

*Dariusz Foszcz**, *Daniel Saramak**,
*Tadeusz Tumidajski**, *Tomasz Niedoba**, *Tomasz Gawenda**

MOŻLIWOŚCI POPRAWY DOKŁADNOŚCI APROKSYMACJI KRZYWYCH SKŁADU MATERIAŁÓW UZIARNIONYCH

1. Wstęp

W analizie przebiegu procesów przeróbki surowców mineralnych kluczową rolę odgrywają krzywe składu, czyli krzywe ilustrujące zależność ilości danej frakcji od cechy tej frakcji. Krzywe składu mogą być nieskumulowane lub skumulowane (krzywe rozkładu, prawdopodobieństwa, gęstości, dystrybuanty). Ze względu sposób wyznaczania krzywych składu (kilka lub kilkanaście punktów pomiarowych), stosuje się zwykle graficzne metody aproksymacji otrzymanych danych. Kolejne punkty kumulowanej krzywej składu (dystrybuanty empirycznej) mają na osi y sumę ilości (wychodów) kolejnych frakcji, co jest powodem kumulacji błędów oraz zmiany rangi punktów mierzonej np. przyrostem wychodów.

Aby właściwie wyeksponować rangę (jakość) punktów wyznaczających dystrybuantę empiryczną w procedurach aproksymacyjnych należy uwzględnić następujące ich kwalifikacje:

- ograniczenie liczby punktów branych pod uwagę w obliczeniach (zastosowanie rozkładów uciętych lub cenzurowanych),
- pominięcie punktów wątpliwych (elementy analizy błędów),
- nadanie punktom wag w zależności od ich wkładu w sumaryczną wartość dystrybuanty.

Prezentowany artykuł jest poświęcony szczegółowej analizie wymienionych wyżej podejść do poszukiwania wzorów aproksymacyjnych dystrybuant rozkładów cech materiałów uziarnionych [1, 2, 7, 9, 11–14, 17, 19, 21].

* Wydział Górnictwa i Geoinżynierii, Akademia Górniczo-Hutnicza, Kraków

2. Zastosowanie rozkładów cenzurowanych — problemy ustalenia granicy cenzurowania

Przy analizie krzywych składu materiałów uziarnionych należy uwzględnić rzeczywisty zakres zmian wartości badanej cechy i stosować, jako wzory aproksymujące funkcje gęstości, czy dystrybuanty, wzory dla rozkładów cenzurowanych lub uciętych. Generalnie, obydwa rodzaje takich rozkładów dotyczą wielkości o ograniczonych przedziałach zmienności danej cechy, przy czym dla rozkładów uciętych stosuje się przeliczenie wartości dystrybuant przy użyciu wartości rozkładu podstawowego (wyznaczonego dla pełnego, nawet nieskończonego, zakresu zmienności) w punkcie jego podziału na części (umownego podziału na produkty), a dla rozkładu cenzurowanego — modyfikację określenia zmiennej (cechy) prowadzącą do rozszerzenia zakresu jej wartości na przedział nieograniczony lub zamknięcia w przedziale $\langle 0, 1 \rangle$.

Równaniami empirycznymi (rozkładami) stosowanymi w praktyce, dla krzywych składu ziarnowego, (po zmianie przedziału $\langle 0, d_{\max} \rangle$ na przedział $\langle 0, 1 \rangle$) są:

— rozkład potęgowy o dystrybuancie (wzór Gaudina-Schuhmanna)

$$F(d) = \left(\frac{d}{d_{\max}} \right)^k, \quad 0 \leq d \leq d_{\max}; \quad (1)$$

— rozkład Gaudina-Meloya o dystrybuancie

$$F(d) = 1 - \left(1 - \frac{d}{d_{\max}} \right)^k, \quad 0 \leq d \leq d_{\max}; \quad (2)$$

— rozkład Harrisa o dystrybuancie

$$F(d) = 1 - \left(1 - \left(\frac{d}{d_{\max}} \right)^s \right)^k, \quad 0 \leq d \leq d_{\max}. \quad (3)$$

gdzie $F(d)$ oznacza kumulowany (sumowany) wychód poszczególnych frakcji zawartych w materiale ziarnistym a d jest cechą ziarna (na przykład jego rozmiar), d_{\max} to maksymalny wymiar ziarna, dla którego $F(d) = 1$; s i k to parametry stałe.

