

PIOTR POGODA, WOJCIECH OCHAŁ, STANISŁAW ORZEŁ

Szerokość pasma jądrowego estymatora rozkładu pierśnic w drzewostanach olszy czarnej (*Alnus glutinosa* (L.) Gaertn.) z zachodniej części Kotliny Sandomierskiej*

Bandwidth of kernel estimator of DBH distribution in black alder (*Alnus glutinosa* (L.) Gaertn.) stands from west part of the Sandomierz Basin

ABSTRACT

Pogoda P., Ochał W., Orzeł S. 2018. Szerokość pasma jądrowego estymatora rozkładu pierśnic w drzewostanach olszy czarnej (*Alnus glutinosa* (L.) Gaertn.) z zachodniej części Kotliny Sandomierskiej. Sylwan 162 (5): 411-421.

Set of 'nonparametric' methods, that don't make *a priori* assumption about functional form of empirical distribution was developed as an alternative to the parametric distribution modeling. The kernel estimators are one of such methods, that can be used to describe the frequency of data representing for example DBH records. Kernel smoothing requires the choice of weighting function and bandwidth also called as smoothing parameter or window. The lack of comprehensive analysis on the applicability of particular bandwidth selection methods to model DBH structure gave an impulse to present investigation aimed at determining value and variability of smoothing parameter in black alder stands. The optimal bandwidth was obtained according to six different variants of plug-in method proposed by Altman and Léger. Presented investigations were based on DBH measurements collected in 163 managed black alder stands aged from 6 to 89 years, growing in the west part of the Sandomierz Basin (S Poland). We measured in total 22,530 black alders, from 48 to 359 in individual stand. Stands were characterized by: age, quadratic mean diameter, basal area, mean height, Reineke's stand density index and standard deviation of DBH. Smoothing parameter was obtained by means of plug-in method with the pilot bandwidth selected by: Silverman's rule of thumb (nrd0), Scott's method (nrd), unbiased cross-validation (ucv), biased cross-validation (bcv), method of Sheather and Jones (sj) and one-stage method of Wand and Jones (onestage). The bandwidth was first obtained to real data, then to 100 bootstrap samples of 5, 10, 15 ... and 100 trees from each stand. Smoothing parameters were characterized by mean and variance. Relationship between values of smoothing parameter and stand characteristics was determined. Finally the influence of sample size on value and variability of bandwidth was assessed. Value and variability of smoothing parameter in black alder stands are determined by stand age, sample size and method of bandwidth choice. There is a close relationship between bandwidth and the mean height (r from 0.75 to 0.83), quadratic mean diameter (r from 0.79 to 0.88) and standard deviation of DBH (r from 0.84 to 0.93). Potentially these stand features can be used to predict smoothing parameter values. Minor changes of bandwidth for samples containing above 50 trees together with persistence of standard error give an objective grounds for defining optimal number of diameters, that are necessary to kernel estimation of DBH distribution.

*Badania wykonano w ramach tematu DS/ZBiPL/2017. W pracy wykorzystano materiał empiryczny zebrany w projektach badawczych: 2 P06L 0-56 26 (KBN 2004-2005) oraz N N309 160638 (NCN 2010-2012).

KEY WORDS

kernel estimator, bandwidth, DBH structure, black alder

ADDRESSES

Piotr Pogoda – e-mail: piotr.pogoda@krakow.lasy.gov.pl

Wojciech Ochał – e-mail: w.ochal@ur.krakow.pl

Stanisław Orzeł – e-mail: s.orzel@ur.krakow.pl

Zakład Biometrii i Produkcji Lasu, Uniwersytet Rolniczy w Krakowie; al. 29 Listopada 46, 31-425 Kraków

Wstęp

Struktura grubości, rozumiana jako podział drzew w drzewostanie na stopnie o określonej szerokości, stanowi nadal inspirację dla kolejnych analiz i rozważań. Matematyczny opis frekwencji pierśnic ma bowiem, oprócz poznawczego, także wymiar czysto pragmatyczny, związany z możliwością określenia liczby drzew w zdefiniowanych przedziałach grubości bez konieczności pomiarów w terenie. Model rozkładu pierśnic może być wykorzystany na potrzeby określania miąższości drzewostanu [Haara i in. 1997; Liu i in. 2002], a także wspierać proces wyboru lokalizacji, rodzaju i intensywności planowanych zabiegów hodowlanych oraz dobór maszyn o optymalnych parametrach do prac przy pozyskaniu drewna [Robinson, Hamann 2011]. Jest również pomocny przy typowaniu do cięć drzewostanów gwarantujących uzyskanie określonych grup sortymentów i koncentracji prac ścinkowo-zrywkowych przy minimalnych kosztach późniejszego transportu surowca drzewnego [Uusitalo i in. 2006].

Modelowe przedstawienie struktury pierśnic daje naukowe podstawy zarządzania zasobami leśnymi. Realizacja przyjętego celu gospodarczego, dotyczącego produkcji określonych sortymentów surowca drzewnego (np. drewna tartacznego), a w pewnych sytuacjach także pełnione przez las funkcje ochronne i krajobrazowe, wymagają rozpoznania struktury pierśnic, a następnie utrzymania konkretnej liczby drzew w poszczególnych stopniach grubości [Hann, Bare 1979]. Pukkala i in. [2010] zwracają uwagę na przydatność modeli frekwencji pierśnic w procesach optymalizacyjnych zmierzających do ustalenia pożądanej liczby drzew w kolejnych stopniach grubości, gwarantującej uzyskanie maksymalnych dochodów. Modele rozkładu pierśnic umożliwiają symulowanie przyszłych struktur grubości oraz zmian warunkowanych przez nie charakterystyk drzewostanu [Pretzsch 2000; Čavlović i in. 2006]. Zdaniem Knighta oraz Wrighta [za Lima i in. 2015] modelowe przedstawienie frekwencji pierśnic dla lasów mieszanych daje możliwość uzyskania informacji o trendach zmian zachodzących w populacji drzew, wynikających na przykład ze specyficznego dla każdego gatunku sposobu odnowienia, umożliwiając jednocześnie retrospekcję dotychczasowych zakłóceń w rozwoju lasu i wnioskowanie dotyczące aktualnego stadium jego rozwoju [Coomes, Allen 2007 za Lima i in. 2015].

