

Modellazione e trattamento di dati geografici: il problema dell'incertezza



Gloria Bordogna

**Consiglio Nazionale delle Ricerche
Istituto per la Dinamica dei Processi
Ambientali**

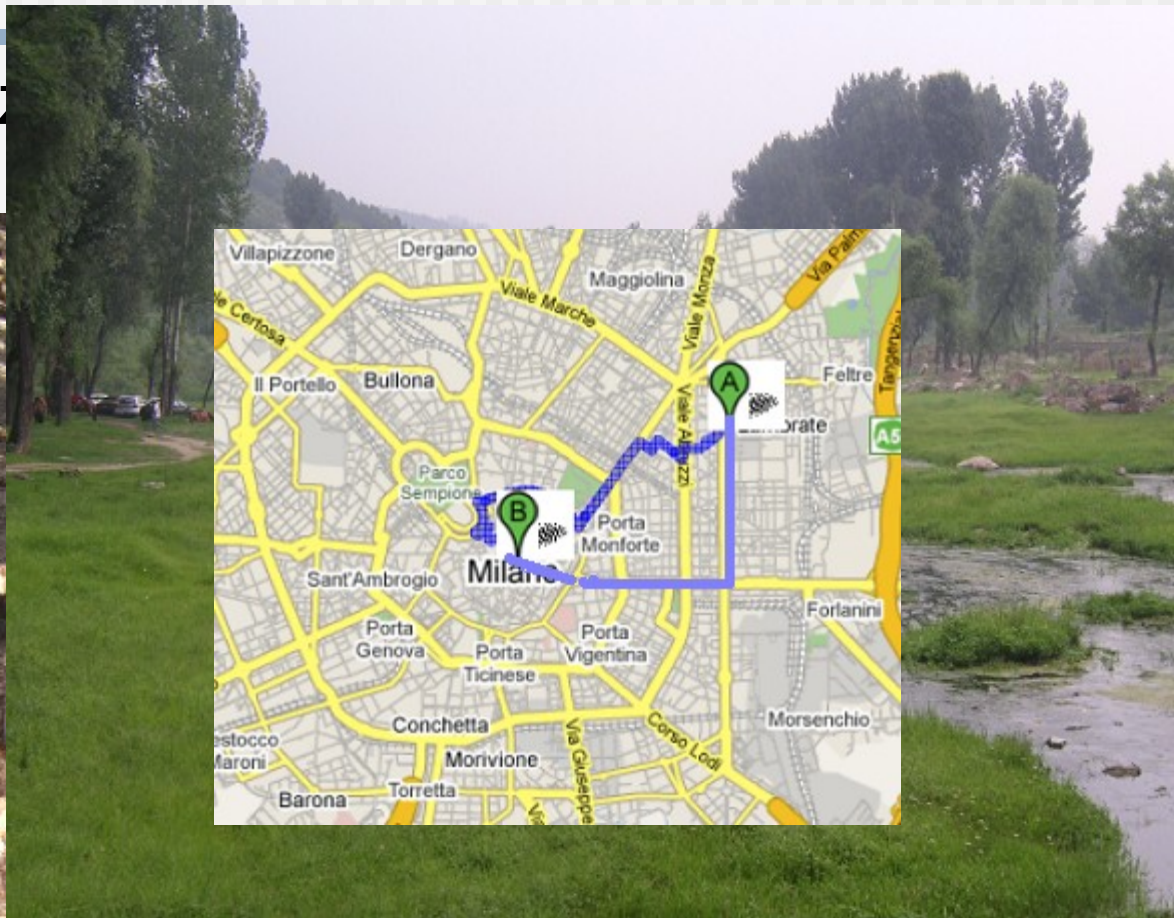
c/o Polo per l'Innovazione Tecnologica,
via Pasubio 5, 24044 Dalmine (BG)
tel: +390356224262

e-mail: gloria.bordogna@idpa.cnr.it



Piano della presentazione

L'incertezza
lato è il



senso

2].

25.08.2008



Scopi della gestione dell'imperfezione dell'informazione geografica

2. Ridurre l'imperfezione e l'errore
3. Rappresentare l'imperfezione dei dati (in un

⇒ Qualità dell'informazione

dell'imperfezione quando si effettuano
operazioni sui dati (nei GIS)

5. Misurare e stimare l'errore nel dato

Comprendere le sorgenti e le cause
dell'imperfezione dell'informazione geografica
[Veregin 1989, Goodchild 1988]



Qualità dell'informazione

- **Affidabilità/credibilità della sorgente;**
- **Corrispondenza alla realtà:**
 - **Ripetibilità delle osservazioni** (è corretta, corrisponde alla realtà se si ottengono gli stessi dati effettuando misurazioni diverse) [Wang and Wand 1996];
 - **Dipendenza dalla qualità del processo** di acquisizione/produzione [Timpf et al. 1996].
- **Appropriatezza per l'uso:** (fitness for use) (ISO 8402);
 - Caratteristiche dell'informazione che la rendono adatta a soddisfare necessità prestabilite;
 - dipendenza dalla qualità delle decisioni prese sulla base dell'informazione



Qualità dell'informazione

Elementi della della Qualità dell'informazione geografica (**quantitativi** e **qualitativi**):

[ISO 19113:202 e ISO 19114:2003]