Powszechnie stosowanymi równaniami do analizy wielkości ziaren są: rozkład Weibulla (wzór Rosina-Rammlera), rozkład logarytmiczno-normalny oraz rozkład logistyczny, oparte na przedziale zmienności $(0, +\infty)$. Transformując wielkość ziarna do wielkości:

$$\xi = \frac{d}{d_{\max}} \quad (4)$$

oraz

$$\eta = \frac{\xi}{1-\xi} = \frac{d}{d_{\max} - d} \quad (5)$$

możemy dystrybuanty rozkładów cenzurowanych zapisać wzorami:

— rozkład Weibulla

$$F(d) = 1 - \exp\left[-\left(\frac{\eta}{\eta_{63,2}}\right)^\alpha\right], \quad 0 \leq d \leq d_{\max}; \quad (6)$$

— rozkład logarytmiczno-normalny

$$F(d) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{t^2}{2}} dt, \quad (7)$$

gdzie:

$$t = \frac{\ln\left(\frac{\eta}{\eta_{50}}\right)}{\sigma},$$
$$\sigma = \frac{1}{2}[\ln(\eta_{84}) - \ln(\eta_{16})];$$

— rozkład logistyczny

$$F(d) = \frac{1}{1 + be^{-c\left(\frac{d}{d_{\max} - d}\right)}}. \quad (8)$$

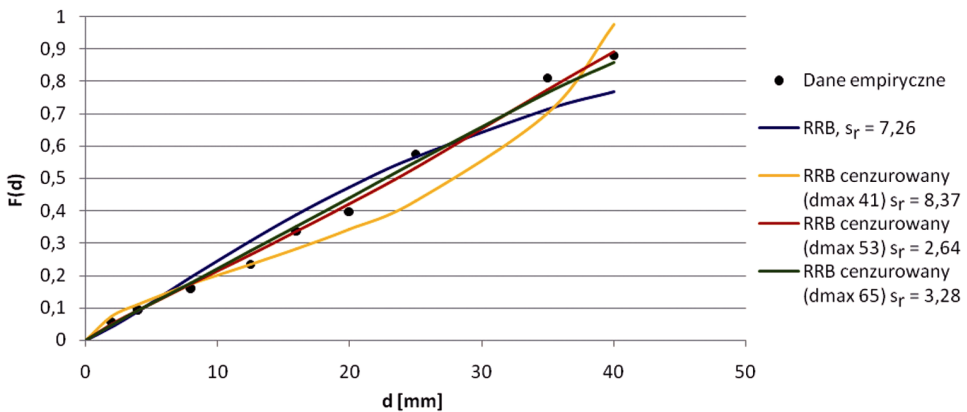
Zagadnienie cenzurowania rozkładu empirycznego polega generalnie na modyfikacji zakresu zmienności zmiennej niezależnej, którą w naszej pracy będzie rozmiar wielkości ziarna (d). W pełnym rozkładzie zmienna d może zmieniać się w granicach $(0, +\infty)$, przy stosowaniu procedury cenzurowania, zaś zakres ten się zmienia na przedział $(0, d_{\max})$. We wzorze opisującym rozkład empiryczny zmienna d przyjmuje postać $\left(\frac{d}{d_{\max} - d}\right)$.

Opisana procedura cenzurowania dotyczy przypadku ograniczenia rozkładu od góry, kończąc go na przyjętej arbitralnie wartości maksymalnej zmiennej niezależnej d wynoszą-

cej d_{\max} . Istnieje także możliwość ocenzurowania rozkładu „od dołu”, czyli ustalenia minimalnej wartości d_{\min} , w rezultacie tego zakres stosowności wzoru będzie w przedziale $(d_{\min}, +\infty)$. Można dokonać też cenzury rozkładu z obu stron, wtedy we wzorze zamiast wartości d występuje wartość $\left(\frac{d - d_{\min}}{d_{\max} - d}\right)$.

