

RAFAŁ PODLASKI, FRANCIS A. ROESCH

Aproksymacja rozkładów pierśnic drzew w dwugeneracyjnych drzewostanach za pomocą rozkładów mieszanych.

III. Estymatory jądrowe a rozkłady mieszane

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

III. Kernel density estimators vs mixture models

ABSTRACT

Podlaski R., Roesch F. A. 2014. Aproksymacja rozkładów pierśnic drzew w dwugeneracyjnych drzewostanach za pomocą rozkładów mieszanych. III. Estymatory jądrowe a rozkłady mieszane. Sylwan 158 (6): 414-422.

Two-component mixtures of either the Weibull distribution or the gamma distribution and the kernel density estimator were used for describing the diameter at breast height (dbh) empirical distributions of two-cohort stands. The data consisted of study 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). Kernel density estimators belong to a class of nonparametric density estimators. Nonparametric estimators have no fixed structure and depend upon all the data points to reach an estimate. In this study the Weibull and the gamma mixture distributions were the most versatile models. The results also support the conclusion that there are only minor differences between the parametric models and the kernel density estimates.

KEY WORDS

two-component models, kernel density estimator, tree diameter distribution, Monte Carlo simulation

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

Modelowanie rozkładów pierśnic drzew w leśnictwie wykorzystywane jest m.in. do budowy tablic miąższości i zasobności, prowadzenia właściwych zabiegów hodowlanych oraz analizy wzrostu drzewostanów [Zasada 1995, 2013; Bruchwald, Zasada 2010; Bruchwald i in. 2011]. Rozkłady pierśnic w dwugeneracyjnych drzewostanach są na ogół dwumodalne, ale często oprócz dwóch głównych lokalnych maksimów, reprezentujących poszczególne generacje, występują inne, mniejsze lokalne maksima. Ta niejednorodność rozkładów pierśnic jest wynikiem przede wszystkim zróżnicowania mikrosiedliskowego drzewostanu oraz zaburzeń towarzyszących rozwojowi kohort drzew. Lepsze warunki glebowe oraz większa ilość światła stymulują przyrost grubości drzew, podobnie zmniejszenie zwarcia w drzewostanie wpływa na ogół korzystnie na okazy

rosnące w pobliżu luk. Przedstawione zjawiska występują zarówno w lasach o charakterze pierwotnym, jak i w drzewostanach gospodarczych [Bernadzki 1994; Jaworski 1997; Jaworski, Podlaski 2007].

Pojedyncze rozkłady teoretyczne, nawet najbardziej elastyczne, w większości przypadków nie oddają dwumodalności rzeczywistych rozkładów pierśnic w drzewostanach dwugeneracyjnych [Gove i in. 2008]. Rozkłady mieszane, złożone z dwóch rozkładów składowych, znacznie dokładniej aproksymują rozkłady pierśnic, a ponadto umożliwiają wyodrębnienie poszczególnych kohort drzew [Zhang i in. 2001; Zasada 2003; Zasada, Cieszewski 2005; Podlaski 2011a, b; Podlaski, Roesch 2013a, b].

Do modelowania rozkładów dwumodalnych – oprócz metod parametrycznych, ograniczających się do konkretnych rozkładów teoretycznych – wykorzystywane są również metody nieparametryczne. W metodach tych nie określa się z góry teoretycznego rozkładu charakteryzującego badane zmienne. Doboru funkcji opisujących rozkład oraz doboru współczynników tych funkcji dokonuje się, stosując odpowiednie kryteria optymalizacyjne [Kulczycki 2005]. Istnieje wiele metod nieparametrycznych umożliwiających modelowanie rozkładów. Najprostsza z nich to histogram, znacznie bardziej złożone metody to m.in. estymatory najbliższego sąsiedztwa, funkcje sklepane i estymatory jądrowe. Z wymienionych metod najczęściej stosowane są histogramy oraz estymatory jądrowe, ponieważ cechuje je naturalność konstrukcji i łatwość analizy. Wyniki otrzymywane przy pomocy estymatorów jądrowych dają możliwość skutecznej interpretacji m.in. lokalnych ekstremów, stopnia asymetrii fragmentów rozkładu oraz zachowania rozkładu dla minimalnych i maksymalnych wartości badanej zmiennej [Pagan, Ullah 1999; Li, Racine 2007].

