

Il ruolo della statistica in un sistema di supporto alle decisioni per la protezione civile

Elisabetta Carfagna
elisabetta.carfagna@unibo.it

Università di Bologna
Departmento di Scienze Statistiche

Il ruolo dell'informazione geo-referenziata nella gestione delle emergenze
Ciclo di seminari «Numeri per decidere»



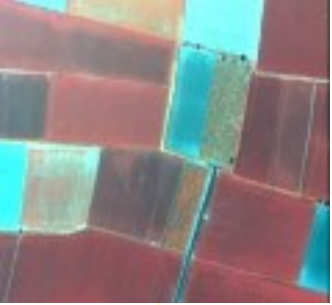
Argomenti della presentazione

- Analisi statistica di dati geo-referenziati in un sistema di supporto alle decisioni per la protezione civile
- Valutazione di impatto del sistema di supporto alle decisioni e dell'intero progetto



Analisi statistica di dati geo-referenziati

- Valutazione del rischio di insorgenza del disastro
- Analisi di procedure emergenziali e piani di intervento sviluppati (con altri partner regionali)
- Definizione di misure di valutazione dell'efficacia e dell'efficienza delle attività di risposta ai disastri
- Barriere non tecnologiche, accettazione sociale, fattori umani, etici e legali
- Accettazione sociale ecc. implica sviluppo di questionari, selezione di utilizzatori finali statisticamente rappresentativi, analisi di dati spesso qualitativi
- Analisi dei dati a sostegno di un sistema di supporto alle decisioni, inclusa l'identificazione dei principali limiti dell'intervento e principali fattori di successo



Implicazioni statistiche

- Combinazione di dati non comparabili
- Diverse unità di misura, campi di variazione ecc.
- Diversi approcci possibili (e.g. ranking, z-scores, re-scaling, categorical scale, etc.) per analizzare i dati e applicare diverse trasformazioni per individuare quella che consente la comparabilità preservando le principali caratteristiche della base di dati
- Analisi multivariata
- Costruzione di indicatori composti che consentano di misurare aspetti multidimensionali



Valutazione d'impatto del sistema

- Valutazione di impatti diretti ed indiretti del sistema di supporto alle decisioni sulla gestione delle emergenze
- Valutazione di impatto dell'intero progetto
- Valutazione dell'impatto delle diverse componenti del Sistema:
- uso di tecnologia avanzata e metodologie innovative
- Organizzazione
- Coordinamento
- Individuazione delle situazione attuale di riferimento (baseline)
- Analisi controfattuale (interventi virtuali, scenari)

Risoluzione spaziale per la quale varie tipologie di dati sono disponibili

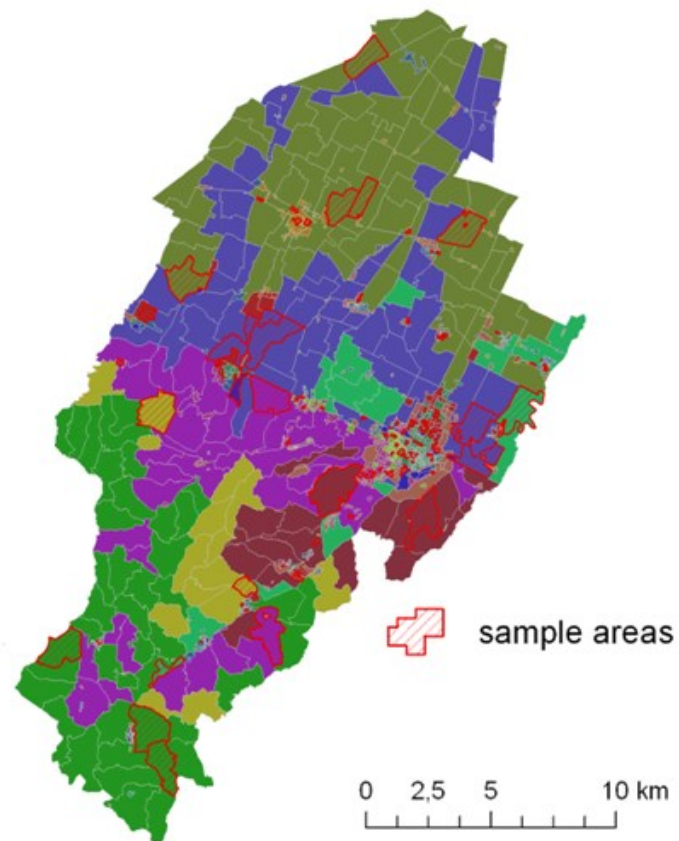


Famiglie
Sezioni di censimento
Aree amministrative

STRATIFIED SAMPLE

Legend

- 1-a'
- 1-a.o
- 1-a.p
- 1-a.h
- 1-b
- 1-c
- 1-c'
- 1-de
- 1p
- 2-a
- 2-b
- 2-cde
- 3-a
- 3-cd
- 4-cd
- 4-e



Dati geografici

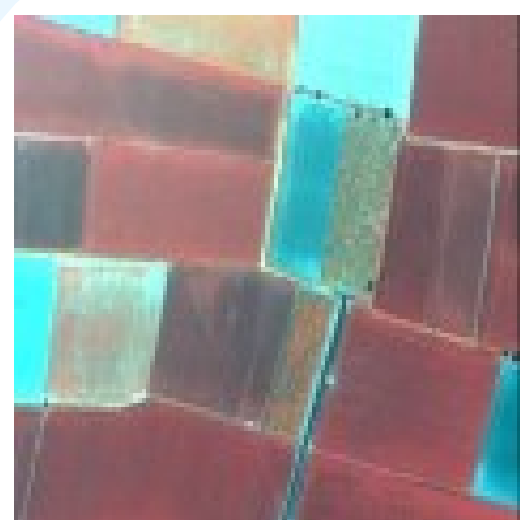
Usò del suolo;
Copertura del suolo;
Pendenza;
Esposizione;
Altre caratteristiche biofisiche e geofisiche



