

OCENA DOKŁADNOŚCI OSZACOWANIA WSKAŹNIKA STANDARYZOWANEGO OPADU (SPI) WYZNACZONEGO ZA POMOCĄ RÓŻNYCH ROZKŁADÓW PRAWDOPODOBIENSTWA

Edward Gąsiorek¹, Elżbieta Musiał¹, Marian Rojek²

¹ Katedra Matematyki, Uniwersytet Przyrodniczy we Wrocławiu, ul. Grunwaldzka 53, 50-357 Wrocław, e-mail: elzbieta.musial@up.wroc.pl

² Instytut Kształtowania i Ochrony Środowiska, Uniwersytet Przyrodniczy we Wrocławiu.

STRESZCZENIE

Stosowanie różnych metod obliczania wskaźnika standaryzowanego opadu (SPI) związane jest z otrzymywaniem różnych jego przybliżeń. Metody oparte na rozkładzie normalnym i jego transformacjach oraz na rozkładzie gamma dają podobne wyniki, stąd można je stosować zamiennie. Metoda oparta na rozkładzie logarytmiczno-normalnym daje wyraźnie różne wartości tego wskaźnika szczególnie dla wartości ekstremalnych SPI. Pojawia się wobec tego problem, która z metod daje rozkład lepiej dopasowany do danych empirycznych. Celem tej pracy jest uporządkowanie powyższych metod ze względu na stopień dopasowania do danych empirycznych z Obserwatorium Agro- i Hydrometeorologii we Wrocławiu Swojcu z wielolecia 1964–2009.

Słowa kluczowe: wskaźnik standaryzowanego opadu, rozkład gamma, rozkład normalny, rozkład log-normalny.

ASSESSMENT OF ACCURACY OF PRECIPITATION INDEX (SPI) DETERMINED BY DIFFERENT PROBABILITY DISTRIBUTIONS

ABSTRACT

The use of different calculating methods to compute the standardized precipitation index (SPI) results in various approximations. Methods based on normal distribution and its transformations, as well as on gamma distribution, give similar results and may be used equally, whereas the lognormal distribution fitting method is significantly discrepant, especially for extreme values of SPI. Therefore, it is problematic which method gives the distribution optimally fitted to empirical data. The aim of this study is to categorize the above mentioned methods according to the degree of approximation to empirical data from the Observatory of Agro- and Hydrometeorology in Wrocław-Swojec from 1964–2009 years.

Keywords: standardized precipitation index, gamma distribution, normal distribution, lognormal distribution.

WSTĘP

Susze meteorologiczne charakteryzowane za pomocą wskaźnika standaryzowanego opadu (SPI) są okresami, w których opady atmosferyczne są znacząco niższe od normy dla danego obszaru. Wskaźnik standaryzowanego opadu (SPI) identyfikujący różne rodzaje susz meteorologicznych można wyznaczać wieloma metodami wykorzystującymi różne rozkłady prawdopodobieństw. Są to metody oparte na rozkładzie normalnym i jego transformacjach, na rozkładzie gamma oraz na rozkładzie logarytmiczno-normalnym. Stosując metody oparte na rozkładzie normalnym i jego transformacjach oraz na rozkładzie gamma otrzymuje się niemal identyczne wartości SPI, stąd metody te można stosować zamiennie [Gąsiorek i in. 2012]. Metoda oparta na rozkładzie logarytmiczno-normalnym daje wyraźnie różniące się wartości wskaźnika SPI, szczególnie dla wartości ekstremalnych SPI. Pojawia się wobec tego problem, która z metod daje rozkład lepiej dopasowany do danych empirycznych.

Celem tej pracy jest uporządkowanie powyższych metod ze względu na stopień dopasowania do danych empirycznych na podstawie materiału empirycznego pochodzącego z Obserwatorium Agro-i Hydrometeorologii we Wrocławiu Swojcu z wielolecia 1964–2009.