Celem pracy jest: (1) porównanie dokładności aproksymacji rozkładów empirycznych pierśnic w dwugeneracyjnych drzewostanach rozkładami mieszanymi i estymatorem jądrowym oraz (2) ocena istotności różnic między rozkładami mieszanymi i estymatorem jądrowym.

Materiał i metody

Analizowany materiał zebrano na powierzchniach badawczych założonych w Polsce, w Górach Świętokrzyskich (50°50'-50°53'N, 20°48'-21°05'E; 11 powierzchni) oraz w Stanach Zjednoczonych, w Appalachach Południowych (34°59'-36°32'N, 78°43'-84°13'W; 19 powierzchni). Na badanych powierzchniach zmierzono pierśnice wszystkich żywych drzew. Dla każdej powierzchni określono liczbę drzew, zagęszczenie, średnią i przeciętną pierśnicę oraz pierśnicowe pole przekroju (tab. 1). Szczegółowy opis prac terenowych oraz dokładna charakterystyka powierzchni badawczych znajdują się w pracach Podlaskiego i Roescha [2013a, b].

Do parametrycznego modelowania empirycznych rozkładów pierśnic zastosowano rozkłady mieszane złożone z dwóch rozkładów składowych:

$$f_{(\bullet)X}(x|\psi) = \pi_1 f_{(\bullet)1}(x|\theta_1) + \pi_2 f_{(\bullet)2}(x|\theta_2) \quad [1]$$

gdzie:

π_1, π_2 – wagi (frakcje) rozkładów składowych Weibulla lub gamma $f_{(1)1}(\cdot), f_{(1)2}(\cdot)$,

θ_1, θ_2 – wektory parametrów dla rozkładów składowych,

ψ – zestaw wszystkich parametrów danego rozkładu mieszanego [Podlaski, Roesch 2013a, b].

Do nieparametrycznego modelowania empirycznych rozkładów pierśnic wykorzystano estymator jądrowy. Metoda ta polega na estymacji empirycznych rozkładów za pomocą znanej funkcji

Tabela 1.

Podstawowe charakterystyki powierzchni badawczych wykorzystanych do modelowania rozkładów pierśnic
Basic characteristics of study plots used for dbh distributions modelling

Zmienna	Min.	Pierwszy kwartyl	Mediana	Trzeci kwartyl	Maks.	Średnia	Odchylenie standardowe
Liczba drzew [szt./pow.]	21,0	31,0	46,0	112,0	188,0	74,0	50,1
Zagęszczenie [szt./ha]	312,3	400,0	468,0	580,0	752,0	494,1	119,7
Średnia pierśnica [cm]	14,7	20,8	23,1	25,8	30,6	23,4	3,9
Przeciętna pierśnica [cm]	18,5	22,5	25,8	29,1	37,2	26,2	4,6
Pierśnicowe pole przekroju [m ² /ha]	10,9	19,8	26,6	31,7	46,2	26,7	10,0

gęstości estymatora jądrowego uśredniającej kolejno wartości badanej zmiennej. Estymator jądrowy gęstości prawdopodobieństwa ma postać:

$$\hat{f}(x|h) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \quad [2]$$

gdzie:

- x_i – poszczególne elementy n -elementowej próby losowej,
- x – uśredniane elementy,
- funkcja $K(\cdot)$ – jądro (ang. kernel; analogia do jądra operatora całkowego),
- h – liczba dodatnia, nazywana szerokością pasma lub współczynnikiem wygładzania.

Zastosowano jądro normalne zdefiniowane wzorem:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad [3]$$

Wybór konkretnej postaci jądra nie jest istotny, ponieważ wszystkie standardowe jądra pozwalają na uzyskanie zbliżonych rezultatów [Li, Racine 2007]. Najwyższą efektywnością cechuje się jądro Epanecznikowa, jednak spadek efektywności w przypadku stosowania innych standardowych jąder jest niewielki. Dla jądra normalnego spadek efektywności nie przekracza 5% [Li, Racine 2007].