Można przyjąć, że cenzurowany rozkład Weibulla ma udokumentowane zastosowania praktyczne [3, 10]. Kołmogorow i Epstein [7, 12] wykazali zasadność stosowania rozkładu logarytmiczno-normalnego do opisu rozdrabniania drobnego. Ustalenie dokładnej wartości d_{\max} jest niemożliwe i należy rozważyć szacowanie tej wartości metodą aproksymacji. Jeżeli przyjmiemy, że któryś ze wzorów opisanych wyżej jest dobrym przybliżeniem dystrybucyjności rozkładu wielkości ziaren w rozpatrywanym zbiorze, to można wskazać d_{\max} jako wartość, która minimalizuje sumę kwadratów odchyleń między wartościami dystrybucyjnymi — rzeczywistymi i aproksymowanymi (metodą kolejnego wprowadzania wartości d_{\max} większych od wymiaru oczka największego używanego sita w analizach).

W pracy [10] zostały przedstawione wyniki kruszenia różnych surowców w kruszarce szcękowej oraz aproksymacje dystrybucyjności rozkładem cenzurowanym Weibulla, wzorem sugerowanym przez zespół badaczy włoskich [3]. Wykorzystując dane z wymienionej pracy dokonano aproksymacji danych empirycznych wzorami określającymi dystrybucyjność rozkładu Weibulla (wzorem Rosina-Rammlera-Bennetta) oraz dystrybucyjności cenzurowanych rozkładów Weibulla (dla d_{\max} równych odpowiednio 45, 52 i 60 mm). Efekty aproksymacji przedstawiono na rysunku 1.



Rys. 1. Aproksymacja danych eksperymentalnych dotyczących składu ziarnowego w postaci wychód skumulowany $F(d)$ jako funkcja rozmiaru oczka większego sita danej klasy ziarnowej d cenzurowanym rozkładem Weibulla (RRB)

Jak łatwo zauważyć, dobór wartości d_{\max} ma bardzo istotny wpływ na dokładność aproksymacji (wartości odchyleń resztowych s_r). Wartości odchyleń resztowych s_r oblicza się z zależności:

Należy podkreślić, że klasyczny rozkład Weibulla raczej nie powinien być stosowany do aproksymacji danych rzeczywiście opisujących skład ziarnowy materiału.

3. Aproksymacja uwzględniająca wkład (wagi) punktów w wartości dystrybuanty

Ze względu na to, że drugie współrzędne punktów wyznaczających dystrybuantę empiryczną badanego rozkładu cechy materiału uziarnionego powstają przez sumowanie wychodów klas lub frakcji ($F(d)$), można rozpatrywać aproksymację dystrybuanty uwzględniającą rolę poszczególnych punktów mierzona przyrostem wartości dystrybuanty (wagę punktu). Jeżeli zadanie minimalizacji ważonej sumy kwadratów odchyłeń zapiszemy w postaci:

$$\sum_{i=1}^k w_i (y_i - \hat{y}_i)^2 = \min,$$

to w przypadku linearyzacji wzoru aproksymującego otrzymujemy klasyczne wzory na parametry prostej a i b , tzn.

