

NIEPEWNOŚĆ DANYCH PRZESTRZENNYCH W SYSTEMACH INFORMACJI GEOGRAFICZNEJ (GIS)

Dariusz Felcenloben

Uniwersytet Przyrodniczy we Wrocławiu

Starostwo Powiatowe w Kłodzku

Streszczenie. Złożoność świata rzeczywistego sprawia, że jego obraz reprezentowany w modelu pojęciowym w postaci danych przestrzennych – nie może być wykonany z dowolnie przyjętą dokładnością. Wiedzy odbiorcy towarzyszy bowiem poczucie niepewności, które może mieć charakter stochastyczny, bezpośrednio związany z dokładnością samego pomiaru opisywanego zjawiska lub epistemiczny, wynikający z niepełnej wiedzy odbiorcy informacji, a także ograniczonych możliwości dokonywania pomiarów z oczekiwaną dokładnością. Poczucie niepewności w odbiorze danych geoinformacyjnych nierozłącznie związane jest z koniecznością stopniowego upraszczania reprezentacji poszczególnych jednostek przestrzennych do postaci obrazu przedstawianego systemach GIS. W artykule przedstawiono sposoby definiowania danych przestrzennych w postaci obiektów dyskretnych i pól, z wykorzystaniem metody wektorowej i rastrowej, pojęcia niepewności z tym związanej, jego rodzaje, metody jej szacowania, pojęcie dokładności i precyzji pomiaru, a także techniki modelowania opisywanych obiektów i zjawisk przestrzennych, odwołujące się zarówno do klasycznych analiz statystycznych, jak i probabilistycznych metod, wykorzystujących teorię zbiorów rozmytych.

Słowa kluczowe: niepewność, modelowanie zjawisk przestrzennych, system GIS

WSTĘP

Za twórcę nowoczesnej teorii informacji uznawany jest powszechnie amerykański uczoney Claude E. Shannon, który jest autorem, tzw. *ilościowej teorii informacji* (zwanej także klasyczną czy matematyczną), opartej w swoich założeniach na zdefiniowanym probabilistycznie pojęciu entropii jako miary prawdopodobieństwa wystąpienia danego sygnału [Shannon 1948, Mazur 1970]. Informacja, według Shannona, mierzona jest jako prawdopodobieństwo zajścia określonego zdarzenia i definiowana jako poziom niepewności odbiorcy co do treści przekazywanego komunikatu. Według autora tej teorii – zdarzenia mniej prawdopodobne dostarczają bowiem odbiorcy więcej informacji.

Adres do korespondencji – Corresponding author: Dariusz Felcenloben, Starostwo Powiatowe w Kłodzku, 57-300 Kłodzko, ul. Okrzei 1, e-mail: felcen@powiat.klodzko.pl

Brak wiedzy co do stopnia określoności układu informacyjnego (zbioru danych) sprawia, że informacja w tym ujęciu traktowana jest jako zmienna losowa. Odbiorca informacji pozostaje bowiem w niepewności, jaką wiadomość otrzyma w przyszłości. Miarą stopnia niepewności odbiorcy wiadomości jest właśnie *entropia informacyjna*, którą interpretuje się jako średnią ilość informacji przypadającą na pojedynczą wiadomość, którą należy dostarczyć odbiorcy, aby znikła niepewność wynikająca z prawdopodobieństwa następstwa układu informacji w zbiorze (redundancja, relewancja).

W ujęciu jakościowym pojęcie informacji i niepewności z tym związanej koncentruje się natomiast na semantycznych i pragmatycznych funkcjach środka przekazu i rozumiane jest zarówno w kontekście znaczenia wypowiedzi (komunikatu, zdania), ale także jako element szeroko pojętej wiedzy [Stefanowicz 2004], na zrozumienie której wpływ ma zarówno kontekst sytuacyjny, w jakim jest ona formułowana, a następnie przekazywana, jak i posiadanie przez jej adresata doświadczenia, pozwalającego na wyciągnięcie właściwych wniosków w kontekście rozpatrywanego problemu.