W obliczeniach wykorzystano szerokość pasma $h=2$ cm, ponieważ podczas parametrycznego modelowania uwzględniano stopnie grubości o szerokości 2 cm. Określenie szerokości pasma ma istotny wpływ na kształt obliczanej funkcji gęstości estymatora jądrowego. Gdy szerokość pasma jest za mała, pojawia się przesadne „wyostrzenie” funkcji gęstości widoczne w postaci wielu ekstremów lokalnych, natomiast gdy szerokość pasma jest za duża, estymator jądrowy cechuje się nadmiernym „rozmyciem” [Łukasik 2008]. Istnieje wiele metod wyboru szerokości pasma [Wand, Jones 1995, Simonoff 1996; Pagan, Ullah 1999]. Najogólniej można powiedzieć, że wybór tego parametru zależy od charakteru danych, a przede wszystkim od koncepcji konkretnej analizy statystycznej.

W celu określenia dokładności aproksymacji rozkładów empirycznych pierśnic rozkładami mieszanymi i estymatorem jądrowym wykorzystano następujące statystyki:

$$B = \frac{1}{l} \sum_{j=1}^l (n_j - \hat{n}_j) \quad [4]$$

$$A = \frac{1}{l} \sum_{j=1}^l |n_j - \hat{n}_j| \quad [5]$$

gdzie:

- \hat{n}_j – odpowiednio oszacowana (dla rozkładu mieszanego Weibulla, gamma i estymatora jądrowego) liczba drzew w j -tym stopniu grubości,
- n_j – rzeczywista liczba drzew w j -tym stopniu grubości,
- l – liczba stopni grubości.

Do oceny wielkości różnic między rozkładami mieszanymi i estymatorem jądrowym zastosowano dwie statystyki bazujące na analizie różnic między funkcją gęstości rozkładu mieszanego $f_{(\cdot),X}(x)$ i estymatora jądrowego $\hat{f}(x)$:

$$BI = \frac{1}{(d_{cmax} - d_{cmin})} \int_{-\infty}^{+\infty} (f_{(\cdot),X}(x) - \hat{f}(x)) dx \quad [6]$$

$$AI = \frac{1}{(d_{cmax} - d_{cmin})} \int_{-\infty}^{+\infty} |f_{(\cdot),X}(x) - \hat{f}(x)| dx \quad [7]$$

gdzie:

- d_{cmin} – środek najmniejszej klasy grubości,
- d_{cmax} – środek największej klasy grubości.

Im różnice między analizowanymi modelami parametrycznymi i modelem nieparametrycznym są mniejsze, tym bardziej statystyki BI i AI są zbliżone do zera.

Testy zgodności między rozkładami mieszanymi i estymatorem jądrowym przeprowadzono, wykorzystując statystyki BI i AI oraz metody Monte Carlo (MC). Metody MC są zaliczane do klasy metod symulacyjnych. W metodach MC analiza danego procesu bazuje na losowaniu ze znanego rozkładu wielkości charakteryzujących badany proces. Podczas testowania zgodności między rozkładami mieszanymi i estymatorem jądrowym, metody MC wykorzystano do symulacji oraz generowania zmiennej losowej z danego rozkładu prawdopodobieństwa. Zastosowane testy zgodności, bazujące na metodach MC, składają się z następujących etapów:

1. numeryczne obliczenie statystyki BI i AI (oznaczanych odpowiednio jako BI_{STAT} i AI_{STAT}) dla n pomierzonych pierśnic na danej powierzchni badawczej;
2. wygenerowanie n -elementowej próby symulacyjnej pierśnic z rozkładu mieszanego $f_{(\cdot),X}(x)$ z wykorzystaniem algorytmu Metropolisa-Hastingsa (generowany ciąg zmiennych losowych jest jednorodnym łańcuchem Markowa);
3. aproksymacja rozkładu n -elementowej próby symulacyjnej pierśnic estymatorem jądrowym;
4. numeryczne obliczenie statystyki BI i AI (oznaczanych odpowiednio jako BI_{SYM_i} i AI_{SYM_i}) dla n -elementowej próby symulacyjnej pierśnic ($i=1$);
5. powtórzenie 999 razy ($i=2, \dots, 1000$) punktów od 2. do 4.;
6. obliczenie granicznego poziomu istotności (p_{BI} oraz p_{AI}) jako: frakcji statystyk BI_{SYM_i} i AI_{SYM_i} nienależących do przedziału $\langle -BI_{STAT}, BI_{STAT} \rangle$ oraz frakcji statystyk $AI_{SYM_i} > AI_{STAT}$.

Wszystkie analizy statystyczne przeprowadzono w środowisku R (www.r-project.org).

