

Tomasz PYTLOWANY¹
Szczepan WOLIŃSKI²

WYBÓR ROZKŁADU ODDZIAŁYWAŃ KLIMATYCZNYCH Z WYKORZYSTANIEM METODY BAYES'A

Wybór typu zmiennej losowej na podstawie wyników badań polega najczęściej na aproksymacji parametrów przyjętego rozkładu. Jednak dostępne wyniki badań lub obserwacji często nie pozwalają na dostatecznie precyzyjne dopasowanie jednego z powszechnie stosowanych typów rozkładu zmiennej losowej. W artykule przedstawiono metodę opartą na wnioskowaniu Bayesa, która umożliwi oszacowanie miary dopasowania i optymalnego wyboru jednego z typowych rozkładów zmiennej losowej oraz wyznaczenie dystrybuanty liniowej kombinacji testowanych rozkładów. Zastosowanie przedstawionej metody zilustrowano na przykładzie oceny i wyboru dystrybuanty rozkładu reprezentującego obciążenie śniegiem gruntu oraz wyznaczenia kombinacji testowanych rozkładów w celu określenia wartości charakterystycznej obciążenia śniegiem dla przyjętego okresu powrotu.

Słowa kluczowe: oddziaływania klimatyczne, obciążenie śniegiem, metoda Bayesa

1. Wstęp

Analizując prognozy oddziaływań klimatycznych oparte na modelach statystycznych, w których wartości oddziaływań są traktowane jako realizacje zmiennej losowej, obserwuje się często znaczne rozbieżności z wynikami pomiarów. Ponadto w zależności od przyjętego typu rozkładu zmiennej losowej uzyskuje się bardzo zróżnicowane wyniki prognoz dla dłuższych okresów powrotu szacowanych oddziaływań. Związane jest to ze stosunkiem wymiaru przestrzeni obserwacji do wymiaru przestrzeni parametrów [1]. W przypadku oddziaływań klimatycznych do dyspozycji są najczęściej obserwacje z kilkunastu lat, a parametry statystyczne modeli standardowych używanych do analiz są

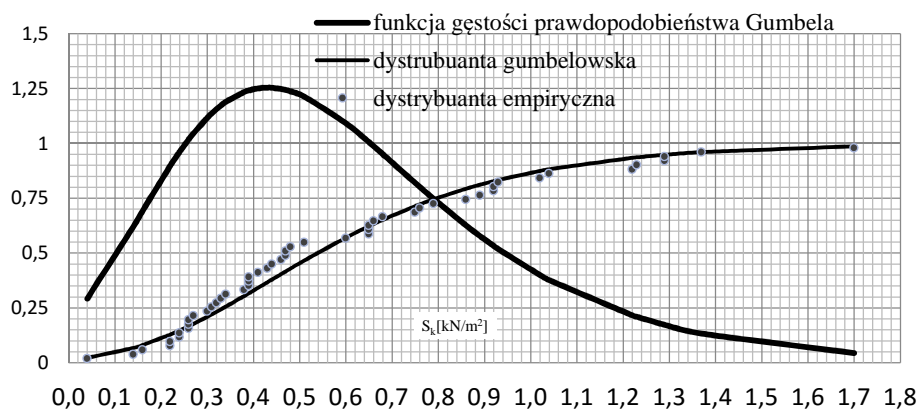
¹ Autor do korespondencji/corresponding author: Tomasz Pytlowany, Instytut Politechniczny, PWSZ im. St. Pigoń w Krośnie, 38-400 Krosno, ul Rynek 1, 134375559, tompyt@pwsz.krosno.pl

² Szczepan Woliński, Wydział Budownictwa Inżynierii Środowiska i Architektury, Politechnika Rzeszowska im. I. Łukasiewicza, 35-084 Rzeszów, ul. Poznańska 2, 178542974, szwolkkb@prz.edu.pl

