Zeszyty Naukowe
Wyższej Szkoły Bankowej we Wrocławiu
Nr 2(34)/2013

Łukasz Kuźmiński
Uniwersytet Ekonomiczny we Wrocławiu

Zastosowanie teorii wartości ekstremalnych w prognozowaniu ostrzegawczym dla ciągu niezależnych zmiennych o rozkładzie normalnym

Streszczenie. Artykuł dotyczy zastosowań teorii granicznych rozkładów dla ekstremów w prognozach ostrzegawczych dla ciągu zmiennych losowych o rozkładzie normalnym. W początkowej części zawiera elementy teorii dotyczącej statystyk pozycyjnych, natomiast w dalszej przedstawione są podstawowe twierdzenia związane z teorią rozkładów typów ekstremalnych i dziedzin przyciągania. Na koniec zaprezentowano badania empiryczne, w których budowany jest model prognoz ostrzegawczych dla charakterystyk hydrologicznych. Wykorzystane w pracy dane dotyczą stanów wód na dwóch wybranych rzekach Dolnego Śląska.

Słowa kluczowe: statystyki pozycyjne, dystrybuanta graniczna, typy rozkładów ekstremalnych, prognoza ostrzegawcza, charakterystyki hydrologiczne.

Wstęp

Ekstremalne wartości określonych charakterystyk w różnych dziedzinach są w większości przypadków zjawiskiem o niekorzystnym działaniu. Charakterystyki meteorologiczne i hydrologiczne, przyjmując ekstremalne wartości, powodują szereg różnego rodzaju zjawisk o działaniu katastroficznym. Na rynkach finansowych ekstremalne wartości pewnych charakterystyk powodują olbrzymie straty finansowe. Tak naprawdę niepożądany efekt oddziaływania wartości ekstremalnych występuje prawie w każdej dziedzinie życia społeczno-gospodarczego.

Wpływu wartości ekstremalnych nie da się powstrzymać. Można jedynie przygotować się odpowiednio wcześnie na efekt ich działania. Aby można przygotować określone działania mające na celu niwelację efektów oddziaływań wartości ekstremalnych określonych charakterystyk lub zabezpieczenie się przed tymi efektami w możliwie jak największym stopniu, trzeba odpowiednio wcześnie uzyskać informacje o czasie wystąpienia tych efektów. Do tego celu wykorzystuje się systemy prognoz ostrzegawczych.

W tym opracowaniu przedstawione zostanie podejście do procesu budowy prognoz ostrzegawczych oparte na granicznych rozkładach ekstremów analizowanej zmiennej. Ekstrema to określony rodzaj statystyk pozycyjnych. W tym przypadku zmienna będzie miała powszechnie znany i stosowany rozkład normalny. Rozpatrywany będzie przypadek, w którym zostało założone, że zmienne losowe są niezależne. Prognozowaną charakterystyką będzie stan wody na rzekach Nysa Kłodzka i Nysa Łużycka w konkretnych punktach pomiarowych, które zostaną podane w dalszej części pracy.

1. Ekstrema jako szczególny rodzaj statystyk pozycyjnych i ich rozkłady

Jeśli zmienne losowe $\xi_1, \xi_2, ..., \xi_n$ ułożymy w kolejności według rzędu wielkości i zapiszemy je jako

$$\xi_{(1)} \leq ... \leq \xi_{(n)}, \tag{1}$$

to $\xi_{(i)}$ będziemy określać jako i-tą statystykę pozycyjną zmiennych losowych, natomiast $F_{(i)}(x)$ oznaczać będzie dystrybuantę i-tej statystyki pozycyjnej.

Zakładamy, że $\xi_1, \xi_2, ..., \xi_n$ będzie ciągiem n zmiennych losowych o identycznych rozkładach prawdopodobieństwa lub inaczej mówiąc o wspólnej dystrybuancie $F(x)$. Przez $x_1, x_2, ..., x_n$ oznaczymy odpowiednio realizacje zmiennych losowych $\xi_1, \xi_2, ..., \xi_n$. Zgodnie z oznaczeniami wcześniejszymi $X_{(n)}$ oznaczać będzie n-tą statystykę pozycyjną, która jest jednym z ekstremów, a mianowicie maksimum, które można zapisać również w postaci

$$X_{(n)} = \max(\xi_1, \xi_2 ..., \xi_n) \tag{2}$$

oraz oznaczyć przez M_n, jak proponują inni autorzy. Podobnie oznaczać będziemy minimum przez $X_{(1)}$, czyli pierwszą statystykę pozycyjną, którą można zapisać jako

$$X_{(1)} = \min(\xi_1, \xi_2 ..., \xi_n) \tag{3}$$

oraz oznaczyć przez m_n, zgodnie z propozycjami innych autorów[1].

[1] H.A. David, H.N. Nagaraja, *Order Statistics*, A John Wiley & Sons, Inc., 2003.

Dystrybuanta największej statystyki pozycyjnej oznaczona będzie jako $F_{(n)}(x)$ i określona następującym wzorem:

$$F_{(n)}(x) = P\{X_{(n)} \leq x\} = P\{\text{wszystkie } X_i \leq x\} = F^n(x), \tag{4}$$

gdzie $F(x)$ jest dystrybuantą zmiennej losowej ξ_i.

