].

Celem pracy jest: (1) określenie wpływu różnych wartości startowych stosowanych podczas inicjowania procedury numerycznej (algorytm EM + metoda Newtona) na prawidłową estymację parametrów badanych rozkładów mieszanych oraz (2) ocena uniwersalności badanych rozkładów mieszanych na podstawie przeprowadzonych aproksymacji rzeczywistych rozkładów pierśnic.

Material i metody

Powierzchnie badawcze zostały założone w Leśnictwie Święta Katarzyna i Święty Krzyż w Świętokrzyskim Parku Narodowym (Góry Świętokrzyskie; 50°50'-50°53'N, 20°48'-21°05'E) oraz w Północnej Karolinie, na obszarze obejmującym m.in. fragment Parku Narodowego Great Smoky Mountains (Appalache Południowe; 34°59'-36°32'N, 78°43'-84°13'W). W Górach Świętokrzyskich dominują lasy z jodłą *Abies alba* i bukiem *Fagus sylvatica*. Na terenach badanych w Appalachach Południowych występują lasy wielogatunkowe z udziałem sosen (*Pinus strobus*, *P. taeda*, *P. virginiana*), dębów (*Quercus falcata*, *Q. alba*, *Q. rubra*, *Q. coccinea*, *Q. prinus*) oraz choiny *Tsuga canadensis*, magnolii *Magnolia fraseri*, ambrowca *Liquidambar styraciflua*, klonu *Acer rubrum*, tulipanowca *Liriodendron tulipifera*, orzesznika *Carya glabra*, czeremchy *Prunus serotina*, brzozy *Betula lenta*, kwaśnodrzewu *Oxydendrum arboreum* i grochodrzewu *Robinia pseudoacacia*.

W Górach Świętokrzyskich założono w latach 2008-2009 11 powierzchni o wielkości 0,25 ha każda. W Appalachach Południowych wykorzystano dane zebrane w latach 2003-2008 na 19 powierzchniach, z których każda miała 0,067 ha powierzchni. Na każdej powierzchni zmierzono pierśnicę wszystkich żywych drzew. W Górach Świętokrzyskich uwzględniono drzewa o pierśnicy

równej i większej od 7 cm, natomiast w Appalalach Południowych za próg pomiaru przyjęto 13 cm.

W celu oceny uniwersalności rozkładów mieszanych podczas modelowania rzeczywistych rozkładów pierśnic zastosowano rozkłady mieszane złożone z dwóch rozkładów składowych Weibulla i dwóch rozkładów składowych gamma. Funkcje gęstości pojedynczego rozkładu Weibulla i gamma mają odpowiednio następującą postać:

$$f_{(Weib)}(x|\alpha, \beta, \gamma) = \frac{\alpha}{\beta} \left(\frac{x-\gamma}{\beta} \right)^{\alpha-1} e^{-\left(\frac{x-\gamma}{\beta}\right)^\alpha} \quad [1]$$

$$f_{(gam)}(x|\alpha, \beta, \gamma) = \frac{\beta^\alpha (x-\gamma)^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta(x-\gamma)} \quad [2]$$

gdzie:

parametr kształtu α oraz parametr skalujący $\beta > 0$, a $x \geq$ od parametru przesunięcia γ oraz $\Gamma(\cdot)$ – funkcja gamma.

Funkcje gęstości rozkładu mieszanego złożonego z dwóch rozkładów składowych Weibulla lub gamma można przedstawić jako:

$$f_{(Weib)}(x|\psi) = \pi_1 f_{(Weib)}(x|\theta_1) + \pi_2 f_{(Weib)}(x|\theta_2) \quad [3]$$

$$f_{(gam)}(x|\psi) = \pi_1 f_{(gam)}(x|\theta_1) + \pi_2 f_{(gam)}(x|\theta_2) \quad [4]$$

gdzie:

π_1 i π_2 są wagami (frakcjami) rozkładów składowych $f_{(1)}(\cdot)$ i $f_{(2)}(\cdot)$, θ_1 i θ_2 oznaczają wektory parametrów dla odpowiednich rozkładów składowych, a ψ to zestaw wszystkich parametrów danego rozkładu mieszanego.