W większości prac poświęconych opisowi struktury grubości drzewostanu zakłada się, że empiryczny rozkład pierśnic jest zgodny z pewnym teoretycznym rozkładem gęstości prawdopodobieństwa. Weryfikacja trafności przyjętej hipotezy sprawdzana jest następnie na podstawie pobranej próby. Alternatywę dla opisanej parametrycznej metody modelowania stanowią metody nieparametryczne, obejmujące szereg narzędzi służących do rozpoznania wzorca struktury i jego modelowania [Duin 1976], zarówno w przypadku organizmów żywych, jak też zagadnień z zakresu ekonomii [Baszczyńska 2014] czy techniki [Kulczycki 2007; Łukasik 2008]. U podstaw nieparametrycznych metod opisu rozkładu gęstości prawdopodobieństwa znajduje się założenie, że

można go przedstawić pewną funkcją, nie należy jednak *a priori* przyjmować jej postaci [Silverman 1986]. Téllez i in. [2008] wskazują dodatkowo, że funkcja ta musi być ciągła i różniczkowalna.

Arbitralnemu założeniu zgodności rozkładu z konkretnym rozkładem Silverman [1986] przeciwstawia właściwe wykorzystanie informacji zawartych w zebranych materiale empirycznym. W podobny sposób do zagadnienia nieparametrycznych metod opisu rozkładów podchodzą Wand i Jones [1995], którzy charakteryzują je często cytowaną w piśmiennictwie formułą „letting the data speak for themselves” (niech dane same przemówią za siebie).

W literaturze z zakresu leśnictwa można znaleźć przykłady nieparametrycznego modelowania struktury pierśnic. Należą do nich metody: percentyli [Borders i in. 1987; Kangas, Maltamo 2000], „k – najbliższych sąsiadów” [Haara i in. 1997; Maltamo, Kangas 1998; Maltamo i in. 2003; Bollandsås i in. 2013], krzywych rozdzielczych (decylowych) [Borowski 1974] oraz estymatorów jądrowych [Uutera, Maltamo 1995; Podlaski 2013, 2017; Podlaski, Roesch 2014].

Estymacja jądrowa funkcji gęstości prawdopodobieństwa bądź dystrybuanty odbywa się lokalnie na podstawie pewnego otoczenia [Téllez i in. 2008], którego zakres wyznacza wartość tzw. pasma, określanego również mianem okna lub parametru wygładzania. Wybór parametru wygładzania decyduje o jakości estymatora jądrowego [Sheather 2004; Zambom, Dias 2012; Guidoum 2015]. W dotychczasowych pracach poświęconych modelowaniu struktury pierśnic z zastosowaniem estymatora jądrowego przeważnie stosowano parametr wygładzania odpowiadający szerokości stopnia grubości używanego podczas parametrycznego modelowania rozkładu [Uutera, Maltamo 1995; Podlaski 2013, 2017; Podlaski, Roesch 2014].

Dostępny w oprogramowaniu statystycznym szereg sposobów określania parametru wygładzania [Sheather 2004] wymaga rozpoznania przydatności poszczególnych metod. Brak szczegółowych analiz ich zastosowania do estymacji rozkładu pierśnic stał się impulsem do podjęcia badań, których celem było określenie wielkości i zmienności szerokości pasm oszacowanych sześcioma sposobami oraz ocena ich związku z wybranymi cechami drzewostanów olszowych.

Material i metody

Do badań wykorzystano wyniki pomiarów pierśnic wykonanych w 163 drzewostanach gospodarczych olszy czarnej w wieku 6-89 lat. Podstawowe charakterystyki analizowanych drzewostanów zestawiono w tabeli 1, zaś opis ich lokalizacji, liczby i wielkości założonych powierzchni oraz zakresu prac pomiarowych podano w innej pracy autorów [Orzeł i in. 2017].

Estymator jądrowy dystrybuanty zaproponowany przez Nadaraya w 1964 roku przedstawia następujący wzór [Zambom, Dias 2012]:

Tabela 1.

Minimum (Min), pierwszy kwartył (Q1), mediana (Me), trzeci kwartył (Q3), maksimum (Max), średnia (M) i odchylenie standardowe (SD) liczby pomierzonych drzew (N), zagęszczenia (Zag [szt./ha]), indeksu Reineke (SDI), przeciętnej pierśnicy (Dg [cm]) i pierśnicowego pola przekroju (G [m²/ha]) badanych drzewostanów olszowych

Minimum (Min), first quartile (Q1), median (Me), third quartile (Q3), maximum (Max), mean (M) and standard deviation (SD) of number of trees (N), stand density (Zag [trees/ha]), stand density index (SDI), quadratic mean diameter (Dg [cm]) and basal area (G [m²/ha]) for analyzed black alder stands