Podobnie jest dla statystyki będącej drugim ekstremum, czyli minimum. Oznaczamy ją jako $X_{(1)}$ i odpowiednio jej dystrybuantę $F_{(1)}(x)$, którą określa następujący wzór:

$$\begin{aligned} F_{(1)}(x) = P\{X_{(1)} \leq x\} = 1 - P\{X_{(1)} > x\} = 1 - P\{\text{wszystkie } X_i > x\} = \\ = 1 - [1 - F(x)]^n. \end{aligned} \tag{5}$$

Statystyki ekstremalne minimum i maksimum stanowią szczególne przypadki ogólnego rezultatu dla dystrybuanty r-tej statystyki pozycyjnej $F_{(r)}(x)$, która określona jest wzorem:

$$\begin{aligned} F_{(r)}(x) = P\{X_{(r)} \leq x\} = P\{\text{przynajmniej } r \text{ z } X_i \leq x\} = \\ = \sum_{i=r}^{n} \binom{n}{i} F^i(x)[1 - F(x)]^{n-1}, \end{aligned} \tag{6}$$

gdyż składnikiem sumy po prawej stronie wzoru jest prawdopodobieństwo rozkładu binominalnego, że dokładnie i spośród $X_1, ..., X_n$ są mniejsze lub równe x. Więcej na temat dystrybuanty $F_{(r)}$ wraz z jej alternatywnymi formami można znaleźć w pracy H.A. Davida[2].

2. Asymptotyczne rozkłady ekstremów – elementy teorii

W rozdziale tym przedstawimy zarys teorii asymptotycznych rozkładów dla ekstremów zmiennych losowych jedynie na potrzeby niniejszego opracowania. Klasyczna teoria wartości ekstremalnych w szczególności określa możliwe postacie dystrybuant granicznych dla ekstremów (maksimów i minimów) w ciągach niezależnych zmiennych losowych o identycznych rozkładach. Z uwagi na asymptotyczny charakter tej teorii, zajmuje się ona rozkładami dla maksimów $X_{(n)}$ (lub M_n), zgodnie z oznaczeniami z poprzedniego podrozdziału oraz w szczególności ich własnościami, gdy $n \to \infty$. Oczywiście wszystkie wyniki uzyskane dla maksimów $X_{(n)}$ (lub M_n) prowadzą do otrzymania wyników dla minimów $X_{(1)}$ przez prostą relację:

---

[2] Ibidem.

$$X_{(1)} = \min(\xi_1, ..., \xi_n) = -\max(-\xi_1, ..., -\xi_n). \tag{7}$$

W skończonych warunkach dystrybuanta statystyki $X_{(n)}$ jest dana wzorem (4). Teoria dla warunków skończonych jest szeroko stosowana. W tym miejscu można sobie postawić pytanie: Po co zgłębiać i stosować teorię rozkładów asymptotycznych dla ekstremów, skoro teoria w warunkach skończonych jest dobrze znana i daje dobre wyniki? Odpowiedź jest bardzo prosta. W teorii asymptotycznej nie musimy znać dystrybuanty rozkładu zbyt precyzyjnie, żeby swobodnie stosować ją do rozwiązywania rzeczywistych problemów. Przykładem może być centralne twierdzenie graniczne Lindeberga-Levy'ego dotyczące rozkładu sum zmiennych losowych o dowolnej dystrybuancie. Podobna sytuacja ma miejsce w teorii rozkładów asymptotycznych dla ekstremów. W rzeczywistości niezdegenerowana dystrybuanta rozkładu $X_{(n)}$ musi należeć do jednej z trzech możliwych rodzin dystrybuant, niezależnie od oryginalnej dystrybuanty $F(x)$ zmiennych ξ_i ($i = 1, ..., n$). Ponadto nie jest potrzebna znajomość szczegółowa dystrybuanty $F(x)$. Wystarczy wiedzieć, do której dziedziny przyciągania (ang. *domain of attrction*) i której z trzech rodzin dystrybuant należy rozpatrywany rozkład ekstremum. Pojęciu dziedziny przyciągania zostanie poświęcony punkt 3 niniejszej pracy. Identyfikacja, do której dziedziny przyciągania należy rozpatrywany rozkład ekstremum, w dużej mierze zależy od zachowania się ogona dystrybuanty $F(x)$ dla dużych wartości x. W związku z tym dużo można powiedzieć o asymptotycznych własnościach maksimum, opierając to na raczej ograniczonej wiedzy o własnościach $F(x)$.