- **Completezza** → (# valori mancanti)
- **Consistenza** → rispetto alla struttura dei dati (# di violazioni)
- **Accuratezza spaziale** rispetto alle specifiche di prodotto (deviazione standard e % confidenza)
- **Accuratezza temporale** rispetto alle specifiche di prodotto (deviazione standard e % confidenza)
- **Accuratezza tematica** relativa alla classificazione degli elementi e dei loro attributi rispetto alle specifiche di prodotto
- **Scopi** : informazioni sugli obiettivi dell'utilizzo dei dati
- **Usò**: applicazioni per i quali i dati sono stati usati
- **Lineage**: storia del dato e del processo di produzione



Imperfezione dell'informazione geografica

Aspetti dell'imperfezione [Plewe, 2002] :

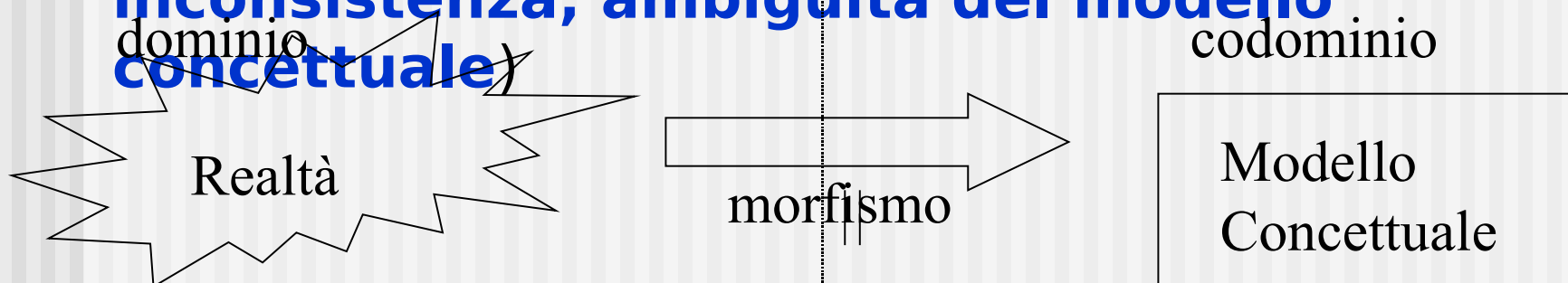
- **Dimensione** [Goodchild 1989]
 - Quale manifestazione dell'entità è imperfetta
 - La componente spaziale, la localizzazione/forma di un fenomeno
 - La componente temporale, la data/intervallo di validità
 - La componente tematica , Il valore della/e proprietà
- **Causa**
 - Quali le sorgenti/processi hanno generato l'imperfezione (strumenti di misura, metodo di misura, modello concettuale, ecc.)
- **Tipo**
 - Imprecisione, vaghezza, indeterminatezza, incompletezza, inconsistenza, incertezza
- **Forma**
 - Linguistica, intervallo di validità, distribuzione di probabilità, ecc.



Cause dell'Imperfezione dell'Informazione geografica

- Le rappresentazioni della realtà non descrivono esattamente la realtà stessa per varie ragioni (spesso necessarie) [Frank, 2006]:

- A causa dell'incapacità/limitatezza/necessità con cui gli umani concettualizzano una realtà complessa (**approssimazione, incompletezza, inconsistenza, ambiguità del modello concettuale**)



- A causa **dell'inaccuratezza/errori/incompletezza delle misure** necessarie per creare le rappresentazioni formali dei modelli concettuali della realtà



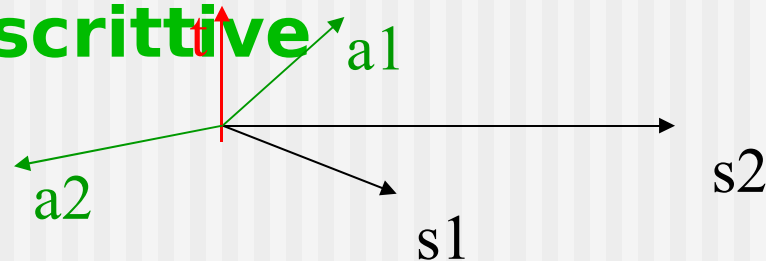
Processi che possono determinare Imperfezione dell'Informazione geografica

- **Processo di acquisizione:** [Aronoff, 1989]:
 - Es. Errore degli strumenti di misura
- **Processo di data entry:**
 - Inaccuratezza della digitalizzazione dovute all'hw o all'operatore;
 - Inaccuratezza dovute alla natura degli oggetti
- **Processo di memorizzazione dei dati:**
 - Insufficiente accuratezza numerica e/o spaziale (risoluzione)
- **Processo di elaborazione dei dati:**
 - Definizione inappropriata delle classi
 - Propagazione aritmetica dell'errore
- **Processo di visualizzazione dei dati:**
 - Inaccuratezza del dispositivo di output
- **Processo di utilizzo dei dati:**
 - Conoscenza insufficiente o utilizzo sbagliato



Propagazione dell'Imperfezione dell'Informazione geografica

- Un'entità ha un identificatore univoco, un'estensione **spazio temporale**, e un insieme **A** di **proprietà descrittive**



- Un **Datum** è una funzione **f** di variabili di controllo [Plewe, 2002]

- Es. la popolazione di una regione in un dato periodo di tempo $\rightarrow \mathbf{p} = \mathbf{f}(\mathbf{s}, \mathbf{t})$
- I confini di una nazione alla data $t \rightarrow \mathbf{s} = \mathbf{f}(\mathbf{n}, \mathbf{t})$
- la data di un incendio $\rightarrow \mathbf{t} = \mathbf{f}(\mathbf{s}, \mathbf{a})$

- C'è simmetria tra le dimensioni nel senso che ciascuna può essere variabile di controllo o datum misurato



