

WPŁYW SUMOWANIA OPADÓW ATMOSFERYCZNYCH NA POSTAĆ DYSTRYBUANTY

Tadeusz Górski

Zakład Agrometeorologii i Zastosowań Informatyki, IUNG
ul. Czartoryskich 8, 24-100 Puławy
e-mail: tgorski@iung.pulawy.pl

Streszczenie. Rozkład statystyczny sum opadów można aproksymować rozkładem gamma. Najprostszym sposobem określenia parametrów empirycznych daje metoda momentów, w której są one obliczane na podstawie wartości średniej i wariancji. W miarę agregowania opadów w czasie, rozkład sum zmierza do rozkładu normalnego, czemu odpowiadają zmiany obu parametrów. Parametry te można ocenić określając na drodze empirycznej zmiany wariancji sum opadów według wzoru:

$$\ln \sigma_s = a + b \ln (n - 0,2),$$

w którym σ_s oznacza odchylenie standardowe sumy zagregowanej, n – liczbę dni w okresie, a i b – parametry empiryczne. Znając dwie wartości wariancji (np. sum dobowych i miesięcznych), można teraz znaleźć wariancję w dowolnym okresie, a co za tym idzie, także parametry rozkładu. Oceny dystrybuanty w granicach prawdopodobieństwa 1-99% uzyskane na tej drodze można uważać za wystarczająco dokładne w zastosowaniach.

Słowa kluczowe: opady atmosferyczne, rozkład gamma, skalowanie wariancji

WSTĘP

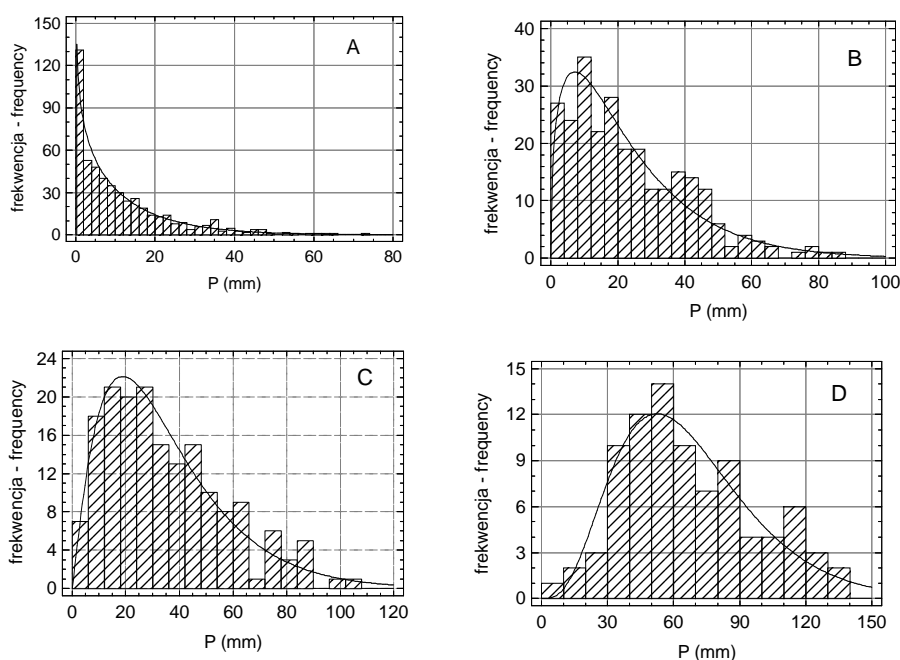
Rozkłady statystyczne sum opadów mogą być przybliżane różnymi rozkładami teoretycznymi, jednakże powszechnie (Kaczmarek 1970, Pruchnicki 1987) stosowany jest rozkład gamma. Parametry rozkładu bywają określane różnymi metodami; najczęściej używana jest metoda momentów, a to ze względu na prostotę i brak ograniczeń w zastosowaniach (Górski i Górski, 2002, 2004). Aby wyznaczyć parametry rozkładu gamma, potrzebna jest znajomość wartości średniej oraz wariancji.

