

## Wybrane problemy stosowania metod statystycznych

*Są trzy rodzaje kłamstw: kłamstwa,  
cholerne kłamstwa i statystyki.  
Mark Twain albo Beniamin Disraeli  
(tylko Pan Bóg wie kto naprawdę)*

### 1. Wstęp

Zakres zastosowań probabilistyki, a w szczególności statystyki jako eksperymentalnego narzędzia wnioskowania, stale się poszerza. Istnieją ku temu co najmniej dwa powody. Jeden z nich – można go nazwać teoretycznym – jest nadal pewnym zaskoczeniem związanym z poszerzaniem się obszaru niepewności. Okazuje się, że znane od dawna deterministyczne równania rządzące np. dynamiką atmosfery, są dla pewnych zakresów parametrów niezwykle czułe na zmianę wartości tych parametrów, co prowadzi do niemożności tworzenia sensownej predykcji deterministycznej (zob. np. Schuster 1995). Drugi powód to zapoczątkowana około połowy wieku temu rewolucja w przetwarzaniu danych związana z komputerami. Możliwości jakie dają komputery pozwoliły z jednej strony na wydatne poszerzenie bazy pomiarowej (m.in. poprzez satelity, automatyczne stacje pomiarowe), a z drugiej – na stosowanie pewnych metod przetwarzania danych silnie związanych ze statystyką, jakie byłyby niemożliwe do szerszego stosowania przedtem.

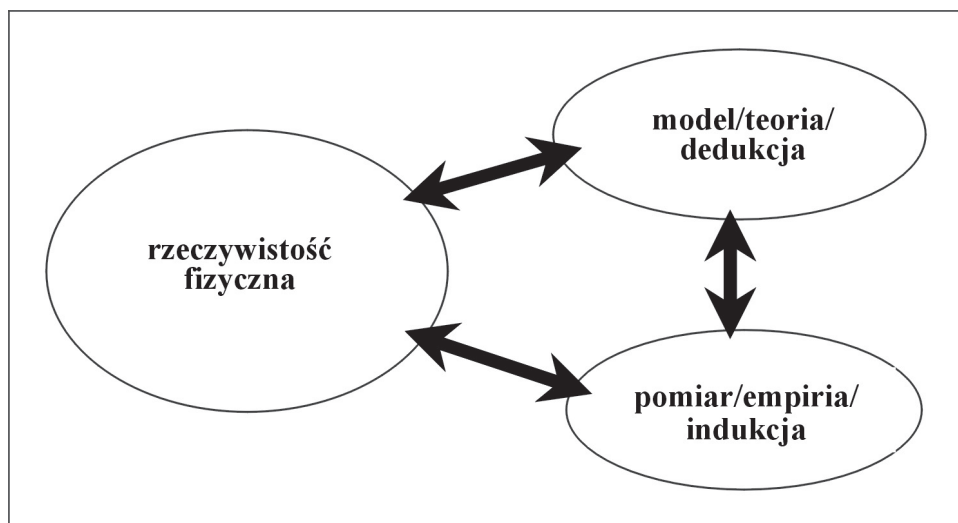
Łatwość korzystania z ogólnie dostępnych narzędzi przetwarzania danych, głównie narzędzi statystycznych, oprócz niewątpliwych i nie do przecenienia zalet, prowadzi może i prowadzi – szczególnie w warunkach niewystarczającego krytycyzmu badawczego – do prymatu narzędzi nad podmiotem narzędzia tego używającym i do nadzwyczajnej proliferacji modeli, co w jakimś sensie już się dzieje. Problem ten widział już przed prawie dwudziestoma laty Klemeš (1986), nazywając go *mathematistry* i zalecając staranny wgląd w istotę badanego zagadnienia. Dzisiaj, w dobie powszechnego stosowania komputerów osobistych i dostępnego powszechnie oprogramowania, problem ten nabiera szczególnego znaczenia a statystyka, jako najczęstsze narzędzie używane w badaniach naukowych, jest szczególnie na to narażona. Posiadanie pewnego zbioru danych empirycznych umożliwia bowiem tworzenie dowolnej liczby modeli empirycznych (statystycznych), które niekoniecznie cokolwiek wyjaśniają w takim sensie, w jakim jest to na ogół oczekiwane od podejścia naukowego.

