Zastosowanie statystycznej analizy szeregów czasowych do krótkoterminowego prognozowania rozwoju dyscyplin naukowych

Łukasz Opaliński
Biblioteka Politechniki Rzeszowskiej im. I. Łukasiewicza
Marcin Jaromin
Zakład Biotechnologii i Bioinformatyki Wydziału Chemicznego Politechniki Rzeszowskiej

Abstrakt

Cel/Teza: Celem pracy jest opis podstaw statystycznej metody analizy szeregów czasowych, zbadanie możliwości jej zastosowania do krótkoterminowego prognozowania rozwoju dyscyplin naukowych w oparciu o dane na temat cytowalności publikacji dyscyplinarnych, wykorzystanie zapewnianej przez metodę możliwości oceny niepewności prognozy, a także interpretacja uzyskanych wyników w kategoriach naukoznawczych.
Koncepcja/Metody badań: W ramach niniejszego opracowania zastosowano metodę analitycznej dekompozycji szeregu czasowego jako metodę wyodrębniania trendu rozwojowego w oparciu o dane na temat dynamiki zjawisk masowych. Zjawiskiem masowym było w tym przypadku cytowanie literatury naukowej w polskojęzycznej sferze nauk o Ziemi. Metoda polega w głównej mierze na sformułowaniu równania opisującego ilościowo przebieg zjawiska z rozróżnieniem trendu głównego oraz składowej odpowiadającej wahaniom przypadkowym. Równanie stanowi podstawę przewidywania przyszłych wartości szeregu czasowego.
Wyniki i wnioski: Uzyskane wyniki przemawiają za możliwością efektywnego zastosowania statystycznej analizy szeregów czasowych do krótkoterminowego prognozowania tempa rozwoju dyscyplin naukowych. W przypadku zrealizowanego badania prognozy okazały się akceptowalne dla dwuletniego horyzontu prognozy, chociaż istnieją podstawy, aby przypuszczać, że skonstruowanie prognoz w postaci przedziałów ufności lub nieznaczne zmodyfikowanie metody mogłoby wydłużyć ten okres. Wyniki zinterpretowano w kategoriach naukoznawczych jako uproszczony indeks natychmiastowości dyscyplin, który jest szczególnie predysponowany do praktycznych zastosowań komparatywnych.
Oryginalność/Wartość poznawcza: W ramach analizy literatury przedmiotu nie natrafiono na przykłady badań, które byłyby prowadzone w sposób zaproponowany w ramach niniejszej publikacji. Analiza szeregów czasowych była wcześniej stosowana jedynie do wyników podziału zbioru publikacji dyscyplinarnych na, nazywane „klastrami", podzbiory prac w założeniu reprezentatywnych dla określonych subdyscyplin. Operacja podziału jest zawsze dokonywana w oparciu o jedno z wielu dostępnych kryteriów podobieństwa publikacji, a w szczególności kryterium oparte na współwystępowaniu słów i terminów kluczowych oraz w oparciu o bardzo liczne i dające różne efekty algorytmy grupujące. Wydaje się w związku z tym, że podejście oparte na bezpośredniej cytowalności publikacji pozwala wyeliminować ze zbioru wyników element pewnej niejednoznaczności (względności) uwarunkowany znaczącym zróżnicowaniem i brakiem uniwersalności w zakresie sposobu działania algorytmów grupujących oraz kryteriów kognitywnego podobieństwa publikacji naukowych.

Słowa kluczowe

Analiza szeregów czasowych. Bibliometria. Czasopisma naukowe. Metody statystyczne w informatologii. Rozwój dyscyplin naukowych.

Otrzymany: 1 marca 2017. Zrecenzowany: 19 lipca 2017. Zaakceptowany: 22 grudnia 2017.

1. Wprowadzenie

Prognozowanie tempa i kierunków rozwoju dyscyplin naukowych jest jednym z tradycyjnych zadań stojących przed bibliometrią. Wspominał o nim już Alan Pritchard w 1969 r. pisząc, że celem bibliometrii jest m.in. „naświetlenie (...) natury i kierunków rozwoju dyscyplin [naukowych] (...) za pomocą zliczania i analizowania różnorakich aspektów komunikacji piśmienniczej"[1] (Pritchard, 1969, 348). Z oceną dynamiki i prognozowaniem dalszego rozwoju dyscyplin naukowych ściśle wiąże się wyodrębnianie nowo powstających i nie w pełni jeszcze ukształtowanych (ang. *emerging*) subdyscyplin (obszarów problemowych czy, inaczej mówiąc, wąskich specjalności dyscyplinarnych) (Skalska-Zlat, 1993, 20; Skalska-Zlat, 1999, 60). Ta ostatnia funkcja działalności bibliometrycznej polega najczęściej na tzw. mapowaniu (ang. *mapping*) nauki, czyli odwzorowywaniu i wizualizowaniu poznawczej (teoretycznej, logicznej czy heurystycznej) struktury różnych dyscyplin naukowych i tworzeniu tzw. topologicznych map nauki (zob. np. Chen, 2006; Garfield, 2004, 119). Niektórzy badacze twierdzą przy tym, że okres najintensywniejszych prac prowadzonych w ramach tego odgałęzienia bibliometrii przypada na ostatnie 15 lat (Ena et al., 2016, 1015; Wang & Ho, 2016, 482), chociaż początków realizacji takich zadań można dopatrywać się już w latach 50. XX w., zaraz po założeniu filadelfijskiego Institute for Scientific Information (Boyack & Klavans, 2014, 671). Obecnie można nawet spotkać się z opinią, że zadanie „mapowania" nauki i rozpoznawanie – mówiąc słowami Dereka de Solli Price'a – „frontów badawczych" (ang. *research fronts*) jest pierwszoplanowym komponentem bibliometrii (Huang & Chang, 2015, 2042; Price, 1967, 114; Wang & Ho, 2016, 482).

Przewidywanie dalszego rozwoju dyscyplin naukowych niesie ze sobą rozliczne korzyści praktyczne. Jest ono na tyle istotne, że istnieją nawet towarzystwa naukowe i cykliczne konferencje, które poświęca się tym zagadnieniom (Small et al., 2014, 1450). Po pierwsze, jest ono cenne z punktu widzenia samych naukowców. Śledzenie najnowszych tendencji dyscyplinarnych i identyfikacja najbardziej perspektywicznej tematyki i pytań badawczych (ang. *hot topics*) pozwoliłyby naukowcom nadążyć za szybkim i wielopłaszczyznowym rozwojem dyscyplin, a także racjonalnie ukierunkować swoje wysiłki badawcze lub ułatwić planowanie przyszłej współpracy naukowej (Takeda & Kajikawa, 2009, 544; Tu & Hsu, 2016, 2016–2017). Drugim obszarem praktycznych zastosowań bibliometrycznego prognozowania rozwoju dyscyplin naukowych, jest ogólnie rozumiana polityka naukowa, w skład której wchodzi administrowanie i zarządzanie działalnością naukową oraz kwestie finansowania badań naukowych z budżetu państwa (zob. np. Braun et al., 2000, 24). Ponieważ badania te są „zaczynem" (ang. *seeds*) innowacyjności przemysłowej, znaczenie prac badawczo-rozwojowych (ang. *R&D – Research and Development*) jest uznawane za fundamentalne w procesie stymulowania innowacyjności i konkurencyjności narodowego przemysłu i technologii (Shibata et al., 2008, 758–759). W kontekście pragmatycznego znaczenia prognoz rozwoju dyscyplin wydaje się też, że mogą one okazać się wartościowe także dla bibliotekarzy, a zwłaszcza dla bibliotekarzy dziedzinowych. Hanning Guo i in. (2011) zauważyli bowiem, że ta grupa zawodowa może odnieść korzyść wynikającą z dostępu do narzędzi ułatwiających tworzenie nowych kategorii tematycznych lub haseł opisu

[1] Oryg.: „(...) to shed light on (...) the nature and course of development of a discipline (...), by means of counting and analysing the various facets of written communication" (Pritchard, 1969, 348).

rzeczowego w trakcie opracowywania zbiorów, uzupełnianie słowników istniejących języków informacyjno-wyszukiwawczych (np. tezaurusów) o nowe terminy, albo wyodrębnianie specjalnych kolekcji zbiorów bibliotecznych, ukierunkowanych na potrzeby informacyjne badaczy prowadzących studia o profilu zgodnym z nowo powstającymi subdyscyplinami (Guo et al., 2011, 422).