METODY BADAŃ

W celu sprawdzenia, czy określony rozkład prawdopodobieństwa dobrze opisuje zmienność danych empirycznych charakteryzujących pewną cechę, np. opad, stosuje się test zgodności. W wyniku przeprowadzenia tego testu w praktyce albo odrzuca się zgodność rozkładu z danymi, albo przyjmuje się, że rozkład ten poprawnie je opisuje. Najczęściej stopień dopasowania rozkładu do danych określa się wizualnie przedstawiając graficznie dystrybuantę empiryczną na tle dystrybuanty dopasowanego rozkładu [Wibig 2012]. Miarą jakości tego dopasowania jest stopień w jakim te wykresy się pokrywają, natomiast za precyzyjną miarę tej rozbieżności przyjmuje się maksymalną różnicę dystrybuant, zwaną statystyką Kołmogorowa-Smirnowa [Kaczmarek 1970, Angelidis i in. 2012], którą w tej pracy zastosowano w celu określenia jakości dopasowania rozkładu do danych empirycznych. Rozkłady ciągów miesięcznych sum opadów z wielolecia były modelowane rozkładami gamma, logarytmiczno-normalnym i normalnym.

Funkcja gęstości rozkładu gamma ma postać:

$$f(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} & \text{dla } x > 0 \\ 0 & \text{dla } x \leq 0 \end{cases}$$

gdzie:

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$$

natomiast dystrybuanta ma postać:

$$F(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^x y^{\alpha-1} e^{-\frac{y}{\beta}} dy$$

W celu wyznaczenia parametrów α i β rozkładu gamma zastosowano estymatory zaproponowane przez Edwardsa i McKee [1977] wyznaczone przez Thoma [1958] metodą największej wiarygodności, postaci:

$$\hat{\alpha} = \frac{1}{4A} \left(1 + \sqrt{1 + \frac{4A}{3}} \right), \hat{\beta} = \frac{\bar{x}}{\hat{\alpha}}$$

gdzie $A = \ln(\bar{x}) - \frac{\sum \ln(x)}{n}$, natomiast \bar{x} oznacza średnią arytmetyczną.

Funkcja gęstości rozkładu logarymiczno-normalnego ma postać:

$$f(x) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} & \text{dla } x > 0 \\ 0 & \text{dla } x \leq 0 \end{cases}$$

natomiast dystrybuanta ma postać:

$$F(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$$

gdzie F oznacza dystrybuantę rozkładu normalnego-standardowego.

Powyższe rozkłady nie są określone w zerze, stąd ostatecznie dystrybuantę $H(x)$, potrzebną do wyznaczenia wskaźnika SPI, oblicza się według poniższego wzoru:

$$H(x) = (1 - q)F(x)$$

gdzie q oznacza prawdopodobieństwo wystąpienia miesiący bezopadowych. Parametr ten estymuje się za pomocą frakcji miesiący bezopadowych w wieloleciu.

Wskaźnik standaryzowanego opadu wyznacza się ze wzoru:

$$SPI = \Phi^{-1}(F(x))$$

Do wyznaczenia tej wartości można zastosować metodę aproksymacji kwantyli rozkładu normalnego standardowego [Abramowitz i in. 1965, str. 933] lub funkcję ROZKŁAD.NORMALNY.S.ODW programu Excell.

W przypadku aproksymacji rozkładu empirycznego za pomocą rozkładu normalnego zastosowano transformacje postaci $\sqrt[3]{x}$, $\sqrt[3]{x+10}$ i $\ln x$. Wskaźnik SPI otrzymano po standaryzacji transformowanych danych.

Estymatory dla parametrów m i s wyznacza się korzystając z metody największej wiarygodności:

$$\hat{\mu} = \frac{\sum \ln x}{n}$$

$$\hat{\sigma}^2 = \frac{\sum (\ln x - \mu)^2}{n}$$

W celu weryfikacji zgodności rozkładu prawdopodobieństwa z rozkładem empirycznym ciągu miesięcznych sum opadowych zastosowano test χ^2 .

Za miarę stopnia dopasowania rozkładu do danych empirycznych przyjęto statystykę Kołmogorowa-Smirnowa:

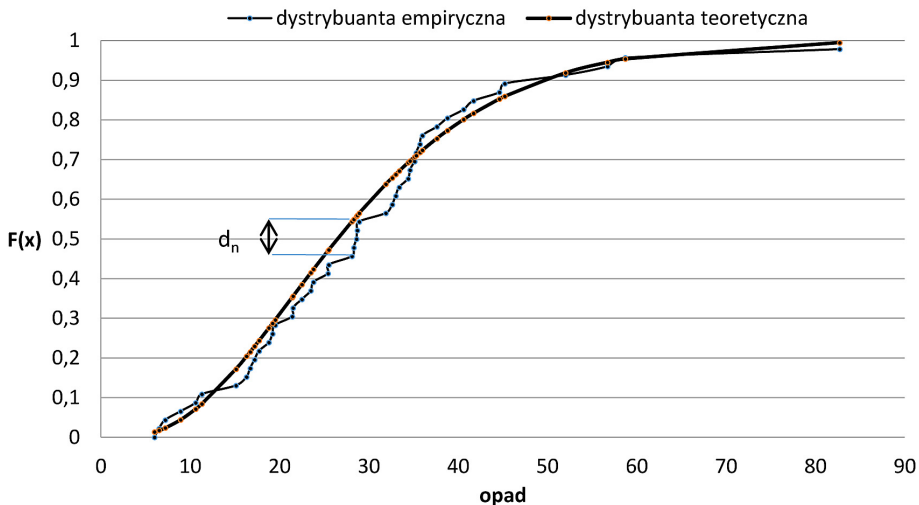
$$D_n = \sup |F_n(x) - F_0(x)|$$

gdzie F_0 oznacza dystrybuantę rozkładu teoretycznego, a F_n dystrybuantę empiryczną.

Wartość d_n statystyki D_n wyznacza się następująco [Krysicki i in. 1997]:

- porządkujemy wynik obserwacji w ciąg rosnący postaci $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$, gdzie $x_{(i)}$ oznacza i -tą statystykę pozycyjną (i -tą obserwację w uporządkowanym ciągu danych);
- obliczamy $\frac{i}{n} - F_0(x_{(i)})$ dla $i = 1, \dots, n$;
- obliczamy $d_n^+ = \max_i |\frac{i}{n} - F_0(x_{(i)})|$;
- obliczamy $F_0(x_{(i)}) - \frac{i-1}{n}$ dla $i = 1, \dots, n$;
- obliczamy $d_n^- = \max_i |F_0(x_{(i)}) - \frac{i-1}{n}|$;
- obliczamy $d_n = \max(d_n^+, d_n^-)$.

Im mniejsza wartość d_n statystyki D_n , tym rozkład teoretyczny jest lepiej dopasowany do rozkładu empirycznego.



Rys. 1. Dystrybuanta empiryczna i dystrybuanta dopasowanego rozkładu gamma dla stycznia z zaznaczoną wartością statystyki d_n

Fig. 1. Empirical and fitted gamma distributions for January with a marked value of d_n statistics

WYNIKI BADAŃ

Pierwszy etap badań to weryfikacja hipotezy o zgodności rozkładu miesięcznych sum opadów w wieloleciu 1964–2009 z rozkładami: logarytmiczno-normalnym, gamma oraz normalnym za pomocą testu χ^2 . Dane dotyczące weryfikacji tej hipotezy zostały zamieszczone w tabeli 1.

Tabela 1. Zgodność rozkładów gamma, normalnego i logarytmiczno-normalnego z empirycznym rozkładem miesięcznych sum opadów w wieloleciu 1964–2009

Table 1. Consistency of gamma, normal and lognormal distributions of monthly precipitation sums in the years 1964–2009 with empirical distribution