	Min	Q1	Me	Q3	Max	M	SD
N	48	101	122	163	359	138	56
Zag	223	430	690	1283	4360	1020	885
SDI	212	437	518	612	873	528	129
Dg	4,5	13,6	22,7	29,5	43,0	22,0	9,4
G	5,48	17,78	25,52	31,74	49,79	25,21	9,11

$$\hat{F}_h(x) = \frac{1}{n} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad [1]$$

gdzie:

- n – liczebność próby,
- x – wartość cechy (pierśnicy), dla której jest szacowane skumulowane prawdopodobieństwo $\hat{F}_h(x)$,
- X_i – poszczególne elementy należące do próby,
- h – parametr wygładzania/pasmo ($h > 0$),
- $K(\cdot)$ – funkcja pierwotna jądra k wyrażona wzorem:

$$K(x) = \int_{-\infty}^x k(t) dt \quad [2]$$

W analizach zastosowano jądro normalne:

$$k(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad [3]$$

Funkcja jądrowa (jądro) ma następujące właściwości [Raykar, Duraiswami 2006; Téllez i in. 2008; Guidoum 2015]:

- jej wartości są dodatnie $k(x) \geq 0$,
- jest symetryczna względem zera, tzn. $k(x) = k(-x)$ dla każdego $x \in \mathbb{R}^n$,
- suma jej wartości obliczona dla całej dziedziny jest równa jedności $\int_{\mathbb{R}^n} k(x) dx = 1$.

Jądro jest funkcją ważącą poszczególne obserwacje i zapewnia estymatorowi ciągłość, będącą warunkiem koniecznym jego różniczkowalności [Parzen 1962]. Jądro normalne, w odróżnieniu od innych (np. Cauchy’ego, Epanecznikowa, jednostajnego, dwuwagowego czy trójkątnego) posiada pochodną dowolnego rzędu [Kulczycki 2005]. Cecha ta ma istotne znaczenie w procedurach optymalizacyjnych poszukujących najodpowiedniejszej szerokości pasma, która decyduje o właściwościach estymatora. Zmiana wartości parametru wygładzania wpływa zarówno na obciążenie (tj. tendencję do przeszacowania lub niedoszacowania „prawdziwej” wartości), jak i wariancję (tj. wielkość oscylacji wokół „prawdziwej” wartości) estymatora [Kulczycki 2005]. Zbyt mała szerokość pasma prowadzi do opracowania estymatora o małym obciążeniu i dużej wariancji, z kolei zbyt duża jego wartość powoduje spadek wariancji i wzrost obciążenia.

Metodom wyboru optymalnej szerokości pasma poświęcono dużo uwagi w literaturze naukowej [Altman, Léger 1995; Loader 1999; Hansen 2004; Raykar, Duraiswami 2005; Quintelal-del-Río, Estévez-Pérez 2012; Guidoum 2015]. Głównym ich zadaniem jest znalezienie parametru wygładzania, który dla każdej wartości cechy wykorzystanej podczas modelowania pozwoli zminimalizować różnicę pomiędzy prawdopodobieństwem rzeczywistym i prawdopodobieństwem określonym za pomocą estymatora. O ile drugie z nich można obliczyć numerycznie, to pierwsze podlega oszacowaniu. W literaturze przedstawiono wiele rozwiązań pozwalających przybliżyć wartość prawdopodobieństwa rzeczywistego [Sheather 2004; Raykar, Duraiswami 2005]. Za optymalną metodę uznaje się sposób podstawiania wykorzystujący empiryczne oszacowania pochodnych rzeczywistej funkcji gęstości prawdopodobieństwa [Altman, Léger 1995].

W pracy optymalną wartość parametru wygładzania określono metodą podstawiania [Altman, Léger 1995], w której drugą pochodną funkcji prawdopodobieństwa rzeczywistego oszacowano sześcioma sposobami [Sheather 2004]:

- nieobciążone krzyżowe uwiarygodnianie (ucv),
- obciążone krzyżowe uwiarygodnianie (bcv),
- Silvermana (nrd0),
- Scotta (nrd),
- Sheathera i Jonesa (sj),
- Wanda i Jonesa (onestage).

Dla uproszczenia przyjęto, że nazwy sposobów ustalania pochodnej prawdopodobieństwa rzeczywistego będą jednocześnie nazwami metod określenia optymalnej wartości parametru wygładzania. Zmienność szerokości pasm ustalonych wyżej wymienionymi sposobami scharakteryzowano zarówno dla ogółu drzewostanów, jak i w obrębie poszczególnych klas wieku. Metodą analizy wariancji zweryfikowano hipotezę o istotności różnic pomiędzy średnimi szerokościami pasm oszacowanych analizowanymi sposobami. Określono także związek pomiędzy wartościami parametru wygładzania a podstawowymi cechami rozpatrywanych drzewostanów, tj. wiekiem, przeciętną pierśnicą i wysokością, odchyleniem standardowym pierśnic, polem przekroju i zagęszczeniem rozumianym jako liczba drzew na jednostce powierzchni oraz wyrażonym indeksem Reineke [Shaw 2006]:

$$SDI = N_{1ha} \left(\frac{d_g}{25} \right)^{1,605} \quad [4]$$

gdzie:

- N_{1ha} – liczba drzew na powierzchni 1 ha,
- d_g – przeciętna pierśnica drzewostanu [cm].

Analizie poddano także wpływ wielkości próby na szerokość pasma. W tym celu ze zbioru pierśnic pomierzonych w każdym drzewostanie olszowym wygenerowano metodą losowania ze zwracaniem po 100 prób o liczebności: 5, 10, 15, ..., 100 elementów. Dla każdej z nich określono analizowanymi sposobami pasmo estymatora jądrowego. Wariancję ogólną pasm obserwowaną w zbiorze wygenerowanych wyników rozdzielono na wariancję między- i wewnątrzgrupową, uzyskując informację o zmienności parametru wygładzania w obrębie drzewostanów wynikającą ze struktury generowanych prób.