2. Literatura przedmiotu

Literatura przedmiotu dotycząca wyodrębniania nowo powstających specjalności w ramach dobrze ugruntowanych teoretycznie dyscyplin oraz prognozowania ich rozwoju została szczegółowo omówiona w dwóch odrębnych opracowaniach pierwszego z autorów niniejszego artykułu (Opaliński, 2017a; Opaliński, 2017b). W związku z tym wydaje się, że w tym miejscu wystarczy przypomnieć jeden z głównych wniosków płynących z analizy literatury przedmiotu, według którego w metodologii zaczerpniętej ze statystyki matematycznej zawarty jest znaczny potencjał, chociaż wydaje się on dotąd nie w pełni wykorzystany. Świadczy o tym niewielka liczba publikacji, w których autorzy posługiwali się *stricte* analizą szeregów czasowych (zob. Opaliński, 2017b). W szczególności zauważalne jest, że w przeciwieństwie do m.in. metody tworzenia klastrów publikacji i analizy głównych ścieżek rozwoju dyscyplin, statystyka matematyczna jako jedyna stwarza możliwość szacowania niepewności prognoz. Ponadto wydaje się, że wyodrębnienie klastrów nie jest dokładnie tym samym co prognozowanie w ścisłym sensie. W kategoriach analiz klastrów publikacji nie podaje się bowiem konkretnych, ilościowych wartości jakichkolwiek wskaźników, które można byłoby uznać za np. przewidywaną liczbę cytowań, jakie otrzymają kluczowe dokumenty, patenty czy czasopisma. Nie określa się też liczbowo tempa rozwoju dyscyplin w postaci np. funkcji matematycznej lub prawdopodobieństwa, że specjalność albo obszar badawczy osiągnie maksimum swojej siły wpływu np. w ciągu pięciu lat. Wymiar prognostyczny metod grupowania zbioru publikacji polega raczej na przyjmowanym implicite założeniu, że klastry najaktywniejsze w chwili badania, wykazujące się najwyższą cytowalnością lub największą liczbą publikacji, utrzymają swoją wiodącą pozycję także w (bliżej nieokreślonej) przyszłości oraz że to na poruszanej w nich problematyce będą koncentrować się dalsze badania, rozwijając ją, modyfikując, testując i poszukując dla niej zastosowań praktycznych. W przypadku statystycznej analizy szeregów czasowych możliwe jest natomiast nie tylko dokonanie oceny niepewności prognozy, ale też podanie konkretnych wartości, które sygnalizują np. to, jak będzie kształtować się przyszła cytowalność zbioru publikacji dyscyplinarnych czy, alternatywnie, ich liczba, jako tempo produkcji piśmienniczej. W związku z tym można postawić tezę, że metody wyróżniania klastrów publikacji oraz identyfikacji głównych ścieżek rozwoju dyscyplin, jako reprezentantów frontów badawczych, służą w większym stopniu ocenie dynamiki rozwoju dyscyplin, podczas gdy ich aspekt prognostyczny jest drugoplanowy. W przypadku metod statystyki matematycznej mamy natomiast do czynienia z sytuacją przeciwną – aspekt prognostyczny dominuje tutaj nad oceną aktualnego stanu rozwoju dyscyplin.

3. Cel pracy

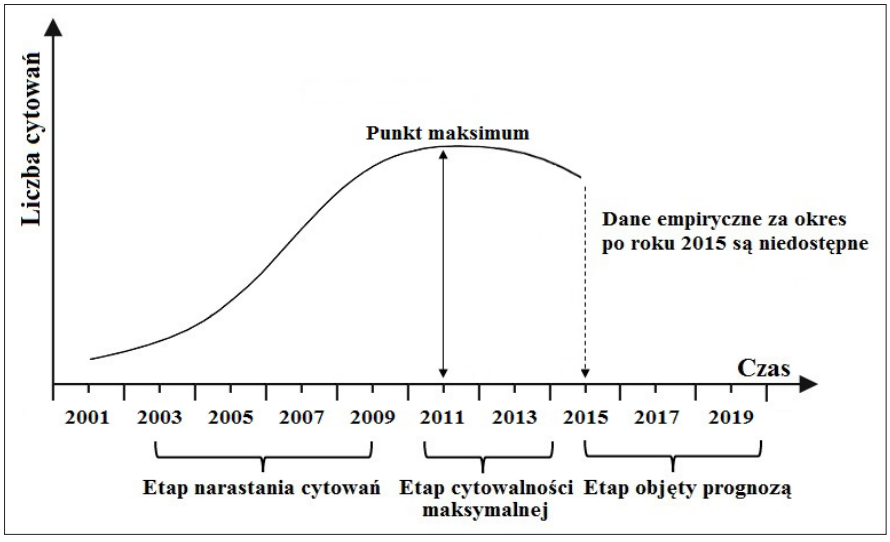
Celem pracy jest opis podstaw statystycznej metody analizy (dekompozycji) szeregów czasowych, zbadanie możliwości jej zastosowania do krótkoterminowego prognozowania tempa rozwoju dyscyplin naukowych w oparciu o dane na temat cytowalności publikacji dyscyplinarnych, wykorzystanie zapewnianej przez metodę możliwości oceny niepewności prognozy, a także interpretacja uzyskanych wyników w kategoriach naukoznawczych.

4. Metoda badań

Metoda analizy szeregów czasowych jest bliższa naukom społecznym i ekonomicznym (np. socjologii, ekonometrii, demografii) niż naukom ścisłym, w przeciwieństwie do większości streszczonych powyżej algorytmicznych metod tworzenia klastrów publikacji wzajemnie powiązanych. Pod nazwą „analiza szeregów czasowych" kryje się cały zespół różnych metod analizy dynamiki zjawisk masowych, których wspólnym celem jest przewidywanie przyszłego zachowania się zjawisk zachodzących w czasie przy pomocy identyfikacji głównych trendów rozwojowych rządzących zdarzeniami, ich sezonowych lub okresowych odchyleń od trendu i oceny niepewności towarzyszącej prognozie. Są to przede wszystkim tzw. metody indeksowe, metody wygładzania wykładniczego oraz metody mechanicznego i analitycznego wyodrębniania trendu. Dekompozycja szeregu czasowego, która należy do analitycznych metod wyodrębniania trendu i zostanie zastosowana w ramach niniejszego opracowania, polega na sformułowaniu równania opisującego ilościowo przebieg zjawiska z rozróżnieniem trendu głównego, składowej wahań cyklicznych lub sezonowych oraz składowej odpowiadającej wahaniom przypadkowym (zob. np. Aczel, 2007, 621–667; Kildiszew & Frenkel, 1976; Krawiec, 2014; Łapkowska-Baster, 2009, 72–92; Sobczyk, 2015, 307–356; Sobczyk, 2008, 35–87).

W niniejszym opracowaniu pominięto wyznaczanie składowej wahań cyklicznych oraz składowej wahań okresowych. Wahania okresowe wywołują regularny wzrost lub spadek wartości badanego zjawiska i występują w miarowych odstępach czasu, np. dobowych, tygodniowych lub miesięcznych. Przykładem może być tu dobowa zmienność w zużyciu energii elektrycznej. Wahania sezonowe charakteryzują się natomiast zawsze cyklem rocznym i wywołują równie regularną sezonowość zjawiska. Przykładem wahań sezonowych jest np. kwartalny wzrost sprzedaży owoców (Aczel, 2007, 629–632; Krzysztofiak & Luszniewicz, 1976, 311; Łapkowska-Baster, 2009, 72). Wydaje się w związku z tym, że w przypadku, gdy rozpatrywanym zjawiskiem jest cykl życia publikacji naukowych, nie może być mowy o tak uregulowanych czynnikach wpływających na poziom cytowalności. Za przyjęciem tego założenia przemawia fakt, że autorzy niniejszej pracy nie natrafili dotąd na żadną publikację, w której pojawiałyby się stwierdzenia czy wnioski sygnalizujące istnienie jakichkolwiek zjawisk wywołujących okresowość lub sezonowość w obrębie cykli życiowych dokumentów naukowych. Z tego powodu w niniejszej pracy uwzględniono jedynie wahania przypadkowe, które są spowodowane działaniem czynników losowych, występują nieregularnie z różnym nasileniem i z reguły nie są znane ich przyczyny (Sobczyk, 2015, 355; Zeliaś et al., 2013, 50–51). Uznaje się też, że uwidaczniają się one we wszystkich rodzajach szeregów czasowych (Sobczyk, 2008, 40).

W ramach niniejszego artykułu przyjęto definicję szeregu czasowego, zgodnie z którą w kontekście cykli życiowych publikacji naukowych jest nim ciąg liczb cytowań publikacji o różnym wieku, które były cytowane w jednym, wybranym roku. W związku z tym podstawą proponowanej w ramach niniejszej pracy metody prognozowania jest obserwacja tzw. synchronicznego (retrospektywnego) cyklu życiowego publikacji naukowych (nt. definicji bibliometrycznej metody synchronicznej i alternatywnej wobec niej metody diachronicznej zob. np. Burrell, 2001, 3; Diodato, 1994, 122–123; Diodato & Smith, 1993, 101; Glänzel, 2004, 512–513; Stinson & Lancaster, 1987, 65–67). Typowy diachroniczny cykl życiowy publikacji składa się z etapu narastania jej cytowalności, osiągnięcia momentu cytowalności maksymalnej oraz następującego po punkcie maksimum spadku cytowalności. Na przedstawionym poniżej rysunku 1 zobrazowano hipotetyczny (i nieco wyidealizowany) przebieg synchronicznego odpowiednika diachronicznego cyklu życiowego pewnego zbioru publikacji, z których wszystkie były cytowane – w różnej liczbie – w 2015 r. Publikacje wydane w latach 2011–2015 były cytowane najczęściej i uformowały one ten fragment krzywej cytowań, który w modelu synchronicznym odpowiada diachronicznej fazie narastania cytowalności. Publikacje wydane w 2011 r. były cytowane z największym nasileniem, w związku z czym można stwierdzić, że znajdują się one w centrum synchronicznego punktu maksimum krzywej. Publikacje wydane przed rokiem 2011 (a więc w latach 2010–2001 i wcześniej) są cytowane (w 2015 r.) coraz rzadziej i w coraz mniejszej liczbie w miarę przesuwania się na osi czasu ku jej przecięciu z pionową osią współrzędnych. Etap objęty prognozą rozpoczyna się natomiast po 2015 r. – w praktyce jego początek to punkt osi czasu oznaczający 2016 r. – ponieważ źródłowym materiałem badawczym autorów niniejszego artykułu były prace wydane w 2015 r. Jakiejkolwiek źródła bibliograficzne ogłoszone po tym czasie z naturalnych przyczyn nie mogły więc już figurować w bibliografiach załącznikowych badanych publikacji (zob. Rys. 1; zob. też Aneks 3).