Dysponując seriami notowań opadów w ujęciu miesięcznym, stosunkowo łatwo można podać średnie sumy opadów dla każdego dnia i dowolnie zagregowanego okresu, stosując analizę harmoniczną opisującą cykl klimatyczny (Górski i Górski 2004). Duże trudności pojawiają się natomiast przy określaniu wariancji.

Zasadniczym celem opracowania jest właśnie zaproponowanie metody określania wariacji sum opadów w dowolnym okresie na podstawie znanych wartości dobowych i miesięcznych.

ROZKŁAD STATYSTYCZNY OPADÓW JAKO ROZKŁAD GAMMA

Prawostronnie asymetryczny rozkład gamma (rozkład Pearsona III typu z ograniczeniem zerowym) szczególnie dobrze nadaje się do aproksymacji sum opadów i znajduje rozliczne zastosowania, także poza hydrologią i klimatologią (np. w ubezpieczeniach – Martin i in. 2001). Asymetryczność rozkładów maleje w miarę sumowania opadów w dłuższym okresie, kiedy – w myśl tzw. centralnego twierdzenia granicznego statystyki matematycznej (Pawłowski 1965) – rozkład dąży do normalnego (gaussowskiego). I tak, suma opadów rocznych w warunkach Polski może być równie dobrze opisywana rozkładem gamma, jak i normalnym, ale sumy miesięczne są jeszcze wyraźnie asymetryczne. Zmiany funkcji gęstości ilustruje rysunek 1, przedstawiający histogramy opadów czerwcowych w ponad 80 letnim ciągu puławskim.



Rys. 1. Histogramy empiryczne i funkcje gęstości opadów (P) czerwca w Puławach (1918-2005). A, B, C, D – kolejno okresy 5, 10, 15, 30 dniowe

Fig. 1. Empirical histograms and density functions of June precipitation (P) in Puławy (1918-2005). A, B, C, D – periods of 5, 10, 15, 30 days.

Przy krótkim okresie sumowania funkcja gęstości jest w całym przedziale malejąca; dopiero przy okresie ok. 10 dniowym modalna odsuwa się od zera, co odpowiada przejściu parametru kształtu (α) przez wartość „1”.

Parametry α (kształtu) i β (skali) obliczane są w funkcji wartości średniej m i wariancji σ^2 :

$$\alpha = m^2 / \sigma^2 \quad \beta = m / \sigma^2$$

Mając wartości obydwu parametrów rozkładu gamma, można przystąpić do określenia dystrybuanty, korzystając z tablic niepełnej funkcji gamma (Kaczmarek 1970), lub algorytmów komputerowych. Odpowiedni algorytm skonstruowano na podstawie tablic Słuckiego (1950) używając nieliniowych równań, w których wartość prawdopodobieństwa z tablic była funkcją parametrów rozkładu. Błędy tak uzyskanej dystrybuanty (w stosunku do tablic) są znikome i nie mają znaczenia dla interpretacji wyników.

SUMOWANIE WARIANCJI

Wariancja opadów podlega oczywiście ogólnym zasadom sumowania wariancji, przy czym jednak sprawę komplikuje skorelowanie opadów w kolejnych dniach. W przypadku występowania korelacji między zmiennymi wariancję sumy łącznej oblicza się następująco (Pawłowski 1965):

$$\sigma_s^2 = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2 + 2 \sigma_1 \sigma_2 r_{1,2} + 2 \sigma_1 \sigma_3 r_{1,3} + \dots + 2 \sigma_{n-1} \sigma_n r_{n-1,n} \quad [1]$$

gdzie σ_s^2 – wariancja sumy agregowanej w ciągu n dni,

$\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$ – wariancje sum dobowych,

$r_{1,2}, r_{1,3}, \dots, r_{n-1,n}$ – współczynniki korelacji sum dobowych.