Z kolei probabilistyka, a w szczególności statystyka, zawierają wiele ważnych pojęć nie całkiem jasnych, określanych różnymi terminami, lub wieloznacznych bądź mylących, co rozpowszechnia bogata literatura zarówno podręcznikowa, jak i publikacje badawcze. Nieświadomość tego faktu może poważnie utrudniać lub czasami wręcz uniemożliwiać poprawne wnioskowanie statystyczne. Dotyczy to m.in. testów statystycznych, których moc weryfikacji jest mniej rozstrzygająca, niż na ogół się sądzi.

Celem niniejszej pracy jest dyskusja tego typu problemów związanych z podstawami poznawczymi statystyki z nadzieją, że okaże się ona pomocna w rozważnym korzystaniu z metod statystycznych. Należy bowiem mieć na uwadze, że metody te pozwalają na wykrywanie wzorców niekoniecznie istniejących w rzeczywistości, a więc – zgodnie ze starożytną sentencją *praemonitus praemunitus* – świadomość takiej możliwości powinna zmniejszyć możliwość porażki.

## 2. Metody badania rzeczywistości oraz zasada niedookreśloności

Za Encyklopedią Multimedialną PWN (2000) przyjmijmy, że nauka to „działalność badawcza zmierzająca do obiektywnego, zasadnego i uporządkowanego poznania rzeczywistości oraz wytwór tej działalności”. Przyjmijmy dalej, że definicja ta jest na tyle oczywista, że nie wymaga żadnych dodatkowych objaśnień (co oczywiście nie jest prawdą, ale dla celów niniejszego tekstu jest wystarczające). Działalność tę – w sposób przybliżony – można podzielić na badania teoretyczne oraz badania empiryczne, jak to schematycznie zostało ujęte na ryc. 1.



Ryc. 1. Możliwe relacje pomiędzy rzeczywistością a jej opisem

Podjęcie teoretyczne (dedukcyjne) oznacza przyjęcie z góry pewnej struktury badanego fragmentu rzeczywistości lub, inaczej mówiąc, przyjęcie pewnego modelu tegoż fragmentu. Model ten czasami tłumaczy rzeczywiste zjawiska lub procesy w kategoriach przyczynowo-skutkowych, a czasami je tylko naśladuje. Stworzony model nie tylko może pełnić rolę eksplikacyjną (objasniać rzeczywistość), ale także może zostać wykorzystany do uzyskiwania pewnych danych ilościowych, np. do symulacji lub do predykcji, która może dotyczyć zarówno przeszłości (rekonstrukcja, postgnoza) jak i przyszłości (prognoza).

Podjęcie empiryczne (indukcyjne) jest w swoich założeniach skromniejsze, jakkolwiek pewna wizja rzeczywistości jest konieczna po to, by pytania stawiane poprzez pomiar i otrzymywane za jego pomocą odpowiedzi znajdowały się w tych samych jednolitych ramach. Oznacza ono stopniowe dochodzenie do pewnego obrazu rzeczywistości poprzez doświadczenie. Zachodzi oczywiste a istotne pytanie na ile sformułowany w końcu model jest zdeterminowany badaniami empirycznymi. Wydaje się, że nie zawsze jest tutaj wystarczająco uświadamiany fakt, że te same wyniki badań empirycznych mogą być dobrze tłumaczone przez różne teorie. Fakt ten – nazwany zasadą niedookreśloności (*undetermination principle*<sup>1</sup>) (Życiński 1993) – oznacza, że doświadczenie nie wyznacza jedynie prawdziwego opisu rzeczywistości. Poważną konsekwencją uznania tej zasady jest to, że doświadczenie nie może być rozstrzygającym dowodem prawdziwości badanej teorii, może ją jedynie wspomagać.

### 3. Niepewność a rachunek prawdopodobieństwa i statystyka matematyczna

Problem opisu rzeczywistości staje się bardziej skomplikowany, gdy ma on uwzględniać niepewność. Ogólnie można mówić o niepewności na poziomie wyboru modelu, niepewności na poziomie oszacowania parametrów wybranego modelu oraz niepewności na poziomie pomiaru. Zwykle nie ocenia się liczbowo tej pierwszej, głównie z powodu trudności, na jakie taka ocena napotyka. Niepewność druga, tj. niepewność modelu powstająca wskutek niepewności jego parametrów, jest oceniana najczęściej (często jest ona minimalizowana przez wybór odpowiednich wartości parametrów), natomiast niepewność pomiarowa – choć na ogół łatwa do ilościowego określenia – jest często zaniedbywana i nie jest brana pod uwagę.