Tipi di Imperfezione dell'Informazione

Un'unità di informazione può essere schematizzata da una quadrupla [Bosc and Prade, 1997]

(*id, dim, val, truth*) Imprecisione/vaghezza/approssimazione

- ***id*** è l'identificatore di un'entità
- ***dim*** è una dimensione dell'entità (es. proprietà)
- ***val*** è il valore di ***dim*** dell'entità ***id***
- ***truth*** è la veridicità della conformità dell'informazione (*id, dim, val*) alla realtà:

Es: (Milano, popolazione, 1M, vero)

“***è vero*** che la popolazione di Milano è di 1 M di abitanti”

incertezza



Modellazione di informazioni imperfette

- La **teoria delle probabilità** permette di sintetizzare misure ripetute di una variabile e modellare fenomeni stocastici:
 - **es: precipitazioni medie in un intervallo temporale,**
- La **teoria degli insiemi fuzzy** [Zadeh, 1965] permette di rappresentare con il concetto di insieme fuzzy una classe di elementi dai confini non ben definiti: è utile per rappresentare proprietà graduali quali i concetti linguistici.
 - **es: l'insieme dei punti di una mappa che appartengono a un ecosistema (confine tra latifoglie e conifere), il concetto di rischio basso, alto, medio.**
- La **teoria delle possibilità** [Zadeh 1978, Dubois & Prade, 1988] permette di rappresentare l'incertezza derivante da considerazioni di natura soggettiva:
 - Es: per sintetizzare poche misure coerenti o considerazioni soggettive
 - **es: precipitazioni intense,**
- La **teoria dei Rough sets** [Pawlak, 1982] permette di rappresentare conoscenze incomplete : è utile per rappresentare l'inconsistenza/ambiguità dell'informazione



Modellazione di sistemi informativi geografici flessibili

Gli strumenti tradizionali per la gestione automatica dell'informazione geografica **GIS** sono carenti per diversi aspetti:

- sono **poco flessibili**:
 - l'informazione geografica deve essere rappresentata in modo preciso;
 - Non forniscono strumenti per modellare la conoscenza "vaga" dei fenomeni per poter elaborare scenari o nuove informazioni
 - necessitano di un'utenza esperta; non sempre forniscono risultati soddisfacenti.



Gestione dell'imperfezione dell'informazione geografica

- Estensione delle basi di dati spaziali con **tipi di dati spaziali imperfetti**
- Integrazione nei GIS di strumenti che permettano all'esperto
 - di **modellare attitudini alle decisioni distinte** (più o meno rischiose)
 - di **codificare la conoscenza (vaga)** dei fenomeni e di elaborare scenari/previsioni
- Definizione di **linguaggi flessibili per l'interazione con i GIS** in modo da trattare i dati memorizzati come rappresentazioni approssimate della realtà (es: specifica di condizioni "tolleranti" di selezione
 - Es: Regioni ad *alta* densità di popolazione e a *10min* da una stazione della metro
- Definizione di **ontologie** per ridurre le ambiguità/errori nell'uso dei dati geografici



Esempi di gestione dell'informazione imperfetta nei GIS

Estensione dei modelli di dati spaziali nei GIS:

Definizioni di modelli dei dati spaziali intermedi tra il modello di dati a oggetti e il modello a campo

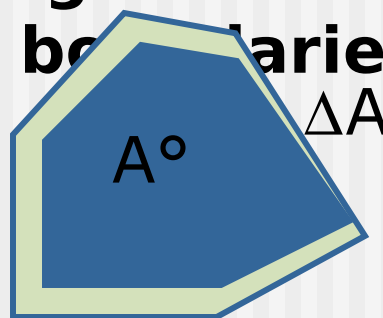




Rappresentazione nei GIS di “Regioni dai confini larghi”

Rappresentazione di regioni dai confini “larghi”
(Broad o indeterminate boundaries)

$$A = A^{\circ} \cup \delta A$$



Caratteristiche

- le proprietà geometriche quali area, distanza ecc, vengono calcolate rispetto a A° o ad A
- Le proprietà topologiche sono estese (es. partially overlapping, nearly covered by)

Modelli :

Es: modello egg-yolk : indeterminate boundaries [Cohn & Gotts, 1994], **modello broad boundaries**: assume $A^{\circ} \gg \Delta A$ [Clementini & Di Felice, 1994], **modello ROSA**, : indeterminate spatial objects , basato su Rough sets [Schneider e Pauly, 2007]; **modello VASA**: [Schneider e Pauly, 2008];



Rappresentazione nei GIS di "Regioni dai confini spessi"

Modello VASA [Schneider e Pauly, 2008];

Le operazioni insiemistiche sono basate su una logica a tre valori

$$A = A^\circ \cup \delta A \quad o = \text{not } A$$

$$A^\circ = k \text{ kernel}$$

$\delta A = b$ boundary

intersezione

| | | | |
|---|---|---|---|
| | k | b | o |
| k | k | b | o |
| b | b | b | o |
| o | o | o | o |

differenza

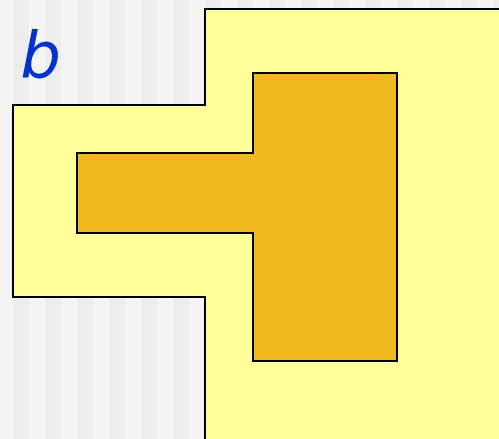
| | | | |
|---|---|---|---|
| | k | b | o |
| k | o | b | k |
| b | o | b | b |
| o | o | o | o |

