



**AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA W KRAKOWIE**

Zagadnienie niedokładności w GIS

Beata Hejmanowska

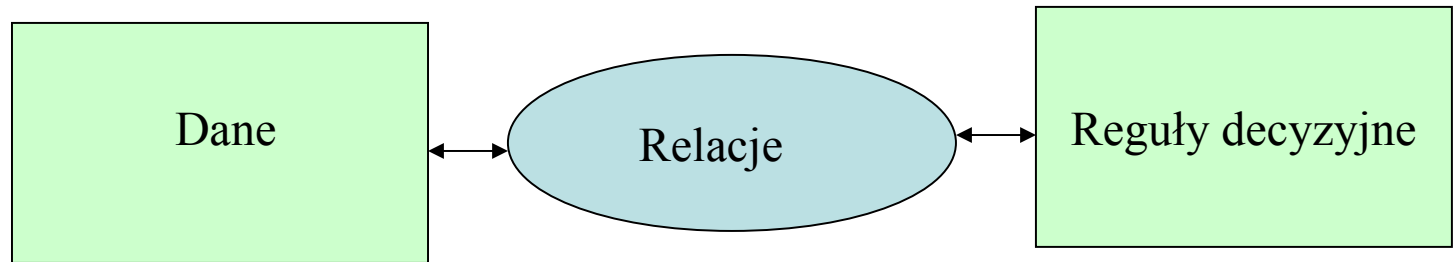
Analiza dokładności procesu wspomaganie decyzji za pomocą narzędzi GIS

Niedokładność danych :

błędy pomiarowe: określanie błędu, szacowanie błędu,
przenoszenie błędów, raporty o jakości danych

Wpływ niedokładności danych w GIS na proces podejmowania decyzji
(reguły decyzyjne) ?????? (*hard, soft decision rule*)

Typy błędów



*Uncertainty
in the Evidence*

*Uncertainty
in the Relation*

*Uncertainty
in the Decision Set*

Dane

Uncertainty in the Evidence

W przypadku danych ilościowych
błąd średni kwadratowy
Root Mean Square (RMS)

W przypadku danych jakościowych
błąd „proporcjonalny”



AGH

Relacje

Uncertainty in the Relation

1. Definicja kryterium – teoria zborów rozmytych *fuzzy set*
(*nachylenie < 3.5 %*, *3.4999% ???*)
2. Dane nie pozwalają na zbudowanie bezpośredniej i jednoznacznej reguły decyzyjnej – (woda absorbuje promieniowanie podczerwone, ale nie tylko ona....) – dane sugerują obecność zjawiska, obiektu na pewnym poziomie ufności - *Bayes and Dempster-Shafer theory*.
3. Błąd modelu - *model specification error* – np. złożenie prostych kryteriów:
 - nachylenia – 0.6
 - odległości od obiektu (drogi, wody....) – 0.7prawdopodobieństwo: 0.42
fuzzy – 0.6
Bayes – 0.78
Dempster-Shafer – 0.88
liniowa kombinacja – 0.65
??

Reguła decyzyjna

Uncertainty in the Decision Set

Ostateczny błąd (niedokładność) reguły decyzyjnej.

Podjęcie decyzji:

- sprecyzowanie elementarnych kryteriów, warunków
- złożenie wyników zgodnie z jakąś regułą.

W takim procesie nastąpi sumowanie błędów, a ostateczna decyzja musi uwzględnić jakiś dopuszczalny próg błędu (*decision risk*)

Database Uncertainty,
Decision Rule Uncertainty
Decision Risk

$$X(m, \sigma)$$

Rozkład normalny zmiennej X

m – średnia

σ – odchylenie standardowe zmiennej X

$$\bar{X}\left(m, \frac{\sigma}{\sqrt{n}}\right)$$

Rozkład normalny średniej arytmetycznej

m – wartość oczekiwana zmiennej \bar{X}

$\frac{\sigma}{\sqrt{n}}$ – odchylenie standardowe zmiennej \bar{X}

$$\bar{\sigma}\left(m, \frac{\sigma}{\sqrt{2n}}\right)$$

Rozkład normalny odchylenia standardowego

m – średnia

$\frac{\sigma}{\sqrt{2n}}$ – odchylenie standardowe zmiennej $\bar{\sigma}$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} nm = m$$

