



Zbigniew Sitek

## MODELE W GEOGRAFICZNYCH SYSTEMACH INFORMACYJNYCH DLA WYBRANYCH ZADAŃ MONITORINGU ROLNICZEGO

### I. Dane dla potrzeb dynamicznego monitoringu rolniczego

Celem uporządkowania pojęć przytaczam za Słownikiem Komputerowym (Chandor i in., 1985) definicje danych i informacji. **Dane** - to ogólne określenie używane do opisu jakiejkolwiek grupy argumentów operacji lub czynników składających się z liczb, znaków alfanumerycznych lub symboli, które oznaczają jakieś warunki, wartości lub stany, tzn. wszystkie wartości i dane opisowe wykorzystywane w programie komputerowym - ale wyłączając z tego sam program. Określenie "dane" jest czasami zestawiane przeciwstawnie z pojęciem "informacji". **Informacja** - niegdyś termin ten oznaczał wiadomości wogóle. Zachowując swoje pierwotne znaczenie, zaczął być używany do określania wiadomości, wiedzy obserwatora o systemie i otoczeniu, w którym dany system funkcjonuje (Słownik, 1985). Informacja jest wynikiem przetwarzania danych, tzn. informację czerpie się z zestawu, analizy bądź podsumowania danych i ich przekształcenia.

Geograficzne systemy informacyjne przetwarzają dane z wykorzystaniem modeli w celu dostarczenia informacji w środowisku zarządzanym komputerem. Dane są faktami świata rzeczywistego.

Można wydzielić dwie kategorie danych: podstawowe i dodatkowe. Do danych podstawowych zalicza się: identyfikatory, dane sytuacyjne (przestrzenne) i dane atrybutowe (nieprzestrzenne). Dane dodatkowe to dane czasowe, parametry jakościowe, wskaźniki, oznaczenia, itp., które są informacjami o faktach i o danych. Są to dane podczas których obserwowano lub mierzono dane podstawowe. Ich znajomość umożliwia przedstawianie i modelowanie dynamicznych zmian zachodzących w świecie lub monitorowanym zjawisku czy procesie.

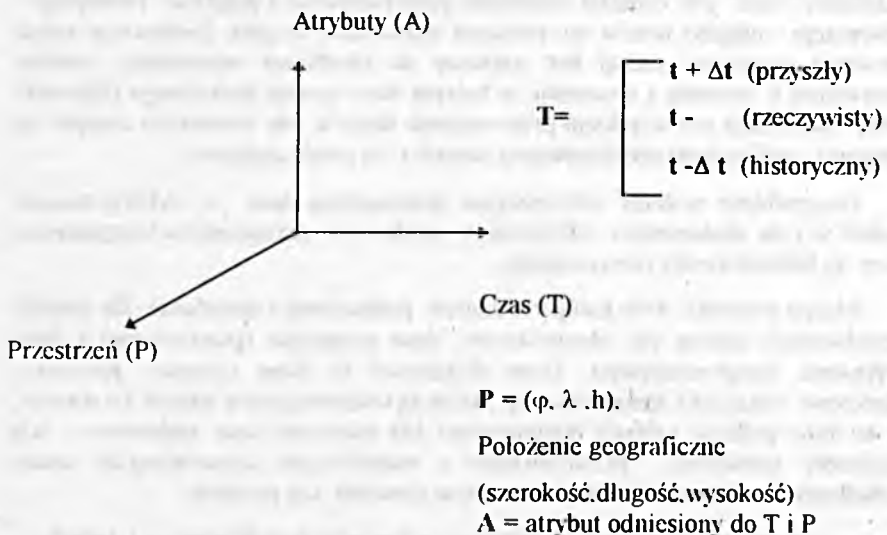
Informacje dla potrzeb monitoringu rolniczego pozyskiwane są z danych:

- k a r t o g r a f i c z n y c h, które przekazują z map informacje o jednostkach administracyjnych, ukształtowaniu terenu itp. Źródłami są mapy regionalnej administracji użytkowania ziemi, mapy gleb, mapy ekologiczne i inne.

- f o t o g r a m e t r y c z n y c h - dostarczają aktualnych danych kartometrycznych i cyfrowych, które mogą także być wykorzystane do kalibrowania danych teledetekcyjnych.
- t e l e d e t e k c y j n y c h - pozyskiwane z różnych systemów satelitarnych takich jak np. Landsat MSS i TM, SPOT, ESA ERS-1, NOAA AVHRR i innych.
- o g ó l n y c h - które zawierają np. dane klimatyczne, dane o gatunkach i rodzajach gleb dane o zaludnieniu, zatrudnieniu, ekonomii, dane o gospodarce w odniesieniu do badanego sektora produkcji, wyniki badań terenowych, statystykę klęsk żywiołowych itp.

Pozyskiwane z tych danych informacje można podzielić na historyczne i aktualne. Celem historycznych baz danych jest udostępnienie użytkownikowi zmian spowodowanych upływem czasu. Wprowadzono dwie definicje czasu: czas światowy i czas systemowy określany inaczej czasem prowadzenia bazy danych (Raafat, Yong, 1992). Można je pokazać na dwóch osiach: na osi czasu światowego zaznacza się zmiany w świecie rzeczywistym, a na osi czasu systemowego - zmiany notowane w bazie danych. Historyczna baza danych ujmuje tylko czas światowy. Czasowa baza danych zawiera zarówno czas światowy jak i czas systemowy.