unione

| | | | |
|---|---|---|---|
| | k | b | o |
| k | k | k | k |
| b | k | b | b |
| o | k | b | o |

complemento

| | | |
|---|---|-------|
| k | b | o |
| o | b | k.... |





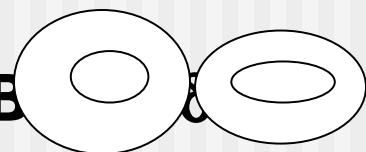
Estensione delle relazioni topologiche

[Schneider e Pauly, 2005],

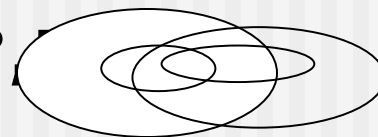
$A = A^\circ \cup \delta A$ Relazioni topologiche a tre valori

Es:

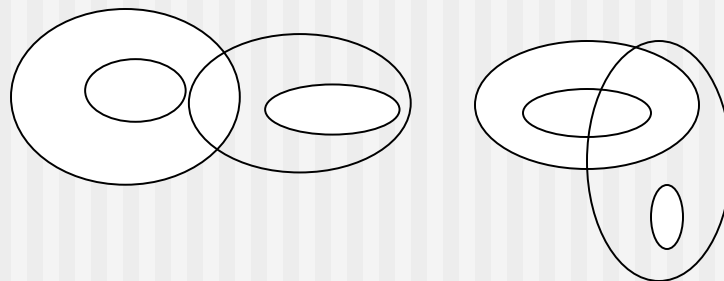
$Disjoint(A;B) = true$, $\Leftrightarrow disjoint(A^\circ \cup \delta A ; B)$



$Disjoint(A;B) = false$, $\Leftrightarrow not\ disjoint(A^\circ \cup \delta A ; B)$



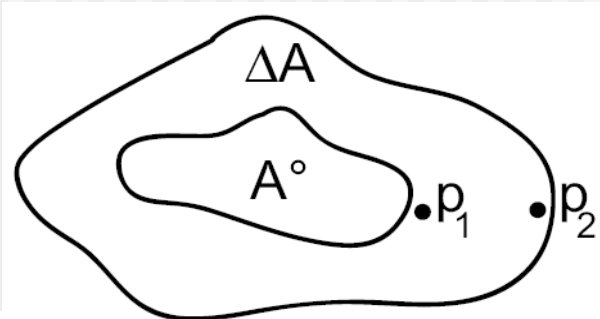
$Disjoint(A;B) = maybe$, $\Leftrightarrow otherwise$





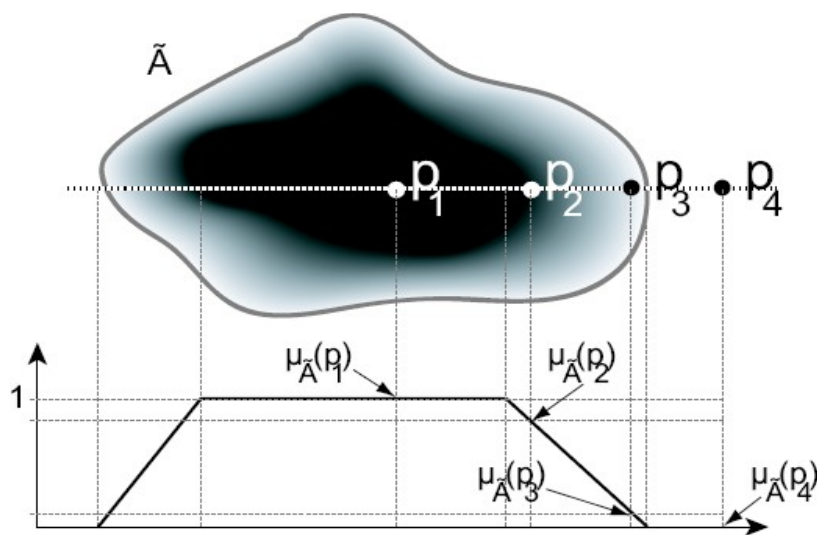
Rappresentazione nei GIS di “Regioni dai confini sfocati”

Rappresentazione di regioni dai confini fuzzy: A° , ΔA e $\neg A$ sono definiti come insiemi fuzzy



$$\tilde{A} = \{(p, \mu_{\tilde{A}}(p)) | p \in U, \mu_{(\tilde{A})}(p) > 0\}$$

$$\begin{aligned} \mu_{\tilde{A}} : U &\rightarrow]0, 1] \\ p &\mapsto \mu_{\tilde{A}}(p) \end{aligned}$$



$$\begin{aligned} \mu_{\tilde{A}}(p_1) &= 1 \\ \mu_{\tilde{A}}(p_2) &= 0.8 \\ \mu_{\tilde{A}}(p_3) &= 0.1 \\ \mu_{\tilde{A}}(p_4) &= 0 \\ &\dots \end{aligned}$$