Dla rozkładów mieszanych Weibulla i gamma wektor parametrów $\theta_j = (\alpha_j, \beta_j)$ jest często zastępowany przez wektor $\theta_j^* = (\mu_j = f_{\mu}(\alpha_j, \beta_j), \sigma_j = f_{\sigma}(\alpha_j, \beta_j))$, gdzie f_{μ} i f_{σ} są funkcjami pozwalającymi na obliczenie średniej μ_j i odchylenia standardowego σ_j przy wykorzystaniu wartości parametrów α_j i β_j [Du 2002]. To przekształcenie umożliwi generowanie wartości startowych z przestrzeni danych rzeczywistych.

Obliczenia parametrów rozkładów mieszanych wykonano przy pomocy MLE, wykorzystując do znalezienia ekstremum funkcji wiarygodności algorytm EM połączony z metodą Newtona [Böhning 2000]. Zastosowana procedura numeryczna ma za zadanie znalezienie minimum globalnego funkcji wiarygodności postaci [Du 2002]:

$$LL(\psi) = -2 \sum_{j=1}^l n_j \log \left(\frac{P_j(\psi)}{O_j} \right), \quad O_j = \frac{n_j}{N} \quad [5]$$

gdzie:

$P_j(\psi)$ – oznacza teoretyczne prawdopodobieństwo zdarzenia, że dana pierśnica należy do j -tego stopnia grubości,

O_j – empiryczna częstość względna dla j -tego stopnia,

l – liczba stopni grubości.

Wartości parametrów, dla których powyższa funkcja wiarygodności osiąga minimum globalne, są szukanymi parametrami rozkładu mieszanego.

W celu analizy wpływu różnych wartości startowych na znalezienie minimum globalnego funkcji wiarygodności wykorzystano następujące metody ich określania [McLachlan, Krishnan 2008; Podlaski 2011a,b]:

- min. k /max. k – procedura numeryczna rozpoczyna estymację parametrów od wartości startowych $\mu_1^0 = \min.k$, $\sigma_1^0 = s$ i $\pi_1^0 = 0,5$ dla pierwszego rozkładu składowego oraz $\mu_2^0 = \max.k$, $\sigma_2^0 = s$ i $\pi_2^0 = 0,5$ dla drugiego (min. k , max. k i s to odpowiednio najmniejsza i największa k -ta pierśnica oraz odchylenie standardowe wszystkich zmierzonych pierśnic), za min. k i max. k przyjęto kolejno: min.1 i max.1, czyli najmniejszą i największą pierśnicę, następnie min.5 i max.5, czyli 5. najmniejszą i 5. największą i wreszcie min.10 i max.10, czyli 10. najmniejszą i 10. największą.
- 0,5/1,5/średnia – estymacja parametrów rozpoczyna się od wartości startowych $\mu_1^0 = 0,5m$, $\sigma_1^0 = s$ i $\pi_1^0 = 0,5$ dla pierwszego rozkładu składowego oraz $\mu_2^0 = 1,5m$, $\sigma_2^0 = s$ i $\pi_2^0 = 0,5$ dla drugiego (m – średnia zmierzona pierśnica).
- gat – estymacja parametrów rozpoczyna się od wartości startowych $\mu_1^0 = m_{sp1}$, $\sigma_1^0 = s_{sp1}$ i $\pi_1^0 = w_{sp1}$ dla pierwszego rozkładu składowego oraz $\mu_2^0 = m_{sp2}$, $\sigma_2^0 = s_{sp2}$ i $\pi_2^0 = w_{sp2}$ dla drugiego (m_{sp1} , m_{sp2} , s_{sp1} , s_{sp2} , w_{sp1} , w_{sp2} to odpowiednio średnia, odchylenie standardowe i frakcja zmierzonych pierśnic dla dwóch gatunków głównych charakteryzujących się największym udziałem pierśnicowego pola przekroju, przy czym $m_{sp1} < m_{sp2}$).
- metoda wielopunktowa w_p – przestrzeń zmierzonych pierśnic zostaje pokryta siatką punktów $u_1 \equiv \min$, $u_2 = u_1 + (u_1 - u_{10})/9$, ..., $u_{10} \equiv \max$ (min i max to odpowiednio najmniejsza i największa zmierzona pierśnica), a estymacja parametrów rozpoczyna się od wartości startowych $\mu_1^0 = u_1$, $\sigma_1^0 = s$ i $\pi_1^0 = 0,5$ dla pierwszego rozkładu składowego oraz $\mu_2^0 = u_2$, $\sigma_2^0 = s$ i $\pi_2^0 = 0,5$ dla drugiego. Po zakończeniu estymacji dla u_1 i u_2 rozpoczyna się ponownie estymację dla u_1 i u_3 , i tak aż do estymacji dla u_9 i u_{10} . Łącznie dla dziesięciu punktów pokrywających przestrzeń danych zostaje wykonanych 45 estymacji. Za maksimum globalne przyjmujemy wartość, która była największa dla funkcji wiarygodności $LL_1(\psi)$, co jest równoznaczne z wartością najmniejszą dla funkcji wiarygodności $LL_2(\psi)$ i przynajmniej dwukrotnie wystąpiła w zbiorze wyników.