Wprowadzając trzy rodzaje informacji (przestrzenne-P, atrybutowe-A i czasowe-T) na trzy osie układu współrzędnych (Lj i in., 1992) otrzymuje się pewien rodzaj trójwymiarowej przestrzeni informacyjnej obejmującej te skomplikowane informacje z przestrzeni geograficznej - jak to ilustruje rys. 1.



Rys. 1. Trójwymiarowa przestrzeń informacyjna

## 2. Modele wykorzystywane przez GIS

Istnieje kilkadziesiąt definicji i klasyfikacji modeli jakie wykorzystuje się do rozwiązywania zadań i problemów w różnych dziedzinach nauki i techniki. Według (Słownika..... 1985) najbardziej precyzyjną i ogólną jest definicja oparta na pojęciach *homeomorfizmu i izomorfizmu*. Powstający u obserwatora obraz obiektu badania jest obiektem o takim samym kształcie i przedstawia uproszczone odwzorowanie rzeczywistości. Obserwator abstrahuje przy tym od nieistotnych według niego właściwości obiektu. Warunkiem koniecznym w takim przypadku jest wykorzystanie metody abstrakcji do każdego badania. Kiedy ukształtowany został obraz odwzorowania rzeczywistości, obserwator formułuje właściwy model, tj. abstrakcyjny lub materialny *system izomorficzny* przystający do otrzymanego wcześniej obrazu obiektu. Taki model jest uproszczonym zbiorem charakterystycznych dla obiektu właściwości lub relacji.

W zależności od wyboru środków modelowania (wyboru systemu) mamy do czynienia z *modelem abstrakcyjnym* (konceptualnym) lub *modelem materialnym* (fizycznym).

Modelem abstrakcyjnym może być *model matematyczny* w postaci układu równań matematycznych opisujących cechy charakterystyczne modelowanego obiektu i ich wzajemne zależności. Modele w których występują liczbowe wartości cech charakterystycznych są nazywane *modelami liczbowymi*. Modele zapisane za pomocą wyrażen logicznych nazywa się *modelami logicznymi*.

W systemach monitoringu procesów rolniczych modele odgrywają bardzo ważną rolę. Model jest narzędziem do analizy informacji zebranych w bazach danych oraz podstawą do działania systemu operacyjnego monitoringu. Modele warunkują i uruchamiają procedury manipulacyjne i analityczne, które na podstawie zgromadzonych w geograficznym systemie informacyjnym danych generują nowe informacje. Zatem modele wykorzystuje się w GIS do przetwarzania danych. Mogą one być rozmyślnie upraszczane a niekiedy nawet złe. Rezultaty działania modeli można sprawdzać, ale wymaga to pracy terenowej. W Geograficznych Systemach Informacyjnych można wyróżnić dwa rodzaje modeli: matematyczne i liczbowe (Ramlal, Drummond, 1992).

Modele matematyczne przekształcają i manipulują zmiennymi typu ciągłego (np. współrzędnymi terenowymi X,Y lub geodezyjnymi - szerokością i długością) oraz wartościami stałymi (np. parametrami modelu, współczynnikami przekształcenia lub półosiami elipsoidy). Modelami matematycznymi mogą być równania służące do przekształcania układów współrzędnych lub równania określające zmiany odwzorowań kartograficznych.

Modele matematyczne są często używane w opracowaniach GIS-owskich. Do nich należy także model oceny jakości rezultatów otrzymanych w wyniku manipulacji i przetwarzania danych zgromadzonych w bazach GIS. Dlatego modele matematyczne powinny być na stałe włączone do pakietów systemu zwłaszcza wtedy, gdy GIS generuje nowe informacje, których jakość i dokładność nie można

sprawdzić inaczej niż przez analizę i ocenę rozkładu wariancji danego modelu matematycznego.

Modele liczbowe manipulują wartościami liczbowymi lub cechami, którym przypisano liczby i zgrupowano te cechy w różnych kategoriach. Mogą to być również zasady planowania w rolnictwie lub moduły umożliwiające modelowanie rozprzestrzeniania się zanieczyszczeń pyłowych albo gazowych w atmosferze. Mogą to też być programy pozwalające na modelowanie skażeń i zanieczyszczeń w glebie lub wodzie albo modele zagrożenia ekologicznego bądź ekotoksykologicznego, czy inne modele charakteryzujące różne procesy lub zjawiska.

Modele logiczne, są zazwyczaj wykorzystywane okazjonalnie, powinny być osobnymi modułami GIS, nie instalowanymi na stałe, gdyż są stosowane sporadycznie i dlatego wprowadzane tylko na czas użytkowania.