$$D^2(\bar{X}) = D^2\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n D^2(X_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

$$D(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Niepewność bazy danych

Ryzyko w podejmowaniu decyzji

Database Uncertainty and Decision Risk

Szacowanie błędu - *Error Assessment*

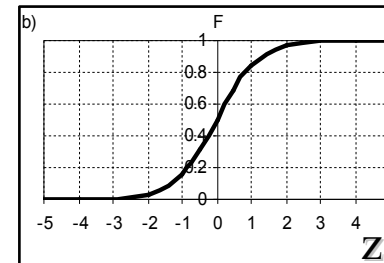
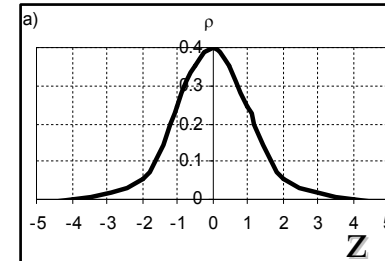
$$\bar{X}(m, \frac{\sigma}{\sqrt{n}}) \quad Z = \frac{\bar{X} - m}{\sigma} \quad N(0,1)$$

$$P\left\{-z_{\frac{\alpha}{2}} < z < +z_{\frac{\alpha}{2}}\right\} = 1 - \alpha$$

$$P\left\{-z_{\frac{\alpha}{2}} < \frac{\bar{X} - m}{\sigma} \sqrt{n} < +z_{\frac{\alpha}{2}}\right\} = 1 - \alpha$$

$$P\left\{\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < m < \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha$$

$$\left(\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$



Zagadnienie minimalnej liczebności próby

$$\left(\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) \quad \left(\bar{\sigma} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{2n}}, \bar{\sigma} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{2n}} \right)$$

$$e = z \frac{\sigma}{\sqrt{n}} \longrightarrow n = \frac{z^2 \sigma^2}{e^2}$$

Dla wartości średniej

$$e = z \frac{\sigma}{\sqrt{2n}} \longrightarrow n = \frac{z^2 \sigma^2}{2e^2}$$

Dla odchylenia standardowego



AGH

Niepewność bazy danych Ryzyko w podejmowaniu decyzji

Database Uncertainty and Decision Risk

Szacowanie błędu - *Error Assessment*

$$e = z \frac{\sigma}{\sqrt{2n}} \longrightarrow n = \frac{z^2 \sigma^2}{2e^2}$$

Błąd CMT

RMS = +/- 3 m

e = +/- 0.5 m

$$n = (1.645^2 * 3^2) / (2 * 0.5^2) = 49$$

$$e = z \sqrt{\frac{pq}{n}} \longrightarrow n = \frac{z^2 pq}{e^2}$$

Błąd interpretacji

q= 0.15, p= 0.85

e = 0.05 m

z=1.645 (poziom ufności 90%)

$$n = (1.645^2 * 0.15 * 0.85) / (0.05^2) = 138$$

Niepewność bazy danych

Ryzyko w podejmowaniu decyzji

Database Uncertainty and Decision Risk

Szacowanie błędu - *Error Assessment* – analiza jakościowa

		True					error
		Conifers	Mixed	Deciduous	Water	Total	
Mapped	Conifers	24	0	0	3	27	0.11
	Mixed	3	36	16	0	55	0.35
	Deciduous	0	0	28	0	28	0.00
	Water	2	0	0	14	16	0.12
	Total	29	36	44	17	126	
	error	0.17	0.00	0.36	0.18		0.19

errors of commission

errors of omission

Figure 2-1 An Error Matrix

Błędy przenoszą się w trakcie wszystkich analiz przestrzennych:

Overlay Add / Subtract
(e.g., $Z=X+Y$ or $Z=X-Y$)

$$S_z = \sqrt{(S_x^2 + S_y^2)}$$

Overlay Multiply / Divide
(e.g., $Z=X*Y$ or $Z=X/Y$)

$$S_z = \sqrt{(S_x^2 Y^2 + S_y^2 X^2)}$$

Scalar Add / Subtract
(e.g., $Z=X+k$ or $Z=X-k$)

$$S_z = S_x$$

Scalar Multiply
(e.g., $Z=X*k$)

$$S_z = S_x * k$$

Scalar Divide
(e.g., $Z=X/k$)

$$S_z = S_x / k$$

Fragment siatki *grid'a* CMT

z_5	z_2	z_6
z_1	z_9	z_3
z_8	z_4	z_7

--- Δh ---

$$S = \arctan\left(\frac{1}{2\Delta h} \sqrt{(z_3 - z_1)^2 + (z_2 - z_4)^2}\right)$$