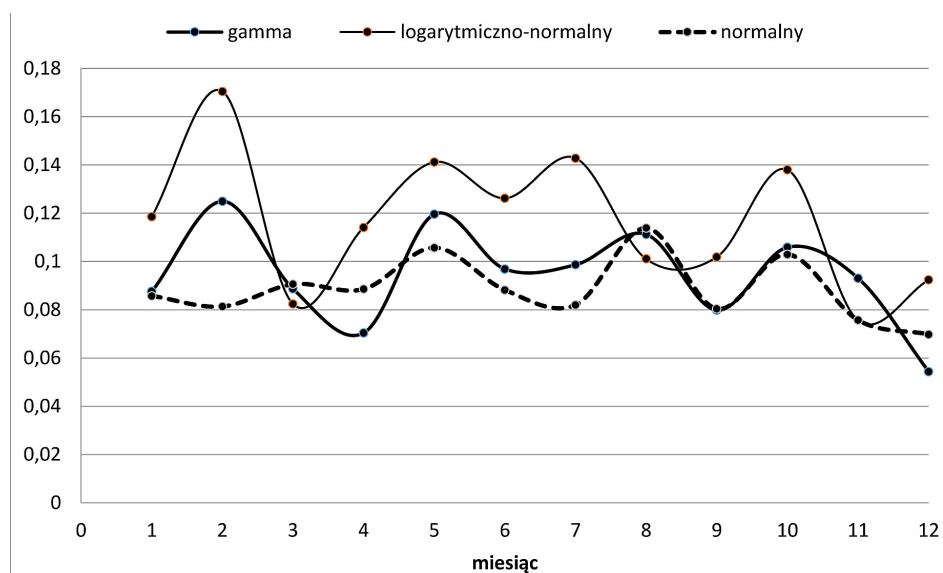
Miesiąc	Rozkład teoretyczny							p-wartość
	Logarytmiczno-normalny			Gamma			Transformacja	
	p-wartość	μ	σ	p-wartość	α	β		
Styczeń	0,246	3,23	0,57	0,66	3,65	7,97	$Y = \sqrt[3]{P}$	0,64
Luty	0,006	3,03	0,75	0,06	2,59	10,17	$Y = \sqrt[3]{P+10}$	0,6
Marzec	0,324	3,31	0,62	0,51	3,29	9,76	$Y = \sqrt[3]{P}$	0,92
Kwiecień	0,197	3,47	0,65	0,63	3,33	11,3	$Y = \sqrt[3]{P+10}$	0,94
Maj	0,236	3,84	0,72	0,06	2,67	21,2	$Y = \sqrt[3]{P+10}$	0,42
Czerwiec	0,322	4,17	0,45	0,51	5,4	13,17	$Y = \sqrt[3]{P+10}$	0,47
Lipiec	0,081	4,24	0,7	0,34	2,57	33,33	$Y = \sqrt[3]{P+10}$	0,81
Sierpień	0,051	4,09	0,67	0,04	2,85	25,46	$Y = \sqrt[3]{P}$	0,16
Wrzesień	0,66	3,56	0,71	0,73	2,53	17,14	$Y = \sqrt[3]{P+10}$	0,55
Październik	0,004	3,43	0,79	0,06	2,13	18,73	$Y = \sqrt[3]{P}$	0,76
Listopad	0,599	3,67	0,4	0,52	6,57	6,49	$\ln P$	0,86
Grudzień	0,433	3,38	0,66	0,51	2,94	11,89	$Y = \sqrt[3]{P+10}$	0,96

W celu porównania wyników dopasowania poszczególnych rozkładów do danych empirycznych, wyznaczone wartości statystyki D_n przedstawiono w tabeli 2. Powyższe wyniki zaprezentowano w formie graficznej (rys. 2).

Z rysunku 1 wynika, że wartości statystyki D_n , charakteryzujące jakość dopasowania rozkładu teoretycznego do rozkładu empirycznego, dla rozkładów normalnego i gamma oscylują wokół siebie. W przypadku kilku miesięcy (styczeń, marzec, kwiecień, sierpień, wrzesień, październik i grudzień) wartości statystyki D_n dla rozkładu normalnego są większe lub bardzo zbliżone do wartości tej statystyki dla rozkładu gamma, natomiast dla innych miesięcy (luty, maj, czerwiec, lipiec, listopad) są mniejsze. Wartości statystyki D_n natomiast dla rozkładu logarytmiczno-normalnego wyraźnie majoryzują wartości tej statystyki dla rozkładów gamma i normalnego.

Tabela 2. Wartości statystyki D_n wyznaczone dla kolejnych miesięcy dla rozkładów gamma, logarytmiczno-normalnego i normalnego**Table 2.** Value of D_n statistics calculated for subsequent months for gamma, lognormal and normal distributions

Miesiąc	Wartości statystyki D_n		
	Rozkład gamma	Rozkład logarytmiczno-normalny	Rozkład normalny
Styczeń	0,08756	0,11854	0,08571
Luty	0,12493	0,17045	0,08131
Marzec	0,08863	0,08244	0,09055
Kwiecień	0,0704	0,11412	0,08853
Maj	0,11959	0,14117	0,10569
Czerwiec	0,09687	0,12619	0,08814
Lipiec	0,09864	0,14281	0,08197
Sierpień	0,11125	0,10108	0,11383
Wrzesień	0,07983	0,10183	0,08036
Październik	0,10589	0,13799	0,10292
Listopad	0,09306	0,07569	0,07569
Grudzień	0,05437	0,09236	0,06978

**Rys. 2.** Wartości statystyki D_n wyznaczone dla kolejnych miesięcy dla rozkładów gamma, logarytmiczno-normalnego i normalnego**Fig. 2.** Value of D_n statistics calculated for subsequent months for gamma, lognormal and normal distributions

Test Shapiro-Wilka pozwala przyjąć założenie o normalności rozkładu statystyki D_n dla rozkładów gamma, logarytmiczno-normalnego i normalnego (p-wartości wynoszą odpowiednio 0,99, 0,94 i 0,52). W celu porównania istotności różnic pomiędzy wartościami średnimi dla statystyki D_n w przypadku poszczególnych rozkładów zastosowano test wartości średnich dla prób zależnych. Z porównania istotności różnic średnich wartości statystyki D_n wynika, że rozkłady gamma i normalny nie różnią się istotnie (p-wartość statystyki testowej równa 0,24), natomiast rozkłady gamma i logarytmiczno-normalny oraz rozkłady normalny i logarytmiczno-normalny różnią się istotnie (p-wartości statystyki testowej dla obu testów wynoszą odpowiednio 0,000350 i 0,000544, co wskazuje na mocne odrzucenie hipotezy o równości średnich).