Podczas estymacji parametrów dla każdego rozkładu stosowano kolejno wszystkie cztery metody określania wartości startowych. Parametry rozkładów obliczano, wykorzystując pakiet *mixdist* [Macdonald, Du 2004] wchodzący w skład środowiska R (<http://www.Rproject.org>).

Wyniki

W Górach Świętokrzyskich zdecydowanie przeważają jodła i buk. Udział tych dwóch gatunków, określony na podstawie pierśnicowego pola przekroju, kształtował się na badanych powierzchniach od 92,1 do 100,0%. W skład poszczególnych generacji wiekowych wchodziły na ogół jodły i buki. Na niektórych powierzchniach w generacji młodszej dominował buk, natomiast w generacji starszej jodła. Pole powierzchni przekroju pierśnicowego wynosiło w przypadku jodły od 3,3533 (ŚPN26) do 31,9877 m²/ha (ŚPN24), natomiast w przypadku buka od 1,5689 (ŚPN30) do 26,4403 m²/ha (ŚPN26). Jodła największym udziałem odznaczała się na 8 powierzchniach, buk zaś – na 3. Dla wszystkich gatunków razem pierśnicowe pole przekroju kształtowało się od 11,5802 (ŚPN30) do 46,1221 m²/ha (ŚPN31). Liczba drzew na powierzchniach badawczych wynosiła od 93 (ŚPN24) do 188 (ŚPN21). W Appalachach Południowych skład gatunkowy badanych drzewostanów był znacznie bardziej zróżnicowany. Udział gatunków głównych, określony na podstawie pierśnicowego pola przekroju, kształtował się na badanych powierzchniach dla generacji młodszej od 2,8 do 47,4%, natomiast dla generacji starszej od 22,4 do 83,8%. Pierśnicowe pole przekroju wynosiło w przypadku generacji młodszej od 0,8339 (*Carya glabra*, powierzchnia 183084) do 10,9484 m²/ha (*Liriodendron tulipifera*, powierzchnia 193028), a w przypadku generacji

starszej od 2,4515 (*Oxydendrum arboreum*, powierzchnia 39065) do 26,2673 m²/ha (*Pinus taeda*, powierzchnia 37022). W drzewostanach wielogatunkowych często występuje kilka lub nawet kilkanaście gatunków o zbliżonym udziale i dlatego pole powierzchni przekroju pierśnicowego gatunków niebędących gatunkami głównymi wynosiło od 2,3055 (powierzchnia 37022) do 22,3932 m²/ha (powierzchnia 99043). Liczba drzew na powierzchniach badawczych kształtowała się od 21 (powierzchnia 39065) do 48 (powierzchnia 193028).

Średnia pierśnica drzew wynosiła w Górach Świętokrzyskich od 14,7 do 30,6 cm, a w Appalachach Południowych od 17,8 do 29,1 cm (tab. 1). Większą zmiennością pierśnicy charakteryzowały się powierzchnie z Gór Świętokrzyskich. Odchylenie standardowe dla tych powierzchni kształtowało się od 8,8 do 21,3 cm, natomiast dla powierzchni z Appalachów przybierało wartości od 5,4 do 17,1 cm (tab. 1). W Górach Świętokrzyskich najgrubsze drzewa cechowały się pierśnicą dochodzącą do 90 cm, natomiast w Appalachach – do 78 cm (tab. 1). Dla wszystkich badanych rozkładów pierśnic drzew współczynnik skośności przyjmował wartości większe od zera (tab. 1).

Tabela 1.