POJĘCIE NIEPEWNOŚCI DANYCH PRZESTRZENNYCH

Przyjęty na etapie pozyskiwania danych przestrzennych stopień generalizacji zapisu poszczególnych encji świata rzeczywistego sprawia, iż dane opisujące obiekty ujawnione w strukturze modelu pojęciowego, obciążone są poczuciem niepewności, wynikającej z braku pewności pozyskanej wiedzy w oczekiwanym zakresie, ograniczonej precyzji ich pomiaru, a także uzyskanych na tej podstawie wyników analiz. Złożoność opisywanych zjawisk, ale także istniejące ograniczenia techniczne i ekonomiczne powodują, iż niemożliwym lub nieuzasadnionym jest ich przedstawianie z dowolną, w domyśle z największą, dokładnością. Potrzeba generalizacji danych, wynikająca z konieczności uproszczenia projektowanego modelu conceptualnego powoduje, iż niepewność z tym związana stanowi nieodłączną cechę systemów geoinformacyjnych. Niemożliwym jest bowiem, co do zasady, przedstawienie w projektowanym systemie nieskończenie dokładnej reprezentacji poszczególnych encji świata rzeczywistego, w którym oprócz rejestrowanych widocznych obrazów ujawnione byłyby wszelkie istniejące relacje pomiędzy nimi i zależności z szeroko rozumianym otoczeniem, a także zachodzące w nim zmiany o charakterze dynamicznym.

Niepewność, o której mowa, związana jest zatem ściśle z dokładnością opisu obiektów w systemach geoinformacyjnych, a jej miarą może być różnica pomiędzy światem rzeczywistym a wyobrażeniem odbiorcy o nim. Subiektywna ocena postrzeganych przez odbiorcę własności obiektów przestrzennych prowadzić może bowiem do rozmycia granic określających zasięg ich występowania.

Świadomość odbiorcy co do realnej możliwości opisu dynamicznego świata rzeczywistego za pomocą statycznych modeli numerycznych stanowić może tym samym uogólnioną definicję niepewności danych przestrzennych. Złożoność świata rzeczywistego sprawia, że przestrzenno-czasowy jego opis, w postaci cyfrowej reprezentacji danych w modelu pojęciowym, nie jest obecnie możliwy do zrealizowania z dowolnie określoną dokładnością. Wiedzy odbiorcy, tego rodzaju danych, towarzyszy zatem poczucie niepewności, które może mieć wymiar stochastyczny, wynikający z samego charakteru opisywanego zjawiska lub epistemiczny, u którego źródła leży poczucie braku pełnej wiedzy w zakresie metod zbierania danych, jak i sposobu ich obrazowania w modelu.

Tak definiowane pojęcie „niepewności”, aby mogło być w jakikolwiek sposób mierzalne i stanowić tym samym miarę jakości gromadzonych danych, winno być dodatkowo określone innymi terminami, takim jak niejednoznaczność czy nieokreśloność, które powinny być analizowane w kontekście podstawowych cech ilościowych i semantycznych decydujących o jakości danych, takich jak [Gaździcki 2002]:

- kompletność i poprawność atrybutów oraz geometrii opisywanych obiektów,
- spójność logiczna,
- dokładność położenia,
- dokładność czasowa,
- dokładność tematyczna,
- kompletność i stopień pokrycia itp.

NIEPEWNOŚĆ WYNIKÓW POMIARÓW I ICH REPREZENTACJI W MODELU

Budując model pojęciowy obiektów i zjawisk przestrzennych, przyjmuje się założenie stanowiące uproszczenie problemu, że przestrzeń wypełniona jest obiektami dyskretnymi o wyraźnie określonych *a priori* granicach lub reprezentowana jest w postaci pól stanowiących ciągłą emanację jednostek przestrzennych, wyrażoną w postaci zmiennych obszarów, których wartość może być określona w dowolnym jego punkcie. W przypadku bowiem wielu danych przestrzennych problematycznym jest jednoznaczne zdefiniowanie granic opisywanych obiektów, np. zasięgu występowania klas gleboznawczej klasyfikacji gruntów, naturalnych granic zbiorników wodnych, terenów oznaczanych umownie na mapach w postaci skarp, budowli ziemnych itp.

Dane przestrzenne, reprezentowane w modelu w postaci zdefiniowanych obiektów lub pól, pozyskiwane są do systemów GIS z wykorzystaniem metody rastrowej lub wektorowej, gwarantujących spełnienie warunku logicznej spójności ich zapisu. Sam wybór metody reprezentacji danych modelu w postaci pola lub obiektów dyskretnych determinuje poziom niepewności danych. Niepewność w tym zakresie ustalana jest bowiem odrębnie dla każdej z tych metod.