Rys.1. Synchroniczny cykl życiowy zbioru hipotetycznych publikacji naukowych

Prognozy liczb cytowań, które odnoszą się do okresu po 2015 r. wymagają w związku z tym odpowiedniej interpretacji, którą autorzy postarali się przedstawić w szóstej sekcji niniejszego artykułu.

Wyjątki od tego wzorca, który określono wyżej jako typowy dla przeciętnych i występujących najliczniej publikacji, mogą stanowić np. publikacje klasyczne (tzw. *citation classics*), które z biegiem czasu są cytowane coraz wyżej i w ogóle nie osiągają punktu maksimum, publikacje cytowane bardzo rzadko lub wcale, albo tzw. publikacje efemeryczne (ang. *flash in the pan*). Ten ostatni rodzaj to prace, które stają się wysoko cytowane natychmiast po opublikowaniu, ale bardzo szybko wytracają swój impet, a ich cytowalność spada niemal do zera (zob. np. Costas et al., 2010, 331; Garfield, 1985, 193–194; Levitt & Thelwall, 2008, 42; Van Dalen & Henkens, 2005, 211). Są to jednak raczej przypadki odosobnione, a zdecydowana większość publikacji zachowuje się pod względem swojej cytowalności tak, jak przykładowy zbiór prac naukowych z rysunku 1. Inaczej mówiąc, w dziedzinie lub w dyscyplinie naukowej, która znajduje się na etapie tzw. nauki normalnej w sensie teorii Thomasa Kuhna (2001), ma się na ogół do czynienia z publikacjami cytowanymi na przeciętnym poziomie, które tworzą gros dziedzinowej/dyscyplinarnej literatury (zob. np. Redner, 1998, 134; Seglen, 1992, 629–630, 635; Shibata et al., 2008, 762–763). Przez „naukę normalną" rozumie się tutaj taki etap jej rozwoju, na którym nie dokonują się akurat żadne „przewroty kopernikańskie", nie przewiduje się żadnych nadchodzących przełomów, zmian paradygmatu ani naukowych rewolucji. Założenie, że wybrana do badania polskojęzyczna sfera nauk o Ziemi podlega aktualnie prawidłowości opisywanej terminem „nauka normalna" jest konieczne w kontekście przyjęcia synchronicznej metodologii badań cyklu życiowego losowej próby publikacji. Ponieważ w ramach niniejszej pracy autorzy dysponują wyłącznie danymi o roczniku 2015 wyselekcjonowanych czasopism (zob. Aneks 1), obraz synchronicznego cyklu życiowego publikacji, zbudowany na podstawie tych danych, jest w pewnym sensie statyczny. To znaczy, że obraz ten odzwierciedla stan dyscypliny taki, z jakim mamy do czynienia w tym właśnie (2015) roku. Prognoza przyszłego rozwoju dyscypliny będzie w związku z tym opisywać przebieg tego rozwoju w warunkach, jakie panowały w zbadanej gałęzi nauki w roku 2015, będzie przez warunki te podyktowana i nimi uwarunkowana.

Ogólnie rzecz biorąc trendy rozwojowe zjawisk można podzielić na liniowe (tzw. trendy addytywne) i nieliniowe (tzw. trendy multiplikatywne). Wśród trendów nieliniowych wyróżnia się trendy wykładnicze, potęgowe, wielomianowe, hiperboliczne i logistyczne (De Gooijer & Hyndman, 2006, 444; Krzysztofiak & Luszniewicz, 1976, 364–366; Łapkowska-Baster, 2009, 73; Sobczyk, 2008, 57–58; Sobczyk, 2015, 345–349). W przypadku cyklu życia zbioru dyscyplinarnych publikacji, w którym występuje tendencja rosnąco-malejąca (zob. Rys. 1; Aneks 2; Aneks 3), zastosowanie znajdują funkcje wielomianowe, a zwłaszcza wielomiany drugiego stopnia, czyli parabole (Krzysztofiak & Luszniewicz, 1976, 365). Należy zaznaczyć, że w literaturze przedmiotu daje się zauważyć polaryzację poglądów na temat oceny przydatności stosowania trendu wielomianowego w zagadnieniach predykcyjnych. Ma ona zarówno swoich zwolenników, jak i przeciwników (szerzej na ten temat zob. np. Purczyński, 2010; Purczyński, 2014, 120–121; zob. też: Żurowska, 2005).

W ramach niniejszej pracy zastosowano analizę szeregów czasowych do publikacji cytowanych w obrębie polskojęzycznej literatury czasopiśmienniczej z pogranicza nauk o Ziemi i nauk biologicznych (ekologii, ochrony środowiska, botaniki, mikrobiologii). Do badań

wybrano czasopisma figurujące w części „B" wykazu MNiSW za 2015 r., które otrzymały pięć lub więcej punktów MNiSW (zob. Aneks 1). Zastosowano ponadto ograniczenie dla zebranych danych o cytowaniach literatury przywoływanej na łamach czasopism wyszczególnionych w Aneksie 1. Dane empiryczne zawężono mianowicie do wyłącznie tych, które obejmują lata 2004–2014 z uwagi na fakt, że wykorzystana w ramach prezentowanego modelu parabola jest funkcją symetryczną w stosunku do swojego wierzchołka (tzw. ogniska paraboli). Inaczej mówiąc, modelowanie krzywej cytowań za pomocą paraboli nie powinno uwzględniać całego etapu narastania cytowalności publikacji (zob. Rys. 1; Aneks 3), a w szczególności jego wczesnego stadium, w którym cytowania są nieliczne. Wydaje się, że aby modelowanie funkcją paraboliczną odpowiadało rzeczywistemu przebiegowi zjawiska kumulowania się cytowań, należy ograniczyć zasięg danych wejściowych tylko do tego etapu cyklu, który zachowuje się jak typowa parabola, tj. cechuje się najwyraźniej przebiegiem rosnąco-malejącym (zob. Aneks 3).

5. Wyniki badań

Ogólna postać paraboli w opisie tendencji rozwojowej zjawisk ilościowych wyraża się następującym równaniem (Sobczyk, 2008, 57):

$$y_t = a_0 + (a_1 \times t) + (a_2 \times t^2) + \varepsilon_t, (a_2 \neq 0, t = 1, 2, 3, ..., n),$$

gdzie „a_0", „a_1" i „a_2" to nieznane (szacowane na podstawie danych empirycznych) parametry strukturalne paraboli, „n" jest długością szeregu czasowego, „ε_t" jest tzw. składnikiem losowym, zaś „y_t" to wartość zjawiska (w tym przypadku liczba cytowań zbioru publikacji) w momencie „t". Zmienną „y_t" można też określić jako analityczną postać trendu. Wartość „n" w rozpatrywanym przypadku wynosi 11 (szereg liczb cytowań publikacji z lat 2004–2014 zawiera 11 wyrazów). Prognozę punktową uzyskuje się podstawiając do wzoru, z oszacowanymi wartościami „a_0", „a_1" i „a_2", wartości „t" równe „n+1", „n+2" itd., czyli wartości momentów czasu, dla których nie dysponuje się już danymi empirycznymi. Parametry strukturalne paraboli wyznacza się na podstawie następującego równania macierzowego[2] oraz przy przyjęciu takiej numeracji jednostek czasu „t", która spełnia warunek $\sum t = 0$ (Sobczyk, 2008, 57–60):

$$a = (T^T T)^{-1} \times T^T \times y, \text{ gdzie: } T = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2^2 \\ \vdots & \vdots & \vdots \\ 1 & n & n^2 \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, a = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}$$

[2] Podstawy rachunku macierzy, takie jak mnożenie macierzy przez liczbę i przez macierz, transpozycję macierzy czy macierze odwrotne omówiono np. w podręczniku Tadeusza Trajdosa (2013) lub Włodzimierza Szkutnika i Agnieszki Przybylskiej-Mazur (2008). W Internecie dostępne są ponadto kalkulatory wykonujące działania na macierzach. Jeden z nich zamieszczony jest np. pod adresem: http://obliczone.pl/start/kalkulatory/924-mnozenie-macierzy-kalkulator