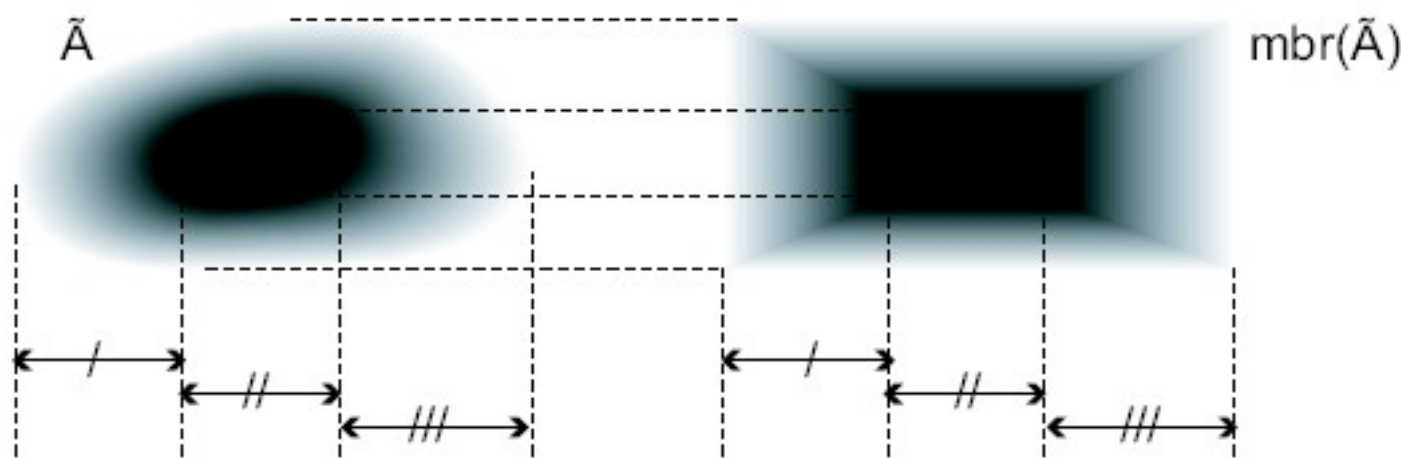
Operazioni su “Regioni dai confini sfocati”

Le operazioni su insiemi sono state estese in ambito fuzzy:

unione $A \cup B = \{p, \mu_{A \cup B}(p) \mid \mu_{A \cup B}(p) = \max(\mu_A(p), \mu_B(p))\}$

intersezione $A \cap B = \{p, \mu_{A \cap B}(p) \mid \mu_{A \cap B}(p) = \min(\mu_A(p), \mu_B(p))\}$
 $\mu_{A \cap B}(p) = \sup\{\alpha \mid \alpha \in]0, 1] \wedge p \in \text{MBR}(A_{\alpha})\}$

complemento $\neg A = \{p, \mu_{\neg A}(p) \mid \mu_{\neg A}(p) = 1 - \mu_A(p)\}$
 $\alpha\text{-cut}(A)$

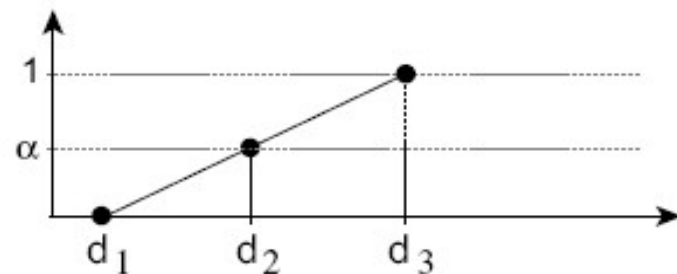
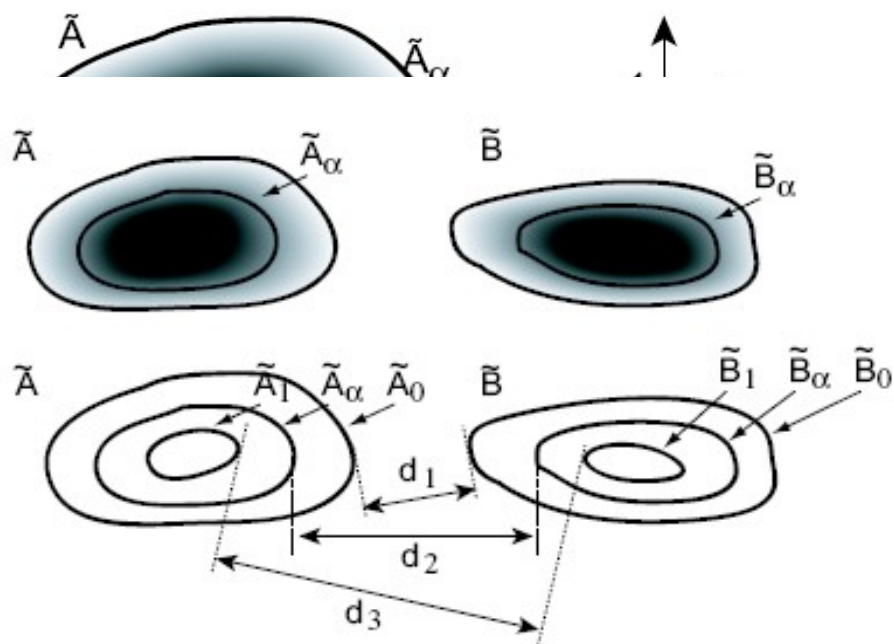


Proprietà geometriche di “Regioni dai confini sfocati”

Le proprietà geometriche sono estese in modo da assumere valori “fuzzy”:

Fuzzy Area:

Fuzzy distance:





Relazioni Topologiche tra “Regioni dai confini sfocati”