Metoda wektorowa, w przeciwieństwie do rastrowej, już na etapie tworzenia modelu konceptualnego wymaga od operatora jednoznacznego (cyfrowego) określenia granic opisywanych jednostek przestrzennych, umożliwiając przedstawienie w sposób jawny relacji topologicznych pomiędzy obiektami przestrzennymi [Longley 2008], np. linii brzegowej oddzielającej łąd od terenów pokrytych wodą. W modelu rastrowym natomiast poszczególne encje świata rzeczywistego reprezentowane są w postaci elementarnych komórek w kształcie kwadratu, trójkąta równobocznego lub sześciokąta foremnego tworzących regularną sieć, które nie mają jednak bezpośredniego odniesienia do opisywanych obiektów przestrzennych. W modelu rastrowym położenie obiektu określa się bowiem za pomocą wielu komórek pokrywających opisywaną jednostkę przestrzenną, które podane dopiero procesowi aproksymacji pozwalają na określenie kształtu definiowanych obiektów i zjawisk przestrzennych [Kwiecień 2004]. Aby określić strukturę zapisanego w modelu rastrowym obiektu czy zjawiska przestrzennego, w tym jego granice, niezbędna jest zatem subiektywna interpretacja operatora systemu lub konieczność posiłkowania się w tym zakresie odpowiednimi narzędziami informatycznymi. W modelach rastrowych nie definiuje się bowiem *a priori* poszczególnych obiektów przestrzennych, lecz stosuje

się metodę klasyfikacyjną polegającą na przyporządkowaniu wyodrębnionych komórek (pikseli) do odpowiedniej klasy obiektów. Problem niepewności w procesie klasyfikacji, bez względu nawet na rozdzielczość samego analizowanego obrazu, pojawia się w momencie, w którym operator musi rozstrzygnąć o przynależności piksela o charakterze mieszanym (zwanym mikselem) do więcej niż jednej z klas wyodrębnionych obiektów w zbiorze danych, np. czy dany punkt jest łądem, czy wodą.

Atrybuty opisowe tak arbitralnie ustalonego piksela winny zatem odzwierciedlać jego niejednorodny obraz spektralny, umożliwiając tym samym jego charakterystykę za pomocą jednej z przyjętych zasad, pozwalających na ustalenie, czy opisywany obiekt [Werner 1992, Kwiecień 2004]:

- wypełnia swoim obszarem większą część komórki rastrowej,
- wypełnia w jakiegokolwiek części komórkę rastrową,
- położony jest w centrum komórki,
- jako liniowy przecina komórkę rastrową.

Niepewność w odbiorze danych dotyczyć może zarówno samej geometrii opisywanej jednostki przestrzennej, jak i atrybutów definiujących jej wewnętrzne właściwości czy też zależności, w jakich pozostaje ona z otoczeniem. Z tego też powodu istotnym jest, aby na etapie przeprowadzanych analiz danych przestrzennych dokonać właściwego doboru narzędzi analitycznych do ich opracowania, np. odpowiednich metod statystycznych, technik kartograficznych pozwalających na obrazowanie obiektów rozmytych czy też metod umożliwiających dokonywanie poprawnej ich klasyfikacji.

Brak standaryzacji pojęć stanowiących podstawę formułowanych kryteriów, na podstawie których klasyfikowane są obiekty przestrzenne, jak i różnorodność samego definiowania obiektów (nazwy) powodują, iż ich relacje topologiczne są niejednoznaczne, a tym samym niepewne. Problem ten dotyczy w szczególności przypadków łączenia zbiorów danych pochodzących z różnych systemów informacyjnych prowadzonych przez niezależne od siebie instytucje, które na swój użytek definiują pojęcia stanowiące podstawę dokonywanych klasyfikacji obiektów przestrzennych czy zjawisk im towarzyszących. Poczucie niepewności danych wynikać może także z niewłaściwego doboru wskaźników statystycznych opisujących przedstawiane w systemach GIS zjawiska przestrzenne, odniesione np. do powierzchni ich występowania.