ściśle związane z długością analizowanego szeregu wyników obserwacji oraz liczbą opóźnień w wymiarze funkcji autokorelacji. Zagadnienie prognozowania sprowadza się w takich sytuacjach do wnioskowania statystycznego opartego na aproksymacji asymptotycznej. Kluczowe znaczenie ma wówczas wnioskowanie w zakresie niepewności co do prawidłowego przebiegu ścieżek prognozy wykonanej z wykorzystaniem różnych modeli (rozkładów prawdopodobieństwa zmiennej losowej). Z punktu widzenia wyboru właściwego modelu statystycznego ważnym zagadnieniem jest problem wyboru metody estymacji parametrów modelu. Najczęściej stosowane są metody największej wiarygodności i metoda momentów, a sporadycznie inne metody. Natomiast powszechnie przyjmowane są standardowe modele w postaci zmiennych losowych: Gumbela, logarytmiczno-normalnych, normalnych, niekiedy gamma, Rayleigha, wykładniczych i innych [1,2]. W artykule skupiono uwagę na dwóch zagadnieniach związanych z problematyką prognozowania oddziaływań klimatycznych. Pierwsze dotyczy uzasadnienia wyboru modeli rozważanych oddziaływań w postaci zmiennej losowej o różnym rozkładzie prawdopodobieństwa i estymacji jej parametrów z wykorzystaniem metody Bayesa [3,4]. Zakładając, że dostępna jest wstępna informacja na temat wartości estymowanych parametrów i znane są dodatkowe wyniki pomiarów, obliczono na podstawie twierdzenia Bayesa ich wartości *a posteriori* i wagi traktowane jako miara dopasowania rozważanych rozkładów zmiennej losowej. Drugie zagadnienie wiąże się z zastosowaniem autorskiej procedury wyznaczenia modelu będącego kompilacją modeli standardowych z wykorzystaniem ich wag obliczonych metodą Bayesa. Rozważania zilustrowano przykładem zastosowania przyjętej metody i procedury obliczeń do wyboru modelu obciążenia śniegiem gruntu na podstawie danych z wybranej stacji meteorologicznej na terenie Polski.

2. Zastosowanie wnioskowania statystycznego do szacowania parametrów standardowych modeli oddziaływań

Rozkład empiryczny w przypadku danych klimatycznych powstaje ze zbioru maksymalnych wartości rocznych, sezonowych, dziennych itp.. W przypadku obciążenia śniegiem gruntu maksima roczne są uszeregowane w ciągu rosnącym. Do aproksymacji stosowane są się różne typy rozkładów teoretycznych. Są to zazwyczaj rozkłady wartości ekstremalnych: Gumbela (typ I), Frecheta i Weibulla lub rozkład logarytmiczno-normalny. Najczęściej stosowanym rozkładem do szacowania oddziaływań klimatycznych jest rozkład Gumbela [2,5]. Na rys. 1 oraz przedstawiono wykres funkcji gęstości prawdopodobieństwa i dystrybuanty rozkładu Gumbela o parametrach oszacowanych metodą największej wiarygodności oraz wartości dystrybuanty empirycznej dla danych ze stacji znajdującej się w pierwszej strefie obciążenia śniegiem (Olsztyn, 1950-2000, [4]).



Rys. 1. Metoda analityczna w szacowaniu obciążenia śniegiem gruntu. Opracowanie własne

Fig. 1 Analytical method for estimating snow loads on the ground. Individual elaboration