Jeśli założymy, że wariancje sum dobowych w pewnym (niezbyt długim) okresie są w przybliżeniu jednakowe:

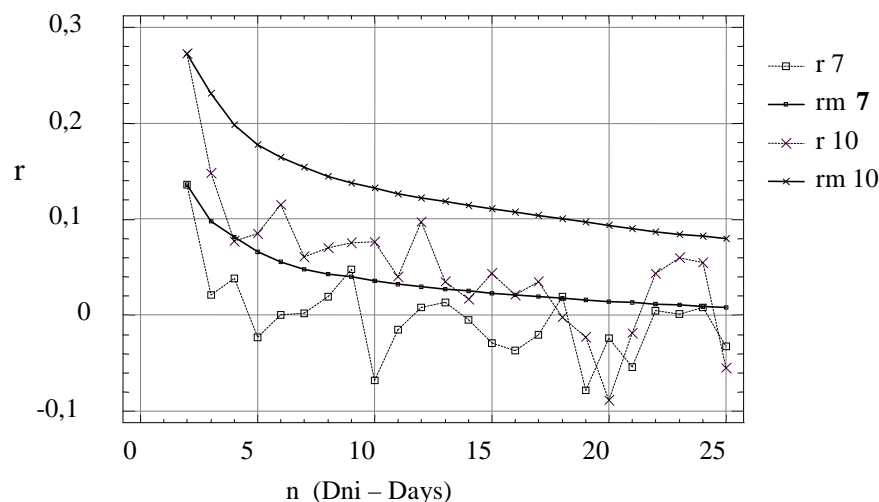
$$\sigma_1^2 \approx \sigma_2^2 \approx \dots \approx \sigma_n^2 \approx \sigma^2$$

to można wzór [1] zapisać w postaci:

$$\sigma_s^2 = n \sigma^2 + n(n-1) \sigma^2 r_m = \sigma^2 [n + (n^2 - n) r_m] \quad [2]$$

gdzie r_m – średni w zbiorze współczynnik korelacji.

Znając wariancję sum dobowych oraz ich autokorelację można teraz określać wariancję sum w dowolnym okresie o długości n dni. W praktyce jednak jest to zadanie dość trudne, ze względu na labilność ocen autokorelacji. Jak to przedstawia rysunek 2, nawet w ponad 80 letnich materiałach puławskich funkcje autokorelacyjne wykazują duże nieregularności i znaczne zróżnicowanie sezonowe, co przenosi się oczywiście także na dokładność ocen średnich współczynników autokorelacji.



Rys. 2. Autokorelacja (r) dobowych sum opadów w Puławach (1919-2005). r_7 – współczynnik korelacji sum odległych o n dni lipca; r_{10} – października; r_m – średni współczynnik korelacji w zbiorze wszystkich n dni.

Fig. 2. Autocorrelation (r) of daily precipitation sums in Puławy (1919-2005). r_7 – correlation coefficient of sums distant n days in July; r_{10} – in October; r_m – mean correlation coefficient in n day sample

Po wielu próbach na ponad 80 letnich materiałach puławskich i 25 letnich danych z dziesięciu innych stacji meteorologicznych (Opady atmosferyczne 1954-1981), okazało się, że wykorzystywanie analitycznego wzoru [2] nie jest celowe, wobec małej efektywności estymatorów r_m przy dużym ich zróżnicowaniu sezonowym i przestrzennym. Zdecydowano się więc na podejście empiryczne, pomijające sprawę autokorelacji.