Charakterystyka pierśnicy dla badanych drzewostanów
Characteristics of breast height diameters of the investigated stands

Powierzchnia	Średnia [cm]	Odchylenie standardowe [cm]	Maksimum [cm]	Współczynnik skośności	Współczynnik spłaszczenia
ŚPN21	21,4	14,3	90	2,7717	9,5660
ŚPN22	23,4	13,7	56	0,5068	-0,9760
ŚPN23	22,3	12,6	55	0,6999	-0,5578
ŚPN24	30,6	21,3	82	0,8641	-0,3620
ŚPN25	22,8	15,8	62	0,6484	-0,9468
ŚPN26	26,2	15,7	71	0,7722	-0,1107
ŚPN27	29,0	14,5	66	0,4148	-0,5883
ŚPN28	21,9	15,2	73	1,7039	2,7817
ŚPN29	17,1	8,8	45	1,3657	1,3139
ŚPN30	14,7	11,2	58	2,3815	5,1172
ŚPN31	25,8	14,0	78	1,1084	1,4168
3028	22,8	9,1	40	0,5979	-1,1276
37022	23,8	10,1	47	0,5818	-0,9525
75045	24,0	9,7	52	1,4665	2,4864
71025	17,8	8,4	56	3,4107	13,7273
183084	20,3	9,0	48	1,7085	2,5667
111039	28,6	14,0	77	1,9452	5,3044
57049	24,1	8,5	41	0,1449	-1,1600
11025	28,8	13,4	68	1,7144	2,7763
23057	22,8	9,5	48	1,2741	0,8142
75028	25,7	11,4	71	2,3049	7,9045
99043	29,1	17,1	78	1,2814	0,9427
71013	19,0	7,9	50	2,5768	6,9774
113079	20,6	7,9	45	1,7697	2,7008
115045	24,2	13,8	67	1,8357	3,1242
199016	26,9	13,3	53	0,7588	-0,8065
23046	22,1	8,5	41	0,8995	-0,3141
27013	25,6	13,0	54	0,8890	-0,3608
39065	19,7	7,7	49	2,8710	10,5005
193028	19,6	5,4	32	0,6676	-0,6174

Największe asymetryczności wykazywały rozkłady z powierzchni ŚPN21 dla Gór Świętokrzyskich oraz z powierzchni 71025 dla Appalachów (tab. 1). W porównaniu do rozkładów z Appalachów rozkłady z Gór Świętokrzyskich były na większości powierzchni bardziej spłaszczone (mniejsze wartości współczynnika spłaszczenia, tab. 1).

Metoda $\min.k/\max.k$ umożliwiła prawidłową estymację parametrów dla 25 powierzchni w przypadku każdego z badanych modeli (tab. 2). Metoda 0,5/1,5/średnia pozwoliła na znalezienie globalnego minimum funkcji wiarygodności dla 19 i 15 powierzchni w przypadku rozkładu mieszanego Weibulla i gamma, natomiast metoda gat dla 18 i 16 powierzchni. Z kolei metoda \wp umożliwiła prawidłową estymację parametrów dla 29 i 30 powierzchni. Jeżeli dana metoda wyboru wartości startowych zawodziła, to wynikiem jej zastosowania był albo brak zbieżności procedury numerycznej, albo znalezienie minimum lokalnego. Tego typu rezultaty uzyskano

Tabela 2.

Pierwsza z metod wyboru wartości startowych algorytmu estymującego parametry modeli, dla której osiągnięto globalne minimum funkcji wiarygodności

The first out of applied methods of choosing the initial values for the algorithm estimating the parameters of models for which global minimum of likelihood function was reached