Uwzględnienie niepewności w opisie rzeczywistości oznacza przeważnie wykorzystanie podejścia probabilistycznego. W kategoriach ryciny 1, podejście dedukcyjne uwzględniające niepewność oznacza stosowanie modeli probabilistycznych, a podejście indukcyjne – stosowanie metod statystyki matematycznej, czyli wykonywanie szeroko rozumianych pomiarów i analizę uzyskanych wyników. Godzi się tutaj przypomnieć, że rachunek prawdopodobieństwa opisuje badaną wielkość (zmienną losową) bez wykonywania doświadczenia losowego (dostarcza modeli

---

<sup>1</sup>„żaden zespół danych empirycznych nie określa w sposób jednoznaczny interpretacji teoretycznej, która jest w stanie wyjaśnić te dane”.

zjawisk/procesów), natomiast statystyka matematyczna opisuje badaną wielkość na podstawie wyników wielokrotnie wykonanego doświadczenia losowego (czyli na podstawie tzw. próby losowej). Warto też mieć na uwadze, że parametry modelu probabilistycznego są często utożsamiane w wartościami rzeczywistymi a metody statystyczne dostarczają szacunkowych wartości tych wielkości. Jeśli np. tworzymy model pomiaru pewnej stałej wielkości, to zwykle utożsamiamy prawdziwą wartość tej wielkości z wartością oczekiwaną  $EX$  zmiennej losowej  $X$  (rozumianej jako zbiór wszystkich możliwych wyników pomiaru; zwykle też przyjmuje się normalny rozkład tej zmiennej), której odpowiada nieskończenie wiele różnych wartości pomiaru. Innymi słowy, rachunek prawdopodobieństwa daje jedną wartość, statystyka matematyczna natomiast – nieskończenie wiele oszacowań tej wartości.

Nie ma wątpliwości, że – chociaż w różnym stopniu – niepewność jest obecna praktycznie wszędzie, toteż praktycznie każdy opis powinien ją uwzględniać. Jednakże, gdy z powodu niejasnych i najprawdopodobniej również skomplikowanych powiązań przyczynowo-skutkowych, badacz – aby uzyskać jakiś wgląd w problem – jest praktycznie skazany na pomiar, wtedy podejście statystyczne jest z reguły jedynym środkiem uzyskania informacji. Wydaje się, że wiele problemów klimatologii jest właśnie tego typu.

Podejście empiryczne oznacza pomiar, a więc tworzenie pewnych zbiorów danych, które następnie, wraz z przyjętymi założeniami, będą podstawą wnioskowania, niekoniecznie statystycznego, o badanym problemie. Stąd ważna jest świadomość posiadania lub nieposiadania przez ten zbiór pewnych charakterystyk. W przypadku podejścia statystycznego zbiór ten nazywany jest *próbą losową* i tak będzie on zwany w dalszym ciągu niniejszego tekstu. Próba losowa jest pewnym podzbiorem (w ogólności wielowymiarowym) zbioru możliwych wyników pomiaru nazywanym w statystyce (wielowymiarową) zmienną losową (można go też nazywać populacją).

Podstawową konieczną charakterystyką próby losowej jest jej *reprezentatywność*. Termin ten oznacza, że próba ta zawiera istotne informacje o zmiennej losowej, pozwalając wnioskować o własnościach tej zmiennej. Niestety, reprezentatywność próby jest na ogół postulowana i nie podlega weryfikacji. W większości zastosowań przyrodniczych przyjmuje się, że własność ta będzie zachowana jeśli zapewnimy próbie losowość.

Znaczenie terminu *losowy* (użyteczne synonimy: *przypadkowy*, *stochastyczny*) nie jest jasne. Cramér (1958) wręcz twierdzi, że nie można podać dokładnej definicji tego określenia. Hooke i Shaffer (1969) uważają, że losowy to (i) zawierający element przypadkowości i (ii) pochodzący z rozkładu równomiernego. Można powiedzieć, że w ogólnym sensie *losowy* oznacza *bez wzorca lub celu* lub *nieprzewidywalny*, natomiast w sensie szczegółowym znaczenie tego słowa zależy od konkretnego problemu. Przykładem mogą być tutaj testy losowości, gdzie znaczenie tego terminu zależy od postaci sformułowanej hipotezy lub hipotezy alternatywnej i najczęściej oznacza, iż testowana próba losowa pochodzi z populacji jednorodnej, a sama próba losowa jest próbą prostą. Praktycznie wyraża się to przez testowanie braku trendu lub wahań cyklicznych. Można tu wymienić testy autokorelacji, testy kombinatoryczne (uporządkowania), korelacji rangowej czy testy białości ciągu (Zieliński 1979; Domański 1990).