Wyniki

Wartości statystyki B są najmniejsze dla estymatora jądrowego (na poszczególnych powierzchniach wynoszą od 0,0004 do 0,0166), co dowodzi, że funkcja gęstości estymatora jądrowego cechuje się najmniejszym obciążeniem, symetrycznie „nad i pod” aproksymując rozkłady empiryczne (tab. 2). Większe obciążenie występuje w przypadku aproksymacji rozkładami mieszanymi.

Statystyka B przyjmuje wartości odpowiednio od 0,0004 do 0,1023 dla rozkładu mieszanego Weibulla oraz od 0,0003 do 0,1166 dla rozkładu mieszanego gamma (tab. 2). Wartości statystyki A są miarą „elastyczności” badanych rozkładów. Kształtowały się one od 0,2629 do 1,9880 (tab. 2). Najmniejsze wartości statystyki A występowały dla rozkładu mieszanego Weibulla i gamma, a największe w przypadku estymatora jądrowego (tab. 2). Dane empiryczne z największą dokładnością aproksymowano za pomocą rozkładów mieszanych, estymator jądrowy cechował się nieznacznie mniejszą precyzją (ryc. 1, 2).

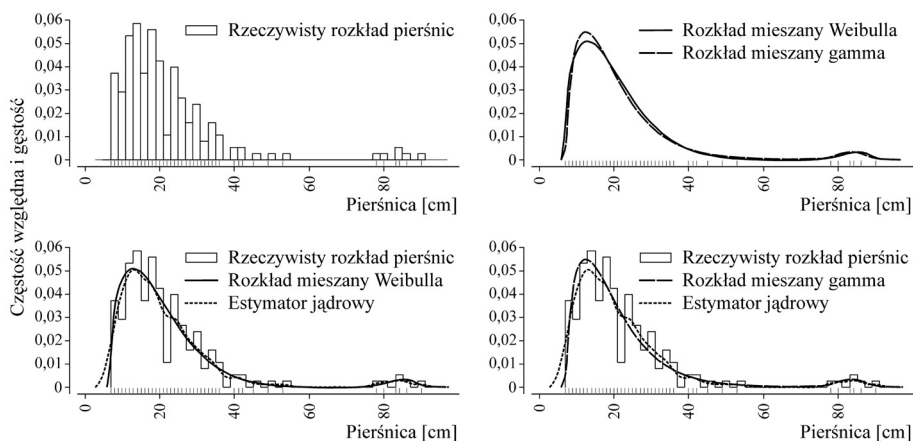
Testy zgodności nie pozwoliły na odrzucenie hipotez o równości odpowiednich funkcji gęstości rozkładu mieszanego $f_{(-),X}(x)$ i estymatora jądrowego $\hat{f}(x)$ dla żadnej z analizowanych powierzchni (wszystkie wartości p były większe od 0,05; tab. 3). Zbliżone do zera wartości staty-

Tabela 2.

Wartości statystyk (B , A) dla rozkładu mieszanego Weibulla, rozkładu mieszanego gamma i estymatora jądrowego

Values of statistics (B , A) of the mixture Weibull model, the mixture gamma model and the kernel density estimator

Zmienna	Min.	Pierwszy kwartył	Mediana	Trzeci kwartył	Maks.	Średnia	Odchylenie standardowe
Rozkład mieszaný Weibulla – B	0,0004	0,0029	0,0080	0,0121	0,1023	0,0113	0,0184
Rozkład mieszaný gamma – B	0,0003	0,0041	0,0106	0,0177	0,1166	0,0160	0,0221
Estymator jądrowý – B	0,0004	0,0017	0,0028	0,0052	0,0166	0,0040	0,0036
Rozkład mieszaný Weibulla – A	0,3040	0,5238	0,6657	0,8851	1,9500	0,8120	0,4186
Rozkład mieszaný gamma – A	0,3061	0,5031	0,6871	0,8843	1,9550	0,8154	0,4239
Estymator jądrowý – A	0,2629	0,6042	0,8207	1,1440	1,9880	0,9188	0,4234