Logarytmując [2] otrzymuje się:

$$\ln \bar{\sigma}_s = \ln \bar{\sigma} + 0,5 \ln [n + (n^2 - n) r_m]$$

ponieważ wyraz w nawiasach jest w sposób oczywisty silnie skorelowany z n , można spodziewać się, że zmiany odchylenia standardowego sum opadów w miarę ich agregowania w czasie dobrze aproksymuje równanie liniowe:

$$\ln \sigma_s = \ln \sigma + b \ln n \quad [3]$$

gdzie b – parametr empiryczny.

Jak się okazuje, rzeczywiście w całej Polsce równanie to stosunkowo dokładnie opisuje rzeczywiste stosunki, przynajmniej w zakresie sumowania od doby do miesiąca (rys. 3). Przy bliższej analizie danych z długich serii pomiarowych, za-

uważa się jednak pewną nieliniowość związku [3], dającą mały systematyczny błąd, który można wyeliminować stosując nieco zmienioną postać równania:

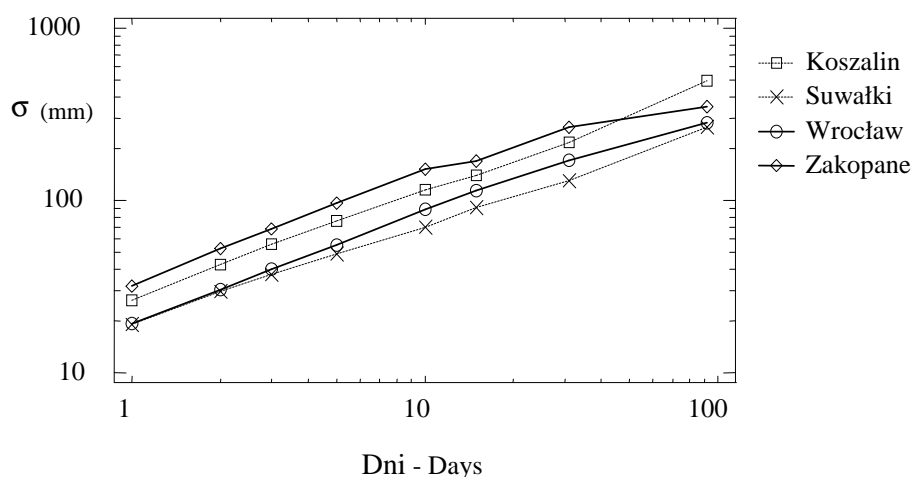
$$\ln \sigma_s = a + b \ln (n - k) \quad [4]$$

gdzie a i b parametry empiryczne w równaniu regresji, określone po uprzednim wyznaczeniu trzeciego parametru k .

Parametr k ustalono na podstawie długich serii, szukając największego współczynnika korelacji dla związku odchylenia standardowego z liczbą dni. Wykorzystywano sumy 1,2,3,5,10,15 i 30 dobowe we wszystkich miesiącach. Najściślejsze związki otrzymano dla $k = 0,2$ (tab. 1). Ten dodatkowy parametr prawdopodobnie rekompensuje częściowo wady założenia o równości wariancji sum dobowych w zbiorze.

Ostatecznie więc równanie [4] przybiera postać:

$$\ln \sigma_s = a + b \ln (n - 0,2) \quad [5]$$



Rys. 3. Odchylenie standardowe sum opadów (σ) stycznia (1956-1980), kolejno w okresach 1, 2, 3, 5, 10, 15, 31, 92 dni

Fig. 3. Standard deviations of precipitation sums (σ) in January (1956-1980), from left: in 1, 2, 3, 5, 10, 15, 31, 92 day periods

Charakter związku [5] podlega cyklowi rocznemu, który może być opisany analizą fourierowską (Górski i Górka 2004). W celach praktycznych można w obrębie miesiąca stosować stałe wartości parametrów.