Na początku powinniśmy zainteresować się warunkami, przy których, dla odpowiednio znormalizowanych stałych $a_n > 0$ i b_n,

$$P\{a_n(M_n - b_n) \leq x\} \xrightarrow{w} G(x), \tag{8}$$

(gdzie $\xrightarrow{w}$ oznacza, że zbieżność pojawia się przy ciągłych punktach G, chociaż tak naprawdę zgłębiając asymptotyczną teorię dla rozkładów ekstremów okazuje się, że wszystkie analizowane funkcje G są ciągłe). Teoria ta ukazuje również, że niezdegenerowane dystrybuanty G, które mogą pojawić się jako ograniczenia w (8) tworzą dokładnie klasę rozkładów **max-stabilnych** i odwrotnie, każdy **max-stabilny** rozkład G ma jedną z trzech parametrycznych form podanych poniżej i nazywanych **rozkładami wartości ekstremalnych**[3].

$$\text{Typ I:} \quad G(x) = \exp(-e^{-x}), \qquad -\infty < x < \infty,$$

$$\text{Typ II:} \quad G(x) = \begin{cases} 0, & x \leq 0, \\ \exp(-x^{-\alpha}) & \text{dla pewnego } \alpha > 0,\ x > 0, \end{cases} \tag{9}$$

[3] R. Leadbetter, G. Lindgren, H. Rootzen, *Extremes and related properties of random sequences and processes*, Springer – Verlag New York Heidelberg, New York 1983.

$$\text{Typ III: } G(x) = \begin{cases} \exp(-(-x)^{\alpha}) & \text{dla pewnego } \alpha > 0,\ x \le 0, \\ 1, & x > 0. \end{cases} \tag{9 cd.}$$

Powyższe stwierdzenia dokładnie opisuje twierdzenie, które pojawi się w dalszej części pracy.

Wykorzystując (4) i (8) możemy zapisać, że:

$$F^n(a_n^{-1}x + b_n) \xrightarrow{w} G(x), \tag{10}$$

gdzie ponownie $\xrightarrow{w}$ oznacza zbieżność tylko przy ciągłych punktach funkcji granicznej. Jeśli (10) jest spełnione dla pewnych stałych $\{a_n > 0\}$, $\{b_n\}$, to oznacza, że F należy do obszaru przyciągania (dla maksimów) funkcji G, co zapisuje się jako $F \in D(G)$. Więcej szczegółowych informacji na temat tej teorii można znaleźć w pracach R. Leadbettera i J. Galambosa[4].

3. Ogólna teoria dziedzin przyciągania

Rozdział ten rozpoczniemy od przedstawienia dwóch istotnych twierdzeń związanych z dziedzinami przyciągania. Pierwsze z nich mówi, że dystrybuanta określonego rozkładu jest max-stabilna jeśli i tylko jeśli jest takiego samego typu jak jedna z dystrybuant rozkładów wartości ekstremalnych.

Twierdzenie 1[5]. Każdy max-stabilny rozkład jest typem wartości ekstremalnych (ang. *extreme value type*), np. równa się $G(ax + b)$ dla pewnych $a > 0$ i b, gdzie dla

$$\text{Typ I: } G(x) = \exp(-e^{-x}), \qquad -\infty < x < \infty,$$

$$\text{Typ II: } G(x) = \begin{cases} 0, & x \le 0, \\ \exp(-x^{-\alpha}) & \text{dla pewnego } \alpha > 0,\ x > 0, \end{cases} \tag{11}$$

$$\text{Typ III: } G(x) = \begin{cases} \exp(-(-x)^{\alpha}) & \text{dla pewnego } \alpha > 0,\ x \le 0, \\ 1, & x > 0. \end{cases}$$

Odwrotnie, każdy rozkład typu wartości ekstremalnych jest max-stabilny.

Szczegółowy dowód tego twierdzenia można znaleźć w pracy R. Leadbettera[6].

---

[4] R. Leadbetter, op. cit.; J. Galambos, *The asymptotic Theory of Extreme Order Statistics*, Wiley, New York 1978.

[5] Por. R. Leadbetter, op. cit.

[6] Ibidem.

Drugie twierdzenie dotyczy typów ekstremalnych i nosi nazwę twierdzenia o typach ekstremalnych.

Twierdzenie 2. Niech $M_n = \max(\xi_1, \xi_2, ..., \xi_n)$, gdzie ξ_i są niezależnymi zmiennymi losowymi o identycznych rozkładach. Jeśli dla pewnych stałych $a_n > 0$, b_n mamy:

$$P\{a_n(M_n - b_n) \leq x\} \xrightarrow{w} G(x) \tag{12}$$

dla pewnej niezdegenerowanej dystrybuanty G, to wtedy G jest jedną z trzech dystrybuant granicznych typów wartości ekstremalnych opisanych powyżej. Odwrotnie, każda dystrybuanta G typów wartości ekstremalnych może występować jako ograniczenie w (12), faktycznie występuje, kiedy G sama sobie jest dystrybuantą każdej zmiennej ξ_i[7].