Korzystając z danych zamieszczonych w Aneksie 2 uzyskujemy następujące zależności:

$$y = \begin{bmatrix} 150 \\ 143 \\ 152 \\ 166 \\ 169 \\ 175 \\ 204 \\ 179 \\ 170 \\ 168 \\ 152 \end{bmatrix}, T = \begin{bmatrix} 1 & -5 & 25 \\ 1 & -4 & 16 \\ 1 & -3 & 9 \\ 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix}$$

$$T^TT = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -5 & -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 & 5 \\ 25 & 16 & 9 & 4 & 1 & 0 & 1 & 4 & 9 & 16 & 25 \end{bmatrix} \times \begin{bmatrix} 1 & -5 & 25 \\ 1 & -4 & 16 \\ 1 & -3 & 9 \\ 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix} \times \begin{bmatrix} 11 & 0 & 110 \\ 0 & 110 & 0 \\ 110 & 0 & 1958 \end{bmatrix}$$

Wyznacznik macierzy „T^TT", oznaczany symbolem „$\det(T^TT)$", jest w tym przypadku równy 1038180. Macierz odwrotna do macierzy „T^TT", oznaczana symbolem „$(T^TT)^{-1}$" jest iloczynem odwrotności wyznacznika macierzy „T^TT" oraz transponowanej macierzy dopełnień algebraicznych dla macierzy „T^TT". Macierz odwrotna do „(T^TT)" ma więc postać:

$$(T^TT)^{-1} = \frac{1}{1038180} \times \begin{bmatrix} 215380 & 0 & -12100 \\ 0 & 9438 & 0 \\ -12100 & 0 & 1210 \end{bmatrix} = \begin{bmatrix} 0.20746 & 0 & -0.01166 \\ 0 & 0.00909 & 0 \\ -0.01166 & 0 & 0.001166 \end{bmatrix}$$

Iloczyn macierzy $T^T \times y$ wynosi:

$$T^Ty = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -5 & -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 & 5 \\ 25 & 16 & 9 & 4 & 1 & 0 & 1 & 4 & 9 & 16 & 25 \end{bmatrix} \times \begin{bmatrix} 150 \\ 143 \\ 152 \\ 166 \\ 169 \\ 175 \\ 204 \\ 179 \\ 170 \\ 168 \\ 152 \end{bmatrix} = \begin{bmatrix} 1828 \\ 225 \\ 17177 \end{bmatrix}$$

Wektor ocen parametrów strukturalnych paraboli ma wobec tego postać:

$$a = (T^T T)^{-1} \times T^T y = \begin{bmatrix} 0.20746 & 0 & -0.01166 \\ 0 & 0.00909 & 0 \\ -0.01166 & 0 & 0.001166 \end{bmatrix} \times \begin{bmatrix} 1828 \\ 225 \\ 17177 \end{bmatrix} = \begin{bmatrix} 178.95 \\ 2.045 \\ -1.286 \end{bmatrix}$$

Parabola wyraża się zatem następującym wzorem:

$y_t = 178.95 + (2.045 \times t) - (1.286 \times t^2) + \varepsilon_t$, przy zachowaniu warunku $\sum t = 0$.

Precyzję przybliżenia tendencji rozwojowej badanego zjawiska (w tym przypadku procesu zmian cytowalności publikacji) za pomocą otrzymanej funkcji, szacuje się za pomocą tzw. współczynnika zmienności resztowej (V_e), współczynnika zbieżności (φ^2), współczynnika determinacji (R^2) oraz średniego błędu oceny parametrów strukturalnych ($D(a_0)$, $D(a_1)$ i $D(a_2)$) (Krzysztofiak & Luszniewicz, 1976, 362–369; Sobczyk, 2008, 45–47, 61–64; Sobczyk, 2015, 339–342; Zeliaś et al., 2013, 78–81). Współczynnik V_e informuje, jaką częścią średniego poziomu analizowanego zjawiska są przypadkowe odchylenia danych od teoretycznej funkcji trendu (paraboli). Parabola tym bliżej odpowiada danym doświadczalnym, im wartość V_e jest bliższa zeru. Współczynnik φ^2 określa, jaka część ogólnej zmienności wartości „y" (w tym przypadku zaobserwowanych liczb cytowań) nie została wyjaśniona przez uzyskaną funkcję trendu. Współczynnik R^2 – przeciwnie, wskazuje jaka część zmienności y została wyjaśniona przez funkcję trendu (wartość $R^2 = 1 - \varphi^2$). Średnie błędy szacunku parametrów strukturalnych mówią z kolei o tym, o ile średnio mylimy się szacując te parametry w oparciu o wyniki próby losowej (czyli zebrane dane empiryczne). Im mniejsze błędy szacunku, tym lepiej funkcja opisuje zebrane dane. Przyjmuje się, że błędy przekraczające 50% przekreślają wartość poznawczą otrzymanych ocen. Wartości liczbowe tych błędów zapisuje się w nawiasach pod odpowiednimi parametrami strukturalnymi równania, które jest w tym przypadku równaniem paraboli (Sobczyk, 2008, 46–47).

Tabela 1. Wyznaczone wartości współczynników określających stopień precyzji dopasowania paraboli do danych doświadczalnych

L.p.	Nazwa współczynnika	Wartość współczynnika
1	Współczynnik zmienności resztowej (V_e)	7.34%
2	Współczynnik zbieżności (φ^2)	0.41 (41%)
3	Współczynnik determinacji (R^2)	0.59 (59%)
4	Średni błąd oceny parametru strukturalnego a_0 paraboli ($D(a_0)$)	5.55
5	Średni błąd oceny parametru strukturalnego a_1 paraboli ($D(a_1)$)	1.16
6	Średni błąd oceny parametru strukturalnego a_2 paraboli ($D(a_2)$)	0.42
7	Odchylenie standardowe składnika resztowego paraboli (S_e)	12.19

Aby uniknąć przedstawiania wszystkich obliczeń, których również dokonuje się w zapisie macierzowym i które zajęłyby dużo miejsca, poniżej podano jedynie wartości wskaźników

wyliczone w oparciu o dane empiryczne[3] przedstawione w Aneksach 2 i 3. Przejrzystym sposobem prezentacji wydaje się zebranie ich w tabeli (zob. Tab. 1).

Krótkoterminowa prognoza punktowa dla lat T=2015, T=2016, T=2017, T=2018 i T=2019 jest skonstruowana przy użyciu wskazanego powyżej równania paraboli. Np. dla roku 2015, do równania podstawia się w miejsce zmiennej „t" wartość „6", ponieważ rok 2015 jest rokiem szóstym z kolei, jeżeli zachowany jest warunek „$\sum t = 0$ ". Bezwzględny błąd *ex ante* prognozy to z kolei oczekiwana różnica między wartością prognozowaną a rzeczywistą realizacją zmiennej prognozowanej (wartością empiryczną zjawiska lub pewnej jego cechy, którą się prognozuje, w okresie nadchodzącym, tj. takim, za który nie dysponuje się już danymi doświadczalnymi). O błędzie „bezwzględnym" mówi się kiedy jednostka, w jakiej wyraża się błąd, jest tą samą jednostką, w której podaje się wartości zjawiska (w tym przypadku jest nią liczba cytowań). O błędzie „względnym" – kiedy jego wartość jest niemianowana (procentowa). Wartość błędu względnego pozwala ocenić, czy prognoza jest dopuszczalna. Jeżeli błąd względny przekracza 10% prognozę kwalifikuje się na ogół jako niedopuszczalną, chociaż przedział ten (0% – 10%) może być uznany za relatywny. Duże znaczenie ma tutaj bowiem rodzaj, znaczenie i interpretacja zmiennej prognozowanej (Krawiec, 2014, 10; Sobczyk, 2008, 27–28; Zeliaś, 1997, 71–72; Zeliaś et al., 2013, 49–50). Na podstawie odpowiednich wzorów (Sobczyk, 2008, 48–49, 58, 63) uzyskano następujące wartości prognozowane oraz wartości błędu *ex ante*, jakim są one obarczone (zob. Tab. 2).

Tab. 2. Przewidywane liczby cytowań dla lat z przedziału 2015–2019 oraz szacunek ich niepewności

Rok, dla którego obowiązuje prognoza	**Wartość „t"**	**Prognozowana wartość liczby cytowań publikacji**	**Bezwzględny błąd *ex ante* prognozy punktowej**	**Względny błąd prognozy punktowej**	**Empirycznie stwierdzona wartość liczby cytowań publikacji (tylko rok 2015)**
2015	6	145 (+/ – 12)	18.1	12.49%	109
2016	7	134 (+/ – 12)	22.18	16.55%	
2017	8	113 (+/ – 12)	27.46	24.3%	
2018	9	93 (+/ – 12)	33.84	36.38%	
2019	10	71 (+/ – 12)	41.24	58.08%	

Jak wynika z tabeli 2, tylko prognozy na lata 2015–2016 wydają się akceptowalne przy założeniu, że można tu zastosować nieco wyższy (od dziesięcioprocentowego) próg dopuszczalności. Przyczyną takiego posunięcia jest przede wszystkim fakt istnienia pewnej płynności granicy dopuszczalności prognoz, tj. uwarunkowanie jej konkretnym kontekstem praktycznym i zależnością wymaganej precyzji prognozy od indywidualnych potrzeb badacza i charakteru objaśnianych zmiennych (Sobczyk, 2008, 27–28; Zeliaś, 1997,

[3] Na życzenie osób zainteresowanych szczegółowe wyliczenia mogą zostać przesłane na wskazany adres mailowy po skontaktowaniu się z pierwszym z autorów niniejszego artykułu. Ten sam komentarz odnosi się również do przedstawionych poniżej wartości wskaźników oceny dokładności prognoz.