Tabela 1. Współczynniki korelacji w równaniu regresji [4] przy różnych wartościach k w miesiącach. Puławy 1919-2005

Table 1. Correlation coefficients from regression equation [4] with different k values in 12 months. Puławy 1919-2005

Miesiąc – Month	$k = 0$	$k = -0,2$	$k = -0,4$
Styczeń – January	0,9991	0,9998	0,9996
Luty – February	0,9997	0,9996	0,9984
Marzec – March	0,9989	0,9994	0,9977
Kwiecień April	0,9941	0,9961	0,9974
Maj – May	0,9985	0,9989	0,9982
Czerwiec – June	0,9975	0,9989	0,9995
Lipiec – July	0,9991	0,9995	0,9989
Sierpień – August	0,9993	0,9998	0,9993
Wrzesień – September	0,9991	0,9997	0,9994
Październik – October	0,9998	0,9992	0,9973
Listopad – November	0,9991	0,9998	0,9995
Grudzień – December	0,9996	0,9994	0,9980
Średnia – Mean	0,9987	0,9992	0,9986

Jeśli znane jest odchylenie standardowe sum dobowych $\bar{\sigma}$ i sum miesięcznych $\bar{\sigma}_m$ to odchylenie sum $\bar{\sigma}_s$ w dowolnym okresie miesiąca można obliczać przy pomocy następujących wzorów, wynikających z wzoru [5] po prostych przekształceniach:

$$\ln \bar{\sigma}_s = \ln \bar{\sigma}_m - 0,938 (\ln \bar{\sigma}_m - \ln \bar{\sigma}) + 0,276 (\ln \bar{\sigma}_m - \ln \bar{\sigma}) \ln(n-0,2)$$

w miesiącu 30 dniowym, lub

$$\ln \bar{\sigma}_s = \ln \bar{\sigma}_m - 0,939 (\ln \bar{\sigma}_m - \ln \bar{\sigma}) + 0,274 (\ln \bar{\sigma}_m - \ln \bar{\sigma}) \ln(n-0,2)$$

w miesiącu 31 dniowym

DOKŁADNOŚĆ UZYSKIWANYCH WYNIKÓW

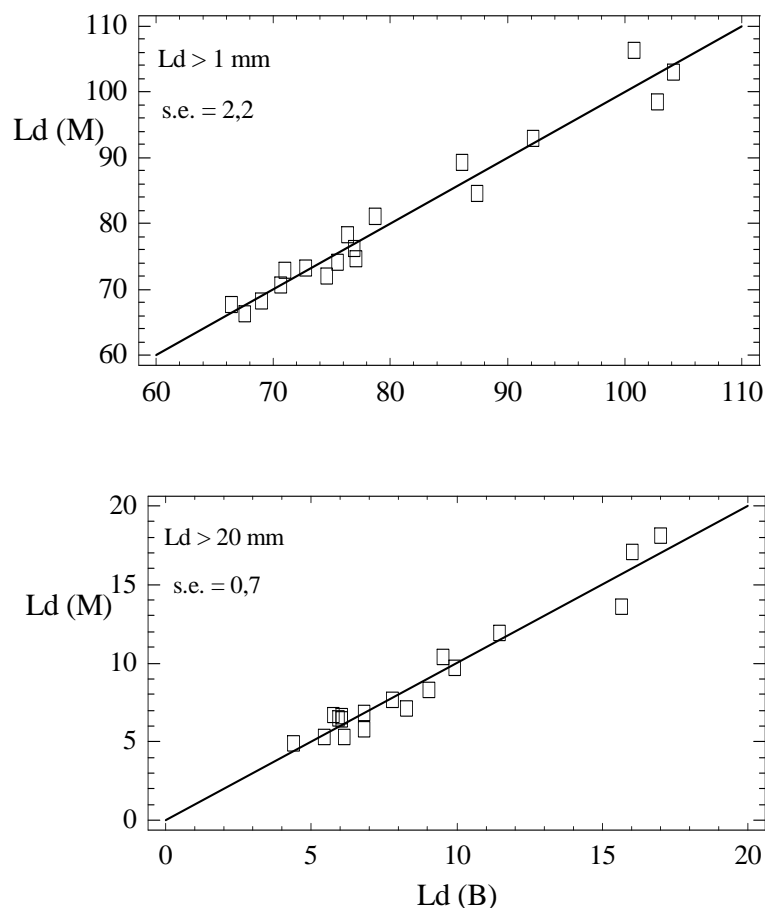
Błędy odchyłeń standardowych $\bar{\sigma}_s$ wyliczonych z równania [5] są nieznaczne, nawet w regionach klimatycznie skrajnych, przynajmniej w granicach miesiąca. Ekstrapolacja do okresu 3 miesięcy (po jednym miesiącu przed i po miesiącu rozpatrywanym), przynosi już błędy, których zaakceptować nie można (tab. 2).