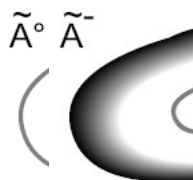
Il modello
[0,1] di s
rapprese
A°, il con

$$\begin{aligned} \tilde{A}^\circ \cap \Delta B &= \emptyset \\ \Delta A \cap \tilde{B}^\circ &= \emptyset \end{aligned} \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & b & a_1 \\ 1 & a_2 & 1 \end{pmatrix} \quad \begin{matrix} \tilde{A} \\ \tilde{B} \end{matrix}$$

where $b \in]0, 1], a_1 \in]0, 1], a_2 \in]0, 1]$

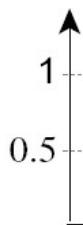
$$\begin{aligned} \tilde{A}^\circ \cap \Delta B &= \emptyset \\ \Delta A \cap \tilde{B}^\circ &\neq \emptyset \end{aligned} \quad \begin{pmatrix} 0 & c_1 & 1 \\ 0 & b & a_1 \\ 1 & a_2 & 1 \end{pmatrix} \quad \begin{matrix} \tilde{A} \\ \tilde{B} \end{matrix}$$

where $b \in]0, 1], c_1 \in]0, 1[, a_1 \in]0, 1], a_2 \in]0, 1]$



$$\begin{aligned} \tilde{A}^\circ \cap \Delta B &\neq \emptyset \\ \Delta A \cap \tilde{B}^\circ &= \emptyset \end{aligned} \quad \begin{pmatrix} 0 & 0 & 1 \\ c_2 & b & a_1 \\ 1 & a_2 & 1 \end{pmatrix} \quad \begin{matrix} \tilde{A} \\ \tilde{B} \end{matrix}$$

where $b \in]0, 1], c_2 \in]0, 1[, a_1 \in]0, 1], a_2 \in]0, 1]$



$$\begin{aligned} \tilde{A}^\circ \cap \Delta B &\neq \emptyset \\ \Delta A \cap \tilde{B}^\circ &\neq \emptyset \end{aligned} \quad \begin{pmatrix} 0 & c_1 & 1 \\ c_2 & b & a_1 \\ 1 & a_2 & 1 \end{pmatrix} \quad \begin{matrix} \tilde{A} \\ \tilde{B} \end{matrix}$$

where $b \in]0, 1], c_1 \in]0, 1[, c_2 \in]0, 1[, a_1 \in]0, 1], a_2 \in]0, 1]$

Gradi in
ologiche
o
sfocate



Altre applicazioni della teoria degli insiemi fuzzy alla gestione di informazione geografica imperfetta

Modellazione di attività Decisionali → operatori fuzzy

Ragionamento approssimativo → logica fuzzy

Classificazione → fuzzy clustering

Elicitazione della conoscenza → fuzzy data mining



Integrazione di informazioni imperfette

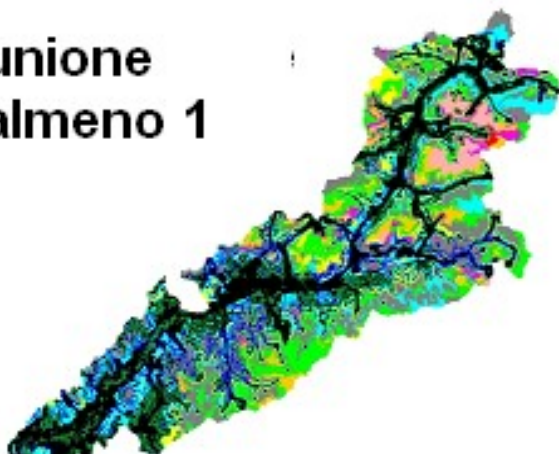
Integrazione di 21 mappe di impatto ambientale

mediante operatori fuzzy

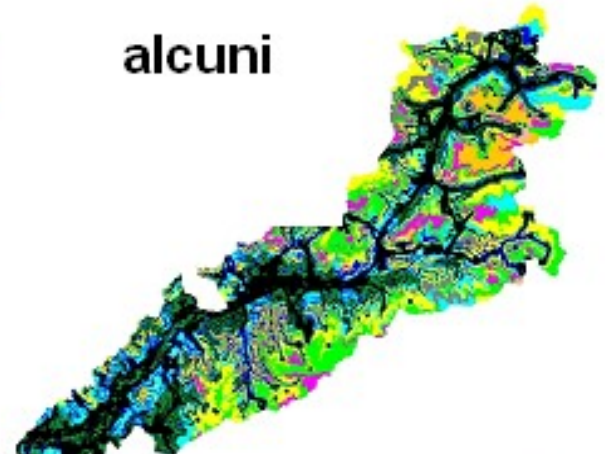
(definiti da quantificatori



unione almeno 1



alcuni



- Aggregano valori "fuzzy"
- Affidabilità, reputazione dell'esperto -> importanza
- Stimare il Consenso tra gli esperti -> misura di accordo
- Tenere conto del contesto spaziale (locale)



istemi basati su regole fuzzy pe il ragionamento approssimativo

Base di fatti: collezione di proposizioni elementari precise o vaghe

X è A, dove **A=alto** oppure **A=u₀** preciso;

Es: **Il versante X ha una pendenza *mediamente elevata*, litologia è *ghiaiosa* e esposizione è *alta***

Base di regole: collezione di regole fuzzy che codificano la conoscenza vaga su un fenomeno

“se X è A allora Y è B” dove sia A sia B possono essere vaghi

Es: **regola-1: Se il versante ha una pendenza *elevata*, litologia *ghiaiosa con argilla*, e l'esposizione è *280°* allora ha un'*alta* franosità**

• **regola-2:.....**

Regole d'Inferenza (Modus Ponens Generalizzata)

Permette di dedurre un fatto relativo a Y partendo dalle premesse

“Y è B”