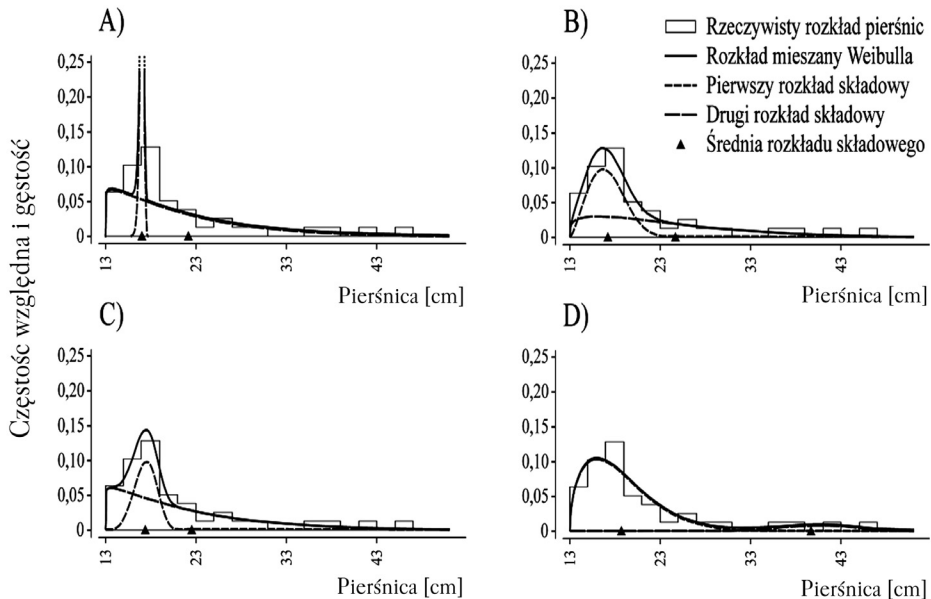
Powierzchnia	Rozkład mieszany Weibulla		Rozkład mieszany gamma	
	Metoda wyboru	Minimum globalne	Metoda wyboru	Minimum globalne
ŚPN21	$\min.k/\max.k$	38,14	$\min.k/\max.k$	40,74
ŚPN22	$\min.k/\max.k$	11,72	$\min.k/\max.k$	12,38
ŚPN23	$\min.k/\max.k$	42,13	$\min.k/\max.k$	43,61
ŚPN24	$\min.k/\max.k$	33,12	$\min.k/\max.k$	34,67
ŚPN25	$\min.k/\max.k$	36,23	$\min.k/\max.k$	33,09
ŚPN26	$\min.k/\max.k$	40,52	$\min.k/\max.k$	39,16
ŚPN27	$\min.k/\max.k$	42,21	$\min.k/\max.k$	44,61
ŚPN28	$\min.k/\max.k$	69,78	$\min.k/\max.k$	70,01
ŚPN29	$\min.k/\max.k$	10,44	$\min.k/\max.k$	10,91
ŚPN30	$\min.k/\max.k$	22,26	$\min.k/\max.k$	23,18
ŚPN31	gat	53,95	$\min.k/\max.k$	51,93
3028	$\min.k/\max.k$	11,77	$\min.k/\max.k$	12,34
37022	$\min.k/\max.k$	17,62	$\min.k/\max.k$	17,46
75045	gat	21,88	gat	22,91
71025	$\min.k/\max.k$	19,63	$\min.k/\max.k$	19,65
183084	$\min.k/\max.k$	11,73	$\min.k/\max.k$	11,76
111039	x	x	$\min.k/\max.k$	20,61
57049	$\min.k/\max.k$	17,64	$\min.k/\max.k$	15,84
11025	$\min.k/\max.k$	26,22	\wp	27,38
23057	$\min.k/\max.k$	17,80	$\min.k/\max.k$	18,96
75028	$\min.k/\max.k$	14,13	0,5/1,5/średnia	14,13
99043	$\min.k/\max.k$	24,57	$\min.k/\max.k$	24,92
71013	$\min.k/\max.k$	11,46	$\min.k/\max.k$	11,64
113079	\wp	9,30	$\min.k/\max.k$	8,93
115045	$\min.k/\max.k$	22,97	$\min.k/\max.k$	22,47
199016	$\min.k/\max.k$	16,34	$\min.k/\max.k$	15,79
23046	\wp	20,42	\wp	20,21
27013	$\min.k/\max.k$	32,09	$\min.k/\max.k$	32,29
39065	$\min.k/\max.k$	14,26	\wp	14,24
193028	$\min.k/\max.k$	7,35	$\min.k/\max.k$	8,04

x – brak zbieżności procedury numerycznej; lack of convergence of numerical procedure

np. dla powierzchni 113079 (ryc.). W tym przypadku zastosowanie kolejno metod $\min.k/\max.k$, $0,5/1,5/\text{średnia}$ i gat doprowadziło tylko do znalezienia minimum lokalnego o wartości 10,18 i dopiero wykorzystanie metody $w\hat{p}$ pozwoliło na zlokalizowanie minimum globalnego o wartości 9,30 (ryc.).