Mając na uwadze specyfikę opisu obiektów przestrzennych, reprezentowanych w modelach pojęciowych w postaci dyskretnych zbiorów wartości charakterystycznych dla nich atrybutów, wyróżnić można w analizowanym kontekście – dane:

- niejednoznaczne – z uwagi na przyjęty sposób definiowania obiektów i ich własności (stanowiących cechy wyróżniające w zbiorze);
- niepełne – reprezentowane przez dane, które nie stanowią wystarczającej ich indywidualnej charakterystyki;
- niepewne – z uwagi na brak pewności dokonanych ustaleń;
- nieprecyzyjne – z uwagi na brak możliwości ustalenia jednoznacznych wartości, jakie mogą przyjmować atrybuty opisywanych obiektów.

DOKŁADNOŚĆ DANYCH PRZESTRZENNYCH

Przedstawiając szeroko rozumiany problem niepewności danych w kontekście dokładności pomiaru, warto zwrócić uwagę, iż pojęcie to należy rozróżnić od innego,

związanego z samą precyzją pomiaru. Uogólniając stwierdzić można, że dokładność ustala zgodność zewnętrzną wyniku pomiaru z wartością rzeczywistą, podczas gdy precyzja określa zgodność wewnętrzną wyników kolejnych pomiarów tego samego obiektu w porównywalnych i powtarzalnych warunkach. Przyjmuje się, że wynik pomiaru jest dokładny, gdy wolny jest od błędów o charakterze systematycznym.

Precyzyjnym określa się natomiast taki wynik pomiaru, którego błędy przypadkowe dążą do minimum. Dokładność pomiaru określa zatem w miarę zbliżenia mierzonej wielkości do prawdziwej lub przyjętej za taką, natomiast precyzja pomiaru ustala stopień skupienia (wzajemnej bliskości) pomiarów tej samej wielkości. W tym ujęciu niepewność może być definiowana jako przedział, wewnątrz którego mieszczą się błędy pomiarowe [Arendarski 2003].

Uwzględniając złożoność opisywanego świata rzeczywistego, pojęcie dokładności danych przestrzennych, ujawnianych w modelu konceptualnym, przedstawić można w kontekście [Gaździcki 2002]:

- *dokładności tematycznej danych* (semantycznej) – odzwierciedlającej wierność utworzenia w modelu pojęciowym opisywanych encji świata rzeczywistego, w postaci odpowiednio zdefiniowanych i sklasyfikowanych obiektów i zjawisk przestrzennych wyrażonych w formie ustalonych atrybutów ilościowych i jakościowych, a także przypisanych im wartości;
- *dokładności położenia* – wyrażonej dokładnością (względną lub bezwzględną) ustalenia współrzędnych punktów wyznaczających położenie opisywanych obiektów przestrzennych w modelu, w odniesieniu do wartości prawdziwych lub uznanych za takie;
- *dokładności czasowej* – pozwalającej na uwzględnienie zachodzących zmian danych przestrzennych w czasie rzeczywistym ich pomiaru, poprawności uporządkowania zdarzeń w odpowiedniej sekwencji czasowej, a także konieczności ich aktualizacji.

ROZDZIELCZOŚĆ DANYCH PRZESTRZENNYCH

Zakładając, że nie istnieje nieskończenie precyzyjny system pomiarowy, przyjąć należy założenie o konieczności definiowania dla każdego indywidualnego przypadku pojęcia precyzji pomiaru odniesionej do wymaganej lub możliwej do uzyskania rozdzielczości obrazu, która stanowić winna element specyfikacji danych, pozwalających określić ich przydatność w odniesieniu do konkretnego zastosowania.

Uogólniając stwierdzić można, że rozdzielczość definiowana może być jako ilość (liczba) szczegółów, jakie można zauważyć i opisać w przestrzeni w odpowiednim czasie lub zakresie tematycznym. Pojęcie rozdzielczości danych przestrzennych rozumiane może być także jako średnia odległość pomiędzy identyfikowalnymi punktami pomiarowymi odniesionymi do zdefiniowanej jednostki powierzchniowej lub jako wektor określający minimalny rozmiar jednostki odwzorowania. Przedstawiając zagadnienie rozdzielczości danych przestrzennych, podkreślić trzeba także, że pojęcie to odróżniać należy od czynności związanych z ustaleniem gęstości (częstotliwości) pobierania próbek pomiarowych, na podstawie których ustalane są wartości danych. Częstotliwość próbkowania (pomiaru) określa bowiem odległości pomiędzy elementami (próbkami), podczas gdy rozdzielczość definiuje ich minimalny rozmiar.