## 4. Pojęcie często źle interpretowane: zależność

Terminem, który w statystyce jest często jest rozumiany niewłaściwie, jest *zależność* (używana na ogół bez bardzo przydatnego określnika *stochastyczna*) dwu lub więcej zmiennych, której funkcyjnym wyrazem jest *regresja* zmiennych.

Zależność w sensie statystycznym jest brakiem *niezależności*, która w przypadku dwu zmiennych  $X$ ,  $Y$  może zostać wyrażona np. w ten sposób: dla każdej pary wartości  $(x, y)$  tych zmiennych zachodzi równość  $P(Y < y \mid X = x) = P(Y < y)$  (zapis  $\mid X = x$  oznacza „pod warunkiem, że  $X = x$ ”) a więc, że prawdopodobieństwo dotyczące zmiennej  $Y$  nie zależy od wartości przyjmowanych przez zmienną  $X$ . Wniosek z tego jest taki, że istnienie zależności stochastycznej dwu zmiennych oznacza, że rozkład prawdopodobieństwa jednej z nich zależy (w sposób funkcyjny) od wartości przyjmowanych przez drugą.

Pojęciem pokrewnym zależności stochastycznej, a stosowanym niezwykle często, jest *korelacja* zmiennych, która w sensie statystycznym oznacza związek zmiennych słabszy niż zależność stochastyczna, gdyż niezależność zmiennych oznacza brak korelacji. Nie jest jednak na odwrót: zmienne skorelowane nie muszą być niezależne. Należy przy tym pamiętać, że badając zależność najczęściej badamy właśnie korelację oraz że najczęściej jest to korelacja liniowa, a więc dość szczególnie przypadek współzależności. Liczbową miarą korelacji jest współczynnik korelacji, który często bywa również miarą współzależności zmiennych.

Konsekwencją zależności zmiennych w podanym wyżej przykładzie dwu zmiennych  $X$ ,  $Y$  jest istnienie zależności średniej warunkowej (a dokładnie wartości oczekiwanej),  $E(Y \mid x)$ , zmiennej  $Y$  od wartości zmiennej  $X$ . Zależność tę nazywamy *regresją* zmiennej  $Y$  względem zmiennej  $X$  (jeśli  $X$  to zmienna losowa) lub  $x$  (jeśli  $x$  to zwykła zmienna, najczęściej czas).

Zadziwiająca niezrozumienie istoty zależności i regresji można znaleźć nawet w podręcznikach, gdzie pojęcia te są rozumiane czasami jako relacje przyczynowo-skutkowe. Przykładem może być Norcliffe (1986), który na str. 182 pisze: „o ile w regresji zakłada się zależność przyczynową między  $X$  i  $Y$  o tyle w korelacji zakłada się asocjacje”. Oba te zdania – zarówno nadrzędne jak i podrzędne – wprowadzają czytelnika w błąd, a zdanie pierwsze szczególnie. Rzecz bowiem nie w tym, co zakłada się w regresji (nie ma to żadnego znaczenia dla stosowanej metody statystycznej), a w tym, co regresja oznacza. W świetle tego, co powiedziano wyżej, zależność stochastyczna (a przez to regresja) oznacza tylko *współzmiennność* badanych zmiennych. A współzmiennność oznacza wyłącznie synchroniczność realizacji (wartości) badanych zmiennych i nic poza tym. Inną sprawą jest natomiast przyczyna tej współzmienności, która jest w ogólności trójaka: (i) współzmiennność jest efektem przypadku (a więc tylko synchroniczność, nic więcej, nie ma przyczyny); (ii) współzmiennność jest skutkiem działania ukrytej zmiennej rządzącej obiema badanymi zmiennymi; (iii) współzmiennność jest wynikiem zależności przyczynowo-skutkowej badanych zmiennych. Należy wyraźnie podkreślić, że statystyka w żaden sposób nie rozstrzyga o przyczynowości czegokolwiek – problem ten leży w ogóle poza matematyką i należy do jej interpretacji.