WNIOSKI

Ocena jakości dopasowania rozkładów teoretycznych do rozkładu empirycznego otrzymanego na podstawie miesięcznych sum opadów w wieloleciu 1964–2009 ze stacji Wrocław-Swojec, oparta na statystyce Kołmogorowa-Smirnowa, wykazała, że:

- dla rozkładów gamma i normalnego różnice wartości średnich statystyk D_n są nieistotne, zatem oba rozkłady podobnie przybliżają rozkład empiryczny.
- dla rozkładów gamma i logarytmiczno-normalnego oraz normalnego i logarytmiczno-normalnego wartości średnie statystyk D_n istotnie się różnią.
- najlepsze dopasowanie otrzymuje się stosując przybliżenie rozkładu empirycznego rozkładem gamma lub normalnym z zastosowaniem odpowiedniej transformacji danych.
- najłabsze dopasowanie do rozkładu empirycznego daje zastosowanie rozkładu logarytmiczno-normalnego

LITERATURA

1. Abramowitz M., Stegun I.A., 1965. Handbook of Mathematical Formulas, Graphs and Mathematical Tables. Dover Publications: New York, s. 1250.
2. Angelidis P., Maris F., Kotsovinos N., Hrisanthou V., 2012. Computation of drought index SPI with alternative distribution function. *Water Resource Manage* 26, 2453–2473, DOI 10.1007/s11269-012-0026-0.
3. Edwards D.C., McKee T.B., 1997, Characteristics of 20th century drought in the United States at multiple scales. *Climatology Report 97-2*, Department of Atmospheric Science, Colorado State University, Fort Collins.
4. Gąsiorek E., Grządziel M., Musiał E., Rojek M., 2012, Porównanie wskaźnika standaryzowanego opadu (SPI) wyznaczonego za pomocą rozkładu gamma i rozkładu normalnego dla miesięcznych sum opadów. *Infrastruktura I Ekologia Terenów Wiejskich PAN 3/III 2012*, 197–209.

5. Kaczmarek Z., 1970. Metody statystyczne w hydrologii i meteorologii. Warszawa, WKiŁ s. 312.
6. Krishnamoorthy K., Mathew T., Mukherjee S., 2008. Normal-based methods for a Gamma Distribution: prediction and tolerance intervals and stress-strength reliability. *Technometrics*, February 2008, vol. 50, no. 1, 69–78.
7. Kryszicki W., Bartos J., Dyczka W., Królikowska K., Wasilewski M., 1997. Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach. Wydawnictwo Naukowe PWN, Warszawa.
8. Łabędzki L., 2006a. Susze rolnicze. Zarys problematyki oraz metody monitorowania i klasyfikacji. *Woda Środowisko Obszary Wiejskie, Rozprawy naukowe i monografie*, Nr 17, ss. 107.
9. Łabędzki L., 2006b. O dwóch metodach wyznaczania wskaźnika suszy meteorologicznej SPI. W: *Mater. XXXVI Seminarium Zastosowań Matematyki*, 41–47.
10. McKee T.B., Doesken N.J., Kleist J., 1993. The relationship of drought frequency and duration to time scales. *Proc. 8th Conf, Applied Climatology*, 17-22 January 1993, Anaheim, California, 179–184.
11. McKee T.B., Doesken N.J., Kleist J., 1995. Drought monitoring with multiple time scales. *Preprints 9th Conf. Applied Climatology*, 15-20 January 1995, Dallas, Texas, 233–236.
12. Thom H.C.S., 1958. A note on the gamma distribution. *Monthly Weather Review* 86, 117–122.
13. Wibig J., 2012. Warunki wilgotnościowe w Polsce w świetle wskaźnika standaryzowanego klimatycznego bilansu wodnego. *Woda-Środowisko-Obszary Wiejskie*, t. 12 z. 2(38), 329–340.