Można przypuszczać, że jeżeli $\xi_1, \xi_2, ..., \xi_n$ nie są koniecznie niezależne, ale $M_n = \max(\xi_1, \xi_2, ..., \xi_n)$ ma asymptotycznie dystrybuantę G w sensie wyrażenia (12), to zbieżność

$$F_n(a_{nk}^{-1}x + b_n) \xrightarrow{w} G^{1/k}(x), \text{ gdy } n \to \infty \tag{13}$$

jest prawdziwa dla $k = 1$, gdzie F_n jest dystrybuantą zmiennej losowej M_n. Powyższa zbieżność jest podstawą następującego twierdzenia:

Twierdzenie 3. (i) Niezdegenerowana dystrybuanta jest max-stabilna, jeśli i tylko jeśli istnieje ciąg $\{F_n\}$ dystrybuant i stałych $a_n > 0$ i b_n taki, że:

$$F_n(a_{nk}^{-1}x + b_n) \xrightarrow{w} G^{1/k}(x), \text{ gdy } n \to \infty \tag{14}$$

dla każdego[8] $k = 1, 2, \ldots$

(ii) W szczególności jeśli G jest niezdegenerowana, $D(G)$ jest niepusty, jeśli i tylko jeśli G jest max-stabilna. Wtedy również $G \in D(G)$. Stąd klasa niezdegenerowanych dystrybuant, która pojawia się jako prawa graniczne w (8) (dla ciągu niezależnych zmiennych losowych o identycznych rozkładach $\xi_1, \xi_2, ..., \xi_n$) pokrywa się z klasą dystrybuant max-stabilnych.

Skoro można pokazać, że jeśli (14) jest spełnione dla $k = 1$, to jest też spełnione dla wszystkich k, to z tego będzie wynikać, że G jest max-stabilna z twierdzenia 3 i że G jest typem wartości ekstremalnych (dystrybuantą graniczną dla ekstremów). Tak więc kiedy będziemy rozważać przypadek dla zmiennych zależnych, będzie łatwo pokazywać, że przy odpowiednich założeniach prawdziwość wyrażenia (14) dla $k = 1$ implikuje prawdziwość tego wyrażenia dla wszystkich k, z którego twierdzenie o typach ekstremalnych znowu wynika.

Zauważmy teraz, że twierdzenie 2 zakłada, iż $a_n(M_n - b_n)$ ma niezdegenerowaną dystrybuantę graniczną i wtedy dowodzi, że G musi mieć jedną z trzech

[7] R. Leadbetter, op. cit.
[8] Ibidem.

określonych form. Można skonstruować ciągi niezależnych zmiennych losowych o identycznych rozkładach $\{\xi_n\}$, dla których taka dystrybuanta graniczna G nie istnieje. Zobaczmy prosty przykład. Wygodne będzie tutaj (i później również) użycie oznaczenia x_F dla prawego punktu końcowego dystrybuanty F, tzn.

$$x_F = \sup\{x; F(x) < 1\} \quad (\leq \infty), \tag{15}$$

co oznacza, że $F(x) < 1$ dla $x < x_F$ i $F(x) = 1$ dla wszystkich $x \geq x_F$.

Załóżmy teraz, że każda ξ_n ma dystrybuantę F taką, że $x_F < \infty$ i że F ma skok w punkcie x_F, tzn. $F(-x_F) < 1 = F(x_F)$. Wtedy wynika z łatwością, że jeśli $\{u_n\}$ jest pewnym ciągiem i $P\{M_n \leq u_n\} \to \rho$, to $\rho = 0$ lub 1. Stąd jeśli $P\{a_n(M_n - b_n) \leq x\} \to G(x)$, to biorąc $u_n = x/a_n + b_n$ otrzymujemy, że $G(x) = 0$ lub 1 dla każdego x tak, że $G(x)$ jest zdegenerowana. Przykładem takiego ciągu, który nie ma granicznej dystrybuanty dla ekstremów jest przypadek, gdzie każda ξ_n ma rozkład Poissona.

Praktycznym aspektem w teorii dziedzin przyciągania jest to, żeby wiedzieć, której z trzech dystrybuant granicznych określonych wzorem (9) użyć, kiedy każda zmienna losowa ξ_n ma rozkład opisany przez dystrybuantę F podkreślając, że postać dystrybuanty nie musi być precyzyjnie określona. Znane są różne warunki potrzebne i wystarczające dla każdego z trzech typów dystrybuant granicznych. Z uwagi na to, iż dowody tych warunków są zbyt długie i mało przydatne na potrzeby tego opracowania, pominiemy je.

W poniższych twierdzeniach przedstawimy pewne proste i użyteczne, a zarazem wystarczające warunki, które stosuje się, kiedy dystrybuanta rozkładu F ma funkcję gęstości prawdopodobieństwa f, do sprawdzenia, do której z trzech dziedzin przyciągania należy dystrybuanta maksimum rozpatrywanego ciągu $\{\xi_n\}$.

Twierdzenie 4[9]. Zakładamy, że dystrybuanta F zmiennych losowych w ciągu niezależnych zmiennych losowych o identycznych rozkładach $\{\xi_n\}$ jest całkowicie ciągła z funkcją gęstości prawdopodobieństwa f. Wtedy wystarczające warunki przynależności dystrybuanty F do trzech możliwych dziedzin przyciągania są następujące:

Typ I: f ma ujemną pochodną f' dla wszystkich x w pewnym przedziale (x_0, x_F), $(x_F \leq \infty)$, $f(x) = 0$ dla $x \geq x_F$ i

$$\lim_{t \uparrow x_F} \frac{f'(t)(1-F(t))}{f^2(t)} = -1 . \tag{16}$$

[9] Por. R. Leadbetter, op. cit.