Szacowanie wartości oddziaływań o określonym prawdopodobieństwie przewyższenia jest z punktu widzenia zastosowań zagadnieniem najważniejszym i umożliwia wyznaczenie wartości maksymalnych obciążeń klimatycznych oraz opracowanie map ryzyka dla obszarów narażonych na zdarzenia katastrofalne spowodowane żywiołami powietrza i wody. W praktyce są one estymowane z wykorzystaniem maksymalnych wartości rocznych. Przy doborze odpowiedniego rozkładu szczególną uwagę należy zwrócić na estymację „ogona” rozkładu po stronie wartości większych od parametru pozycji rozkładu. To w nim zawarta jest informacja o oddziaływaniach mających długi okres powtarzalności. W przypadku szacowania oddziaływań za pomocą rozkładów: Gumbela (typ I), Frecheta i Weibulla należy zwrócić uwagę na parametr skali, który przyjmuje różne wartości dla wymienionych rozkładów ekstremalnych (EVT). W opracowaniach naukowych dotyczących metod estymacji maksymalnych wartości rocznych odnaleźć można kilka testów do weryfikacji hipotez dotyczących rozróżnienia pomiędzy przypadkami gdy parametr skali jest równy zero oraz gdy przyjmuje wartości powyżej bądź poniżej zera. Ma to znaczenie przy odpowiednim doborze rozkładów z rodziny EVT [9]. W niniejszym artykule skupiono uwagę na niepewności statystycznej wynikającej z niepełnej informacji związanej z wynikami obserwacji. Zgodnie z bayesowską koncepcją wnioskowania założono, że znana jest wstępna informacja (*a priori*) dotycząca wartości cech estymowanych parametrów rozkładów. Wiedza ta jest następnie modyfikowana na podstawie wyników obserwacji. Na podstawie twierdzenia Bayesa, z którego można określić rozkład parametrów *a posteriori* uwzględniający początkowe przekonania *a priori* w stosunku do posiadanych danych empirycznych [3].

3. Zastosowanie wnioskowania Bayesa do estymacji parametrów i wyboru modelu oddziaływań

Podstawowym założeniem charakteryzującym wnioskowanie Bayesa jest uznanie ocenianych parametrów modeli statystycznych za zmienne losowe o rozkładach prawdopodobieństwa ustalanych *a priori*, z wykorzystaniem niepewnej wiedzy eksperckiej, które są następnie modyfikowane na podstawie wyników badań. Wnioskowanie Bayesa umożliwia ponadto formalne porównanie ułatwiające wybór miarodajnego typu rozkładu, a także utworzenie modeli będących kombinacją różnych rozkładów tworzącą rozkład najlepiej dopasowany do wyników badań doświadczalnych lub obserwacji.

Twierdzenie Bayesa przystosowane do oceny rozkładu prawdopodobieństwa parametrów $\theta = (\theta_1, \dots, \theta_d)$ można zapisać w postaci [3,4]:

$$\pi(\theta|x) = \frac{l(x|\theta)\pi(\theta)}{\int_b l(x|\theta)\pi(\theta)d\theta} = \frac{l(x|\theta)\pi(\theta)}{\pi(x)} \quad (1)$$

gdzie: $l(x|\theta)$ - funkcja wiarygodności ze zbioru obserwacji $x=(x_1, \dots, x_n)$, gdy wektor parametrów $\theta=(\theta_1, \dots, \theta_d)$ jest podany, $\pi(\theta)$ - zbiór prawdopodobieństw *a priori* $\theta=(\theta_1, \dots, \theta_d)$ przed uzyskaniem obserwacji $x=(x_1, \dots, x_n)$, $\pi(\theta|x)$ -zbiór prawdopodobieństw *a posteriori* $\theta=(\theta_1, \dots, \theta_d)$ po analizie wyników obserwacji $x=(x_1, \dots, x_n)$, $\pi(x)$ - brzegowe gęstości obserwacji $x=(x_1, \dots, x_n)$.