71–72; Zeliaś et al., 2013, 49–50). W ramach teorii zachowań w zakresie cytowań jako determinantów procesów decyzyjnych leżących u podstaw samego zjawiska wstawiania przypisów bibliograficznych często podnosi się kwestie z pogranicza psychologii i nauk społecznych (zob. np. Bornmann, Daniel, 2008), co przybliża niektóre aspekty bibliometrycznego prognozowania cytowalności do ogólnego obszaru nauk społecznych, w których osiągalny poziom dokładności szacunków i statystyk wydaje się być z konieczności niższy niż w naukach ekonometrycznych, ścisłych i przyrodniczych.

Ponadto warto podkreślić, że, ponieważ autorzy dysponują empiryczną wartością liczby cytowań za rok 2015, można stwierdzić, że prognoza jest tylko nieznacznie niedoszacowana. To znaczy, prognozowane 145 cytowań pomniejszone o wartość względnego błędu prognozy (145 – (12.5% × 145) = 145 – 18.125 = 126.875) i ponownie pomniejszone o szacowaną wartość składnika losowego (126.875 – 12.19 = 114.685) daje w wyniku liczbę 115 cytowań, co przekracza wartość empiryczną tylko o sześć cytowań. Należy tu dodatkowo zaznaczyć, że w tym przykładzie niedoszacowanie może wynikać z faktu, że liczba publikacji wydanych i zarazem cytowanych w 2015 r. jest przypuszczalnie nieco zaniżona ze względu na to, że nie wszystkie prace wydane w tym roku były dostępne dla autorów cytujących w trakcie przygotowywania artykułów. Jeżeli np. artykuł powstawał w pierwszych miesiącach 2015 r., dokumenty wydane pod koniec tego roku nie miały już szans na wejście w skład jego bibliografii załącznikowej. Dokładnie z tego samego powodu z zakresu danych wejściowych dla konstrukcji modelu wykluczono liczby cytowań zaobserwowane dla prac wydanych w 2015 r.

Dodatkową ewentualnością, istniejącą w ramach metodologii statystycznej, jest wyznaczenie prognoz przedziałowych zamiast prognoz punktowych. Ponieważ jednak w tym przypadku konieczna jest znajomość rozkładu zmiennej prognozowanej, o którym zazwyczaj zakłada się, że jest on tzw. rozkładem normalnym (rozkładem Gaussa – zob. np. Sobczyk, 2008, 64; Zeliaś et al., 2013, 45), możliwość ta nie zostanie wykorzystana w ramach niniejszego opracowania. Rozkład liczby cytowań publikacji o różnym wieku jest bowiem wyraźnie skośny (zob. Aneks 3) i założenie o normalności takiego rozkładu byłoby nieuprawnione. Zagadnienie zbadania charakteru tego rozkładu, chociaż wydaje się istotne z teoretycznego i praktycznego punktu widzenia, wykracza poza ramy niniejszej pracy i zasługuje na osobne opracowanie. Podobnie rzecz ma się ze zjawiskiem tzw. autokorelacji składnika resztowego (zob. np. Krzysztofiak & Luszniewicz, 1976, 390–391), które – jeżeli istnieje – zakłóca model wahań w czasie i utrudnia prognozowanie. Wymaga ono sformułowania odpowiedniej hipotezy zerowej i zbadania jej przy pomocy specjalnej statystyki testowej[4], co również jest kwestią pozostającą poza nawiasem głównego celu niniejszej pracy.

6. Interpretacja wyników i wnioski

Autorzy niniejszej pracy proponują następującą interpretację wartości prognozowanych liczb cytowań, które uzyskano na podstawie analizy synchronicznej. Przede wszystkim z naturalnych przyczyn wartości te nie mogą być interpretowane jako liczby cytowań pojawiające się w tym samym roku, którego dotyczą zebrane dane doświadczalne (tj. w 2015 r.). W 2015 r. nie można bowiem cytować prac, które dopiero zostaną opublikowane. Z drugiej

[4] Jest to tzw. statystyka Durbina-Watsona.

strony, ostatnią wartością szeregu czasowego dostępną podczas badania była liczba cytowań publikacji wydanych i zarazem cytowanych w 2015 r. W związku z tym wydaje się, że znaczenie wartości prognozowanych (tzn. liczb cytowań dla lat 2015–2019) można wyrazić analogicznie, jako liczby cytowań prac zarazem opublikowanych i cytowanych w latach 2015–2019. Przykładowo, liczba 134 cytowań dla 2016 r. (zob. Tabela 2) oznaczałaby więc, że w 2016 r. w dyscyplinie usytuowanej na pograniczu nauk o Ziemi i nauk biologicznych, zostaną zacytowane 134 publikacje, których rokiem wydania jest ten sam 2016 r. Byłby to wobec tego swego rodzaju uproszczony indeks natychmiastowości dyscypliny (ang. *immediacy index*) określający, jak szybko badacze w tej dyscyplinie nauki reagują na najnowsze informacje i publikacje dyscyplinarne, jak szybko dostrzegają i zapoznają się z najaktualniejszymi pracami i włączają je w nurt własnych dociekań. Im szybciej spadałby tak rozumiany indeks (wartości prognozowane dla kolejnych lat), tym wolniej postępowałby rozwój dyscypliny. Konieczna jest tutaj akceptacja wspomnianego już założenia o pewnej stabilności rozwoju dyscypliny. Innymi słowy, koncepcja indeksu natychmiastowości dałaby się utrzymać przy przyjęciu, że tempo rozwoju w dyscyplinie nie zmienia się dowolnie z roku na rok. Tylko wtedy można byłoby domniemywać, że wartości indeksu uzyskane w oparciu o dane np. za 2015 r. będą obowiązywać przynajmniej w dwóch-trzech kolejnych latach, o ile w dyscyplinie nie nastąpi żaden wyraźny przełom, który wywołałby np. dezaktualizację większości treści opublikowanych (i akceptowanych przez społeczność naukowców) przed 2015 r.

Z powyższej analizy wypływają wnioski, które można zawrzeć w następujących stwierdzeniach.

Po pierwsze, statystyczna metodologia prognostyczna wydaje się znajdować zastosowanie w ramach bibliometrycznej analizy tempa rozwoju dyscyplin naukowych. Świadczą o tym akceptowalne wskaźniki zgodności dopasowania trendu teoretycznego do wartości rzeczywistych (zob. Tab. 1), stwierdzonych doświadczalnie w oparciu o stosunkowo dużą próbę losową.

Po drugie, w przypadku przebadanej próby dopasowanie funkcyjne wielomianu drugiego stopnia (paraboli) do danych doświadczalnych cechowało się wprawdzie dopuszczalnym poziomem dokładności, jednak wartości prognozowane w okresie dłuższym niż dwuletni (tj. dla lat 2017–2019), były obarczone znacznymi błędami.

Po trzecie, z uwagi na powyższy fakt należy uznać za wskazane rozpatrzenie możliwości dalszego dopracowania metody prognostycznej. Mogłoby ono przybrać postać analizy i określenia typu rozkładu zmiennej prognozowanej oraz skonstruowania prognoz przedziałowych dla wybranych lat (tzw. przedziały ufności). Alternatywnie, metoda mogłaby zostać zmodyfikowana poprzez zbadanie większej próby losowej, poprzez wybór próby losowej na podstawie cytowań literatury w innej niż rozpatrzona dyscyplinie czy poprzez wybór innych (dłuższych albo krótszych) zakresów danych wejściowych, jako podstawy sformułowania równania opisującego trend rozwojowy. W odniesieniu do tej ostatniej ewentualności należy jednak nadmienić, że autorzy przetestowali dokładność dopasowania funkcji parabolicznej do kilku szerszych przedziałów czasowych obejmujących zakres danych wejściowych służących sformułowaniu równania paraboli. Testy te wykazały, że przyjęty ostatecznie przedział jedenastoletni (2004–2014) przełożył się na najwyższą precyzję dopasowania, która została zmierzona za pomocą współczynników V_e, φ^2 i R^2.

Po czwarte, próby udoskonalenia przyjętej metodologii mogłyby przebiegać również w kierunku przyjęcia innej niż parabola, funkcji spełniającej rolę predykcyjną. Ponieważ

jednak typowy charakter cyklu życia publikacji naukowych (zob. Rys. 1) posiada swój specyficzny kształt (przebieg rosnąco-malejący) wydaje się, że inne typowe dla statystycznych analiz trendów rozwojowych funkcje (tj. funkcja liniowa, wykładnicza, potęgowa, logistyczna i hiperboliczna) nie spełniłyby się dobrze w roli predyktorów. Należałoby wobec tego poszukiwać innych prawidłowości funkcyjnych, np. za pomocą specjalistycznego oprogramowania w rodzaju pakietu „Origin" lub „MATLAB".