Obliczenia i analizy wykonano w środowisku R (www.r-project.org), wykorzystując pakiety base, stats oraz sROC [Wang 2012].

Wyniki

Średnia szerokość pasma oszacowanego poszczególnymi metodami w analizowanych drzewostanach olszowych zawiera się w przedziale od 2,84 cm (ucv) do 3,50 cm (onestage), zaś jej skrajne wartości wynoszą od 0,18 cm (ucv) do 8,75 cm (onestage) (tab. 2). Stwierdzone różnice są statystycznie istotne (test analizy wariancji: $F=5,645$; $p=0,00004$), co oznacza, że szerokość pasma zależy od sposobu jego określania. Istotne różnice (test Tukeya) występują pomiędzy średnimi wartościami parametru wygładzania ustalonymi sposobami: nrd – onestage, nrd0 – onestage, sj – onestage, ucv – onestage (ryc. 1). Parametr wygładzania określony sposobem Silvermana (nrd0) cechuje się najmniejszą (odchylenie standardowe 1,20 cm; rozstęp 6,46 cm), zaś sposobem Wanda i Jonesa (onestage) największą (odchylenie standardowe 1,48 cm; rozstęp 8,07 cm) zmiennością bezwzględną. Podobny poziom zmienności bezwzględnej, nieznacznie wyższy od określonego dla sposobu nrd0, cechuje parametr wygładzania ustalony sposobami nrd, bcv oraz sj. Największą zmienność względną (49,4%) zaobserwowano wśród wyników otrzymanych sposobem ucv.

Tabela 2.

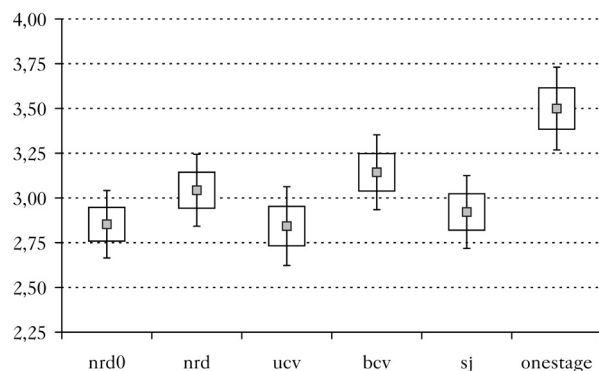
Średnia (M), minimum (Min), maksimum (Max), odchylenie standardowe (SD), rozstęp (R) i współczynnik zmienności (CV [%]) szerokości pasma określonego analizowanymi metodami

Mean (M), minimum (Min), maximum (Max), standard deviation (SD), range (R) and coefficient of variability (CV [%]) of bandwidth estimated by the analyzed methods

	nrd0	nrd	ucv	bcv	sj	onestage
M	2,85	3,04	2,84	3,14	2,92	3,50
Min	0,55	0,59	0,18	0,63	0,50	0,67
Max	7,01	7,48	7,69	7,68	7,55	8,75
SD	1,20	1,28	1,40	1,34	1,30	1,48
R	6,46	6,89	7,51	7,06	7,04	8,07
CV	42,16	42,17	49,42	42,47	44,40	42,19

nrd0 – metoda Silvermana, nrd – metoda Scotta, ucv – nieobciążone krzyżowe uwiarygodnianie, bcv – obciążone krzyżowe uwiarygodnianie, sj – metoda Sheathera i Jonesa, onestage – jednoetapowa metoda Wanda i Jonesa

nrd0 – Silverman's rule of thumb, nrd – Scott's method, ucv – unbiased cross-validation, bcv – biased cross-validation, sj – method of Sheather and Jones, onestage – one-stage method by Wand and Jones



Ryc. 1.

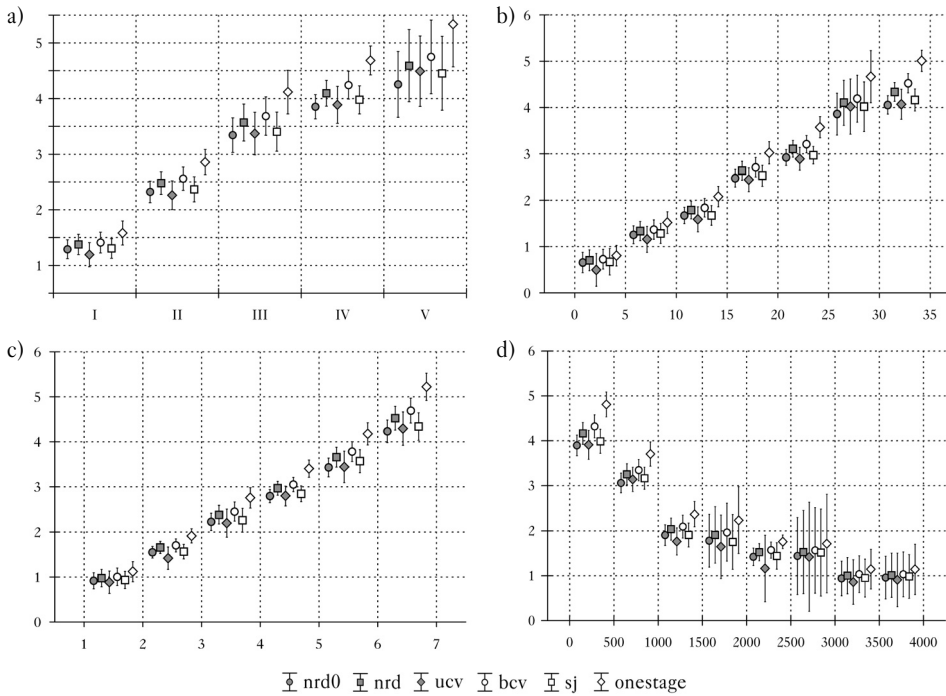
Średnia wartość parametru wygładzania oszacowanego analizowanymi metodami