Funkcja wiarygodności $l(x|\theta)$ reprezentuje nieodłączną niepewność związaną ze zmienną losową X dla danego wektora θ [5], podczas gdy zbiory $\pi(\theta)$ oraz $\pi(\theta|x)$ reprezentują statystyczną niepewność poszczególnych parametrów rozkładów prawdopodobieństw w wektorze θ [3, 4]. Jeśli dowolna zmienna losowa X o gęstości prawdopodobieństwa $l(x|\theta)$ jest zależna od wektora parametrów θ , to funkcję wiarygodności dla przyjętego rozkładu $l(x_1, \dots, x_n|\theta)$ niezależnych obserwacji $x=(x_1, \dots, x_n)$ można zapisać w następującej formie [3,4]:

$$l(x|\theta) = l(x_1, \dots, x_n|\theta) = \prod_{i=1}^n l(x_i|\theta) \quad (2)$$

Twierdzenie Bayesa w postaci (1) można wykorzystać do określenia wag, umożliwiających ocenę dopasowania przyjętych rozkładów prawdopodobieństwa do wyników obserwacji przy uwzględnieniu niepewności statystycznych związanych z jego wyborem. Oznaczając przez H hipotezę, że przyjęty *a priori* rozkład prawdopodobieństwa zmiennej losowej X i zbiór wyników obserwacji

θ są zgodne, brzegowy rozkład prawdopodobieństwa zmiennej $\pi(x|H)$ można aproksymować za pomocą wzoru [3]:

$$\log(\pi(x|H)) \approx \frac{d}{2} \log(2\pi) - \frac{d}{2} \log(n) + \log(l(x|\hat{\theta}, H)) \quad (3)$$

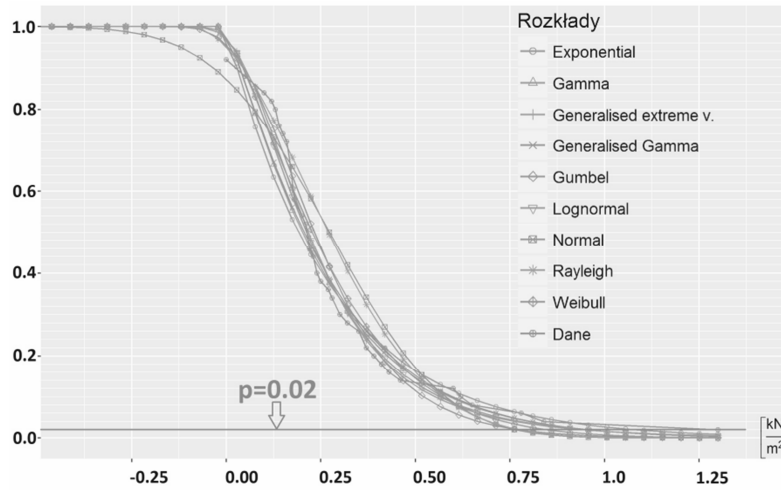
A wartości wag dla przyjętych rozkładów prawdopodobieństwa można obliczyć według formuły:

$$w[H|l(x|\theta)] = \left(\frac{n}{2\pi}\right)^{\frac{d}{2}} l(x|\hat{\theta}, H) \quad (4)$$

gdzie: n – liczba wyników obserwacji, d – liczba estymowanych parametrów, $\hat{\theta}$ – ocena estymowanego parametru.

4. Przykład

Przedstawiona w artykule metoda i procedura obliczeń mogą być zastosowane do oszacowania wagi, dopasowania przyjętego rozkładu prawdopodobieństwa do danych obserwowanych z uwzględnieniem niepewności statystycznych oraz do wyznaczenia dystrybuanty rozkładu będącego kombinacją przyjętych rozkładów standardowych uwzględnieniem niepewności modelowej. Na podstawie procedur przedstawionych w punkcie 3 niniejszego artykułu napisano program w języku programowania statystycznego R [5]. Program oblicza wagi, które określają stopień dopasowania rozważanego rozkładu do wyników obserwacji, z uwzględnieniem niepewności statystycznych i współrzędne rozkładu będącego kombinacją rozważanych rozkładów, z uwzględnieniem niepewności modelowej. Na rys.2 przedstawiono wykresy dystrybant 9 przyjętych rozkładów prawdopodobieństwa o parametrach oszacowanych metodą największej wiarygodności z wykorzystaniem środowiska do obliczeń statystycznych R i procedurę *optim* do minimalizacji funkcji wiarygodności [5], a także dystrybantę empiryczną obciążenia śniegiem gruntu dla wyników obserwacji ze stacji meteorologicznej zlokalizowanej w pierwszej strefie obciążenia śniegiem (Wrocław, 1950-2000 [4]).