Es: **Il versante X ha una franosità *medio-alta***

PREMESSE

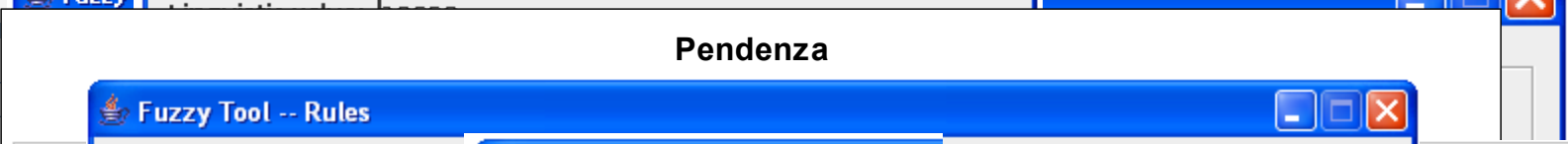
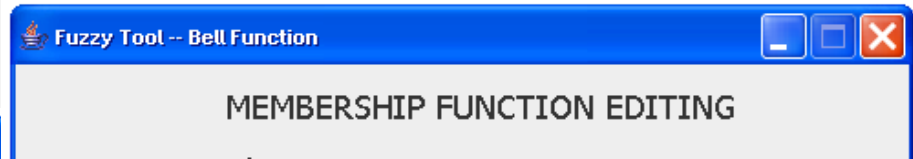
Conclusione



Esempio: stima della pendenza dalle frane

VARI

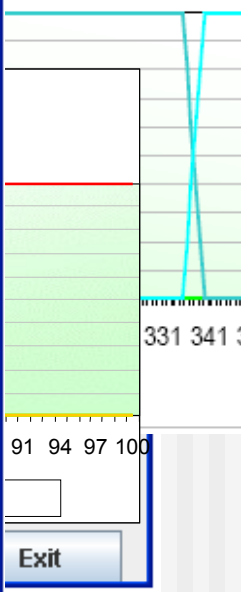
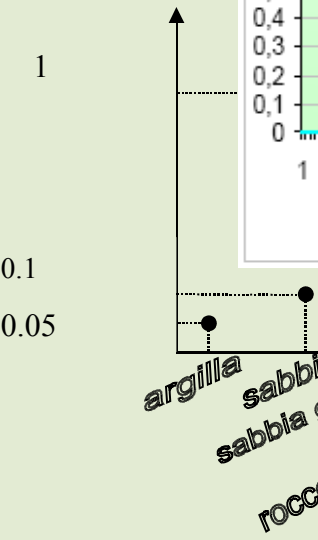
- P
- I
- e