Po piąte należy zauważyć, że obiecującą ewentualnością wydaje się w tym kontekście możliwość zastosowania innej, niż przyjęta w ramach niniejszej pracy, organizacji danych empirycznych. To znaczy, że dane te można zagregować w sposób skumulowany wzorem Roberta Burtona i Richarda Keblera (1960) i podjąć próbę dopasowania zupełnie innej funkcji do tak ustrukturyzowanych danych. Kumulacja liczb cytowań (również w ujęciu synchronicznym) oznacza tutaj sumowanie liczb cytowań prac wydanych w kolejnych latach. Np. cytowania prac wydanych w 1950 r., które pojawiły się np. w roku 2015 dodaje się do liczby cytowań prac wydanych w 1951 r. (i cytowanych w tym samym roku 2015). Następnie otrzymaną sumę powiększa się o liczbę cytowań prac ogłoszonych w 1952 r. itd. Funkcja prognostyczna mogłaby w tym ujęciu przewidywać przyszłą liczbę cytowań skumulowanych, z której możliwe byłoby wyznaczenie liczby cytowań nieskumulowanych (tj. dotyczących tylko prac wydanych w jednym, wybranym roku). Funkcja Burtona i Keblera (1960) była jednak bardzo specyficzna, wymagała wstępnego wyznaczenia dwóch parametrów (stałych) dla każdej poszczególnej dyscypliny i w niczym nie przypominała typowych funkcji wykorzystywanych w ramach prognozowania statystycznego. Wydaje się w związku z tym, że konstrukcja takiej funkcji byłaby zadaniem znacznie trudniejszym.

Po szóste, zaproponowany indeks natychmiastowości opisujący tempo rozwoju dyscyplin mógłby znaleźć praktyczne zastosowania przy akceptacji założenia o tzw. normalnym, bieżącym rozwoju wybranych dyscyplin w sensie, w jakim o rozwoju tym wypowiadał się Kuhn (2001). Wydaje się, że indeks ten w szczególności mógłby znaleźć zastosowania porównawcze. Jego charakter i jednostka, w jakiej wyrażane są jego wartości sprawiają bowiem, że jest on predysponowany do zastosowań porównawczych w większym stopniu, niż do zastosowań polegających na ocenie tempa rozwoju indywidualnych dyscyplin (lub obszarów problemowych) funkcjonujących w obrębie wybranych dziedzin współczesnej nauki.

Po siódme, w przypadkach, w których do danych doświadczalnych, cechujących się dużą nieregularnością, nie udałoby się zadowalająco dopasować żadnej funkcji matematycznej (takiej jak np. parabola lub funkcja wielomianowa stopnia wyższego niż drugi), możliwe byłoby wykorzystanie prognostycznych metod indeksowych lub tzw. mechanicznych metod dekompozycji szeregów czasowych. Polegają one na badaniu średniego tempa zmian poziomu zjawiska w czasie przy użyciu m. in. średniej geometrycznej, średniej ruchomej („kroczącej") prostej lub scentrowanej, wag harmonicznych albo prostego bądź adaptacyjnego wygładzania (wyrównywania) wykładniczego. Modele wygładzania adaptacyjnego to przede wszystkim wygładzanie Browna, Holta, Wintersa oraz model tzw. trendu pełzającego, zwanego też segmentowym (zob. np. Aczel, 2007, 645, 647; De Gooijer & Hyndman, 2006; Gelper et al., 2007, 285–286; Hirschey, 2009, 219–222; Krawiec, 2014; Snarska, 2011, 212–244; Sobczyk, 2008, 97–98, 138; Yaffee & McGee, 2000, 23–44; Zeliaś et al., 2013, 140–153).

Aneks 1

Wykaz objętych badaniami czasopism wydawanych w języku polskim, należących do zakresu z pogranicza nauk o Ziemi i nauk biologicznych (ekologii, ochrony środowiska, botaniki, mikrobiologii), występujących w części „B" wykazu czasopism naukowych MNiSW z 2015 roku. Online: Figshare repository, [21.12.2017], https://figshare.com/s/5a746333fb21f9e9a4e3

L.p.	Tytuł	ISSN	Punkty	Tematyka*	Dziedzina	Język publikacji
1.	Soil Science Annual	2300–4967	13	Gleboznawstwo	Nauki o Ziemi / nauki biologiczne	Angielski, polski
2.	Acta Scientiarum Polonorum Formatio Circumiectus	1644–0765	10	Kształtowanie środowiska, ekologia, geologia, hydrologia	Nauki o Ziemi / nauki biologiczne	Polski, angielski
3.	Gospodarka Wodna	0017–2448	9	Hydrologia, hydrogeologia, ochrona środowiska	Nauki o Ziemi / nauki biologiczne	Polski
4.	Landform Analysis	1429–799X	8	Geomorfologia	Nauki o Ziemi / nauki biologiczne	Polski, angielski
5.	Problemy Ekologii Krajobrazu	1899–3850	8	Geoekologia, geografia fizyczna	Nauki o Ziemi / nauki biologiczne	Polski, angielski
6.	Przegląd Przyrodniczy	1230–509X	6	Ochrona przyrody, fizjografia	Nauki o Ziemi / nauki biologiczne	Polski
7.	Zeszyty Naukowe Południowo-Wschodniego Oddziału Polskiego Towarzystwa Inżynierii Ekologicznej z siedzibą w Rzeszowie i Polskiego Towarzystwa Gleboznawczego Oddział w Rzeszowie	1642–3828	6	Gleboznawstwo, ekologia	Nauki o Ziemi / nauki biologiczne	Polski
8.	Acta Geographica Silesiana	1897–5100	5	Geografia, ekologia	Nauki o Ziemi / nauki biologiczne	Polski, rosyjski, angielski
9.	Rocznik Świętokrzyski seria B – Nauki Przyrodnicze	1427–5929	5	Geografia, biologia, ekologia	Nauki o Ziemi / nauki biologiczne	Polski
10.	Studia Limnologica et Telmatologica	1897–645X	5	Ekologia, geografia, geologia	Nauki o Ziemi / nauki biologiczne	Polski, angielski

* Tematyka poruszana na łamach poszczególnych czasopism została tu wskazana na podstawie oceny autora niniejszej pracy dokonanej w oparciu o deklaracje redakcji czasopism, treści publikowanych w nich artykułów i afiliacje autorów artykułów, kategorie tematyczne wyszczególnione dla każdego z tytułów w Polskiej Bibliografii Naukowej (https://pbn.nauka.gov.pl/), jak również w odniesieniu do Rozporządzenia MNiSW z dnia 8 sierpnia 2011 roku w sprawie obszarów nauki, dziedzin nauki i sztuki oraz dyscyplin naukowych i artystycznych.