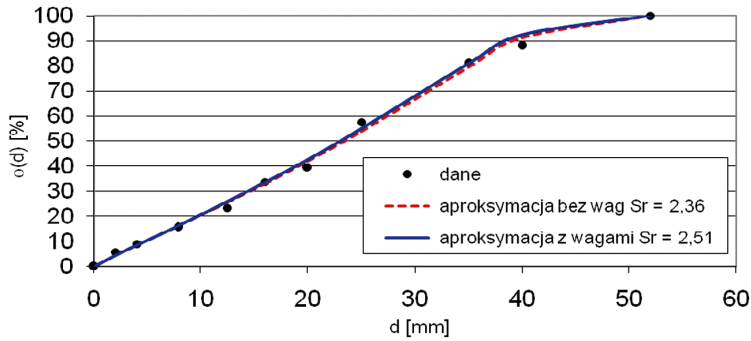
$$a = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2},$$

$$b = \bar{y} - a\bar{x},$$

gdzie nadkreślenia oznaczają uśrednienia względem wagi więc np.

$$\overline{xy} = \frac{\sum_{i=1}^k w_i x_i y_i}{\sum_{i=1}^k w_i},$$

Wyniki aproksymacji z użyciem wag, które są przyrostami wychodów, przedstawiono na rysunku 2. Zmiana wartości s_r dla aproksymacji jest spowodowana tym, że krzywa dostosowuje się do wybranych (bardziej obciążonych punktów), zwiększając różnice w punktach pozostałych. Zasadność tego typu aproksymacji jest dyskusyjna i powinna być komentowana w każdym przypadku jej stosowania.



Rys. 2. Aproxymacja danych empirycznych przy uwzględnieniu wag

4. Nieparametryczne metody aproxymacji dystrybuant rozkładów cech materiałów uziarnionych

Przedstawione metody aproxymacji bazowały na konkretnych wzorach określających dystrybuantę rozkładu badanej cechy materiału uziarnionego, zwłaszcza wielkości ziarna. W wielu przypadkach nie ma żadnych podstaw do preferowania określonego rozkładu cechy, który może być rozkładem absolutnie nietypowych, nie stwarzającym możliwości linearyzacji wzoru określającego dystrybuantę. W takich przypadkach można, dla aproxymacji dystrybuanty zastosować tzw. metody nieparametryczne sprowadzające się do wykorzystania rozwinięcia funkcji w szereg Fouriera [5, 23, 24] lub metod wykorzystujących funkcje jądrowe. W przypadku wykorzystywania metody Fouriera celem jest znalezienie funkcji gęstości (lub dystrybuanty — zależne jest to od postaci zestawu danych empirycznych) o postaci:

$$f(x) = \sum_{j=0}^{\infty} Q_j \varphi_j(x) \tag{9}$$

gdzie Q_j oznacza współczynniki Fouriera zadane wzorem:

$$Q_j = \int_0^1 f(x) \varphi_j(x) dx \tag{10}$$

natomiast

$$\varphi_j(x) = \begin{cases} 1 & \text{dla } j = 0 \\ \sqrt{2} \cos \pi j x & \text{dla } j = 1, 2, \dots \end{cases}$$

Aby to zrealizować, zadany zbiór danych empirycznych musi zostać sprowadzony do przedziału $[0, 1]$, za pomocą transformacji liniowej $X = \frac{d-a}{b-a}$ (w naszym przypadku $b = d_{\max}$, a $a = 0$). Tym samym otrzymujemy ciąg $\{\varphi_n(x)\}$, który jest ciągiem funkcyjnym ortonormalnym na przedziale $[0, 1]$.

Estymatorem gęstości badanego rozkładu jest suma częściowa szeregu Fouriera

$$f_J(x) = \sum_{j=0}^J Q_j \varphi_j(x) \quad (11)$$

gdzie J nazywamy parametrem obcięcia.

Jako estymator współczynników Fouriera \hat{Q}_j przyjmuje się:

$$\hat{Q}_j = \frac{1}{n} \sum_{i=1}^n \varphi_j(x_i) \quad (12)$$

Estymator ten jest estymatorem nieobciążonym i mocno zgodnym.

Za estymator dla parametru obcięcia \hat{J} można przyjąć:

$$\hat{J} = \arg \min_{0 \leq J \leq J_n} \sum_{j=0}^J \left(\frac{2\hat{Q}_0}{n} - \hat{Q}_j^2 \right) \quad (13)$$

dla którego $\arg \min_{0 \leq J \leq J_n} (a_s)$ przyjmuje wartości równe współczynnikowi s najmniejszego elementu spośród (a_0, \dots, a_s) , natomiast J_n oznacza część całkowitą liczby $4 + 0,5 \ln n$.