Typ II: $f(x) > 0$ dla wszystkich $x \geq x_0$ skończonych i

$$\lim_{t\to\infty} \frac{tf(t)}{1-F(t)} = \alpha > 0. \tag{17}$$

Typ III: $f(x) > 0$ dla wszystkich x w pewnym skończonym przedziale (x_0, x_F), $f(x) = 0$ dla $x > x_F$ i

$$\lim_{t\uparrow x_F} \frac{(x_F - t)f(t)}{1-F(t)} = \alpha > 0. \tag{18}$$

Twierdzenie 5[10]. Potrzebne i wystarczające warunki dla dystrybuanty F zmiennych losowych w ciągu niezależnych zmiennych losowych o identycznych rozkładach $\{\xi_n\}$ przynależności do każdego z trzech typów dystrybuant granicznych są następujące:

Typ II: $x_F = \infty$ i

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-\alpha},\ \alpha > 0 \text{ dla każdego } x > 0. \tag{19}$$

Typ III: $x_F < \infty$ i

$$\lim_{h\downarrow 0} \frac{1-F(x_F - xh)}{1-F(x_F - h)} = x^{\alpha},\ \alpha > 0 \text{ dla każdego } x > 0. \tag{20}$$

Typ I: Istnieje pewna ściśle dodatnia funkcja $g(t)$ taka, że:

$$\lim_{t\uparrow x_F} \frac{1-F(t+xg(t))}{1-F(t)} = e^{-x}, \text{ dla wszystkich rzeczywistych } x. \tag{21}$$

Można wykazać, że $\int_0^\infty (1-F(u))du < \infty$, kiedy utrzymany jest typ I ograniczenia i odpowiedni wybór funkcji g, takiej że: $g(t) = \int_t^{x_F} (1-F(u))du / (1-F(t))$ dla $t < x_F$.

Wniosek z powyższego twierdzenia podany poniżej ma istotne znaczenie praktyczne, ponieważ zawiera równości, za pomocą których można otrzymać stałe normujące a_n i b_n ze zbieżności danej wzorem (8).

Wniosek 1. Stałe a_n, b_n w zbieżności $P\{a_n(M_n - b_n) \leq x\} \to G(x)$ mogą być wzięte dla każdego przypadku (typu) w postaci:

[10] Por. R. Leadbetter, op. cit.

$$\begin{aligned} &\text{Typ II:}\quad a_n = \gamma_n^{-1},\ b_n = 0, \\ &\text{Typ III:}\ a_n = (x_F - \gamma_n)^{-1},\ b_n = x_F, \\ &\text{Typ I:}\quad a_n = [g(\gamma_n)]^{-1},\ b_n = \gamma_n, \end{aligned} \tag{22}$$

gdzie $\gamma_n = F^{-1}(1 - 1/n) = \inf\{x;\ F(x) \geq 1 - 1/n\}$. (Ciąg γ_n musi być taki, że $1 - \mathrm{F}(\gamma_n) = n^{-1}$).

Jeżeli określony rozkład prawdopodobieństwa ma dystrybuantę graniczną dla ekstremum, to w myśl twierdzenia 2 jest ona jedną z trzech typów danych wzorem (9). W niniejszej pracy będziemy się zajmować danymi o powszechnie znanym rozkładzie normalnym. Poniższe twierdzenie pokazuje, do którego typu należy dystrybuanta graniczna rozkładu maksimum i jak wyznaczyć stałe normujące a_n i b_n w przypadku, gdy zmienne losowe mają standardowy rozkład normalny.

Twierdzenie 6. Jeśli $\{\xi_i\}$ jest ciągiem niezależnych zmiennych losowych o standardowym rozkładzie normalnym, to asymptotyczna dystrybuanta zmiennej losowej $M_n = \max(\xi_1, \xi_2, ..., \xi_n)$ jest typu I. Szczegółowo:

$$P\{a_n(M_n - b_n) \leq x\} \to \exp(-e^{-x}), \tag{23}$$

gdzie

$$a_n = (2\log n)^{1/2} \tag{24}$$

i

$$b_n = (2\log n)^{1/2} - \frac{1}{2}(2\log n)^{-1/2}(\log\log n + \log 4\pi)\,. \tag{25}$$

Szczegółowy dowód powyższego twierdzenia znajduje się w pracy R. Leadbettera[11].

4. Prognozy ostrzegawcze w hydrologii

W niniejszym paragrafie wprowadzone zostanie pojęcie prognozy ostrzegawczej jako zapowiedzi wystąpienia zdarzenia traktowanego przez odbiorcę jako niekorzystne. Bardziej precyzyjnie prognozę ostrzegawczą określa definicja 1, która jest jedną z wielu, jakie można spotkać w literaturze przedmiotu, ale najbardziej pasuje do problematyki niniejszego opracowania.

[11] Ibidem.

Definicja 1. **Prognozą ostrzegawczą** jest stan zmiennej w momencie lub okresie należącym do przyszłości, gdy przewiduje się niekorzystne kształtowanie się kontrolowanej zmiennej na podstawie informacji dostarczonej przez szereg czasowy.