Dyskusja

W Górach Świętokrzyskich jodła i buk często tworzą drzewostany zbudowane z dwóch generacji wiekowych. W XX wieku intensywne zaburzenia małej i średniej wielkości spowodowały powstanie luk i zmniejszenie zwarcia drzewostanów. W tych warunkach wykształciły się grupy i kępy odnowienia jodłowego i bukowego, które utworzyły młodszą generację drzew [Podlaski 2010a, b]. W Appalachach występują drzewostany wielogatunkowe, złożone z drzew odznaczających się różnymi wymaganiami względem światła, często tworzących dwie generacje wiekowe. Na dynamikę tych drzewostanów istotny wpływ mają zaburzenia o różnej wielkości, od wielkopowierzchniowych pożarów i gradacji owadów po niewielkie oddziaływania związane m.in. z lokalnymi przesuszeniami gleby, infekcjami grzybowymi i silnymi wiatrami [Olano, Palmer 2003; Rentch i in. 2010]. Duże zróżnicowanie wielkości i tekstury powstałych w ten sposób luk umożliwia odnawianie się gatunków o różnych wymaganiach względem światła w małych lukach położonych blisko siebie, co z kolei prowadzi do powstawania dwugeneracyjnych drzewostanów o zwarcu schodkowym, zbudowanych m.in. z gatunków cienioznośnych w górnym piętrze (np. *Acer rubrum*) i światłożądnych w dolnej warstwie (np. *Betula lenta*). W analizowanych drzewostanach można wyodrębnić dwie generacje drzew o różnym przedziale wiekowym, co jest związane z długim procesem odnowienia w lukach oraz pod okapem drzew, tworzących górne piętro. Rozkłady



Ryc.

Estymacja parametrów rozkładu mieszanego (powierzchnia 113079) w zależności od przyjętych wartości startowych

Parameter estimation for mixture model (plot 113079) for various starting values assumed

Funkcja wiarygodności osiągnęła następujące wartości: 10,47 (A), 10,44 (B), 10,18 (C), 9,30 (D; minimum globalne)

Likelihood function reached the following values: 10.47 (A), 10.44 (B), 10.18 (C), 9.30 (D; global minimum)

pierśnic w dwugeneracyjnych drzewostanach nie zawsze są dwumodalne i dlatego rozkłady mieszane mogą być bardzo przydatne do wyróżniania generacji w tego typu drzewostanach. Analogiczne rezultaty uzyskali inni autorzy prowadzący badania w dwugeneracyjnych lasach o różnym składzie gatunkowym [Zhang i in. 2001; Liu i in. 2002; Zasada, Cieszewski 2005].

Należy zwrócić uwagę na duże różnice między średnimi rozkładów składowych dla minimum lokalnego wynoszącego 10,18 i minimum globalnego o wartości 9,30 (ryc.). Obliczenie średnich dla rozkładów składowych umożliwia wyróżnienie generacji wiekowych w drzewostanach dwugeneracyjnych. W analizowanym przypadku zatrzymanie procesu estymacji na etapie znalezienia minimum lokalnego o wartości 10,18 powoduje powstanie dużego błędu obliczenia jednej ze średnich rozkładów składowych. Przedstawiony przykład dowodzi konieczności zlokalizowania minimum globalnego funkcji wiarygodności w trakcie estymacji parametrów rozkładów mieszanych.

Bardzo istotnym warunkiem poprawnego wykorzystania rozkładów mieszanych jest znalezienie minimum globalnego funkcji wiarygodności. Szczególnie starannej analizy wymagają dane zebrane na małych powierzchniach badawczych, gdzie często występują duże różnice między liczbą drzew w sąsiednich stopniach grubości [Hessenmoller, von Gadow 2001]. Przy niewielkiej liczbie drzew prowadzi to do powstania ekstremów mających duży wpływ na przebieg procesu estymacji parametrów rozkładów mieszanych. W tej sytuacji szczególnego znaczenia nabiera wybór odpowiedniej strategii określania wartości startowych dla procedur numerycznych [McLachlan, Krishnan 2008].

Podsumowanie

Zastosowane metody wyboru wartości startowych umożliwiły prawidłową estymację parametrów rozkładu mieszanego Weibulla i gamma w większości analizowanych przypadków. Przeprowadzone aproksymacje rzeczywistych rozkładów pierśnic w dwugeneracyjnych drzewostanach z Polski i Stanów Zjednoczonych dowodzą uniwersalności badanych modeli mieszanych.

W zależności od liczby drzew rosnących na danej powierzchni badawczej proponuje się przyjąć następujący sposób estymacji parametrów rozkładów mieszanych:

- ♣ dla liczby zmierzonych pierśnic mniejszej od 80 należy zastosować metodę wielopunktową wp , uwzględniając przynajmniej dziesięć punktów pokrywających przestrzeń danych;
- ♣ dla liczby zmierzonych pierśnic równej lub większej od 80 należy najpierw zastosować metodę $min.k/max.k$ (dla $k=1, 5, 10$) i $0,5/1,5/średnia$, a dopiero w przypadku gdy otrzymane rezultaty są różne, dodatkowo wykorzystać metodę wielopunktową wp .