Mean value of smoothing parameter estimated by different methods

oznaczenia jak w tabeli 2; denotes as in table 2

Szerokość pasma, niezależnie od sposobu określania, jest powiązana z cechami analizowanych drzewostanów. Silną dodatnią zależność stwierdzono w przypadku wieku ($r_{\epsilon} < 0,74-0,81 >$, ryc. 2a), przeciętnej pierśnicy ($r_{\epsilon} < 0,79-0,88 >$, ryc. 2b), a zwłaszcza odchylenia standardowego pierśnic ($r_{\epsilon} < 0,84-0,93 >$, ryc. 2c). Słabszy związek stwierdzono w przypadku wysokości, pola pierśnicowego przekroju oraz wskaźnika zagęszczenia drzewostanu SDI. Największy wzrost średniej wartości pasma (od 79,7 do 89,2%) stwierdzono pomiędzy drzewostanami I i II klasy wieku, zaś najmniejszy (od 10,4 do 15,7%) pomiędzy drzewostanami IV i V klasy wieku (ryc. 2a). We wszystkich klasach wieku największe wartości szerokości pasma uzyskano sposobem onestage (ryc. 2a).

Szerokość pasma maleje ze wzrostem zagęszczenia drzewostanów wyrażonym liczbą drzew występujących na jednostce powierzchni (ryc. 2d). Jest to zależność krzywoliniowa. Wielkość pasma maleje także ze wzrostem liczebności próby, na podstawie której została ustalona (ryc. 3a). Niezależnie od sposobu określenia duże wartości pasm dla małych prób (liczebność od 5 do 10) ulegają dość gwałtownemu zmniejszeniu ze wzrostem liczebności prób. Począwszy od prób złożonych z 30 drzew dynamika zmniejszania szerokości pasma z dalszym powiększaniem liczebności próby ulega wyraźnemu spowolnieniu. Największe zmiany wartości pasma stwierdzono w przypadku metody ucv. Przeciętna wartość parametru wygładzania dla prób o liczebności 100 drzew była ponad 4,5 razy mniejsza w porównaniu do szerokości pasma ustalonego w przypadku prób złożonych z pięciu drzew. Stosunek ten, charakteryzujący dynamikę zmian szerokości pasma w zależności od liczebności prób, wahał się od 1,8 (sj) do 2,7 (bcv). Podobne spadkowe tendencje ze wzrostem liczebności prób wykazuje błąd standardowy oszacowania średniej szerokości pasma.



Ryc. 2.

Zmiana wartości parametru wygładzania w zależności od wieku drzewostanu (a), przeciętnej pierśnicy (b), odchylenia standardowego pierśnic (c) i zagęszczenia drzewostanu (d)

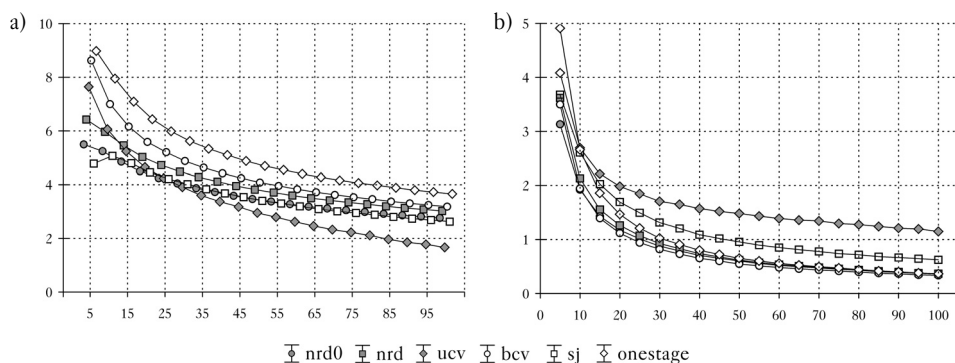
Relationship between estimated value of smoothing parameter and age of the stand (a), quadratic mean diameter (b), standard deviation of DBH (c) and stand density (d)

oznaczenia jak w tabeli 2; denotes as in table 2

kości pasm (ryc. 3b). Początkowo (dla prób o liczebności od 5 do 20-30 drzew) charakterystyczne jest szybkie zmniejszanie się zmienności oszacowań szerokości pasm, później zaś wzrost wielkości próby w niewielkim stopniu wpływa na wielkość błędu średniego. Największą zmiennością w całym zakresie liczebności prób charakteryzują się wyniki otrzymywane metodą ucv – błąd średni waha się od 4,09 cm (n=5) do 1,15 cm (n=100). Najmniejszą zmienność stwierdzono w przypadku pasm określonych metodą bcv – błąd średni zmieniał się od 3,51 cm (n=5) do 0,34 cm (n=100). Równie niskie błędy obserwowano także dla metody nrd0. Dla prób o liczebności powyżej 50 drzew zbliżone błędy stwierdzono dla czterech metod: nrd0, nrd, bcv i onestage. Obserwowana stabilizacja oszacowań szerokości pasm ze wzrostem wielkości próby daje obiektywne podstawy do ustalenia optymalnej liczby drzew niezbędnej do modelowania struktury grubości metodą estymatorów jądrowych.