Tabela 2. Odchylenie standardowe sum opadów (mm) rzeczywiste (a) i obliczone (m) wg równania [5]. Dane w wierszu „92” dni pochodzą z ekstrapolacji. Puławy w latach 1919-2003, pozostałe stacje w latach 1956-1980

Table 2. Standard deviations of precipitation sums (mm), actual (a) and modelled (m) after equation [5]. Data in “92” days row were extrapolated. Puławy 1919-2003, other localities 1956-1980

Styczeń January		Suwałki		Koszalin		Wrocław		Zakopane		Puławy	
Dni Days	a	m	a	m	a	m	a	m	a	m	
1	19	19	26	27	19	19	32	33	20	20	
2	30	29	42	43	30	31	52	53	31	32	
3	37	37	56	55	40	40	69	68	41	40	
5	49	49	76	75	55	56	97	93	55	54	
10	70	71	116	113	88	87	151	141	82	80	
15	90	89	139	143	114	111	170	179	99	100	
31	130	131	217	218	171	174	265	273	148	149	
92	267	232	493	410	282	338	350	515	292	272	

Lipiec July		Suwałki		Koszalin		Wrocław		Zakopane		Puławy	
Dni Days	a	m	a	m	a	m	a	m	a	m	
1	61	59	66	61	74	76	121	130	62	62	
2	95	93	92	96	114	115	198	196	93	93	
3	114	118	118	122	145	143	247	245	116	116	
5	156	159	154	164	195	188	341	322	151	153	
10	230	235	243	243	277	269	487	463	230	221	
15	281	295	323	304	327	330	599	571	267	272	
31	475	442	455	455	461	477	746	829	394	396	
92	796	805	652	828	695	826	1222	1445	595	690	



Rys. 4. Porównanie liczby dni w okresie kwiecień – październik z opadem przekraczającym sumę 1 i 20 mm, obliczonych z modelu – Ld (M), z liczbami w pracy Olechnowicz-Bobrowskiej (1988) – Ld (B). s.e. – błąd standardowy w dniach

Fig. 4. Number of days in April – October period with precipitation sums exceeding 1 and 20 mm, modelled – Ld (M) and after Olechnowicz-Bobrowska (1988) – Ld (B). s. e. – standard error (days)

Liczne próby porównywania dystrybuanty określonej przy użyciu formuły [5] z kumulowanymi histogramami empirycznymi wykazały, że różnice nie są większe, niż w przypadku różnic między kolejnymi 50 letnimi seriami pomiarów opadów, które przecież uważać można za próby z rozkładu stacjonarnego (Kowalczak i Szymański 2000, Zawora i Ziernicka 2003). Tylko na zewnątrz przedziału prawdopodobieństwa 1-99% różnice mogą być znaczne. W serii nawet 100 letniej

jeden ekstremalny przypadek może całkowicie zmienić interpretacje. Wydaje się, że rację ma Bardsley (1994) kiedy podważa zasadność określania skrajnych kwantyli w zagadnieniach hydrologicznych.