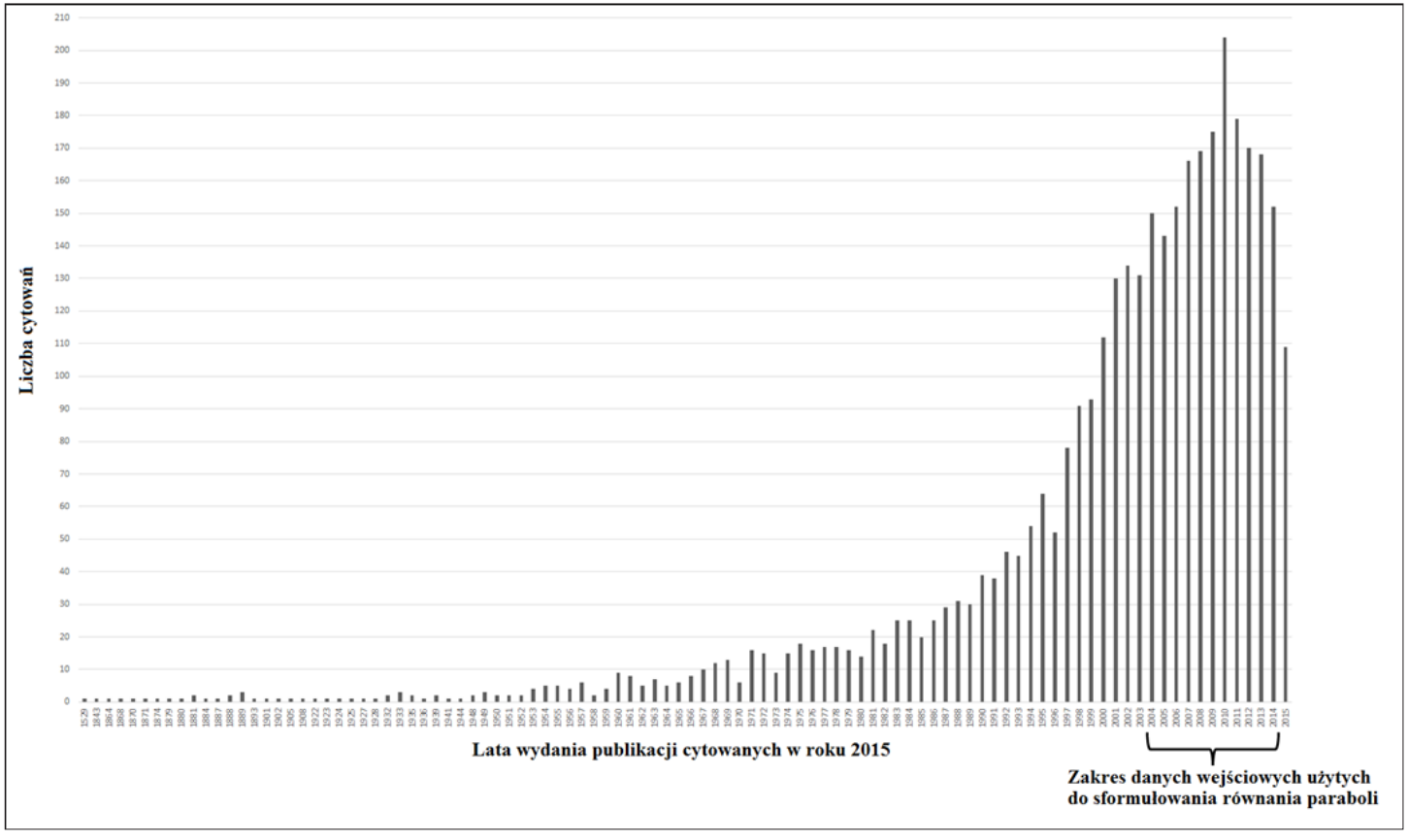
Aneks 2

Cytowania publikacji naukowych, o wszelkich formach wydawniczych, opublikowanych w 2015 r. i latach uprzednich, zaobserwowane w badaniu empirycznym rocznika 2015 czasopism wymienionych w Aneksie 1. Online: Figshare repository, [21.12.2017], https://figshare.com/s/ad0defa403c7dd9300bc

Rok wydania	Liczba cytowań	Rok wydania	Liczba cytowań	Rok wydania	Liczba cytowań	Rok wydania	Liczba cytowań
2015	109	1987	29	1959	4	1901	1
2014	152	1986	25	1958	2	1893	1
2013	168	1985	20	1957	6	1889	3
2012	170	1984	25	1956	4	1888	2
2011	179	1983	25	1955	5	1887	1
2010	204	1982	18	1954	5	1884	1
2009	175	1981	22	1953	4	1881	2
2008	169	1980	14	1952	2	1880	1
2007	166	1979	16	1951	2	1879	1
2006	152	1978	17	1950	2	1874	1
2005	143	1977	17	1949	3	1871	1
2004	150	1976	16	1948	2	1870	1
2003	131	1975	18	1944	1	1868	1
2002	134	1974	15	1941	1	1864	1
2001	130	1973	9	1939	2	1843	1
2000	112	1972	15	1936	1	1529	1
1999	93	1971	16	1935	2		
1998	91	1970	6	1933	3		
1997	78	1969	13	1932	2		
1996	52	1968	12	1928	1		
1995	64	1967	10	1927	1		
1994	54	1966	8	1925	1		
1993	45	1965	6	1924	1		
1992	46	1964	5	1923	1		
1991	38	1963	7	1922	1		
1990	39	1962	5	1908	1		
1989	30	1961	8	1905	1		
1988	31	1960	9	1902	1		

Aneks 3

Graficzna prezentacja rozkładu cytowań publikacji zaobserwowanego w badaniu empirycznym rocznika 2015 czasopism wymienionych w Aneksie 1 (dane z Aneksu 2). Online: Figshare repository, [21.12.2017], https://figshare.com/s/0a620b9e3d37f480623b



Bibliografia

Aczel, A.D. (2007). *Statystyka w zarządzaniu*. Warszawa: PWN.

Bornmann, L., Daniel, H-D. (2008). What Do Citation Counts Measure? A Review of Studies on Citing Behavior. *Journal of Documentation*, 64(1), 45–80.

Boyack, K.W., Klavans, R. (2014). Creation of Highly Detailed, Dynamic, Global Model and Map of Science. *Journal of the Association for Information Science and Technology*, 65(4), 670–685.

Braun, T., Schubert, A., Kostoff, R. (2000). Growth and Trends of Fullerene Reserch as Reflected in Its Journal Literature. *Chemical Reviews*, 100(1), 23–37.

Burrell, Q. (2001). Stochastic Modelling of the First-Citation Distribution. *Scientometrics*, 52(1), 3–12.

Burton, R.E., Kebler, R.W. (1960). The 'Half-Life' of Some Scientific and Technical Literatures. *American Documentation*, 11(1), 18–22.

Chen, Ch. (2006). CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature. *Journal of the American Society for Information Science and Technology*, 57(3), 359–377.

Costas, R., van Leeuwen, T.N., van Raan, A.F.J. (2010). Is Scientific Literature Subject to a 'Sell-By--Date'? A General Methodology to Analyze the 'Durability' of Scientific Documents. *Journal of the American Society for Information Science and Technology*, 61(2), 329–339.

De Gooijer, J.G., Hyndman, R.J. (2006). 25 Years of Time Series Forecasting. *International Journal of Forecasting*, 22(3), 443–473.

Diodato, V. (1994). Dictionary of Bibliometrics. New York: London: Norwood: The Haworth Press.

Diodato, V., Smith, F. (1993). Obsolescence of Music Literature. *Journal of the American Society for Information Science*, 44(2), 101–112.

Ena, O., Mikova, N., Saritas, O., Sokolova, A. (2016). A Methodology for Technology Trend Monitoring: The Case of Semantic Technologies. *Scientometrics*, 108(3), 1013–1041.

Garfield, E. (1985). The Articles Most Cited in the SCI from 1961 to 1982. Another 100 Citation Classics: The Watson-Crick Double Helix Has Its Turn [online]. *Essays of an Information Scientist*, 8, 187–196, [21.12.2017], http://garfield.library.upenn.edu/essays/v8p187y1985.pdf

Garfield, E. (2004). Historiographic Mapping of Knowledge Domains Literature. *Journal of Information Science*, 30(2), 119–145.

Gelper, S., Fried, R., Croux, Ch. (2007). Robust Forecasting with Exponential and Holt–Winters Smoothing. *Journal of Forecasting*, 29(3), 285–300.

Glänzel, W. (2004). Towards a Model for Diachronous and Synchronous Citation Analyses. *Scientometrics*, 60(3), 511–522.

Guo, H., Weingart, S., Börner, K. (2011). Mixed-Indicators Model for Identifying Emerging Research Areas. *Scientometrics*, 89(1), 421–435.

Hirshey, M. (2009). *Fundamentals of Managerial Economics*. Mason, OH: South-Western Cengage Learning.

Huang, M-H., Chang, C-P. (2015). A Comparative Study on Detecting Research Fronts in the Organic Light-Emitting Diode (OLED) Field Using Bibliographic Coupling and Co-Citation. *Scientometrics,* 102(3), 2041–2057.

Kildiszew, G., Frenkel, A. (1976). *Analiza szeregów czasowych i prognozowanie*. Warszawa: PWE.

Krawiec, S. (2014). *Adaptacyjne modele wygładzania wykładniczego jako instrumenty prognozowania krótkoterminowego zjawisk ilościowych*. Gliwice: Wydaw. Politechniki Śląskiej.

Krzysztofiak, M., Luszniewicz, A. (1976). *Statystyka*. Warszawa: PWE.

Kuhn, T.S. (2001). *Struktura rewolucji naukowych*. Warszawa: Fundacja Aletheia.

Levitt, J.M., Thelwall, M. (2008). Patterns of Annual Citation of Highly Cited Articles and the Prediction of Their Citation Ranking: A Comparison Across Subjects. *Scientometrics*, 77(1), 41–60.

Łapkowska-Baster, B. (2009). *Miary współzależności i dynamiki zjawisk w statystyce opisowej. Przykłady i zadania*. Kraków: Wydaw. Uniwersytetu Jagiellońskiego.

Opaliński, Ł. (2017a). Bibliometryczna metodologia prognozowania i oceny rozwoju dyscyplin naukowych. Analiza piśmiennictwa. Część 1: Publikacje pionierskie, metoda powiązań bibliograficznych, metoda współcytowań i metoda współwystępowania specjalistycznej terminologii naukowej. *Zagadnienia Informacji Naukowej. Studia Informacyjne*, 109(1), 34–65.

Opaliński, Ł. (2017b). Bibliometryczna metodologia prognozowania i oceny rozwoju dyscyplin naukowych. Analiza piśmiennictwa. Część 2: badania porównawcze, hybrydowe, statystyczne, analizy dokumentów patentowych, ścieżek rozwoju dyscyplin oraz pozostałe oryginalne podejścia metodologiczne. *Zagadnienia Informacji Naukowej. Studia Informacyjne*, 110(2), 74–106.

Price, D. de Solla (1967). *Mała Nauka – Wielka Nauka*. Warszawa: PWN.

Pritchard, A. (1969). Statistical Bibliography or Bibliometrics? *Journal of Documentation*, 25(4), 348–349.

Purczyński, J. (2010). Wybrane problemy stosowania trendu wielomianowego w prognozowaniu gospodarczym. *Zeszyty Naukowe Uniwersytetu Szczecińskiego. Ekonomiczne Problemy Usług*, 60, 324–339.

Purczyński, J. (2014). Wybrane aspekty prognozowania z wykorzystaniem klasycznych modeli trendu. *Zeszyty Naukowe Uniwersytetu Szczecińskiego. Studia i Prace Wydziału Nauk Ekonomicznych i Zarządzania*, 36(2), 119–130.