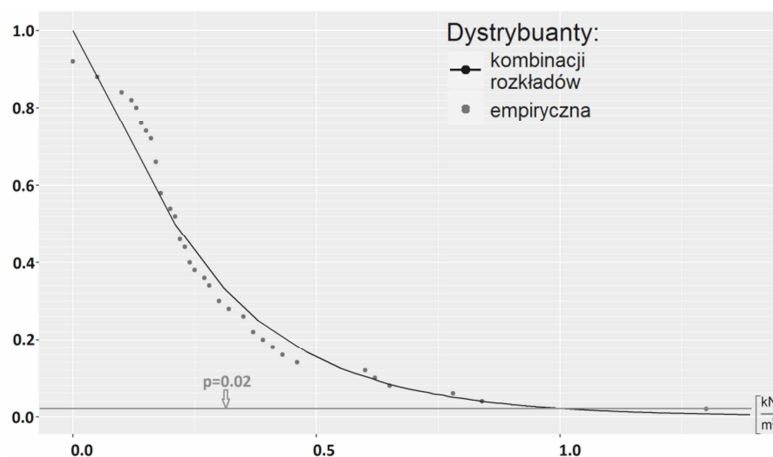
Rys.2. Dystrybuanty analizowanych rozkładów prawdopodobieństwa i dystrybuanta empiryczna

Fig. 2. Cumulative distribution of the analyzed probability distributions and empirical cumulative distribution

Wykres dystrybuanty *a posteriori* uzyskanej jako kombinacja dziewięciu przyjętych *a priori* rozkładów standardowych zmiennych losowych przedstawiono na tle dystrybuanty empirycznej na rys.3. Obliczenia wykonano przy założeniu, że wagi rozkładów *a priori* są jednakowe i równe $1/9$, a wagi rozkładów *a posteriori* można oszacować dla $j = 9$, według wzoru (4), który po przekształceniu ma następującą postać:

$$w[H_j | l(x|\theta)] = \frac{\frac{1}{9} \left(\frac{n}{2\pi}\right)^{d_j/2} l_j(y)}{\sum_{j=1}^9 \frac{1}{9} \left(\frac{n}{2\pi}\right)^{d_j/2} l_j(y)} = \frac{\left(\frac{n}{2\pi}\right)^{d_j/2} l_j(y)}{\sum_{j=1}^9 \left(\frac{n}{2\pi}\right)^{d_j/2} l_j(y)} \quad (5)$$

gdzie: $l_j(y)$ – maksymalna wartość funkcji wiarygodności.