Dyskusja

W przypadku nieparametrycznego opisu struktury grubości realizowanego za pomocą estymatorów jądrowych podstawowym zagadnieniem jest ustalenie optymalnej szerokości pasma. Zdaniem Kulczyckiego [2005], Zambona i Diasa [2012] oraz López-de-Ullibarri [2015] wybór wartości parametru wygładzania w każdym przypadku stanowi kompromis pomiędzy błędem i wariancją wykorzystującego go estymatora jądrowego. Leżący u podstaw koncepcji estymatorów jądrowych brak arbitralnych założeń dotyczących postaci rozkładu pozwala uwidocznić jego charak-



Ryc. 3.

Zależność średniej szerokości pasma (a) i błędu standardowego (b) od liczebności próby

Relationship between mean value (a) or mean square error (b) of estimated bandwidth and size of sample used to computation

oznaczenia jak w tabeli 2; denotes as in table 2

terystyczne cechy, które są generalizowane lub pomijane w przypadku wykorzystania metody parametrycznej [Kulczycki 2007]. Stosowanie estymatorów jądrowych jest zalecane zwłaszcza do modelowania zjawisk nowych i nietypowych, gdy brakuje dostatecznej ilości informacji do przyjęcia założenia o konkretnej postaci rozkładu [Baszczyńska 2014] czy też do opisu danych pochodzących z trudnych do powtórzenia eksperymentów [Łukasik 2008]. Optymalizacyjny charakter rozpatrywanej metody nieparametrycznej wymaga pewnego stopnia ostrożności oraz ciągłej obserwacji uzyskiwanych rezultatów [Podlaski 2013].

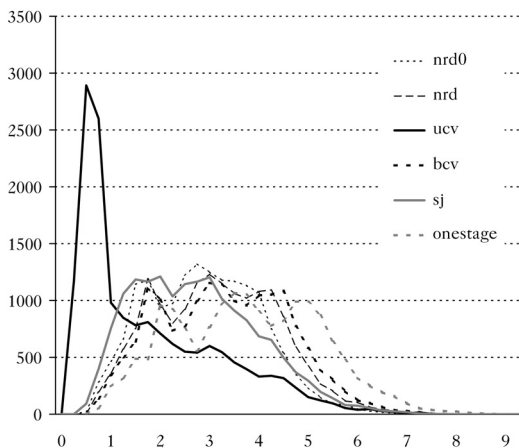
Analizując średnie wartości parametru wygładzania, nietrudno zauważyć, że są one mocno zróżnicowane (ryc. 1) i zależą od podstawowych cech drzewostanu (ryc. 2). W drzewostanach I klasy wieku średnia szerokość pasma, w zależności od sposobu jego oszacowania, przyjmuje wartości od 1,2 do 1,6 cm, by już w drzewostanach II klasy wieku przekroczyć wartość 2 cm, a w najstarszych drzewostanach przyjąć poziom od 4 do 6 cm. Duży wpływ na oszacowania szerokości pasma poszczególnymi metodami ma liczebność próby (ryc. 3a). Spośród analizowanych metod najmniej czułe na zmianę liczebności próby okazały się sposoby sj i nrd0, a największy wpływ stwierdzono w przypadku metody ucw. Podkreślić należy, że największe zmiany oszacowań parametru mają miejsce dla prób małych, o liczebności poniżej 30-40 drzew, natomiast w przypadku większych liczebności obserwuje się pewną stabilizację wielkości pasma, zwłaszcza dla metod: nrd0, sj, nrd i bcw (ryc. 3a). Wielkość próby decyduje jednocześnie o błędzie standardowym uzyskiwanych oszacowań (ryc. 3b). W przypadku małych prób błąd jest bardzo duży, później jego wielkość gwałtownie maleje, a następnie powyżej liczebności 40-50 drzew nie ulega większym zmianom. Metody nrd0, nrd, bcw i onestage charakteryzuje bardzo podobna wielkość błędów standardowych, szczególnie w zakresie liczebności prób od 50 do 100 drzew. Większe wartości błędów standardowych obserwowane są przy zastosowaniu metod sj, a zwłaszcza ucw, w przypadku której błąd jest prawie trzykrotnie większy w porównaniu do pierwszej wyżej wymienionej grupy metod. Podsumowując, należy stwierdzić, że zastosowana metoda automatycznego szacowania szerokości pasma ma wpływ zarówno na oszacowanie jego wartości, jak również na zmienność uzyskiwanych wyników. Spośród analizowanych metod wyróżnia się sposób ucw, który daje przeciętnie najmniejsze wartości pasma przy jednocześnie największej ich zmienności w obrębie prób pobieranych z tego samego drzewostanu. Odwrotne zależności obserwuje się w przypadku metody onestage, przy zastosowaniu której otrzymuje się przeciętnie najwięk-

sze wartości parametru i jednocześnie niską zmienność uzyskiwanych wyników. Największą stabilnością wyników (najmniejsze zmiany pasma ze względu na liczebność próby i jednocześnie najmniejsza jego zmienność w obrębie losowanych prób w drzewostanie) charakteryzuje się metoda nrd0. Wybór metody szacowania parametru wygładzania może decydować więc o jakości opracowywanego modelu. Niskie wartości parametru mogą powodować nadmierne odwzorowywanie struktury próby i tym samym generować model obciążony zarówno dużymi błędami systematycznymi, jak i dużą wariancją. Niekorzystną cechą metod szacowania pasma jest duża zmienność wartości generowanych w kolejnych próbach. Modele struktury pierśnic uzyskiwane z wykorzystaniem takich metod mogą charakteryzować się nadmierną dokładnością odwzorowywania struktury próby lub zbyt dużym jej uogólnieniem. Spośród testowanych metod zastrzeżenia te odnoszą się szczególnie do sposobu ucw, który na ogół generuje najmniejsze wartości pasm, na dodatek obciążone dużą zmiennością (ryc. 4).