Inną możliwością nieparametrycznej aproksymacji są metody jądrowe oparte na tzw. funkcji jądrowej, której postać może być różnorodna [4, 6, 8, 16, 18]. Wówczas dla zadanej prostej próbki losowej X_1, \dots, X_n estymator jądrowy gęstości definiuje się jako:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \quad (14)$$

gdzie $K(x)$ oznacza funkcję jądrową, spełniającą warunek:

$$\int_{-\infty}^{\infty} K(x) dx = 1 \quad (15)$$

x_1, \dots, x_n oznacza realizację próbki X_1, \dots, X_n natomiast $h > 0$ jest parametrem zwanym szerokością pasma lub parametrem wygładzającym.

Dla naszych danych zostały wykorzystane funkcje jądrowe: jądro Epanechnikowa oraz jądro Gaussa, określone wzorami:

$$K(x) = \begin{cases} \frac{3}{4\sqrt{5}b} \left(1 - \frac{1}{5b^2} x^2\right) & \text{dla } |x| \leq b\sqrt{5} \\ 0 & \text{dla } |x| > b\sqrt{5} \end{cases} \quad (16)$$

gdzie b — parametr skali $\left(\int_{-\infty}^{\infty} t^2 K(t) dt = b^2\right)$ oraz

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right), x \in R \quad (17)$$

Optymalna szerokość pasma zdefiniowana jest wzorami:

— dla jądra Epanechnikowa

$$h_{\text{opt}} = 1,056\sigma n^{-\frac{1}{5}} \quad (18)$$

— dla jądra gaussowskiego

$$h_{\text{opt}} = 1,066\sigma n^{-\frac{1}{5}} \quad (19)$$

gdzie:

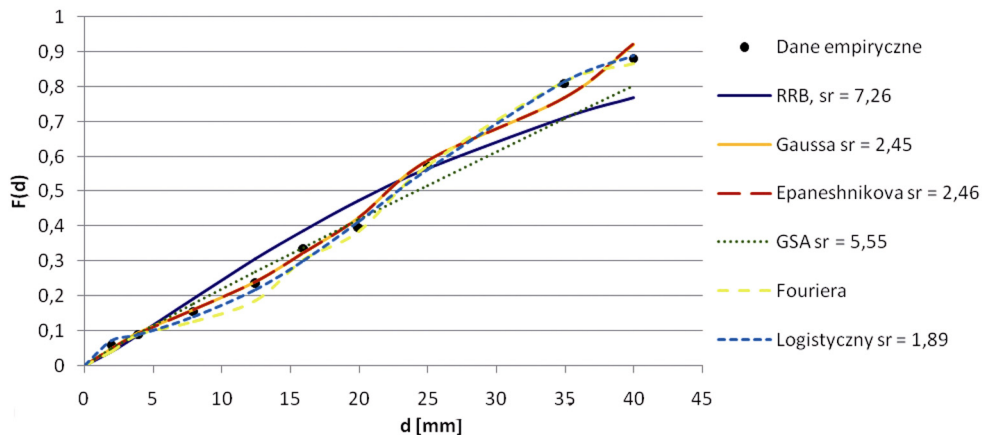
n — liczebność próbki;

σ — odchylenie standardowe.

W praktyce σ zastępujemy estymatorem s obliczonym na podstawie próbki losowej

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

Na rysunku 3 zamieszczono wykresy aproksymant otrzymanych za pomocą metod nieparametrycznych. Wyniki otrzymywane metodami jądrowymi są do siebie bardzo zbliżone i można powiedzieć, że ich dokładność jest w zasadzie jednakowa. Nieco gorsze efekty dała metoda Fouriera. W porównaniu do tzw. rozkładów klasycznych, jedynie rozkład logistyczny dał lepsze wyniki dopasowania.