Dystrybuantę wykorzystać można także do wyznaczenia liczby dni z opadem przekraczającym założoną sumę. Rysunek 4 przedstawia porównanie liczby dni z opadem ponad 1 mm i ponad 20 mm określonych przez Olechnowicz-Bobrowską (1988) w Karpatach, z liczbami uzyskanymi na podstawie przedstawianego modelu (w miejscowościach o dostatecznie długich seriach pomiarów).

Jeśli dane o wariancji opadów dobowych są trudno dostępne, można wykorzystywać pewne związki, jakie istnieją między charakterystykami miesięcznymi i dobowymi wariancji a współrzędnymi geograficznymi (Górski i Górską 2004).

PIŚMIENNICTWO

- Bardsley W.E., 1994. Against objective analysis of hydrological extremes. *J. Hydrology*, 162, 429-431.
- Górski T., Górską K., 2002. O rozkładzie statystycznym sum opadów atmosferycznych. *Mat. 32 Seminarium Zastosowań Matematyki. AR Wrocław-Kobyła Góra*, 39-43.
- Górski T., Górską K., 2004. Model sum opadów atmosferycznych w Polsce. *Mat. 34 Seminarium Zastosowań Matematyki. AR Wrocław— Kobyła Góra*, 65-75.
- Kaczmarek Z., 1970. *Metody statystyczne w hydrologii i meteorologii*. WKiŁ.
- Kowalczak P., Szymański K., 2000. Remarks on the stepping in the Wielkopolska region. In: *Int.Conf. Present and future requirements for agrometeorological information. Poznań-Sielinko. IMGW-IMS*, 43-47.
- Martin S.W., Barnett B.J., Coble K.H., 2001. Developing and pricing precipitation insurance. *J.Agric. Resour. Econ.*, 26, 261-274.
- Olechnowicz-Bobrowska B., 1988. Liczba dni z opadem w okresie wegetacyjnym w polskich Karpatach Zachodnich. *Zesz. Nauk. Akademii Rolniczej w Krakowie*, 225, Mel. 13, 85-102.
- Opady atmosferyczne. *Roczniki 1954-1981. PIHM i IMGW*, Warszawa.
- Pawłowski Z., 1965. *Wstęp do statystyki matematycznej*. PWN, Warszawa.
- Pruchnicki J., 1987. *Metody opracowań klimatologicznych*. PWN, Warszawa.
- Sluckij E.E., 1950. *Tablice dlja vyčislenija niepolnoj Γ -funkcii i funkcii wierojatnosti χ^2* . Izd. Akademii Nauk SSSR, 1950.
- Zawora T., Ziernicka A., 2003. Precipitation variability in time in Poland in the light of multi-annual mean values (1891-2000). *Acta Univ. Wr., St. Geogr.* 75, 123-128.

IMPACT OF THE AGGREGATION OF PRECIPITATION SUMS ON THE DISTRIBUTION FUNCTION

Tadeusz Górski

Department of Agrometeorology and Applied Informatics,
Institute of Soil Science and Plant Cultivation
ul. Czarzoryskich 8, 24-100 Puławy
e-mail: tgorski@iung.pulawy.pl

Abstract. Statistical distribution of atmospheric precipitation sums may be approximated with gamma function. An easy way of determining the empirical parameters is the method of moments, where they are calculated on the basis of mean value and variance. When precipitation sums are aggregated in longer periods, the parameters change and the distribution function approaches a normal distribution. The changes of variance may be determined after the formula:

$$\ln \bar{\sigma}_s = a + b \ln (n - 0.2),$$

where $\bar{\sigma}_s$ – standard deviation of aggregated sums, n – length of period (days) a and b – empirical parameters. Knowing two values of variance (e.g. daily and monthly), the variance and both gamma parameters may be evaluated in any chosen period. Values of probability in the limits 1-99% determined this way may be assessed as sufficiently precise for practical purposes.

Key words: atmospheric precipitation, gamma distribution, variance scaling