Redner, S. (1998). How Popular is Your Paper? An Empirical Study of the Citation Distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*, 4(2), 131–134.

Seglen, P.O. (1992). The Skewness of Science. *Journal of the American Society for Information Science*, 43(9), 628–638.

Shibata, N., Kajikawa, Y., Takeda, Y., Matsushima, K. (2008). Detecting Emerging Research Fronts Based on Topological Measures in Citation Networks of Scientific Publications. *Technovation*, 28(11), 758–775.

Skalska-Zlat, M. (1993). *Bibliometryczne badania rozwoju dyscypliny naukowej*. Seria: Bibliotekoznawstwo XVII. Wrocław: Wydaw. UWr.

Skalska-Zlat, M. (1999). Bibliografia jako odbicie rozwoju nauki. *Zagadnienia Naukoznawstwa*, 1(139), 57–63.

Small, H., Boyack, K.W., Klavans, R. (2014). Identifying Emerging Topics in Science and Technology. *Research Policy*, 43(8), 1450–1467.

Snarska, A. (2011). *Statystyka. Ekonometria. Prognozowanie. Ćwiczenia z Excelem 2007*. Warszawa: Placet.

Sobczyk, M. (2008). *Prognozowanie. Teoria, przykłady, zadania*. Warszawa: Placet.

Sobczyk, M. (2015). *Statystyka*. Warszawa: PWN.

Stinson, E.R., Lancaster, F.W. (1987). Synchronous Versus Diachronous Methods in the Measurement of Obsolescence by Citation Studies. *Journal of Information Science*, 13(2), 65–74.

Szkutnik, W., Przybylska-Mazur, A. (2008). *Algebra liniowa dla studiów ekonomicznych: metody i zadania, algebra macierzy*. Katowice: Wydaw. Uczelniane Akademii Ekonomicznej im. Karola Adamieckiego.

Takeda, Y., Kajikawa, Y. (2009). Optics: A Bibliometric Approach to Detect Emerging Research Domains and Intellectual Bases. *Scientometrics*, 78(3), 543–558.

Trajdos, T. (2013). *Matematyka. Cz. 3, Liczby zespolone, wektory, macierze, wyznaczniki, geometria analityczna i różniczkowa*. Warszawa: WNT.

Tu, Y-N., Hsu, S-L. (2016). Constructing Conceptual Trajectory Maps to Trace the Development of Research Fields. *Journal of the Association for Information Science and Technology*, 67(8), 2016–2031.

Van Dalen, H.P., Henkens, K. (2005). Signals in Science – On the Importance of Signaling in Gaining Attention in Science. *Scientometrics*, 64(2), 209–233.

Wang, Ch-Ch., Ho, Y-S. (2016). Research Trend of Metal-Organic Frameworks: A Bibliometric Analysis. *Scientometrics*, 109(1), 481–513.
Yaffee, R.A., McGee, M. (2000). *Introduction to Time Series Analysis and Forecasting: With Applications of SAS and SPSS*. San Diego: Academic Press.
Zeliaś, A. (1997). *Teoria prognozy*. Warszawa: PWE.
Zeliaś, A., Pawełek, B., Wanat, S. (2013). *Prognozowanie ekonomiczne. Teoria, przykłady, zadania*. Warszawa: PWN.
Żurowska, J. (2005). Zastosowanie modelu trendu wielomianowego do opisu kształtowania się wskaźnika motoryzacji w Polsce w latach 1980–2004. *Transport Miejski i Regionalny*, 12, 15–21.

Application of Statistical Time Series Analysis in Short-Term Forecasting of the Development of Scientific Disciplines

Abstract

Purpose/Thesis: The aim of this paper is to: describe the basics of statistical time series analysis, investigate the possibility of using this approach in short-term forecasting of the development of scientific disciplines based on data showing the citedness of research papers, assess the uncertainty of the forecasts (feature available with this method) and read the results of the research in terms of science studies.

Approach/Methods: The analytical method of time series decomposition was used to isolate so-called development trends based on the observations of the dynamics of mass phenomena. In this case, the mass phenomenon was the citedness of scientific literature in the field of Polish earth science. The method consists of the formulation of an equation which quantitatively describes the course of the phenomenon and distinguishes the main trend from the accidental variability (fluctuations). The equation serves as a keystone in the process of predicting future values of the time series.

Results and conclusions: The results confirm the possibility of an effective application of the approach in short-term forecasting of the scientific discipline development rate. In the case of the research in question, the forecasts proved acceptable as regards two-year horizon of the forecast. However, it is assumed that building the forecast in the form of confidence intervals or slightly modifying the approach could contribute to the lengthening of the horizon. The results were read as a simplified immediacy index, available in particular for comparative applications.

Originality/Value: The authors were unable to find any examples of research which was conducted in a manner presented in this paper. Time series analysis was earlier applied to so-called clusters only, which are a result of dividing disciplinary publications into narrow groups. The groups are usually considered to be representative for certain scientific subdisciplines. Clustering is always done on the basis of one out of many available similarity criteria, in particular the one called a co-word similarity. Furthermore, there is a wide variety of clustering algorithms, which typically yield considerably differing results. As a result, it seems that the approach based on direct citedness of scientific publications helps to eliminate the redundant vagueness and the inevitable lack of universality which appear when one analyzes the results of an algorithmic clustering.

Keywords

Bibliometrics. Development of scientific disciplines. Scientific periodicals. Statistical methods in information science. Time series analysis.

ŁUKASZ OPALIŃSKI uzyskał tytuł magistra w zakresie nauki o języki i komunikacji (jako specjalności wyodrębnionej z kierunku filozofia) na Uniwersytecie Marii Curie-Skłodowskiej w Lublinie w 2005 r. Ukończył studium podyplomowe „Technologie informacyjne i bibliotekoznawstwo" w Wyższej Szkole Informatyki i Zarządzania z siedzibą w Rzeszowie w 2009 r. oraz studium podyplomowe „Bibliotekarz dziedzinowy" w Instytucie Informacji Naukowej i Bibliotekoznawstwa Uniwersytetu Jagiellońskiego w 2012 r. Pracuje w Oddziale Informacji Naukowej Biblioteki Politechniki Rzeszowskiej na stanowisku bibliotekarza. Najważniejsze publikacje: Ł. Opaliński (2012). Niealfabetyczne systemy piśmiennicze a komunikacja językowa w Internecie [online]. Podkarpackie Studia Biblioteczne, 1/2012, 1–19, [25. 02. 2017], http://repozytorium.ur.edu.pl/handle/item/180; Ł. Opaliński (2013). Wybrane aspekty metodologii badań cyklu życiowego publikacji naukowych. Przegląd Biblioteczny, 81(2), 152–171; Ł. Opaliński, M. Jaromin, J. Wikiera (2015). Problem stabilności zachowań naukowców w zakresie cytowań w kontekście metodologii badań starzenia się publikacji naukowych i możliwość jego ujęcia ilościowego. Zagadnienia Informacji Naukowej, 53(2/106), 65–83.

Rola w przygotowaniu artykułu: Opracowanie części teoretycznej, analiza literatury przedmiotu, opracowanie wykresów, tabel i aneksów, zebranie danych empirycznych i interpretacja wyników badania.
Udział: 60%.

Kontakt z autorem:
lopa@prz.edu.pl
Oddział Informacji Naukowej Biblioteki Politechniki Rzeszowskiej
al. Powstańców Warszawy 12
35–959 Rzeszów

MARCIN JAROMIN pracuje na stanowisku asystenta w Zakładzie Biotechnologii i Bioinformatyki Wydziału Chemicznego Politechniki Rzeszowskiej. Tytuł magistra inżynieria uzyskał w 2004 r. na Wydziale Chemicznym Politechniki Rzeszowskiej oraz, równolegle, w 2005 r. na Wydziale Elektrotechniki i Informatyki Politechniki Rzeszowskiej. Specjalizuje się w dziedzinie bioinformatyki i statystyki matematycznej. W dniu 11 stycznia 2012 r. Rada Wydziału Chemicznego podjęła decyzję o otwarciu jego przewodu doktorskiego. Najważniejsze publikacje: B. Dębska, M. Jaromin (2006). Komputerowo wspomagany dobór materiałów inżynierskich w procesie projektowania. W: B. Dębska, G. Fic (red.), Systemy Informacyjne w Chemii. Tom 3. Rzeszów: Oficyna Wydaw. Politechniki Rzeszowskiej, 37–46.; B. Dębska, M. Jaromin, P. Peszko (2006). Nowe metody kształcenia studentów na kierunku Technologia Chemiczna. Przemysł Chemiczny, 85(8–9), 1171–1174; A. Bocian, K. Hus, M. Jaromin, M. Tyrka, A. Łyskowski (2017). Identification of Proteins Differentially Accumulated in Enterococcus Faecalis Under Acrylamide Exposure. Turkish Journal of Biology, 41(1), 166–177.

Rola w przygotowaniu artykułu: analiza statystyczna.
Udział: 40%.

Kontakt z autorem:
mjaromin@prz.edu.pl
Wydział Chemiczny Politechniki Rzeszowskiej
al. Powstańców Warszawy 6
budynek H, pokój H-242
35–959 Rzeszów