### 3. Przykłady modeli stosowanych w monitoringu rolniczym

Przedstawione niżej przykłady dotyczą modeli właściwości pastwiskowej w Holandii oraz modeli oceniających produktyjność terenów łąkowych w Chinach. Dane dotyczące właściwości pastwiskowej były wykorzystane w projektach scaleniowych i relokalizacyjnych. Do określania wartości gospodarstw rolnych wykorzystuje się tam kilka modeli, w tym także model oceniający właściwości pastwiskowe gospodarstwa. Przytoczone dane pochodzą z publikacji (Ramlal i Durmond, 1992). Natomiast modele określające produktyjność łąk zaczerpnięto z pracy (Li, Shi i Lin, 1992).

#### 3.1. Model właściwości pastwiskowej

Model służy do określenia trzech klas właściwości pastwiskowej wykorzystując 3 następujące zestawy informacji:

- 1 - stan odwodnienia (drenażu) gleby ujęty w 5 klasach,
- 2 - zdolność wytrzymałościową gleby (warstwy uprawnej), zestawioną w 3 klasach,
- 3 - zdolność dostarczania wilgoci do gleby (podzielonej na 5 klas).

Model liczbowy właściwości pastwiskowej przedstawiono w formie tabelarycznej (trzy klasy właściwości pastwiskowej zestawiono w prostokacie o pogrubionych liniach).

Tablica 1

|                    |   |   |   |   |   |   |   |   |   |   |   |
|--------------------|---|---|---|---|---|---|---|---|---|---|---|
| Stan odwodnienia   | 1 |   | 2 |   | 3 |   | 4 |   |   | 5 |   |
| Wytrzymałość gleby | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 3 | 2 | 3 |

|  |   |   |   |   |   |   |   |   |   |   |   |
|--|---|---|---|---|---|---|---|---|---|---|---|
| Zdolność dostarczenia wilgoci do gleby | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 3 | 2 | 3 |
|  | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 3 | 2 | 3 |
|  | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 |
|  | 4 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
|  | 5 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |

Jak korzystać z tego modelu objaśniają następujące przykłady. Jeśli stan odwodnienia należy do klas 1, 2 lub 3 a wytrzymałość i wilgotność jest w klasie 1 lub 2 wtedy właściwość pastwiskową określa się jako pierwszą (1). Jeśli stan odwodnienia jest w klasie 4, wytrzymałość w klasie 2 a wilgotność w klasie 1 lub 2, to jest to właściwość pastwiskowa druga (2). Jeśli odwodnienie należy do klasy 5, wytrzymałość do 2 lub 3 a wilgotność do klas 3,4 lub 5 to właściwość pastwiskowa należy do klasy trzeciej (3).

Sprawdzenie terenowe określonych tym modelem właściwości pastwiskowych stwierdziło poprawność w 95%.

Oprócz takiego modelu do rozwiązania zadania potrzebne są dane dotyczące gleb rozmieszczonych w rozważanych wielobokach oraz granice działek wraz z charakterystyką glebową i parametrami umożliwiającymi ocenę jakościową zestawionych danych. Muszą one być wprowadzone do bazy danych geograficznego systemu informacyjnego wraz z modelem właściwości pastwiskowej. Korzystając z tych danych i procesów przetwarzania modelu oraz możliwości programowych stosowanego GIS (np. ILWIS) można opracować mapy:

-określające właściwości pastwiskowe w dwóch klasach (z prawdopodobieństwem > 50% i prawdopodobieństwem < 50%),

-określające te właściwości w 3 klasach (z prawdopodobieństwem słabym, średnim i dobrym) lub też

-przedstawiające właściwości pastwiskowe w 5 klasach (z prawdopodobieństwem < 10%, 10-30%, 30-40%, 40-50% i 50-60%).

### Kryteria doboru klas w zestawach informacji

Stan odwodnienia gleby jest uzależniony od wysokości lustra wody a w szczególności od jego najwyższego średniego poziomu (Tablica 2). Sprawdzenie terenowe pozwoliło na określenie odchyłki standardowej poziomu lustra wody na  $\delta = \pm 14$  cm. Korzystając z oszacowania z wykorzystaniem przedziałów ufności można określić prawdopodobieństwo przynależności danej działki do określonej klasy odwodnienia na podstawie pomierzonej wartości poziomu lustra wody. Na przykład, gdy lustro wody jest na głębokości 60 cm, to prawdopodobieństwo, że działka należy do klasy 2 stanu odwodnienia - wynosi 85%. Takiego obliczenia można dokonać na podstawie rozkładu normalnego błędu pomiaru lustra wody dla zadanej głębokości 60 cm i podanej odchyłki standardowej  $\pm = 14$  cm.

Trzy klasy wytrzymałości warstwy uprawnej odniesiono do 5 gatunków gleb i do głębokości zalegania lustra wody. Rodzaj gleby powiązано jak pokazano w Tablicy 3 z zawartością gliny i składników organicznych.