Zadaniem prognozy ostrzegawczej jest dostarczenie na czas informacji o wystąpieniu w przyszłości niekorzystnej wartości monitorowanej charakterystyki. Jak z tego wynika, prognozy ostrzegawcze stanowią specyficzny rodzaj przewidywania, nie dotyczą bowiem wyznaczania przyszłej wartości monitorowanej zmiennej, lecz tylko faktu, że wartość monitorowanej zmiennej przekroczy wyznaczony poziom lub będzie niższa od danego poziomu w zależności od charakteru badanego zjawiska. W takim kontekście prognoza ostrzegawcza jest prognozą jakościową, ponieważ dotyczy zdarzenia lub sytuacji, które można zapisać w arytmetyce binarnej (+), (–) lub 0, 1[12].

Ograniczając szerokie zastosowanie prognoz ostrzegawczych jedynie do dziedziny hydrologii, należy również ograniczyć zbiór zmiennych kontrolowanych w procesie prognozowania ostrzegawczego. W hydrologii jednym z wielu zadań prognoz ostrzegawczych jest monitorowanie takiej charakterystyki jak stany wód w rzekach, w celu ochrony przeciwpowodziowej. Charakterystykę wymienionej zmiennej prezentuje definicja 2.

Definicja 2. **Stan wody** jest to wzniesienie zwierciadła wody w cieku ponad umowny poziom odniesienia (co nie jest równoznaczne z głębokością cieku). Należy rozróżnić pojęcia stan wody i poziom wody. Są to te same wielkości fizyczne, jednak podawane względem różnych odniesień. Poziomy terenu liczymy od przyjętego poziomu morza, dlatego wysokość, na której znajdują się obiekty na Ziemi, wyrażamy w metrach nad poziomem morza. W Polsce sieć wodowskazowa odniesiona jest obecnie do poziomu morza w Kronsztadzie w Rosji. Dla uproszczenia zapisu wzniesienie zwierciadła wody liczymy od ustalonego „zera" wodowskazu. Taki pomiar nazywamy stanem wody, w odróżnieniu od poziomów liczonych względem przyjętego zera niwelacji. W praktyce zera ustalane są poniżej najniższego stanu wody w celu uniknięcia wartości ujemnych wynikających z możliwej erozji dennej pogłębiającej dno np. rzeki. Rzędna zera każdego wodowskazu określona jest w odniesieniu do państwowej sieci niwelacyjnej, dlatego mając tę informację jesteśmy w stanie wyznaczyć poziom wody. Na podstawie wieloletnich pomiarów można określić charakterystyczny rozkład stanów wody dla danej rzeki w danym miejscu. Wyznacza się wówczas następujące strefy stanów wody:

– strefę stanów niskich,
– strefę stanów średnich,
– strefę stanów wysokich,

[12] U. Siedlecka, *Prognozy ostrzegawcze w gospodarce*, PWE, Warszawa 1996.

– stan ostrzegawczy,
– stan alarmowy.

Wymienionym powyżej opisom dla danej rzeki w określonym miejscu przyporządkowane są konkretne wartości liczbowe, według których identyfikuje się rodzaj stanu wody w punkcie pomiaru na podstawie zaobserwowanego stanu w danym czasie *t*.

W tym opracowaniu prognoza będzie budowana dla stanu wody obserwowanego na rzece Nysa Łużycka w stacji Zgorzelec oraz na rzece Nysa Kłodzka w stacji Nysa. Więcej na temat budowy prognoz zawarto w kolejnym paragrafie.

5. Budowa modelu prognoz ostrzegawczych na podstawie danych empirycznych

W niniejszym opracowaniu analizie poddane zostały stany wód na dwóch polskich rzekach: Nysie Kłodzkiej i Nysie Łużyckiej. Wykorzystamy dzienne dane dotyczące stanów wód uzyskane ze stacji na wodach powierzchniowych w Nysie i Zgorzelcu, pochodzące z Instytutu Meteorologii i Gospodarki Wodnej w Warszawie. Na potrzeby budowy prognoz ostrzegawczych dysponujemy dziennymi stanami wód za okres 1.1981-12.2011 r. (11 322 obserwacji) ze stacji Nysa oraz za okres 1.1982-12.2011 r. (10 951 obserwacji) ze stacji w Zgorzelcu.

Z uwagi na to, że w tej pracy zajmujemy się przypadkiem, w którym analizowane są ciągi niezależnych zmiennych losowych do budowy prognoz ostrzegawczych wzięte zostały stany wód pochodzące z co dziesiątego dnia. Taki zabieg miał na celu zminimalizowanie prawdopodobieństwa wystąpienia zależności pomiędzy danymi z sąsiadujących okresów. Oczywiste jest, że im większa odległość pomiędzy następującymi po sobie danymi, tym prawdopodobieństwo wystąpienia zależności pomiędzy nimi jest coraz mniejsze. Na potrzeby tej pracy pomijamy fakt występowania ewentualnej długoterminowej zależności pomiędzy danymi, zwanej inaczej efektem długiej pamięci.