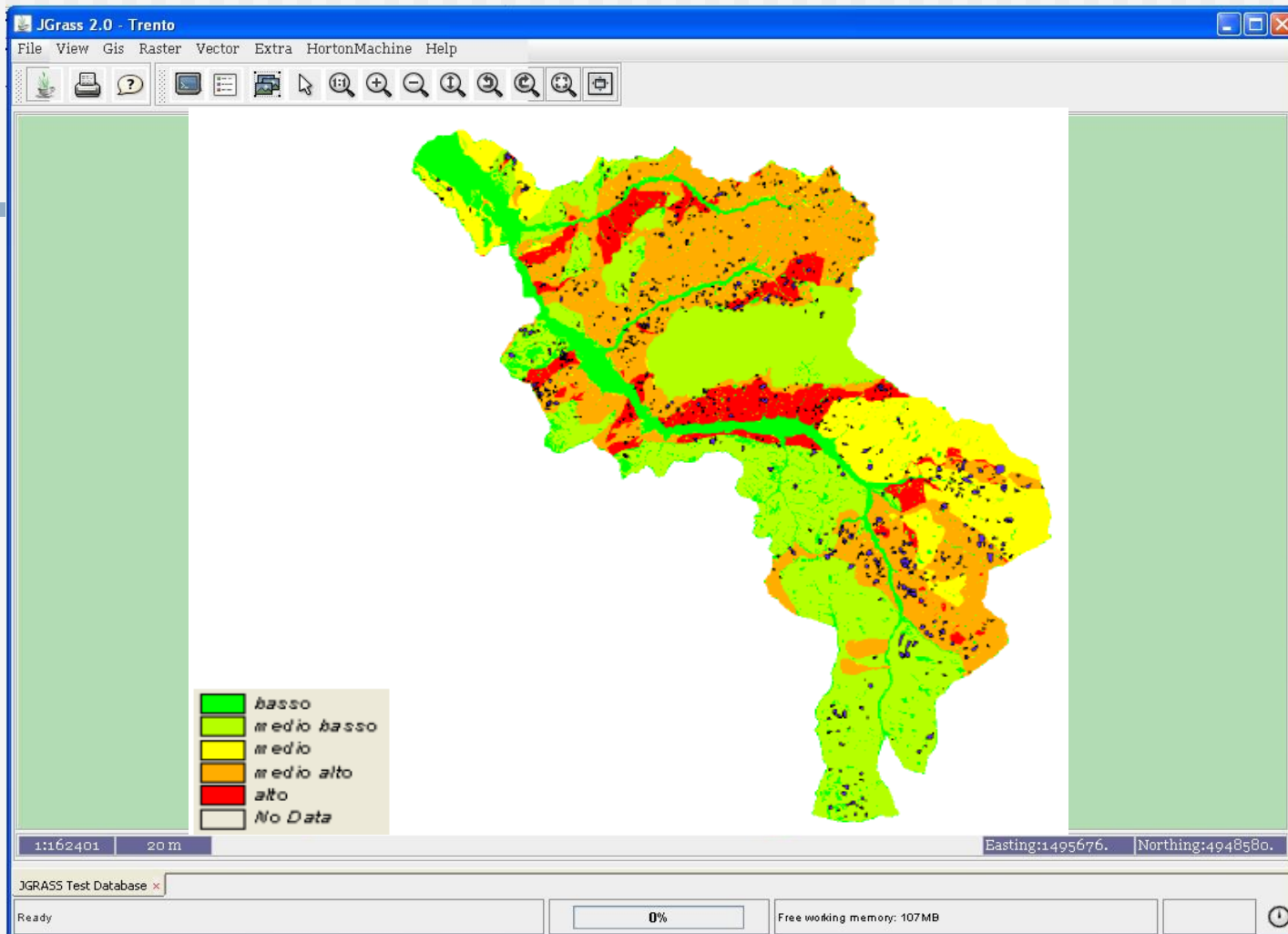


JGrass 2.0 - Trento

- File
- Fuzzy Modeling Tool
- View
- Gis
- Help

- Fuzzy Modeling Tool
- Configuration
- Linguistic Variables
- Rules
- Settings
- Run...





Sviluppi futuri

Implementazioni :

- Es. estensione di GRASS 2004: Dilo A. et Al, ITC Eschede, The Netherlands
- Es. Tagger, Evans, e Waters, 2008 Univ. Leeds UK, <http://www.ccg.leeds.ac.uk/software/tagger/>

Bibliografia

- **Aronoff S., 1989**, Geographic Information Systems: A Management Perspective. WDL Publications, Ottawa.
- **Bordogna G., Chiesa S., Geneletti D., 2006** "Linguistic modelling of imperfect spatial information as a basis for simplifying spatial analysis, Information Sciences, Elsevier, (176) 366-389, 2006.
- **Bosc, P., Prade, H., 1997** "An Introduction to the Fuzzy Set and Possibility Theory-based Treatment of Flexible Queries and Uncertain or Imprecise Databases", in: *Uncertainty Management in Information Systems*, A. Motro, P. Smets (eds.), Kluwer Academic Publishers, 285-324
- **Bosisio, R., Guerrieri V., tesi di laurea specialistica in Informatica, Univ. Milano Bicocca, 2006**
- **Burrough P.A., Frank A.U. eds, 1996**. Geographic Objects with Indeterminate Boundaries, in GISDATA series, Taylor & Francis,
- **Carrara P., Bordogna G., Stroppiana D., Brivio P.A., Nelson A., Boschetti M. 2008** A flexible multi-source spatial data fusion system for environmental status assessment at continental scale, Journal of Geographic Information Science, Vol.22, No.7, 2008.
- **Codd M., Petry F., Robinson V. eds, 2000** Uncertainty in Geographic Information Systems and Spatial data, special issue of Fuzzy Sets and Systems, 113,(1), 1-159.
- **de Caluwe R., de Trè G., Bordogna G. eds, 2004** Spatio Temporal Databases: flexible querying and reasoning, Springer Verlag, 2004.
- **Dubois, D. e Prade, H. 1988**. *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. New York: Plenum Press.
- **Dilo A., Kraipeerapun P., Bakker W., de By R. A., 2004** "Storing and handling vague spatial objects", in Proc. of the 15th Int. Workshop on Database and Expert Systems Applications (DEXA'04)
- **Frank A. 2006** Incompleteness, error, approximation and uncertainty: an ontological approach to data quality, 2006 in **A Morris eds** Geographic Uncertainty in Environmental security, Springer Verlag , NATO advanced research Workshop, Kiev, Luglio 2006
- **Goodchild M.F. and Gopal S. eds 1989** Accuracy of spatial databases, London Taylor and Francis
- **Jankowski P., Nyerges T., 2001**, Geographic Information Systems for Group Decision Making, Taylor and Francis Pub., UK.
- **Malczewski J. 2006** GIS-based multicriteria decision analysis: a survey of the literature, Int. J. of Geographical Information Science, 20(7), 703-726
- **Morris A, Kokhan S. eds 2006** Geographic Uncertainty in Environmental security, Springer Verlag , NATO advanced research Workshop, Kiev, Luglio 2006
- **Pawlak, Z.: 1982 Rough Sets. Int. Journal of Computer and Information Sciences 11 (1982) 341-356**
- **Plewe B., T(2002)** The nature of uncertainty in historical geographic information, Trans. In GIS, 431-456
- **R. Yager and L.A. Zadeh (eds.), 1992** "An Introduction to Fuzzy Logic Applications in Intelligent Systems" Kluwer Academic Publisher, Dordrecht, 1992
- **Robinson, V.B., 2003** A perspective on the fundamentals of fuzzy sets and their use in Geographic Information Systems. Transactions in GIS 7(1), 3-30.
- **Schneider M., Pauly A., 2007** ROSA: An Algebra for Rough Spatial Objects in Databases, J.T. Yao et al. (Eds.): RSKT 2007, LNAI 4481. Springer-Verlag Berlin Heidelberg, pp. 411-418,
- **Timpf S., Raubal M., Kuhn W., 1996** Experiences with metadata 7th int. Symp. On Spatial Data Handling SDH'96, Delft, The Netherlands.
- **Veregin H. 1989** Error modeling for the map overlay operation in Goodchild M.F. and Gopal S. eds 1989 Accuracy of spatial databases
- **Verstraete J., 2007** Fuzzy modelling of Spatial Information, Tesi di dottorato, Univ. Gand, 2007

Grazie per l'attenzione!

Domande?