W przypadku, gdy badane są związki pomiędzy więcej niż dwoma zmiennymi, problem się komplikuje z powodu szybko rosnącej liczby możliwości oraz wymagań

Tab. 1. Rosnąca liczba zmiennych wymaga do pełnej analizy potęgowo rosnącej liczby danych (częściowo za Lindquistem 2002). Za wystarczającą liczebność dwuwymiarowej próby losowej przyjęto 30

Liczba zmiennych	2	3	4	5	6
Liczba par	1	3	6	10	15
Liczba wszystkich możliwych związków	1	4	11	26	57
% wszystkich możliwych związków	100%	75%	55%	38%	26%
liczba par	1	3	6	10	15
liczba trójek	–	1	4	10	20
liczba czwórek	–	–	1	5	15
liczba piątek	–	–	–	1	6
liczba szóstek	–	–	–	–	1
Równoważna liczba danych	30	$30^2 = 900$	$30^3 = 27\ 000$	$30^4 = 810\ 000$	$30^5 = 24\ 300\ 000$

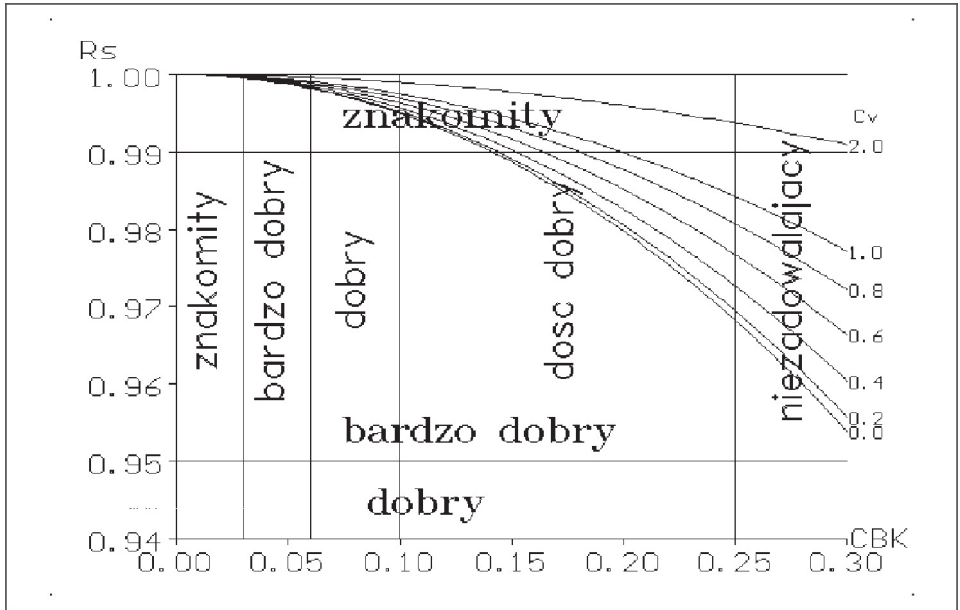
dotyczących liczebności próby losowej, co ilustruje tabela 1.

## 5. Weryfikacja hipotez statystycznych – wiara w przeciętność

Procedura testowania hipotez statystycznych polega na (i) przyjęciu prawdziwości postawionej hipotezy  $H_0$ , (ii) zmierzeniu różnicy pomiędzy tym, co daje hipoteza, a tym, co daje próba losowa i (iii) stwierdzeniu, czy uzyskana różnica (tzw. wartość statystyki testowej) znajduje się w ustalonym wcześniej obszarze krytycznym (lub inaczej: obszarze odrzucenia). Ponieważ przyjęte z góry prawdopodobieństwo znalezienia się wartości statystyki testowej w obszarze odrzucenia jest małe (zwykle rzędu 5%), to uznaje się, że zaistnienie takiego faktu jest wskazówką przeciwko prawdziwości hipotezy. Należy pamiętać o słabości logicznej takiego wnioskowania. Fakt znalezienia się wartości statystyki testowej w obszarze odrzucenia nie jest równoważny fałszywości hipotezy  $H_0$ , bo hipoteza dopuszcza go z góry zadany prawdopodobieństwem. Z kolei, przypadek przeciwny – znalezienia się wartości statystyki testowej poza obszarem odrzucenia – nie jest równoważny prawdziwości hipotezy  $H_0$ , chociaż przewidywała go ona z dużym prawdopodobieństwem, bo zgodnie z zasadą niedookreśloności teorii nie można wykluczyć innej prawdziwości hipotezy. Poprawność wnioskowania statystycznego dotyczącego fałszywości badanej hipotezy jest wspierana przyjmowanym *implicite* założeniem, że pomiar dostarczył wartości typowych, tj. bardzo prawdopodobnych. Ponieważ hipoteza  $H_0$  uznaje je za mało prawdopodobne, jest fałszywa.