Istotnym, a zarazem pierwszym punktem w praktycznym zastosowaniu teorii dystrybuant granicznych dla ekstremów jest ustalenie, z populacji o jakim rozkładzie pochodzą analizowane dane. Tak naprawdę dobre dopasowanie teoretycznego rozkładu do danych empirycznych to problem, z jakim borykają się od wielu lat statystycy i naukowcy innych dziedzin, wykorzystujący w praktyce metody statystyczne. Sposobów na rozwiązanie tego problemu jest wiele, ale nie ma jednego uniwersalnego.

W tej pracy przyjmujemy, że dwa badane ciągi zmiennych losowych $\xi_1, \xi_2, ..., \xi_{n_1}$ oraz $\zeta_1, \zeta_2, ..., \zeta_{n_2}$ są ciągami niezależnych zmiennych losowych o rozkładach normalnych. Zmienne losowe ξ_s i ξ_{s+1} są oddalone od siebie w czasie o 10 dni. To samo dotyczy zmiennych ζ_i (i = 1, 2, ..., n_2). Z uwagi na fakt, iż wstępne ciągi zmiennych losowych dotyczących dziennych stanów wód zostały na potrzeby analizy zastąpione przez ciągi zmiennych losowych przedstawiające stany wód co 10 dni, liczba zmiennych losowych w obu ciągach zmniejszyła się 10-krotnie. Dla ciągu zmiennych losowych $\xi_1, \xi_2, ..., \xi_{n_1}$ $n_1 = 1132$, natomiast dla ciągu $\zeta_1, \zeta_2, ..., \zeta_{n_2}$ $n_2 = 1095$.

Hipotezy o normalności dla dwóch analizowanych ciągów zmiennych losowych zostały poddane testowaniu za pomocą dwóch dobrze znanych testów na normalność: testu zgodności chi-kwadrat i testu zgodności Kołmogorowa. Wartości p-value dla obu testów i ciągów zmiennych losowych przedstawione są w tabeli 1.

Tabela 1. Wartości p-value dla testów na normalność rozkładu

Ciąg zmiennych losowych	Test zgodności chi-kwadrat	Test zgodności Kołmogorowa
$\xi_1, \xi_2, ..., \xi_{n_1}$	$p_v = 0{,}0273$	$p_v = 0{,}0389$
$\zeta_1, \zeta_2, ..., \zeta_{n_2}$	$p_v = 0{,}0121$	$p_v = 0{,}0238$

Ź r ó d ł o: opracowanie własne.

Wyniki z tabeli 1 pokazują, że na poziomie istotności $\alpha = 0{,}01$ możemy przyjąć, iż stany wód w obu badanych stacjach mają rozkład normalny. W związku z tymi wynikami do obliczenia prognoz ostrzegawczych dla badanych punktów pomiarowych wykorzystamy graniczną dystrybuantę typu I dla standardowego rozkładu normalnego, zgodnie z twierdzeniem 5. Przekształcając w prosty sposób wzór (23) otrzymujemy następujące wyrażenie:

$$P\left\{M_n \le \frac{x}{a_n} + b_n\right\} \to \exp\left(-e^{-\left(\frac{x}{a_n}+b_n\right)}\right), \tag{26}$$

z którego łatwo wyznaczymy prawdopodobieństwa dla określonych wartości maksimów badanych zmiennych. Prawdopodobieństwa osiągania przez badane zmienne określonych wartości stanowić będą prognozy ostrzegawcze. Stałe normujące a_n i b_n w wyrażeniu (26) dla obu ciągów zmiennych losowych wyznaczamy wykorzystując wzory (24) i (25). Wartości stałych normujących dla obu ciągów zmiennych losowych przedstawione są w tabeli 2.

Prognozy ostrzegawcze będą prawdopodobieństwami przekroczenia stanu ostrzegawczego i alarmowego przez maksima zmiennych ξ_i i ζ_j w ciągach

Tabela 2. Stałe normujące dla badanych ciągów

Ciąg zmiennych losowych	a_n	b_n
$\xi_1, \xi_2, ..., \xi_{n_1}, n_1 = 1132$	2,471	2,346
$\zeta_1, \zeta_2, ..., \zeta_{n_2}, n_2 = 1095$	2,465	2,340

Ź r ó d ł o: opracowanie własne.

o liczebnościach n_1 i n_2. Stany ostrzegawcze i alarmowe oraz ich wartości standaryzowane dla obu badanych stacji pomiarowych podane są w tabeli 3. Do standaryzacji wykorzystane zostały parametry z prób: średnia i odchylenie standardowe z analizowanych ciągów danych.

Tabela 3. Stany ostrzegawcze i alarmowe wraz z wartościami standaryzowanymi

Stany wód	Stacja Nysa	Stacja Zgorzelec
Stan ostrzegawczy	400 cm	340 cm
Wartość standaryzowana dla stanu ostrzegawczego	5,52	4,2
Stan alarmowy	530 cm	400 cm
Wartość standaryzowana dla stanu alarmowego	8,86	5,73

Ź r ó d ł o: opracowanie własne.