Rezultaty przeprowadzonych analiz stawiają również pod znakiem zapytania praktykę przyjmowania stałej wartości parametru wygładzania dla prób pochodzących z różnych drzewostanów oraz zmuszają do refleksji nad powszechnie stosowaną praktyką wyboru 2-centymetrowych stopni grubości w parametrycznym modelowaniu frekwencji pierśnic. Pomijany na ogół w opisie struktury grubości etap związany z ustaleniem szerokości stopnia grubości może wpływać zarówno na wybór zastosowanego modelu rozkładu, jak i jakość estymacji jego parametrów, co ma miejsce zwłaszcza w przypadku rozkładów skośnych [Taubert i in. 2013].

Nieuzasadnione wydaje się również określanie szerokości pasma na podstawie bardzo licznych prób. Z badań wynika bowiem, że dwukrotne zwiększenie wielkości próby, np. z 50 do 100 wylosowanych pierśnic, skutkuje tylko nieznacznym pomniejszeniem błędu standardowego (ryc. 3b).

Wyniki przeprowadzonych analiz uwiadcniają skalę trudności związanych z określeniem właściwej szerokości pasma w nieparametrycznej metodzie modelowania struktury grubości drzewostanu wykorzystującej estymator jądrowy. Słuszna w tym miejscu wydaje się uwaga Loadera [1999], zgodnie z którą nie powinniśmy polegać na jednej metodzie ustalenia parametru wygładzania. Również Podlaski [2013] zaleca przetestowanie kilku wartości szerokości pasma, zaczynając od szerokości stopnia grubości wykorzystanego w empirycznym opisie rozkładu pierśnic. Przedstawione wyniki, a zwłaszcza stwierdzone związki pasma z cechami drzewostanu mogą być pomocne przy wyborze wstępnych wartości jego szerokości bądź też zakresu testowanych wartości w trakcie modelowania struktury grubości. Wykazane dotychczas problemy aplikacyjne towarzyszące sto-



Ryc. 4.

Rozkład szerokości pasma oszacowanego analizowanymi metodami dla próbek o liczebności 100 drzew

Distribution of bandwidth values determined by means of analyzed methods to samples containing 100 trees

oznaczenia jak w tabeli 2; denotes as in table 2

sowaniu estymatorów jądrowych nie powinny bezpośrednio uzasadniać ich krytyki, a w szerszym kontekście także krytyki metod nieparametrycznych, lecz dostarczać motywacji dla kolejnych badań poszerzających zakres narzędzi umożliwiających i wspierających modelowanie struktury grubości.

Wnioski

- ✦ Szerokość oraz zmienność pasma estymatora jądrowego determinowane są wiekiem drzewostanu, wielkością użytej próby oraz zastosowaną metodą oszacowania.
- ✦ Ścisły związek pasma z wiekiem, przeciętną pierśnicą, a szczególnie z odchyleniem standardowym pierśnic wskazuje na potencjalne możliwości wykorzystania tych cech do ustalania jego szerokości.
- ✦ Arbitralne przyjmowanie stałej wartości parametru wygładzania może być powodem nieścisłości w modelowaniu rozkładu pierśnic drzewostanów olszowych metodą estymatorów jądrowych.
- ✦ Nieznaczne zmiany szerokości pasm dla prób o liczebności powyżej 50 wylosowanych pierśnic, a zwłaszcza utrzymywanie się błędu standardowego ich oszacowania na podobnym poziomie, dają obiektywne podstawy do ustalenia optymalnej liczby drzew niezbędnej do modelowania struktury grubości drzewostanów olszowych metodą estymatorów jądrowych.

Literatura

- Altman N., Léger C. 1995. Bandwidth selection for kernel distribution function estimation. *Journal of Statistical Planning and Inference* 46 (2): 195-214.
- Baszczyńska A. 2014. Metoda jądrowa w analizie finansowych szeregów czasowych. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu* 371: 23-31.
- Bollandsås O. M., Maltamo M., Gobakken T., Naeset E. 2013. Comparing parametric and non-parametric modelling of diameter distributions on independent data using airborne laser scanning in a boreal conifer forest. *Forestry* 86 (4): 493-501.
- Borders B. E., Souter R. A., Bailey R. L., Ware K. D. 1987. Percentile – based distributions characterize forest stand tables. *Forest Science* 33 (2): 570-576.
- Borowski M. 1974. *Przyrost drzew i drzewostanów*. PWRiL, Warszawa.
- Čavlović J., Božić M., Boncina A. 2006. Stand structure of an uneven-aged fir-beech forest with an irregular diameter structure: modelling the development of the Belevine forest, Croatia. *European Journal of Forest Research* 125 (4): 325-333.
- Duin R. P. W. 1976. On the choice of smoothing parameters for Parzen estimators of probability density functions. *IEEE Transactions on Computers*.
- Guidoum A. C. 2015. *kedd: Kernel estimator and bandwidth selection for density and its derivatives*. R package version 1.0.3. <http://CRAN.R-project.org/package=kedd>
- Haara A., Maltamo M., Tokola T. 1997. The k-nearest-neighbour method for estimating basal-area diameter distribution. *Scandinavian Journal of Forest Research* 12 (2): 200-208.
- Hann D. W., Bare B. B. 1979. *Uneven-aged forest management: State of the Art (or Science?)*. USDA Forest Service, General Technical Report INT-50.
- Hansen B. E. 2004. *Bandwidth selection for nonparametric distribution estimation*. Manuskrypt. Uniwersytet Wisconsin.
- Kangas A., Maltamo M. 2000. Percentile based basal area diameter distribution models for Scots pine, Norway spruce and birch species. *Silva Fennica* 34 (4): 371-380.
- Kulczycki P. 2005. *Estymatory jądrowe w analizie systemowej*. Wydawnictwo Naukowo-Techniczne, Warszawa.
- Kulczycki P. 2007. Analiza danych z użyciem estymatorów jądrowych w zastosowaniu do diagnostyki systemów. W: Korbicz J., Patan K., Kowal M. [red.]. *Diagnostyka procesów i systemów*. EXIT, Warszawa. 231-238.
- Lima R. A. F., Batista J. L. F., Prado P. I. 2015. Modeling tree diameter distributions in natural forests: an evaluation of 10 statistical models. *Forest Science* 61 (2): 320-327.
- Liu C., Zhang L., Davis C. J., Solomon D. S., Gove J. H. 2002. A Finite Mixture Model for characterizing the diameter distributions of mixed-species forest stands. *Forest Science* 48 (4): 653-661.
- Loader C. R. 1999. Bandwidth selection: classical or plug-in? *The Annals of Statistics* 27 (2): 415-438.