W ujęciu czasowym rozdzielczość określona jest natomiast interwałem czasu, jaki jest niezbędny do pobrania próbki o oczekiwanych właściwościach, podczas gdy częstotliwość pomiaru (gęstość) determinowana jest ustalonym powtarzalnym cyklem pomiarowym (np. raz dziennie).

Rozdzielczość w kontekście tematycznym przedstawiana może być z kolei jako systemowa zdolność rozróżniania elementarnych kategorii danych w hierarchicznym układzie semantycznym.

SPÓJNOŚĆ DANYCH PRZESTRZENNYCH

Z założenia, każda baza danych przestrzennych stanowić powinna logicznie spójny model prawidłowo odwzorowujący fragment opisywanych obiektów i zjawisk świata rzeczywistego. Pojęcie logicznej spójności odnoszone może być w tym kontekście zarówno do wewnętrznej zgodności zgromadzonych danych (spójność wewnętrzna), jak i ich zgodności ze stanem opisywanego świata zewnętrznego (spójność zewnętrzna), stanowiąc tym samym miarę poprawności konstrukcji przyjętego modelu konceptualnego.

W przypadku danych przestrzennych pojęcie spójności analizować można w kontekście [Veregin 1998]:

- *przestrzennym* – odniesionym do topologicznej zgodności zdefiniowanych relacji w bazie danych;
- *czasowym* – wynikającym z ograniczeń reprezentacji, w tym samym miejscu i czasie jedynie jednego zdarzenia przestrzennego;
- *tematycznym* – odniesionym do poprawności (braku sprzeczności, nadmiarowości) zdefiniowanych atrybutów stanowiących podstawę klasyfikacji obiektów i zjawisk w modelu konceptualnym.

Spójność bazy danych rozpatrywana może być na poziomie:

- *semantycznym* – w zakresie zgodności zgromadzonych danych z opisywanym światem zewnętrznym i przyjętymi w modelu ograniczeniami w postaci tzw. węzłów integralności;
- *integralności* – polegającej na zapewnieniu współdziałania pomiędzy zgromadzonymi danymi, w sposób niekolidujący z przyjętymi ograniczeniami, a także konieczności utrzymania ich integralności, w sytuacji zaistniałej awarii systemu.

Baza danych przestrzennych jest zatem spójna, jeżeli ujawnione w niej dane są zgodne ze stanem reprezentowanego fragmentu rzeczywistości, a także kiedy są spełnione wszystkie zdefiniowane ograniczenia w postaci węzłów integralności.

KOMPLETNOŚĆ DANYCH PRZESTRZENNYCH

Pojęcie kompletności bazy danych może być podobnie jak i zagadnienie spójności, dokładności czy rozdzielczości przedstawiane w kontekście:

- *przestrzennym* – w zakresie kompletności danych odnoszących się do określonego terytorium;
- *czasowym* – odniesionym do obiektów ujawnianych w bazie danych w ściśle określonym czasie;

- *tematycznym* – pozwalającym na zdefiniowanie i zgromadzenie danych dotyczących wskazanych dziedzin, obiektów czy zjawisk przestrzennych, umożliwiającym ustalenie, na ile zgromadzone dane i ich wartości zgodne są z ich specyfikacją określoną w modelu koncepcyjnym, przyjętą dla konkretnego zastosowania, ustalającą pożądany stopień uogólnienia i abstrakcji w rozumieniu selektywnej eliminacji zbędnej lub nieistotnej ich części.

Bez względu zatem na przyjęty poziom uogólnienia każda baza danych powinna być uznana za kompletną, jeżeli zgodna będzie z określoną w modelu pojęciowym strukturą, tzn. zawierać będzie wszystkie wymienione tam elementy (obiekty, atrybuty, wartości, relacje).