Wyznaczone prognozy za pomocą wzoru (26) i danych z tabel 2 i 3 przedstawione są w tabeli 4.

Tabela 4. Prognozy ostrzegawcze dla stacji Nysa i Zgorzelec

Prognozy stanu wód	Stacja Nysa	Stacja Zgorzelec
Prognoza stanu ostrzegawczego	$P(M_{n_1} > 400) – 0,0102$	$P(M_{n_2} > 340) – 0,0174$
Prognoza stanu alarmowego	$P(M_{n_1} > 530) – 0,0026$	$P(M_{n_2} > 400) – 0,0094$

Ź r ó d ł o: opracowanie własne.

Analizując otrzymane wyniki widzimy, że w stacji Nysa jest 1,02% szans na to, iż stan ostrzegawczy zostanie przekroczony, a 0,26%, że przekroczony zostanie stan alarmowy. W stacji Zgorzelec jest 1,74% szans na przekroczenie stanu ostrzegawczego, a 0,94% szans na przekroczenie stanu alarmowego. Oczywiście wraz ze zmianą n_1 i n_2 w analizowanych ciągach zmiennych losowych prognozy w prosty sposób będą aktualizowane. Powoduje to, że przedstawiony model prognoz ostrzegawczych ma charakter dynamiczny i na bieżąco może być aktualizowany, dając rzeczywisty obraz sytuacji. Faktem zasługującym na uwagę jest to, że jakość wyznaczanych prognoz jest wprost proporcjonalna do długości ciągu analizowanych zmiennych. Mówiąc wprost, im dłuższym szeregiem czasowym dysponujemy, tym uzyskiwane wyniki są lepsze. Oznacza to, że im dłużej prowadzimy monitoring, tym lepszej jakości prognozy uzyskujemy.

Podsumowanie

W pracy przedstawione zostało podejście do prognozowania ostrzegawczego oparte na granicznych rozkładach dystrybuant ekstremów obserwowanych charakterystyk. Wyznaczone zostały prognozy dotyczące stanów wód w dwóch wybranych stacjach pomiarowcy na rzekach Dolnego Śląska. Zwróciliśmy uwagę na to, że jedną z najważniejszych rzeczy przy takim podejściu do prognozowaniu ostrzegawczego jest odpowiednie dopasowanie rozkładu teoretycznego do analizowanych danych empirycznych. W rozpatrywanych przykładach poddaliśmy analizie ciągi zmiennych losowych o powszechnie znanym i stosowanym rozkładzie normalnym dla przypadku, w którym zmienne losowe są niezależne. Rozkłady inne niż normalne oraz przypadki dla zmiennych zależnych rozpatrywane będę w kolejnych pracach związanych z tą tematyką.

Istotne jest, że zbudowany model prognoz ostrzegawczych ma charakter dynamiczny. Oznacza to, że może być na bieżąco aktualizowany. Daje to możliwość efektywnego monitoringu analizowanych charakterystyk i odpowiednio wczesnego reagowania na sygnały alarmujące. Oczywiście prezentowany model można stosować do budowy prognoz ostrzegawczych dla dowolnych charakterystyk z różnych dziedzin, np. do prognozowania danych finansowych na rynkach finansowych, charakterystyk związanych z jakością w różnych działach ekonomii i wielu innych.

Literatura

Cieślak M., *Prognozowanie gospodarcze. Metody i zastosowania*, WN PWN, Warszawa 2002.
Czekała M., *Statystyki pozycyjne w modelowaniu ekonometrycznym. Wybrane problemy*, Wyd. Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław 2001.
David H.A., Nagaraja H.N., *Order Statistics*, A John Wiley & Sons, Inc., New York 2003.
Fisz M., *Rachunek prawdopodobieństwa i statystyka matematyczna*, PWN, Warszawa 1967.
Galambos J., *The asymptotic Theory of Extreme Order Statistics*, Wiley, New York 1978.
Jakubowski J., Sztencel R., *Wstęp do teorii prawdopodobieństwa*, SCRIPT, Warszawa 2004.

Leadbetter R., Lindgren G., Rootzen H., *Extremes and related properties of random sequences and processes*, Springer – Verlag Heidelberg, New York 1983.
Magiera R., *Modele i metody statystyki matematycznej*, Oficyna Wydawnicza GiS, Wrocław 2002.
Siedlecka U., *Prognozy ostrzegawcze w gospodarce*, PWE, Warszawa 1996.
Weron A., Weron R., *Inżynieria finansowa*, Wyd. Naukowo-Techniczne, Warszawa 1998.

The application of the extreme values theory in warning predictions for random variables sequences with normal distribution

Summary. The work refers to the application of the theory of limit distributions for extremes in warning predictions for random variables sequences with normal distribution. The beginning of the work contains theory components concerning order statistics. In the next part of the work basic theorems connected to the theory of types of distributions and gravity areas are presented. Ultimately, empirical research is given, in which a model of warning predictions for hydrological characteristic is built. Details used in the work concern water conditions at two selected rivers of Lower Silesia.

Key words: order statistics, the limiting distribution function, types of extreme distributions, warning forecast, hydrological characteristics