- López-de-Ullibarri I. 2015. Bandwidth selection in kernel distribution function estimation. *The Stata Journal* 15 (3): 784-795.
- Łukasik S. 2008. Identyfikacja rozkładu w systemach rzeczywistych za pomocą estymatorów jądrowych. *Czasopismo Techniczne Politechniki Krakowskiej* 17: 3-13.
- Maltamo M., Kangas A. 1998. Methods based on k-nearest neighbour regression in the prediction of basal area diameter distribution. *Canadian Journal of Forest Research* 28 (8): 1107-1115.
- Maltamo M., Malinen J., Kangas A., Härkönen S., Pasanen A. M. 2003. Most similar neighbour-based stand variable estimation for use in inventory by compartments in Finland. *Forestry* 76 (4): 449-463.
- Orzel S., Pogoda P., Ochał W. 2017. Ocena przydatności wybranych funkcji do modelowania rozkładu pierśnic w drzewostanach olszy czarnej (*Alnus glutinosa* (L.) Gaertn.). *Sylvan* 161 (2): 101-113.
- Parzen E. 1962. On estimation of a probability density function and mode. *Annals of Mathematical Statistics* 33 (3): 1065-1076.
- Podlaski R. 2013. Parametryczna i nieparametryczna aproksymacja rozkładów pierśnic w drzewostanach o różnej budowie pionowej. *Acta Agraria et Silvicultura series Silvestris* 51: 27-44.
- Podlaski R. 2017. Forest modeling: the gamma shape mixture model and simulation of tree diameter distributions. *Annals of Forest Science* 74 (2): 1-10.
- Podlaski R., Roesch F. A. 2014. Aproksymacja rozkładów pierśnic drzew w dwugeneracyjnych drzewostanach za pomocą rozkładów mieszanych. III. Estymatory jądrowe a rozkłady mieszane. *Sylvan* 158 (6): 414-422.
- Pretzsch H. 2000. From yield tables to simulation models for pure and mixed stands. *Journal of Forest Science* 46 (3): 97-113.
- Pukkala T., Lähde E., Laiho O. 2010. Optimizing the structure and management of uneven-sized stands of Finland. *Forestry* 83 (2): 129-142.
- Quintela-del-Río A., Estévez-Pérez G. 2012. Nonparametric kernel distribution function estimation with kerdist: An R package for bandwidth choice and applications. *Journal of Statistical Software* 50 (8): 21.
- Raykar V. C., Duraiswami R. 2005. Very fast optimal bandwidth selection for univariate kernel density estimation. In CS-TR-4774, Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, Collegepark.
- Robinson A. P., Hamann J. D. 2011. *Forest analytics with R. An introduction*. Springer New York Dordrecht Heidelberg London.
- Shaw J. D. 2006. Reineke's Stand Density Index: Where are we and where do we go from here? *Proceedings: Society of American Foresters 2005 National Convention*, Oct. 19-23.
- Sheather S. J. 2004. Density estimation. *Statistical Science* 19 (4): 588-597.
- Silverman B. W. 1986. *Density estimation for statistics and data analysis*. Chapman and Hall, London.
- Taubert F., Hartig F., Dobner H. J., Huth A. 2013. On the challenge of fitting tree size distributions in ecology. *PLoS ONE* 8 (2): e58036.
- Téllez A. E., Grande Ortiz M. Á., González García C., Martín Fernández Á. J., García A. I. 2008. The kernel estimation in biosystems engineering. *Journal of Systemics, Cybernetics and Informatics* 6 (2): 23-27.
- Uusitalo J., Puustelli A., Kivinen V.-P., Nummi T., Sinha B. K. 2006. Bayesian estimation of diameter distribution during harvesting. *Silva Fennica* 40 (4): 663-671.
- Uutera J., Maltamo M. 1995. Impact of regeneration method on stand structure prior to first thinning. Comparative study North Karelia, Finland vs. Republic of Karelia, Russian Federation. *Silva Fennica* 29 (4): 267-285.
- Wand M. P., Jones M. C. 1995. *Kernel smoothing*. Chapman and Hall, London, New York.
- Wang X. F. 2012. sROC: Nonparametric smooth ROC curves for continuous data. R package version 0.1-2. <https://CRAN.R-project.org/package=sROC>
- Zambom A. Z., Dias R. 2012. A review of kernel density estimation with applications to econometrics. arXiv preprint arXiv:1212.2812.