POJĘCIE NIEPEWNOŚCI W MODELOWANIU OBIEKTÓW I ZJAWISK

Podkreślić należy, że niezależnie od technologicznych ograniczeń związanych z możliwością wiernego odwzorowania świata rzeczywistego w projektowanych modelach przestrzennych, także z powodów natury ekonomicznej pozyskiwanie danych do systemów GIS odbywa się w praktyce z wykorzystaniem metod statystycznych, bazujących na reprezentatywnych dla poszczególnych opisywanych jednostek przestrzennych – próbkach, które stanowią mogą podstawę projektowanego abstrakcyjnego uproszczonego modelu z wykorzystaniem metod interpolacji czy aproksymacji [Moller, Waagepetersen 2003].

Tworząc tego rodzaju bazy danych, zakłada się tym samym istnienie dyskretnych obiektów przestrzennych w postaci pól, charakteryzujących się równomiernym rozkładem badanej zmiennej w zdefiniowanym obszarze lub w postaci wielu dyskretnych elementów o niewyraźnie określonych wartościach. Pole stanowi zatem ciągłą, w jego granicach, reprezentację obrazu opisywanej jednostki przestrzennej, wyrażoną w postaci zmiennej, której wartość może być ustalona w dowolnym jego punkcie [Kusz, Marciniak 2009].

W klasycznym modelu statystycznym dane o charakterze czasowo-przestrzennym zbierane są w wybranych uprzednio reprezentatywnych punktach pomiarowych opisywanej jednostki, a następnie poddawane procesowi interpolacji, w wyniku czego powstaje mapa przedstawiająca charakterystykę mierzonej zmiennej w sposób ciągły. Zakłada się tym samym, iż zmiana wartości zmiennej ma charakter ciągły i liniowy, co niekoniecznie musi się pokrywać ze stanem faktycznym. Niepewność w odbiorze przedstawionego w modelu pojęciowym obrazu przestrzennego wynika zatem z przyjętego założenia ciągłości zmiany wartości zmiennych, które w rzeczywistości mogą charakteryzować się występowaniem punktów nieliniowych, a także braku rozróżnienia punktów, w których dokonano pomiaru rzeczywistych wartości w określonym czasie a punktami, dla których wartość zmiennej określona została w wyniku interpolacji czy aproksymacji.

PROBABILISTYCZNE MODELOWANIE OBIEKTÓW PRZESTRZENNYCH

Cechą charakterystyczną reprezentowanych w modelu pojęciowym poszczególnych obiektów i zjawisk świata rzeczywistego jest ich kognitywna czasowo-przestrzenna nieokreśloność odniesiona zarówno do granic zewnętrznych, zakreślających obszar ich

występowania, jak i wewnętrzna różnorodność, która prezentowana może być w projektowanym abstrakcyjnym modelu pojęciowym w postaci pól losowych [Dalang i in. 2002]. Niepewność, jaka towarzyszy procesowi identyfikacji poszczególnych obiektów przestrzennych, sprawia, iż w procesie przyporządkowania elementów tych zbiorów i określania ich unikatowych właściwości (atrybutów) wykorzystuje się podejście probabilistyczne oparte na subiektywistycznej koncepcji prawdopodobieństwa odniesionej do przybliżonej oceny postrzeganych zjawisk, którym przypisywany jest ustalony wskaźnik pewności [Marciniak 2009].