Rys. 3. Aproxymacja danych empirycznych nieklasycznymi rozkładami

5. Podsumowanie

Prezentowane w artykule zagadnienia związane z matematycznym opisem (aproxymacją) empirycznych krzywymi charakteryzującymi rozkłady cech materiałów uziarnionych pokazują różnorodność realizacji oraz interpretacji wyników. Okazuje się, że najważniejszym argumentem przy doborze sposobu aproxymacji są dalsze cele jej zastosowań. Jeżeli kumulowane wychody (równania dystrybuant) frakcji (klas ziarnowych) są podstawą modelowania pracy kruszarek czy klasyfikatorów, to odpowiedni wzór (dystrybuanta) gra kluczową rolę. Uzależniając parametry wzoru od charakterystyk techniczno-technologicznych urządzeń można prowadzić symulację jego pracy oraz poszukiwać warunków optymalnych przy przyjętych kryteriach oceny (np. ekonomicznych) [20]. Kwestia dokładności aproxymacji dystrybuanty rozkładu cenzurowanego (uzależnionej od produkowanego d_{\max}) w takich badaniach nabiera bardzo istotnego znaczenia. Należy jeszcze raz podkreślić, że przy proponowaniu postaci dystrybuanty powinno się kierować chociażby minimalnymi uzasadnieniami teoretycznymi jej postaci.

W wielu przypadkach użycie wychodów klas jest niezbędne przy modelowaniu przebiegu procesów, które nie jest nakierowane na określenie składu materiału. W takich przypadkach użycie nieparametrycznych metod aproxymacji dystrybuant jest uzasadnione i umotywowane dużą dokładnością tej metody (pod warunkiem niezmienności przebiegu dystrybuanty rzeczywistej). Nieparametryczna metoda aproxymacji dystrybuant była używana także do generowania dystrybuant rozkładów wielowymiarowych cech materiału po przyjęciu, że cechy te są stochastycznie niezależne [15].

Na podstawie przeprowadzonych badań można sformułować wnioski:

- 1) Przy opisie kumulowanych krzywych składu ziarnowego powinno się preferować rozkłady cenzurowane, zwłaszcza rozkłady Weibulla, logarytmiczno-normalny i logistyczny.

Dobór wartości cenzurowania powinien być dokonywany dla każdego przypadku metodą kolejnych przeliczeń.

- 2) Warto stosować, w uzasadnionych przypadkach, nieparametryczne metody estymacji, które pozwalają zwiększyć jej dokładność.
- 3) Dokładność aproksymacji krzywych składu ziarnowego może być oceniana tylko na podstawie odchylenia resztowego s_r , wyznaczanego dla wartości dystrybuant, a nie ich przekształceń funkcyjnych.

W tym miejscu warto odnotować także jedną istotną uwagę. Jeżeli przyjmuje się wzór aproksymujący daną dystrybuantę rozkładu cechy materiału uziarnionego (uzasadniony lub nie), to dokładność aproksymacji nie zależy od przyjętej metody minimalizacji odchyień resztowych określanych jako suma (lub jej przetworzenie) kwadratów różnic między wartością rzeczywistą a estymowaną. Wyznaczane wartości parametrów mogą różnić się minimalnie w zależności od stosowanej metody obliczeń (linearyzacja zależności, nieliniowe modele, zastosowanie technik ewolucyjnych).