Tablica 2

| Klasy stanu odwodnienia | Poziom lustra wody poniżej powierzchni gruntu (w cm) |
|-------------------------|--|
| 1                       | >80  |
| 2                       | 40-80  |
| 3                       | 25-40  |
| 4                       | 12-25  |
| 5                       | <15  |

Tablica 3

| Rodzaj gleby                 | Zawartość składników organicznych | Zawartość gliny |
|------------------------------|-----------------------------------|-----------------|
| 1. Torf                      | 15-100 %                          | 0-8%            |
| 2. Gлина na podłożu torfowym | 22-70%                            | 8-100 %         |
| 3. Gлина                     | 0-15%                             | 25-100%         |
| 4. Piasek gliniasty          | 0-2.5%                            | 8-25%           |
| 5. Piasek                    | 0-2.5%                            | 0-8%            |

Dla tak określonych rodzajów gleb i głębokości lustra wody zestawiono 3 klasy wytrzymałości warstwy uprawnej. Można je określić z tablicy 4.

Tablica 4

| Rodzaj gleby                 | 1 | 2 | 3 | 4 | 5 |
|------------------------------|---|---|---|---|---|
| Głębokość lustra wody w (cm) |   |   |   |   |   |
| 0-12                         | 3 | 3 | 3 | 3 | 3 |
| 13-24                        | 3 | 3 | 3 | 3 | 2 |
| 25-33                        | 3 | 2 | 2 | 3 | 2 |
| 34-40                        | 2 | 1 | 3 | 2 | 1 |
| 41-60                        | 2 | 2 | 2 | 2 | 1 |
| 61-80                        | 1 | 1 | 2 | 2 | 2 |
| 80-140                       | 1 | 1 | 1 | 2 | 1 |

Ze względu na to, że klasy wytrzymałości warstwy uprawnej określono na podstawie głębokości lustra wody, zawartości składników organicznych i zawartości gliny w glebie, to jakość tych trzech składowych powinna być znana. Zawartość składników organicznych i zawartość gliny sprawdzano doświadczalnie i ich poprawność wynosi 98%.

Prawdopodobieństwo poprawnego określenia głębokości lustra wody (jak podano wyżej) wynosi 85 %. Zatem prawdopodobieństwo określenia poprawnie klas wytrzymałości gleby wynosi:

$$P_{kgw} = 0.85 \times 0.98 \times 0.98 = 0.82 = 82\%$$

**Zdolność dostarczania wilgoci do gleby** jest rejestrowana w milimetrach opadów atmosferycznych i obliczono ją wielomianem o 20 współczynnikach i trzech zmiennych (głębokości korzeniowej, średniej najniższej głębokości lustra wody oraz średniej głębokości wiosennego lustra wody). W taki sposób określono 5 klas zestawionych w Tablicy 5.

Według Holenderskiego Instytutu Badania gleb odchyłka standardowa określania dostarczania wilgoci wynosi 17 mm. Można zatem podobnie jak poprzednio, korzystając z przedziałów ufności, obliczyć prawdopodobieństwo przypisania określonej działce właściwą klasę. Jeśli np. działce na której pomierzono zdolność zaopatrzenia w wilgoć równą 166 mm przypisano klasę 2, to prawdopodobieństwo poprawności tego wynosi 81 %.

Tablica 5

| Klasy dostarczania wilgoci do gleby | Zdolność zaopatrzenia w wilgoć<br>(w mm opadu) |
|-------------------------------------|--|
| 1                                   | >200   |
| 2                                   | 150-200  |
| 3                                   | 100-150  |
| 4                                   | 50-100   |
| 5                                   | <50  |

Posiadając dane oceniające jakość poszczególnych składowych procesu modelowania:

- modelu właściwości pastwiskowych (95%),
- stanu odwodnienia działki (85%),
- wytrzymałości warstwy uprawnej (82%),
- zdolności dostarczania wilgoci (81%)

można w oparciu o teorię mnogości oszacować prawdopodobieństwo P klasyfikacji właściwości pastwiskowej dla danej działki gruntu:

$$P = 0.95(0.85 \times 0.82 \times 0.81) = 0.54 = 54\%$$

### 3.2. Modele do oceny produktywności łąk

Modele do oceny produktywności łąk, które są stosowane w Chinach zaczerpnięto z pracy (Li i in., 1992). Niektóre z nich wykorzystują także zobrazowania satelitarne i tworzą wraz z GIS system monitoringu dynamicznego dla potrzeb rolniczych.

**Model klimatyczny (geograficzny)** do oceny wydajności produkcyjnej terenów łąkowych umożliwia obliczenie biomasy **W** znajdującej się ponad terenem.

Można go zapisać :

$$W = f(P, T, L, H)$$

gdzie:

**P** - opady w czasie sezonów wegetacji,



$T_j$ - akumulacja temperatury ( $> 10^{\circ}\text{C}$ ) w czasie  $j$ -ego sezonu wegetacji.

$T_p$ - suma średniej miesięcznej temperatury w czasie  $p$ -tego sezonu wegetacji.

$P$ ,  $T_j$ ,  $T_p$  - są funkcją położenia geograficznego ( ..h) - długość, szerokość i wysokość.

$L$  - rodzaj gleby.

$H$  - poziom zarządzania gospodarstwem.

Kiedy znane są  $L$  i  $H$  to biomasa jest określana z

$$W_j = f(P, T_j) \quad \text{i odpowiednio} \quad W_p = f(P, T_p)$$

**Model teledetekcyjny** (optyczny) do oceny wydajności produkcyjnej terenów łąkowych wymaga określenia w terenie indeksów wegetacyjnych  $g_i$  odpowiednio dla modelu obliczeniowego  $i = 1, 2$  (dwa sposoby obliczenia modelu).

$$g_i = f_i(k1, k2)$$

gdzie:

$k1$  - terenowe określenie widma optycznego kanału 1.