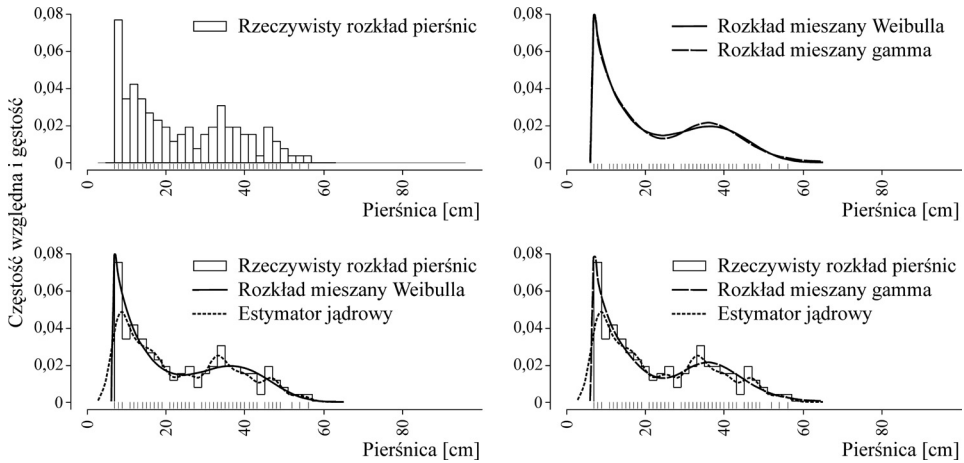


Ryc. 1.

Aproksymacja rozkładów empirycznych pierśnic w drzewostanach dwugeneracyjnych (powierzchnia ŚPN21) rozkładami mieszanymi i estymatorem jądrowym – małe różnice między analizowanymi modelami

Approximation of the dbh empirical distributions of two-cohort stands (plot ŚPN21) by the mixture models and the kernel density estimator – small differences between models analysed

styk BI i AI wskazują na niewielkie różnice między analizowanymi modelami parametrycznymi i modelem nieparametrycznym (tab. 3). Świadczy to o podobnej precyzji aproksymacji wykonanych przy pomocy rozkładów mieszanych Weibulla i gamma oraz estymatora jądrowego.



Ryc. 2.

Aproksymacja rozkładów empirycznych pierśnic w drzewostanach dwugeneracyjnych (powierzchnia ŚPN22) rozkładami mieszanymi i estymatorem jądrowym – duże różnice między analizowanymi modelami
 Approximation of the dbh empirical distributions of two-cohort stands (plot ŚPN22) by the mixture models and the kernel density estimator – large differences between models analysed

Tabela 3.

Wartości statystyk (BI , AI) oraz graniczny poziom istotności (p_{BI} , p_{AI}) dla rozkładu mieszanego Weibulla i rozkładu mieszanego gamma (testy Monte Carlo)
 Values of statistics (BI , AI) and a boundary significance level (p_{BI} , p_{AI}) of the mixture Weibull model and the mixture gamma model (Monte Carlo tests)

Zmienna	Min.	Pierwszy kwartył	Mediana	Trzeci kwartył	Maks.	Średnia	Odchylenie standardowe
Rozkład mieszany Weibulla – BI	-0,0056	-0,0029	-0,0016	-0,0008	-0,0003	-0,0020	0,0014
Rozkład mieszany gamma – BI	-0,0054	-0,0028	-0,0015	-0,0006	-0,0002	-0,0018	0,0014
Rozkład mieszany Weibulla – p_{BI}	0,0920	0,2425	0,3195	0,4495	0,6370	0,3437	0,1447
Rozkład mieszany gamma – p_{BI}	0,1020	0,2758	0,3940	0,5500	0,8390	0,3992	0,1756
Rozkład mieszany Weibulla – AI	0,0008	0,0034	0,0047	0,0082	0,0135	0,0058	0,0034
Rozkład mieszany gamma – AI	0,0013	0,0033	0,0051	0,0088	0,0121	0,0058	0,0031
Rozkład mieszany Weibulla – p_{AI}	0,4420	0,7170	0,8810	0,9300	1,0000	0,8354	0,1548
Rozkład mieszany gamma – p_{AI}	0,5670	0,6570	0,8710	0,9440	1,0000	0,8296	0,1504