Literatura

- Bernadzki E. 1994. Półnaturalna hodowla lasu jest nadal aktualna. Las Pol. 4: 7-10.
- Böhning D. 2000. Computer-Assisted Analysis of Mixtures and Applications. Chapman & Hall/CRC, Boca Raton.
- Böhning D., Dietz E., Schaub R., Schlattmann P., Lindsay B. 1994. The distribution of the likelihood ratio for mixtures of densities from the one-parameter exponential family. Ann. Inst. Statist. Math. 46: 373-388.
- Bruchwald A. 2001. Möglichkeiten der Anwendung von Wachstumsmodellen in der Praxis der Forsteinrichtung. Beitr. Forstwirtsch. u. Landschaftsökol 3: 118-122.
- Bruchwald A., Dmyterko E., Wojtan R. 2011. Model wzrostu dla modrzewia europejskiego (*Larix decidua* Mill.) wykorzystujący cechy taksacyjne drzewostanu. Leś. Pr. Bad. 72: 77-81.
- Bruchwald A., Zasada M. 2010. Model wzrostu modrzewia europejskiego (*Larix decidua* Mill.). Sylwan 154 (9): 615-624.
- Dempster A. P., Laird N. M., Rubin D. B. 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). J. Roy. Statist. Soc. B 39: 1-38.
- Du J. 2002. Combined Algorithms for Constrained Estimation of Finite Mixture Distributions with Grouped Data and Conditional Data. Praca magisterska, McMaster University.

- Gove J. H., Ducey M. J., Leak W. B., Zhang L. 2008. Rotated sigmoid structures in managed uneven-aged northern hardwood stands: a look at the Burr Type III distribution. *Forestry* 81: 161-176.
- Hessenmoller D., von Gadow K. 2001. Beschreibung der Durchmessererverteilung von Buchenbeständen mit Hilfe der bimodalen WEIBULLfunktion. *Allg. Forst- und Jagdzeitung* 172: 46-50.
- Jaworski A. 1997. Karpackie lasy o charakterze pierwotnym i ich znaczenie w kształtowaniu proekologicznego modelu gospodarki leśnej w górach. *Sylvan* 141 (4): 33-49.
- Jaworski A., Podlaski R. 2007. Structure and dynamics of selected stands of primeval character in the Pieniny National Park. *Dendrobiology* 58: 25-42.
- Jaworski A., Podlaski R. 2012. Modelling irregular and multimodal tree diameter distributions by finite mixture models: an approach to stand structure characterization. *J. For. Res.* 17: 79-88.
- Korpeľ Š. 1995. Die Urwälder der Westkarpaten. G. Fischer-Verlag, Stuttgart.
- Liu C., Zhang L., Davis C. J., Solomon D. S., Gove J. H. 2002. A finite mixture model for characterizing the diameter distribution of mixed-species forest stands. *For. Sci.* 48: 653-661.
- Macdonald P. D. M., Du J. 2004. *mixdist: Mixture Distribution Models*.
- McLachlan G. J., Krishnan T. 2008. *The EM algorithm and Extensions*. Wiley, Hoboken.
- Olando J. M., Palmer M. W. 2003. Stand dynamics of an Appalachian old-growth forest during a severe drought episode. *For. Ecol. Manage.* 174: 139-148.
- Podlaski R. 2010a. Diversity of patch structure in Central European forests: are tree diameter distributions in near-natural multilayered *Abies-Fagus* stands heterogeneous? *Ecol. Res.* 25: 599-608.
- Podlaski R. 2010b. Two-component mixture models for diameter distributions in mixed-species, two-age cohort stands. *For. Sci.* 56: 379-390.
- Podlaski R. 2011a. Modelowanie rozkładów pierśnic drzew z wykorzystaniem rozkładów mieszanych. I. Rozkłady mieszane: definicja, charakterystyka, estymacja parametrów. *Sylvan* 155 (4): 244-252.
- Podlaski R. 2011b. Modelowanie rozkładów pierśnic drzew z wykorzystaniem rozkładów mieszanych. II. Aproksymacja rozkładów pierśnic w lasach wielopiętrowych. *Sylvan* 155 (5): 293-300.
- Rentch J. S., Schuler T. M., Nowacki G. J., Beane N. R., Ford W. M. 2010. Canopy gap dynamics of second-growth red spruce-northern hardwood stands in West Virginia. *For. Ecol. Manage.* 260: 1921-1929.
- Runkle J. R. 1985. Disturbance regimes in temperate forests. W: Pickett S. T. A., White P. *The ecology of natural disturbance and patch dynamics*. Academic Press, Orlando. 17-33.
- Rymer-Dudzińska T., Dudzińska M. 1999. Analiza rozkładu pierśnic w drzewostanach bukowych. *Sylvan* 143 (8): 5-24.
- Rymer-Dudzińska T., Dudzińska M. 2001. Rozkład pierśnic drzew w nizinnych drzewostanach bukowych. *Sylvan* 145 (8): 13-22.
- Seidel W., Mosler K., Alker M. 2000a. A cautionary note on likelihood ratio tests in mixture models. *Ann. Inst. Statist. Math.* 52: 481-487.
- Seidel W., Ševčíková H., Alker M. 2000b. On the Power of Different Versions of the Likelihood Ratio Test for Homogeneity in an Exponential Mixture Model. *Diskussionsbeiträge zur Statistik und Quantitativen Ökonomik* 92. Universität der Bundeswehr, Hamburg.
- Siekierski K. 1991. Three methods of estimation of parameters in the double normal distribution and their applicability to modeling tree diameter distributions. *SGGW-AR For.* 42: 13-17.
- Westphal C., Tremer N., von Oheimb G., Hansen J., von Gadow K., Härdtle W. 2006. Is the reverse J-shaped diameter distribution universally applicable in European virgin beech forests? *For. Ecol. Manage.* 223: 75-83.
- Zasada M. 1995. Ocena zgodności rozkładów pierśnic w drzewostanach jodłowych z niektórymi rozkładami teoretycznymi. *Sylvan* 139 (12): 61-69.
- Zasada M. 2003. Możliwość zastosowania rozkładów mieszanych do modelowania rozkładów pierśnic drzew w naturalnych klasach biosocjalnych. *Sylvan* 147 (9): 27-37.
- Zasada M., Cieszewski C. J. 2005. A finite mixture distribution approach for characterizing tree diameter distributions by natural social class in pure even-aged Scots pine stands in Poland. *For. Ecol. Manage.* 204: 145-158.
- Zhang L. J., Gove J. H., Liu C., Leak W. B. 2001. A finite mixture of two Weibull distributions for modeling the diameter distributions of rotated-sigmoid, uneven-aged stands. *Can. J. For. Res.* 31: 1654-1659.
- Zhang L. J., Liu C. 2006. Fitting irregular diameter distributions of forest stands by Weibull, modified Weibull, and mixture Weibull models. *J. For. Res.* 11: 369-372.