$k2$  - terenowe określenie widma optycznego kanału 2.

Produkcyjność łąki z badań terenowych  $W_g$  (biomasę) można określić:

$$W_g = f_i(g_i)$$

Z teledetekcyjnych danych satelitarnych dla  $i$ -tego modelu obliczeniowego indeks wegetacyjny  $G_i$  :

$$G_i = f_i(g_i)$$

Oszacowanie produkcyjności terenów łąkowych z danych teledetekcyjnych  $W_s$  można określić i-tym modelem obliczeniowym

$$W_s = f_i(G)$$

Badanie modelu przeprowadza się na podstawie estymatorów według określonego testu statystycznego przy poziomie istotności  $\alpha = 0.05$ .

**Model ogólny** oceny produkcyjności terenów łąkowych powstaje przez połączenie modelu klimatycznego z modelem teledetekcyjnym. Taki model, który zadawalająco oszacowuje biomasę można określić równaniem:

$$W_o = aW_g + bW_s$$

gdzie:

**a** i **b** współczynniki wagowe.

Model ten podnosi precyzję oceny i zwiększa wymagania operacyjnego systemu dynamicznego, gdyż łatwo nie ulega wpływom czynników metodologicznych jak model teledetekcyjny i dostarcza informacji w określonym czasie a nie średni stan biomasy jak model klimatyczny.

**Model wykorzystywania pastwiska** służy do obliczenia bilansu paszy przy ustalonej liczbie bydła. Umożliwia obliczenie dostępnej części biomasy pastwiska, która jest uzależniona od rodzaju bydła, typu pastwiska i jego sezonowego użytkowania. Model można zapisać równaniem określającym współczynnik wykorzystania  $U$  w określonym czasie  $t$ , typu pastwiska  $i$  przez zwierzęta  $j$

$$U = f(S_i, A_j, W_{it}, U_{max})$$

gdzie:

$S_i$  - powierzchnia pastwiska typu  $i$ .

$A_j$  - liczba zwierząt  $j$ .

$t$  - sezon lub czas.

$W_{it}$  - wznosząca się ponad gruntem biomasa pastwiska  $i$  w określonym czasie  $t$ .

$U_{max}$  - współczynnik maksymalnego wykorzystania określony ze stosunku zwierząt  $j$  do typu pastwiska  $i$ .

**Model równowagi** określający kompensatę paszy i bydła można wyrazić równaniem

$$B = f(W, R, t) - (U, A, t)$$

gdzie:

**W** - produktywność (biomasa) trawy.

**R** - intensywność przyrostu trawy.

**U** - ilościowe wykorzystanie trawy przez trawożernych.

**A** - liczba trawożernych.

**t** - czas.

#### 4. Ocena jakości modelu

W geograficznych systemach informacyjnych podobnie jak przy opracowaniu map i pracach pomiarowych standardy dokładności umożliwiają realizację zadań z określonym stopniem doskonałości i niezawodności. Staje się konieczne wprowadzenie nawyku oceny dokładności danych wejściowych, aby można było ujawnić ich wpływ na rezultatach uzyskanych w wyniku zastosowania modułów GIS. Dlatego w tablicach bazy danych (relacjach) - danym podstawowym powinny towarzyszyć dane pomocnicze zarówno ilościowe jak i jakościowe - takie jak np. pokazane w Tablicy 6. zaczerpniętej z pracy (Ramlal i Drummond, 1992).

Tablica 6

| GLEBI<br>D  | ODSŁ | ODST<br>W | WA1  | WNZ | PRAW<br>D | Jedn1 | WA2  | WCZ | ODST.<br>G.K | Jedn2 | OPIS      |
|---|------|-----------|------|-----|-----------|-------|------|-----|--------------|-------|-----------|
| 1254  | 25   | 16        | Hn35 | N   | 0.74      |       | 1.55 | C   | 0.10         | m     | UWA<br>GI |
| 1255  | 25   | 16        | Hn33 | N   | 0.65      |       | 1.80 | C   | 0.10         | m     |           |
| 6   | 25   | 16        | Hn31 | N   | 0.82      |       | 1.70 | C   | 0.10         | m     |           |
| 7   | 25   | 16        | Ez35 | N   | 0.87      |       | 2.00 | C   | 0.10         | m     |           |
| 8   | 25   | 16        | Zg35 | N   | 0.76      |       | 1.30 | C   | 0.10         | m     |           |
| 9   | 25   | 16        | Hn35 | N   | 0.58      |       | 1.55 | C   | 0.10         | m     |           |
| 1260  | 25   | 16        | Za35 | N   | 0.76      |       | 1.60 | C   | 0.10         | m     |           |
| <p>GLEBID Identyfikator dla powierzchni glebowych</p> <p>ODSŁ Odchyłka standardowa współrz. x,y na łuku</p> <p>ODSTW Odchyłka standardowa współrz. x,y na węzłach</p> <p>WA1 System klasyfikacji glebowej (oznaczenia)</p> <p>WNZ N Wskaźnik nieciągłości zmiennych</p> <p>PRAW D Prawdopodobieństwo poprawności klasy gleb</p> <p>Jedn.1 Jednostka</p> <p>WA2 Wartość 2 - głębokość korzeniowa gleby</p> <p>WCZ C- Wskaźnik ciągłości zmiennych</p> <p>ODST.G.K Odchyłka standardowa pomiaru głębokości korzeniowej</p> <p>Jedn.2 Jednostka pomiaru głębokości korzeniowej</p> <p>OPIS (lub słownik danych) związane z zależnościami obrębów glebowych</p> |      |           |      |     |           |       |      |     |              |       |           |