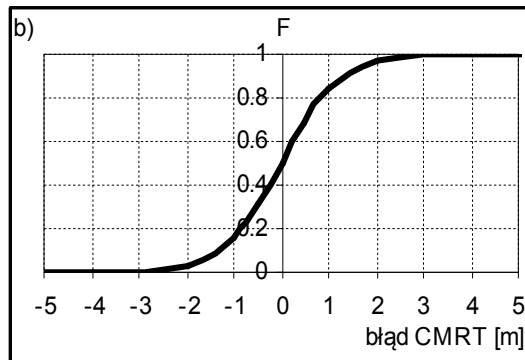
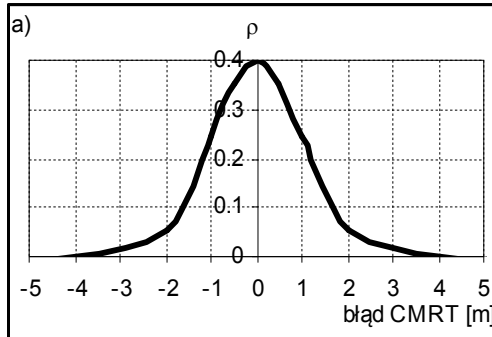
$$A = \arctan\left(\frac{z_1 - z_3}{z_2 - z_4}\right)$$

$$M_S = \frac{\sqrt{2}}{2\Delta h(1 + \operatorname{tg}^2 S)} M_Z$$

$$M_A = \frac{1}{\sqrt{2\Delta h \operatorname{tg} S}} M_Z$$

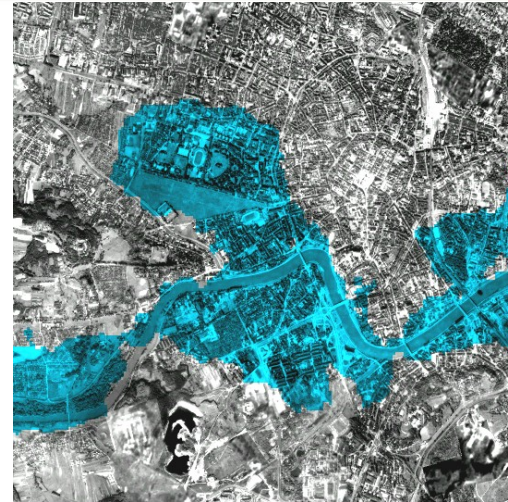
Szacowanie ryzyka związanego z podejmowaną decyzją

Decision Risk

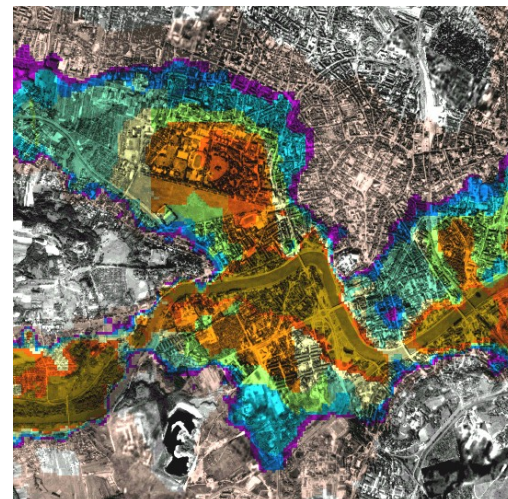


90% błąd CMRT \leq 1.28 m
 95% błąd CMRT \leq 1.645 m

RECLASS



PCLASS



- 0% (grey)
- 0-10% (beige)
- 10-20% (violet)
- 20-30% (blue)
- 30-40% (light blue)
- 40-50% (blue-green)
- 50-60% (green)
- 60-70% (yellow)
- 70-80 (orange)
- 80-90 (red)
- 90-100% (brown)