## 6. Uwaga na niektóre kryteria jakości: każdy może dostać to, co chce

Statystyka dostarcza wielu kryteriów jakości modelu; często kryteria ilościowe jakości (miary jakości modelu) są zamieniane na określenia typu: bardzo dobry, do-



Ryc. 2. Specjalny współczynnik korelacji  $R_s$  w funkcji całkowitego błędów kwadratowych CBK dla niektórych wartości współczynnika zmienności  $C_v$  wraz ze słownymi ocenami jakości modelu opartymi na  $R_s$  i CBK według pracy Sarma i in. (1973)

bry, itp. Określenia te należy, jeśli w ogóle, stosować bardzo ostrożnie, zwłaszcza w przypadku używania kilku na raz, gdyż łatwo okazać się może, że oceniany model jest np. bardzo dobry z punktu widzenia jednego kryterium a nie do przyjęcia z punktu widzenia innego. Rycina 2 posłuży tutaj za przykład stosowania dwu kryteriów jakości jednocześnie (szczegóły w pracy: Węglarczyk 1998). Istniejące ilościowe oceny jakości zostały przez Sarma i in. (1973) przetworzone na oceny słowne bez głębszej analizy ich znaczenia. Wynikające stąd konsekwencje są jasno widoczne.

Jak widać, w skrajnym przypadku model może być jednocześnie bardzo dobry i niezadowolający. Oczywiście, nie sposób uciec od ocen tego rodzaju, trzeba jednak w każdym przypadku zdawać sobie sprawę na jaką charakterystykę ocenianego modelu stosowane kryterium jest czułe.

## 7. Podsumowanie

Stosowanie metod indukcyjnych wnioskowania wymaga stałej świadomości ich ograniczeń, szczególnie w przypadku badania problemów o bardzo skomplikowanej i słabo rozpoznanej strukturze przyczynowo-skutkowej. Dotyczy to zwłaszcza metod statystycznych, które są bardzo często w takich przypadkach stosowane jako główna metoda wnioskowania. Należy też mieć na uwadze konieczność wzmocnienia argu-



mentów statystycznych argumentami pozastatystycznymi, najlepiej przyrodniczymi, aby mieć silniejsze poczucie pewności, że rozwiązanie badanego problemu nie jest artefaktem lecz faktem.

### **Literatura**

- Cramér H., 1958, *Metody matematyczne w statystyce*, PWN, Warszawa.
- Domański Cz., 1990, *Testy statystyczne*, PWE, Warszawa.
- Encyklopedia Multimedialna PWN*, 2000, Warszawa.
- Hooke R., Shaffer D., 1969, *Modele matematyczne a rzeczywistość*, PWN, Warszawa.
- Klemeš V., 1986, *Dilettantism in hydrology: Transition or destiny*, *Water Resour. Res.*, 22, 177-188.
- Lindquist P., 2002, *Stochastic Modeling*, adres internetowy: ([http://www.actuaries.ca/meetings/annual/2002/Handouts/DG11\\_Linquist.pdf](http://www.actuaries.ca/meetings/annual/2002/Handouts/DG11_Linquist.pdf)), data pobrania: 19.06.2004
- Norcliffe G.B., 1986, *Statystyka dla geografów*, PWN, Warszawa.
- Sarma P.B.S., Delleur J.W., Rao A.R., 1973, *Comparison of rainfall-runoff models for urban areas*, *J. Hydrol.*, 18(3/4), 329-347.
- Schuster H.G., 1995, *Chaos deterministyczny. Wprowadzenie*. Wyd. Naukowe PWN, Warszawa.
- Węglarczyk S., 1998, *The interdependence and applicability of some statistical quality measures for hydrological models*, *J. Hydrol.*, (206)1-1, 98-103.
- Zieliński R., 1979, *Generatory liczb losowych*, PWN, Warszawa.
- Życiński J., 1993, *Granice racjonalności. Eseje z filozofii nauki*, Wyd. Nauk. PWN, Warszawa.

Stanisław Węglarczyk  
Instytut Inżynierii i Gospodarki Wodnej  
Politechnika Krakowska  
Kraków