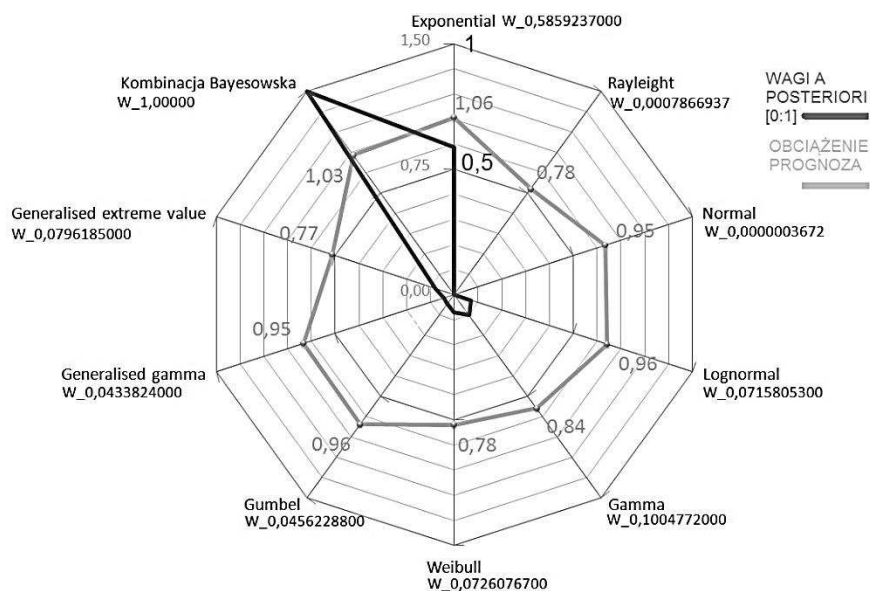


Rys.3. Dystrybuanta kombinacji rozkładów prawdopodobieństwa i dystrybuanta empiryczna

Fig. 3. Cumulative distribution of the combination of probability distributions and empirical cumulative distribution

Należy zauważyć, że pomimo odmiennych kształtów rozkładów apriorycznych, na rozkład aposterioryczny jest związany z funkcją wiarygodności z próby i zależy od informacji uzyskanych drogą empiryczną. W celu ustalenia wartości obciążeń klimatycznych do obliczeń konstrukcji na obiektów budowlanych szacuje się ich statystycznie uzasadnione wartości, które mogą być przewyższone z określonym prawdopodobieństwem, tzn. kwantyle rozkładu maksimów rocznych obliczone na podstawie wyników pomiarów [2,7,8]. W normie [7] zdefiniowano wartość charakterystyczną oddziaływania zmiennego jako kwantyl rzędu p losowych maksimów o okresie powrotu T . Wartość charakterystyczna obciążenia śniegiem gruntu jest kwantylem rzędu $p=0,02$, co koresponduje z okresem powrotu oddziaływania $T=50$ lat [8].

Zakładając równe wartości wag *a priori* dla wszystkich przyjętych rozkładów $w_j = 1/9$, $j=1, 2, \dots, 9$ za pomocą autorskiego programu z wykorzystaniem aplikacji statystycznej wykonanej w programie do analiz statystycznych *R*, obliczono ich wagi *a posteriori*. Na rys. 4 przedstawiono prognozowane wartości obciążenia charakterystycznego śniegiem gruntu, dla poszczególnych rozkładów oraz ich kombinacji bayesowskiej, dla $p=0,02$ i okres powrotu $T=50$ lat, wraz z przypisanymi im wagami.



Rys. 4. Wykres radarowy prognozy wartości charakterystycznej obciążenia śniegiem gruntu z wagami odpowiadającymi poszczególnym rozkładom prawdopodobieństwa

Fig. 4. Radar chart of the forecasts of the characteristic value of the snow load on the ground with weights corresponding to each probability distribution

Największe wagi *a posteriori* uzyskały w kolejności następujące rozkłady: kombinacja bayesowska (1,0), Exponential (0,5859), Gamma (0,1005), Generalised extreme value (0,0796), Weibull (0,0726), Lognormal (0,0716), Gumbela (0,0456), Generalised gamma (0,0434), Rayleigh (0,0008) i Normal (0,0000).

Prognoza wartości obciążenia charakterystycznego dla $p=0,02$ i $T=50$ lat wskazuje, że stopień dopasowania wyników obserwacji do funkcji rozkładu najczęściej stosowanych zmiennych losowych prowadzi do wyników zbieżnych z kombinacją bayesowską jedynie w odniesieniu do rozkładu o wyraźnie dominującej wartości wagi. Dla rozkładu wykładniczego (Exponential) waga wynosi 0,5859, a wartość obciążenia charakterystycznego $S_k=1,06$ kN/m², a dla rozkładu wg kombinacji bayesowskiej 1,03 kN/m². W przypadku pozostałych rozkładów o wartości wag do około 0,10 wartości S_k nie wykazują korelacji z wagami rozkładów.