Dyskusja

Prezentowane wyniki dowodzą wysokiej precyzji rozkładów mieszanych Weibulla i gamma podczas aproksymacji empirycznych rozkładów pierśnic w drzewostanach dwugeneracyjnych. Ich precyzja jest nawet wyższa od dokładności aproksymacji uzyskanej przy użyciu estymatorów jądrowych. Jest to wynik bardzo interesujący, ponieważ estymatory jądrowe cechują się dużą elastycznością [Li, Racine 2007]. Przyczyną uzyskania takich rezultatów jest dokładne dobranie modeli parametrycznych do danych empirycznych. W drzewostanach dwugeneracyjnych empiryczne rozkłady pierśnic są na ogół dwumodalne, o dużych różnicach między ekstremami, zwłaszcza w przypadku stopni grubości grupujących drzewa o mniejszej pierśnicy. Tego typu rozkłady najpierw osiągają pierwsze maksimum lokalne, silnie opadają w dół i następnie rosną, osiągając drugie maksimum lokalne. Modele mieszane Weibulla i gamma, złożone z dwóch rozkładów składowych, dokładnie aproksymują empiryczne rozkłady o przedstawionym przebiegu. Rozkłady składowe precyzyjnie wyrównują fragmenty wokół głównych „ostrych” maksimów. Mniejsze maksima lokalne nie mają istotnego wpływu na jakość aproksymacji. Estymatory jądrowe są mniej precyzyjne podczas wyrównywania „ostrych” ekstremów. Dlatego w drzewostanach dwugeneracyjnych należy korzystać z rozkładów mieszanych, używając ewentualnie estymatorów jądrowych do wstępnych przybliżeń. Metody nieparametryczne mogą okazać się znacznie bardziej przydatne podczas aproksymacji empirycznych rozkładów pierśnic w drzewostanach wielogeneracyjnych i wielopiętrowych.

Estymatory jądrowe są obecnie stosowane do konstrukcji m.in. różnych testów zgodności [Fan 1994, 1998; Ait-Sahalia 1996; Pagan, Ullah 1999; Cao, Lugosi 2005; Li i in. 2005]. W tego typu testach oceniana jest dokładność aproksymacji rozkładów empirycznych modelami parametrycznymi. Przyjmuje się założenie, że funkcja gęstości estymatora jądrowego odpowiada rozkładowi empirycznemu. Następnie testowana jest istotność różnic między parametrycznymi funkcjami gęstości oraz funkcją gęstości estymatora jądrowego. Problem powstaje w sytuacji, gdy modele parametryczne, w porównaniu do estymatora jądrowego, z podobną precyzją aproksymują rozkład empiryczny. W takiej sytuacji możemy testować tylko hipotezy o równości odpowiednich funkcji gęstości, a nie hipotezy dotyczące dokładności aproksymacji rozkładów empirycznych modelami parametrycznymi. Tego typu przykładem jest prezentowana praca, uwzględniająca rozkłady mieszane i estymator jądrowy, które ze zbliżoną precyzją aproksymują dane empiryczne.

Na wykresach funkcja gęstości estymatora jądrowego dla pierśnicy mniejszej od progu pierśnicowania jest w niewielkim przedziale większa od zera (ryc. 1, 2). Gęstość prawdopodobieństwa jest pojęciem operującym na danych będących zmiennymi losowymi i dlatego wykresy funkcji gęstości estymatora jądrowego pokazują wartości przybliżone. Wykorzystanie funkcji jądrowych o nieograniczonym nośniku skutkuje powstaniem niewielkich, większych od zera, wartości poza obszarem dopuszczalnym. Nie zmniejsza to jednak formalnej poprawności otrzymanych wykresów. Ponadto możliwe jest zastosowanie funkcji jądrowych o ograniczonym nośniku i wtedy poza obszarem dopuszczalnym funkcja gęstości estymatora jądrowego przyjmuje wartości równe zero. Należy podkreślić, że estymatory jądrowe umożliwiają interpolacje, ale nie pozwalają na ekstrapolacje gęstości prawdopodobieństwa [Pagan, Ullah 1999].

Podsumowanie

Do modelowania empirycznych rozkładów pierśnic w drzewostanach dwugeneracyjnych zastosowano modele parametryczne (rozkłady mieszane złożone z dwóch rozkładów składowych Weibulla

i gamma) oraz model nieparametryczny (estymator jądrowy z jądrem normalnym i szerokością pasma $h=2$ cm). Najbardziej precyzyjnie dane empiryczne aproksymowano za pomocą rozkładów mieszanych, nieznacznie mniejszą dokładność uzyskano, wykorzystując estymator jądrowy. Różnice między analizowanymi modelami były małe, a zastosowane testy zgodności nie pozwoliły na odrzucenie hipotez o równości odpowiednich funkcji gęstości rozkładów mieszanych i estymatora jądrowego. Do modelowania rozkładów pierśnic w drzewostanach dwugeneracyjnych należy wykorzystywać rozkłady mieszane złożone z dwóch rozkładów składowych Weibulla i gamma, a w dalszej kolejności estymatory jądrowe.