Podają one oprócz identyfikatorów, wartości liczbowych i oznaczeń, również wskaźniki nieciągłości zmiennych, jednostki miar zestawionych wartości, błędy średnie danych, wskaźniki prawdopodobieństwa poprawności zestawionych danych (atrybutów) itp. Jest to szczególnie ważne, kiedy do przetwarzania danych są wykorzystywane modele, gdyż zestawione w tablicach bazy danych błędy i wskaźniki umożliwiają ocenę jakości modelu. Podobnie jak to ma miejsce w geodezji i fotogrametrii również w GIS rozkład wariancji może być wykorzystywany do oszacowania jakości informacji wygenerowanej przez geograficzny system informacyjny, kiedy te informacje powstają przy udziale modeli matematycznych. Procesy matematycznego modelowania zmiennych losowych typu ciągłego (np. współrzędnych) i wartości stałych (np. parametrów modelu) dostarczają nowych informacji. Można przy tym gromadzić w tablicach bazy danych błędy średnie tych nowo wygenerowanych informacji korzystając z rozkładu wariancji danego modelu matematycznego. Rozkład wariancji znany w praktyce jako rozkład błędów umożliwia dla modelu matematycznego  $a = f(b, c)$  obliczenie odchyłki standardowej  $\sigma_a^2$ , jeśli znane są odchyłki  $\sigma_b^2$  i  $\sigma_c^2$  (Mikhail 1976):

$$\sigma_a^2 = \sigma_b^2 + (\delta a / \delta b)^2 + \sigma_c^2 + (\delta a / \delta c)^2 + 2\sigma_b\sigma_c(\delta a / \delta b)(\delta a / \delta c)$$

gdzie:  $b$  i  $c$  - dane zgromadzone w tablicach bazy danych GIS wraz z odchyłkami standardowymi  $\sigma_b$  i  $\sigma_c$ ;  $a$  - nowe informacje; różniczki cząstkowe  $\sigma_a$ ,  $\sigma_b$  i  $\sigma_c$  musi dostarczyć użytkownik albo obliczyć system. Jeśli nie ma korelacji między  $b$  i  $c$  - ostatni człon równania można pominąć.

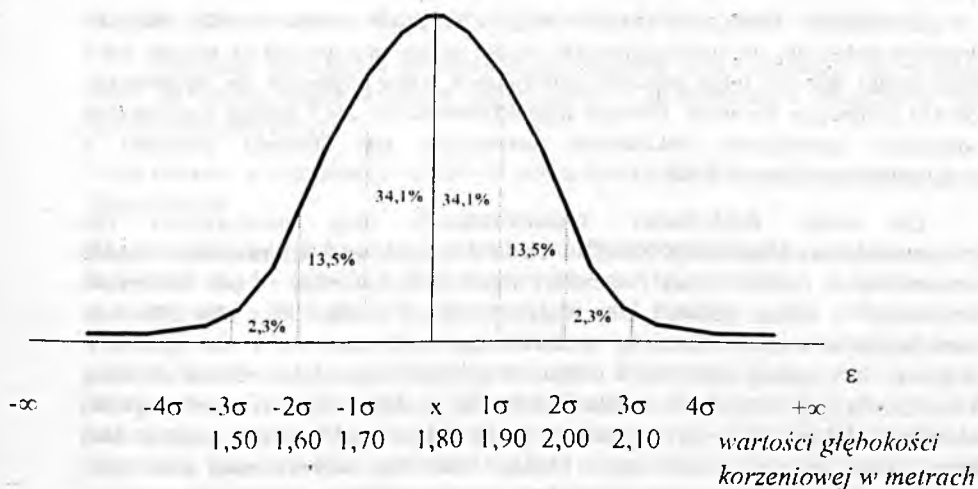
W przypadku kiedy GIS przetwarza dane przy wsparciu modelu liczbowego do oceny jakości otrzymanej informacji można wykorzystać teorię mnogości. W celu zapoznania czytelnika z praktycznym zastosowaniem tej teorii do rozkładu błędów (kiedy rezultaty przetwarzania GIS-em powstały przy wsparciu modelu liczbowego) przytoczę przykład z publikacji (Ramalal i Drummond 1992). Niech modelem liczbowym będzie np. WŁAŚCIWOŚĆ PASTWISKOWA Nr 1 kiedy:

- 1) typ gleby należy do Hn33 (gleby bielcowe) i
- 2) głębokość korzeniowa wynosi 1.50 m do 2.00 m (patrz Tablica 6)

Ograniczymy model do tych dwóch cech chociaż możnaby było go poszerzyć o stan drenażu, urodzajność i zapasy wilgotności. W tabeli bazy danych oprócz identyfikatora obszarów glebowych i wskaźników klasyfikacyjnych gleb podano także odchyłki standardowe współrzędnych  $x, y$  dla węzłów i luków, oznaczenia ciągłości i nieciągłości zmiennych, prawdopodobieństwo poprawności określenia typów gleb i głębokości korzeniowe gleby wraz z jednostkami pomiarowymi i odchyłkami standardowymi ich pomiaru.