LITERATURA

- [1] *Andrejew C.E., Towarow W.W., Pierow W.A.*: Zakonomiornosti izmierzennija i uscislenije charakteristik granulometričeskogo sostawa, Moskwa, 1959
- [2] *Brożek M., Tumidajski T.*: Granulometric characteristics of the product of crushing by compression of single particles, Arch. Min. Sci., vol. 41, pp. 245–258, 1996
- [3] *Cardu M., Clerici C., Morandini A., Ocella E.*: An experimental research on the comminution law and work index in jaw, Proceedings of XVIII International Mineral Processing Congress, Sydney, 1993
- [4] *Efromovich S.*: Nonparametric curve estimation, New York, Springer-Verlag, 1999
- [5] *Efron B., Tibshirani R.*: Using specially designed exponential families for density estimation, Ann. Statist. 24, pp. 2431–2461, 1996
- [6] *Eggermont P.P.B., La Ricia V.N.*: Maximum penalized likelihood estimation, vol. I: Density estimation, Springer, New York, 2001
- [7] *Epstein B.*: Logarithmico-normal distributions in breakage of solids, Ind. Eng. Chem. 40, pp. 2289–2291, 1948
- [8] *Eubank R.L.*: Spline smoothing and nonparametric regression, Marcel Dekker, New York, 1988
- [9] *Gaudin A.M., Meloy T.P.*: Model and comminution distribution equation for single fracture, Trans. AIME, vol. 223, pp. 40–43, 1962
- [10] *Gawenda T.*: Ocena wpływu właściwości fizykochemicznych surowców skalnych i parametrów technologicznych kruszarek szczękowych na efekty rozdrabniania, Rozprawa doktorska, AGH, Kraków, 2004
- [11] *Gawenda T., Niedoba T., Przybycień K., Tumidajski T.*: Zastosowanie algorytmów genetycznych do modelowania procesów przeróbki surowców mineralnych, Górnictwo i Geoinżynieria, vol. 4, pp. 101–111, 2009
- [12] *Kolmogorow A.N.*: O logaritmiczeski normalnom zakonie raspriedielenija razmierow czastic pri drobleniji, Dokl. AN SSSR, 31, pp. 99–101, 1941
- [13] *Malewski J.*: O modelach krzywych składu ziarnowego produktów rozdrabniania, Prace Naukowe Instytutu Górnictwa Politechniki Wrocławskiej, vol. 28, Wrocław, 1978
- [14] *Niedoba T., Tumidajski T.*: Aproksymacja krzywych składu ziarnowego za pomocą różnych metod statystycznych, Zeszyty Naukowe Politechniki Śląskiej „Górnictwo”, vol. 266, pp. 125–138, 2005
- [15] *Niedoba T., Tumidajski T.*: The approximation of grain composition curves by non-parametric statistical methods, Proceedings of XXIII IMPC, vol. 1, pp. 203–208, Istanbul, Turcja, 2006
- [16] *Prakasa Rao B.L.S.*: Nonparametric functional estimation, Academic Press, New York, 1983
- [17] *Saramak D., Tumidajski T.*: Rola i sens aproksymacji krzywych składu ziarnowego surowców mineralnych, Górnictwo i Geoinżynieria, vol. 3/1, pp. 301–314, 2006

- [18] *Silverman B.W.*: Nonparametric density estimation for statistic and data analysis, London New York, Chapman and Hall, 1986
- [19] *Trustrum K., Jayatilaka A.*: Applicability of Weibull analysis for brittle materials, *J. Mat. Sci.*, vol. 18, pp. 2765–2770, 1983
- [20] *Tumidajski T., Foszcz D., Jamróz D., Niedoba T., Saramak D.*: Niestandardowe metody statystyczne i obliczeniowe w opisie procesów przeróbki surowców mineralnych, Wydawnictwo IGSMiE PAN, Kraków, 2009
- [21] *Tumidajski T., Saramak D., Niedoba T.*: Matematyczne aspekty opisu i oceny wzbogacalności rud miedzi, *Górnictwo i Geoinżynieria*, vol. 4, pp. 97–106, 2007
- [22] *Tumidajski T., Saramak D.*: Metody i modele statystyki matematycznej w przeróbce surowców mineralnych, Wydawnictwo AGH, Kraków, 2009
- [23] *Wahba G.*: Data-based optimal smoothing of orthogonal series density estimates, *Ann. Statist.* 9, pp. 146–156, 1981
- [24] *Watson G.S.*: Density estimation by orthogonal series, *Ann. Math. Statist.* 40, pp. 1496–1498, 1969