Literatura

- Aït-Sahalia Y. 1996. Nonparametric pricing of interest rate derivative securities. *Econometrics* 64: 385-426.
- Bernadzki E. 1994. Pónaturalna hodowla lasu jest nadal aktualna. *Las Pol.* 4: 7-10.
- 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.). *Sylvan* 154 (9): 615-624.
- Cao R., Lugosi I. 2005. Goodness-of-fit based on the kernel density estimator. *Scand. J. Statist.* 32: 599-615.
- Fan Y. Q. 1994. Testing the goodness of fit of a parametric density function by kernel method. *Econ. Theory* 10: 316-356.
- Fan Y. Q. 1998. Goodness-of-fit tests based on kernel density estimators with fixed smoothing parameters. *Econ. Theory* 14: 604-621.
- 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.
- Jaworski A. 1997. Karpackie lasy o charakterze pierwotnym i ich znaczenie w kształtowaniu proekologicznego modelu gospodarki leśnej w górach. *Sylvan* 141(4): 33-490.
- 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 characterisation. *J. For. Res.* 17: 79-88.
- Kulczycki P. 2005. Estymatory jądrowe w analizie systemowej. WNT, Warszawa.
- Li Q., Racine J. S. 2007. *Nonparametric Econometrics*. Princeton University Press, New Jersey.
- Li Y., Singh R. S., Sun Y. 2005. Goodness-of-fit tests of a parametric density functions: Monte Carlo simulation studies. *J. Stat. Res.* 39: 103-125.
- Łukasik S. 2008. Identyfikacja rozkładu w systemach rzeczywistych za pomocą estymatorów jądrowych. *Czasop. Techn.* 17: 3-13.
- Pagan A., Ullah A. 1999. *Nonparametric Econometrics*. Cambridge University Press, New York.
- 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.
- Podlaski R., Roesch F. A. 2013a. Aproksymacja rozkładów pierśnic drzew w dwugeneracyjnych drzewostanach za pomocą rozkładów mieszanych. I. Estymacja parametrów. *Sylvan* 157 (8): 587-596.
- Podlaski R., Roesch F. A. 2013b. Aproksymacja rozkładów pierśnic drzew w dwugeneracyjnych drzewostanach za pomocą rozkładów mieszanych. II. Testy zgodności. *Sylvan* 157 (9): 652-661.
- Simonoff J. S. 1996. *Smoothing Methods in Statistics*. Springer, New York.
- Wand M. P., Jones M. C. 1995. *Kernel Smoothing*. Chapman and Hall, London, New York.
- 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. 2013. Evaluation of the double normal distribution for tree diameter distribution modeling. *Silv. Fenn.* 47 (2), article id 956.
- 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.

SUMMARY

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

III. Kernel density estimators vs mixture models

The purposes of this study are (1) to compare two-component mixtures of either the Weibull distribution or the gamma distribution and the kernel density estimator in describing the diameter at breast height (dbh) empirical distributions of two-cohort stands, and (2) to assess the significance of differences between the two-component mixture distributions and the kernel density estimates. The data consisted of study 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).

Kernel density estimation is a nonparametric technique in which the kernel is averaged across the observed data points to create a smooth approximation. In this study a Gaussian density as the kernel and a bandwidth $h=2$ cm were employed. Two goodness-of-fit statistics (B and A) based on the differences between the observed and predicted numbers of trees in the dbh classes were used. The significance of differences between the two-component mixture distributions and the kernel density estimates were tested using (1) statistics (BI and AI) based on the distances between the mixture model probability density functions and the kernel density estimates and (2) Monte Carlo methods.

Most precisely empirical data were approximated using mixture models (statistics A ranged from 0.3040 to 1.9550, the mean A values were less than 0.82), slightly less accurate estimate was obtained with the kernel density estimator (A varied from 0.2629 to 1.9880, the mean A was 0.9188). Differences between analysed models were small, and the Monte Carlo tests applied did not allow for the rejection of hypotheses about the equality of the mixture distribution probability density functions and the kernel density estimate (all p -values were greater than 0.05). For modelling dbh distributions in two-cohort stands one should use two-component mixtures of either the Weibull distribution or the gamma distribution, followed by the kernel density estimates.