Odwołując się do teorii logiki rozmytej, zakłada się, że przynależność opisywanego obiektu przestrzennego do wyróżnionej klasy może być częściowa, co oznacza, że granice pomiędzy klasami, jak i samymi obiektami mogą być nieprecyzyjne i niejednoznaczne (rozmyte). Podobnie wyróżnione w zbiorach elementy atomowe definiowanych zjawisk przestrzennych, w szczególności leżące na ich granicach, przynależą do różnych obiektów i klas. Nieostre kryteria wyróżniające poszczególne opisywane obiekty powodują, że w procesie powstawania tego rodzaju opracowań na podstawie tych samych danych, przy ich interpretacji powstawać mogą wątpliwości co do poprawności ich systemowej klasyfikacji. Potrzeba opisanie w systemach GIS zjawisk i obiektów przestrzennych, które ze swej natury mają charakter wieloznaczny i nieprecyzyjny, leży zatem u podstaw uzasadniających konieczność odwołania się do teorii zbiorów rozmytych, które dają możliwość intuicyjnego zdefiniowania niepewności, jaka towarzyszy operatorowi w procesie klasyfikacji tego rodzaju obiektów, w przeciwieństwie do metod bazujących na klasycznej teorii zbiorów, opartej na logice dwuwartościowej, wyrażonej w postaci 0 (*false*) lub 1 (*true*), która nie jest w stanie rozwiązać występujących w praktyce niejednoznaczności, przy przetwarzaniu rzeczywistych danych. W odniesieniu do klasycznej teorii metoda ta rozszerza przedział wartości, jakie zmienne mogą przybierać, np. pomiędzy wartością określoną jako *prawda* (1) lub *falsz* (0) pojawić może się szereg wartości pośrednich zdefiniowanych jako: *prawie prawda*, *w połowie prawda*, *prawie fałsz*, *w połowie fałsz*, itd., które określają stopień przynależności opisywanego elementu do konkretnego zbioru. O wyborze funkcji przynależności, a tym samym kształcie obiektu i jego klasyfikacji w zbiorze, decyduje przy tym czynnik ludzki w postaci „wiedzy eksperckiej” operatora systemu lub np. odpowiednio przygotowane sieci neuronowe „nauczone” na podstawie wcześniej opracowanych danych wzorcowych.

PODSUMOWANIE

Definiując niepewność jako poczucie braku pewności lub braku wiedzy, pamiętać należy, że pojęcie to nie może być postrzegane jedynie jako synonim dokładności pomiaru. Pojęcie „dokładności” zakłada bowiem pewien poziom niezbędnej wiedzy, podczas gdy niepewność wynika raczej z poczucia jej braku.

Zakładając, że dane przestrzenne pozyskane zostały na podstawie wyników pomiarów wybranych i charakterystycznych cech jednostek przestrzennych, przyjąć można, iż w procesie ustalania błędów pomiarowych, a pośrednio i szacowania ich niepewności wykorzystać można zarówno metody ilościowe, jak i jakościowe, pozwalające na statystyczną ocenę jakości danych i przestrzennego rozmieszczenia ich błędów.

Wynik pomiaru stanowiącego podstawę reprezentacji opisywanych obiektów przestrzennych w systemie GIS jest z założenia wartością przybliżoną, różną od prawdziwej, który interpretować można jako przedstawiający pewien przedział ufności lub niepewność pomiaru, wewnątrz którego znajduje się oczekiwana wartość prawdziwa, charakteryzująca opisywane cechy świata rzeczywistego [Lisiecki, Kłysz 2008].

Niepewność definiowana w ujęciu ilościowym i jakościowym odwołuje się nie tylko do poprawności przyporządkowania poszczególnych encji świata rzeczywistego do wyodrębnionych w zbiorze klas obiektów, lecz odnoszona jest także do samej wartości charakterystycznych cech stanowiących podstawę dokonanej klasyfikacji, wyrażonej w przyjętej skali liczbowej lub względnej [Longley 2008].

Pojęcie niepewności postrzegane jako miara niedokładności pomiaru definiowane może być zatem jako parametr charakteryzujący rozrzut mierzonej wielkości i wyrażone w postaci np. odchylenia standardowego rozkładu wyników pomiarów i ich błędów itp.

W przypadku niepewności stochastycznej do modelowania pól losowych wykorzystywana może być teoria *sieci bayesa* [Marciniak 2009], przy zastosowaniu której istnieje możliwość przedstawienia zależności przyczynowo-skutkowych pomiędzy dowolnymi atrybutami opisywanych obiektów czy też pomiędzy sklasyfikowanymi obiektami lub zjawiskami, a także wnioskowanie o rozkładzie prawdopodobieństwa nieznanymi wartościami cech na podstawie wcześniej ustalonych atrybutów „wzorcowych”. W teorii tej niepewność traktowana jest jako atrybut przypisany obserwatorowi, który stanowi miarę jego niewiedzy na temat opisywanego zjawiska.