Rozpatrzmy obszar glebowy o identyfikatorze np. 1255, w którym prawdopodobieństwo klasy gleby Hn33 (bielcowa) wynosi 0.65, głębokość korzeniowa równa się 1.80 m - tzn. prawdopodobieństwo, że głębokość korzeniowa mieści się w granicach 1.50 m do 2.00m - jest 0.98. Takiego obliczenia można dokonać na podstawie rozkładu normalnego błędu pomiaru głębokości korzeniowej,

dla głębokości 1.80 m i odchyłki standardowej  $\sigma = 0.10$  m (rys.2). wiedząc (co jest również ujęte w tablicy bazy danych), że ma się do czynienia ze zmienną losową typu ciągłego. Z wykresu na rys.2 widać, że dolny kraniec klas (1.50 - 2.00 m) głębokości korzeniowej przypada na  $3\sigma$  (1.50 m) od średniej (1.80 m), natomiast górna granica klas jest na odległości  $2\sigma$  (2.00 m) powyżej średniej. Zatem prawdopodobieństwo, że błąd zawiera się między liczbami 1.50 i 2.00 m (jak pokazano pod osią odciętych będącej osią zasięgu odchyłek standardowych  $\sigma$ ) można wyrazić powierzchnią pola ograniczonego: od dołu osią odciętych  $\varepsilon$ , od góry krzywą prawdopodobieństwa a po bokach rzędnymi  $\varepsilon = 1.50$  m oraz  $\varepsilon = 2.00$  m (która wynosi 97.5%), w stosunku do całego pola ograniczonego osią i krzywą prawdopodobieństwa (100%). Stąd dla przytoczonego przykładu prawdopodobieństwo wynosi 0.98. Przyjmując, że model jest doskonały (prawdopodobieństwo sprawności modelu wynosi 100%), zatem prawdopodobieństwo obszaru glebowego 1255 posiadającego WŁAŚCIWOŚCI PASTWISKOWE Nr 1 wynosi 0.64 (lub  $0.65 \times 0.98$ ). Natomiast gdyby prawdopodobieństwo sprawności modelu wynosiło 80 %, wtedy prawdopodobieństwo tego obszaru glebowego o WŁAŚCIWOŚCIACH PASTWISKOWYCH Nr 1 wynosiłoby 51 % (czyli  $0.64 \times 80$ ). Jest to problem przecięcia teorii mnogości.



Rys.2. Rozkład normalny błędu pomiaru głębokości korzeniowej dla średniej głębokości pomiaru równej 1.80m i odchyłki standardowej = 0.10m. (Ramlal, Drummond, 1992)

Przytoczony przykład dotyczył oceny skuteczności modelu, w którym występowały tylko dwie cechy: obszar glebowy z prawdopodobieństwem występowania tej gleby wynoszącym 0.65 i głębokość korzeniowa. Zarówno w tym jak i w innych modelach liczbowych tych cech może być więcej i jeśli są dla nich podane wskaźniki dotyczące poprawności ich określenia lub odchyłki standardowe pomiarów jakimi je wyznaczono, wtedy również możliwa jest ocena skuteczności generowanych przez model i GIS wyników - stosując przytoczone powyżej jak też inne metody rachunku prawdopodobieństwa.

Analiza ilościowa błędów może być przeprowadzona podczas każdego etapu przetwarzania danych. W zasadzie wartości błędów powinny być określane po każdym etapie tego przetwarzania, ale w opracowaniach z udziałem danych teledetekcyjnych błędy są obliczane po zakończeniu przetwarzania i analiz danych i zazwyczaj są ukierunkowane na dokładność sytuacyjną i tematyczną. Podobnie jak w analizach dokładnościowych prowadzonych w geodezji i fotogrametrii, tak i w opracowaniach GIS - wielkość próbki na której dokonuje się oszacowania wyników odgrywa ważną rolę. Istnieją różne zalecenia i wytyczne do określania wielkości próbki - takiej aby nadawała się do przeprowadzenia analiz statystycznych. Rozmieszczenie badanych punktów pól testowych w próbie spełnia istotną rolę w ocenie dokładności, zwłaszcza, gdy dotyczy to obrazu teledetekcyjnego. Rozmieszczenie to musi być reprezentatywne dla całego klasyfikowanego obrazu.

Dokładność sytuacyjną obrazów teledetekcyjnych można określić błędami średnimi położenia, ale należy pamiętać, że nie ujmuje ona wszystkich pikseli, gdyż błąd średni dotyczy tylko punktów kontrolnych, które posłużyły do wpasowania obrazu cyfrowego do mapy. Dlatego najdokładniejszym, ale i bardzo kosztownym sposobem sprawdzenia dokładności sytuacyjnej jest kontrola terenowa z wykorzystaniem danych GPS.