SUMMARY

Approximation of the breast height diameter distribution of two-cohort stands by mixture models. I. Parameter estimation

Mixture models are an appropriate tool for modelling heterogeneous populations. If the distribution of tree diameters at breast height (dbh) in a stand is treated as a compound of the distributions of trees belonging to different groups one may adopt a mixture distribution approach. The purposes of this study are (1) to compare four methods (min. k /max. k for $k=1, 5, 10, 0.5/1.5$ /mean, gat and wp) for choosing initial values for the numerical procedure for estimating the parameters of mixture distributions, and (2) to analyse a variety of mixture models to approximate empirical dbh distributions. Two-component mixtures of either the Weibull distribution or the gamma distribution were employed. Additionally, a new strategy for maximizing the likelihood during parameter estimation for mixture models is proposed. The data consisted of study plots from the Świętokrzyski National Park (central Poland) and from the Southern Appalachian Mountains (eastern USA). Study plots were established in two-cohort stands.

The methods used for choosing initial values enabled the accurate estimation of the parameters of mixtures of either the Weibull distribution or the gamma distribution in most of the analysed cases. The resultant approximations of empirical dbh distributions in two-cohort stands from Poland and USA show the versatility of the models examined. After analysing the effectiveness and labour-intensity of the tested methods for choosing initial values, the following strategy is recommended for maximizing the likelihood: (1) for small plots (number of measured dbhs<80) the *wp* method should be used – for the two-component mixture Weibull or gamma model select initial values using 10 grid points on the data space (a complete set of initial values consists of 45 subsets), (2) for large plots (number of measured dbhs≥80), the multistage method should be used – 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. When no two solutions are the same, subsequently apply the *wp* method.