Niezależnie jednak od zastosowanej metody śledzenia, nadzorowania i szacowania niepewności danych przestrzennych przyjąć należy jako zasadę, że:

- nie istnieją doskonałe, pozbawione błędów dane – każdy wynik przetwarzania danych obciążony jest błędem;
- należy dokonywać oceny danych pod kątem ich przydatności i jakości – poprawnie zdefiniowane zbiory metadanych mogą służyć takiej wstępnej ocenie;
- wynik analizy może być obciążony znacznie większą niepewnością niż dane wejściowe – przetwarzanie danych odbywa się bowiem często według nieliniowych modeli konceptualnych;
- należy wykorzystywać niezależne źródła danych – pozwoli to na ich weryfikację pod względem ich spójności i przydatności do określonego celu;
- należy informować odbiorcę danych o niepewności i stopniu zaufania do danych i analiz dokonanych na ich podstawie – bez takiego komentarza powstać może wrażenie o bezbłędności danych i uzyskanych na tej podstawie wynikach.

PIŚMIENNICTWO

- Arendarski J., 2003. Niepewność pomiarów. Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa.
- Dalang R., Dozzi M., Russo F., 2002. Seminar on Stochastic Analysis. Random Fields and Applications IV, Centro Stefano Franscini. Ascona. Birkhäuser (Progress in Probability).
- Gążdżicki J., 2002. Internetowy leksykon geomatyczny, dostęp: <http://www.ptip.org.pl/>
- Kusz A., Marciniak A., 2009. Niepewność w reprezentacji zjawisk przestrzennych. Inżynieria Rolnicza nr 5(114), 147–154.

- Kwiecień J., 2004. Systemy Informacji Geograficznej. Podstawy. Wydawnictwo Uczelniane Akademii Techniczno-Rolniczej w Bydgoszczy.
- Lisiecki J., Kłysz S., 2007. Szacowanie niepewności pomiaru. Prace Naukowe ITWL, z. 22.
- Longley P.A., Goodchild M.F., Rhind D.W., 2008. GIS. Teoria i praktyka. PWN, Warszawa.
- Marciniak A., 2009. Probabilistyczne modele zjawisk przestrzennych w rolnictwie. Inżynieria Rolnicza nr 5(114), 193–199.
- Mazur M., 1970. Jakościowa teoria informacji. WNT, Warszawa.
- Mooler J., Waagepetersen R.P., 2003. A Statistical and Simulation for Spatial Point Processes. Monographs on Statistics and Applied Probability. Chapman&Hall/CRC Bayesia 2009. Bayesian Network software. <http://www.bayesia.com>.
- Shannon C.E., 1948. A Mathematical theory of communication. Bell System Techn. J., vol. 27, No. 3–4, 1948, [in:] Mazur M., 1970: Jakościowa teoria informacji. Warszawa.
- Stefanowicz B., 2004. Informacja. Warszawa.
- Veregin H., 1998. Pomiar jakości danych i ich oceny, Podstawy programowe NCGIA w GIScience, dostep: <http://www.ncgia.ucsb.edu/giscc/units/u100/u100.html>
- Werner P., 1992. Wprowadzenie do geograficznych systemów informacyjnych. Wydawnictwo Uniwersytetu Warszawskiego.

UNCERTAINTY OF SPATIAL DATA IN GEOGRAPHIC INFORMATION SYSTEMS (GIS)

Abstract. The complexity of the real world leads to a point where the image represented in the conceptual model in the form of spatial data cannot be made with any accuracy adopted. Recipient of knowledge, therefore, is accompanied by a sense of uncertainty, which may be either stochastic, directly related to the accuracy of the measurement described or cognitive, resulting from the incomplete knowledge of recipient of information and limited possibilities of making measurements with the desired accuracy, which may be characteristic for a statistical sample of the larger population of typical and reproducible objects represented in the model. A sense of uncertainty in the reception of geo-information is inextricably linked with the necessity of gradual simplification of representation of different spatial units to a form of image presented in the Geographic Information Systems (GIS). The article presents ways of defining the spatial data in the form of discrete objects and fields, using the vector and raster method, concepts of uncertainty associated with it, its types, methods of estimation, concepts of accuracy and precision of measurement and modeling techniques of described objects and spatial phenomena, referring both to the classical statistical analysis and probabilistic methods using the theory of fuzzy sets.

Key words: uncertainty, model ling of spatial phenomena, GIS system

Zaakceptowano do druku – Accepted for print: 30.09.2010

Do cytowania – For citation: Felcenloben D., 2010. Niepewność danych przestrzennych w systemach informacji geograficznej (GIS). Acta Sci. Pol. Geod. Descr. Terr., 9(3), 3–12.