Do oceny dokładności wygenerowanych map tematycznych lub przeprowadzonej klasyfikacji obrazów teledetekcyjnych można zastosować metodę zestawienia w postaci rastra (macierzy) kwadratów odchyłek (z pól testowych) uzyskanych z różnic położenia klas wygenerowanych przez GIS i klas aktualnie istniejących w terenie. Kwadraty odchyłek uporządkowane są w taki sposób w macierzy, że kolumny informują o miejscach położenia odchyłek a wiersze określają klasy (cechy) jakie uzyskano z klasyfikacji. Jest to doskonały i efektywny sposób określania dokładności, gdyż umożliwia ocenę dokładności każdej wygenerowanej klasy. Taka macierz zestawionych błędów może być wykorzystana jako dane wyjściowe do różnych oszacowań metodami statystycznymi.

## 5. Uwagi końcowe

W omawianym tutaj systemie monitoringu, podobnie zresztą jak w innych systemach wykorzystywane są różne podsystemy, które odpowiednio do zadań realizują działania monitorowania, rejestracji i przetwarzania informacji na podstawie dostarczonych danych.

W odniesieniu do **podsystemu zbierania danych**, źródła, formy i metody pozyskiwania z nich informacji można podzielić na:

- informacje monitoringu terenowego (polowego) dostarczane z obserwacji fenologicznych i meteorologicznych, z oszacowania produktywności biomasy, monitoringu ekologicznego (pustynnienie i degradacja terenu, erozja, zmiany wilgotności itp.), z monitoringu stanu wykorzystania zasobów, z monitoringu klęsk żywiołowych itp.,
- informacje z zestawień statystycznych i map tematycznych, które pokazują ukształtowanie terenu, użycie ziemi, regionalizację ekologiczną, rodzaje terenów uprawnych i pastwnych, obręby administracyjne itp. Zestawienia statystyczne, informujące o strukturze i liczbie bydła, o zapleczu socjalnym i ekonomicznym itd.,
- informacje teledetekcyjne pozyskiwane z satelitów meteorologicznych i gospodarczych i systemów przetwarzania obrazów.

Zadaniem **podsystemu zarządzania danymi** (DBMS - Data Base Management System) jest ich zebranie i nadzorowanie w obrębie bazy danych. Taka baza powinna być przystosowana do struktury danych i ma tworzyć bazę w ujednoliconym formacie rekordów. Ale musi ona również pomieścić informacje teledetekcyjne. Ponadto ten podsystem powinien umożliwiać oprócz nadzoru, kontrolę, eliminowanie, redagowanie, aktualizację i obliczenia dla rekordów i map.

Natomiast **podsystem przetwarzania danych** przez działania na modelu służy do pozyskiwania wyników użytecznych w podsystemach zarządzania lub przesyłania informacji.

Musi również w tym samym czasie po przeprowadzeniu sortowania zachować je w bazie danych.

W większości geograficznych systemów informacyjnych modele wykorzystywane okazjonalnie (do których należą modele liczbowe) są osobnymi modułami nie instalowanymi na stałe ale wprowadzane do systemu tylko na czas ich wykorzystania. Moduły te powinny posiadać oprogramowania umożliwiające ocenę jakości danych wygenerowanych przez model. Dla oceny błędów modeli należy wykorzystywać rozkład wariancji lub teorię mnogości. W tym celu GIS-y w tabelach baz danych oprócz danych podstawowych powinny zawierać dane ilościowe i jakościowe zestawionych wartości łącznie z błędami średnimi i wskaźnikami prawdopodobieństwa poprawności zestawionych danych. Wtedy pozyskane za pomocą modelu nowe informacje, których niejednokrotnie nie można porównać z danymi rzeczywistymi, będą mogły być ocenione metodami statystycznymi.

**Źródła**

Chandor A., Graham J., Williamson R., 1985. The Penguin Dictionary of Computers. Penguin Books, London.

Li B., Shi P.J., Lin X.Q., 1992. A Research on the Dynamic Monitoring System for Forage Livestock Balance in the Temperature Grassland. Intern. Archives of Ph.&R.S., Vol. XXIX part B4, Washington str.1-4.

Mikhail E.M., 1976. Observations and Least Squares. IEP Series, N.York.

Raafat H., Yong Z., 1992. State and Time for Geographic Information. Intern. Archives of Ph. & R.S., Vol XXIX part B3, Washington, str.155-161.

Ramlal B., Drummond J.E., 1992. A GIS Uncertainty subsystem. Internat. Archives of Ph. & R.S. Vol.XXIX part B3, Washington, str.356-362.

Sitek Z., 1994. Elementy projektowania zintegrowanego GIS dla potrzeb monitoringu środowiska. Archiwum Fotogrametrii, Kartografii i Teledetekcji, Vol.2, Kraków AGH.

Słownik Matematyki i Cybernetyki Ekonomicznej, 1985. Państwowe Wydawnictwo Ekonomiczne, Warszawa