5. Wnioski

Przedstawiona w artykule procedura estymacji oparta na wnioskowaniu bayesowskim umożliwia ocenę i porównanie zasadności wyboru typu rozkładu zmiennej losowej opartego na wnioskowaniu statystycznym oraz tworzenie liniowych kombinacji tych zmiennych uwzględniających niepewności modelo-

wych związane z subiektywnym wyborem typu rozkładów i łączeniem niepewnych informacji z wynikami badań.

Procedura wnioskowania bayesowskiego zakłada, że estymowane parametry przyjętych do analizy rozkładów są zmiennymi losowymi, a niepewności statystyczne można oszacować za pomocą prawdopodobieństw *a priori*, szacowanych z pewną dozą subiektywizmu, na podstawie dostępnych informacji i korygowanych na podstawie aktualnych wyników badań lub obserwacji.

Przedstawione procedury zastosowano do oceny adekwatności wyboru dystrybucyjnego jednego z oddziaływań klimatycznych, jakim jest obciążenie śniegiem gruntu, traktowanego jako zmienna losowa o parametrach estymowanych statystyczną metodą największej wiarygodności. Obliczono wagi odpowiadające założonym rozkładom prawdopodobieństwa i rozkładowi utworzonemu jako liniowa kombinacja tych rozkładów oraz wartości charakterystyczne obciążenia obliczone dla analizowanych rozkładów. Wyniki obliczeń wskazują, że wartości obciążeń charakterystycznych są w dużym stopniu zależne od wyboru typu rozkładu obciążenia, a wybór rozkładu oparty jedynie na wynikach wnioskowania statystycznego może prowadzić do znacznego zaniżenia prognozowanego obciążenia.

Należy podkreślić, że przedstawiona w artykule metoda może być zastosowana również do analizy wielu zagadnień dotyczących niezawodności konstrukcji.

Literatura

- [1] Woliński Sz., Pytlowany T.: Uwagi o szacowaniu wartości oddziaływań za pomocą modelu Gumbela. Materiały konferencyjne 56 KN KILiW PAN oraz KN PZITB Kielce- Krynica 2010.
- [2] Żurański J., Sobolewski A.: Obciążenie śniegiem w Polsce. Wydawnictwo Instytutu Techniki Budowlanej, Seria: Prace Naukowe ITB. Monografie. Warszawa 2009.
- [3] Noortwijk J. M., Kalk H. J., Chhab E. H.: Bayesian estimation of design loads, HERON. Vol. 49, No. 2 (2004).
- [4] Pytlowany T.: Bayesowski model oddziaływań klimatycznych. OW PRz, zeszyt nr 283. nr s.59 3/2 s. 43-50. Rzeszów 2012.
- [5] Steenbergen M. R.: Maximum Likelihood Programming in R.
- [6] Gwóźdź M., Machowski A. Wybrane badania i obliczenia konstrukcji budowlanych metodami probabilistycznymi, Wydawnictwo Politechniki Krakowskiej, Kraków 2010.
- [7] PN-EN 1990:2004. Eurokod: Podstawy projektowania konstrukcji.
- [8] EN 1991-1-3:2005. Eurokod 1: Oddziaływania na konstrukcje. Część 1-3: Oddziaływania ogólne. Obciążenie śniegiem.
- [9] Mitzenmacher M, Upal E.: Metody probabilistyczne i obliczenia WN-T 2009.

SELECTION OF THE DISTRIBUTION FUNCTION FOR CLIMATE ACTIONS USING BAYESIAN METHOD

Summary

Single distribution functions are usually selected based on a best-fit approach theorem but often available random data cannot be accurately described by any of the commonly used types of the random variables. The paper presents a method based on Bayesian approach which solves problems of selecting the single distribution function and combining of probabilities contending different probability functions. The method is illustrated on the selection of single distribution function and application of Bayesian method in combining these functions to determine the characteristic value of snow load for an assumed return period.

Keywords: meteorological action, snow loads, Bayesian method

Przesłano do redakcji: 07.06.2016 r.

Przyjęto do druku: 30.06.2016 r.

DOI: 10.7862/rb.2016.5