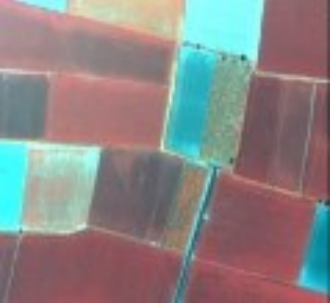
Landsat TM
(30m dimensione pixel)



Sentinel 2
(10m dimensione pixel)



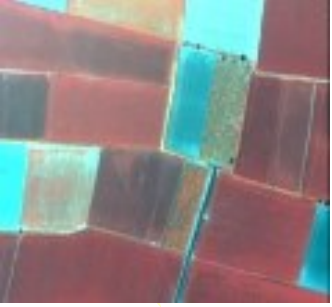
Ikonos
(4m dimensione pixel)



Cambiamento del supporto

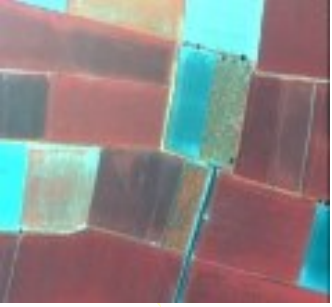
- Necessità di combinare dati osservati su supporti diversi:
 - Osservazioni su punti
 - Linee
 - Aree geometriche (quadrati, rettangoli, ecc.)
 - Aree irregolari (in alcuni casi generate dall'interpolazione di dati puntuali – quali dati meteorologici)
- Quando il supporto del processo spaziale di interesse è diverso da quello dei dati osservati, nasce un problema di cambiamento di supporto

Acronym	Predictors	Model
EBLUP	Empirical Best Linear Unbiased Predictor	$y_{ij} = \beta_0 + \beta_1 x_{ij} + a_i + e_{ij}$
EBLUP_GC	EBLUP plus the geographical coordinates	$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 lon_{ij} + \beta_3 lat_{ij} + a_i + e_{ij}$
GREG	Generalized Regression Estimator	$y_{ij} = \beta_0 + \beta_1 x_{ij} + e_{ij}$
GREG_GC	GREG plus the geographical coordinates	$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 lon_{ij} + \beta_3 lat_{ij} + e_{ij}$
SEBLUP	Spatial EBLUP	$y_{ij} = \beta_0 + \beta_1 x_{ij} + v_i + e_{ij}$
MBDE	Model Based Direct Estimator	$y_{ij} = \beta_0 + \beta_1 x_{ij} + a_i + e_{ij}$
SMBDE	Spatial MBDE	$y_{ij} = \beta_0 + \beta_1 x_{ij} + v_i + e_{ij}$
GWEBLUP	Geographically Weighted EBLUP	$y_{ij} = \beta_0(lon, lat) + \beta_1(lon, lat) x_{ij} + a_i + e_{ij}$
MQ	M-quantile model	$y_{ij} = \beta_{0q} + \beta_{1q} x_{ij} + e_{ij}$
MQ_GC	MQ plus the geographical coordinates	$y_{ij} = \beta_{0q} + \beta_{1q} x_{ij} + \beta_{2q} lon_{ij} + \beta_{3q} lat_{ij} + e_{ij}$
GWR	Geographically Weighted Regression	$y_{ij} = \beta_0(lon, lat) + \beta_1(lon, lat) x_{ij} + e_{ij}$
KRIG	Kriging	$y_{ij} = lon + lat + e_{ij}$



Impact of use of geo-referenced data on statistical yield forecasting models

- Main explanatory variables of statistical yield forecasting models:
 - Trend of historical yields
 - Agro-meteorological models (generally deterministic models like Penman's), which account for water and temperature stress
- Statistical models require variables with common support (polygons or, more frequently raster) obtained through a series of GIS operations:
 - interpolation of point data (e.g. meteorological data, soil data, etc.)
 - disaggregation and aggregation of database layers (land cover, crop masks, remote sensing data etc.)
- Errors in explanatory variables, due to interpolation, location, change of support etc.



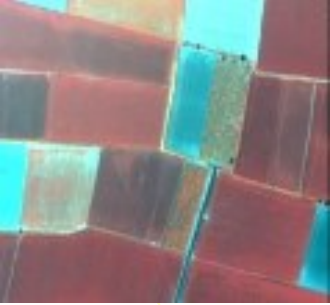
Data combined with other kinds of data in a statistical model

- Uncertainty or bias of certain input data produce a propagated impact on the output
- Four branches related to this issue:
 - methodologies to measure uncertainty induced by some input in final results;
 - operative indications at design stage of the procedure to reduce the impact of those disturbs on the final result;
 - methods to correct the disturb caused by some specific processing procedure on the input data;
 - specific statistical and mathematical procedures adjusted to take into account these disturbs.



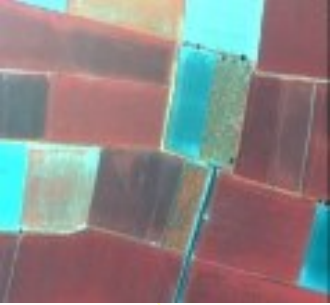
Analytical approach

- Analytical approach, through an extremely detailed “error” modelling and have derived analytically the consequences of the impact of the error on the final model.
- Crosetto and Tarantola (2001): major limitation of the analytical approaches is that it applies only to particular kinds of GIS operations, or to particular types of data.
- Given complexity of relations among different information layers, analytical approach has not been followed extensively.



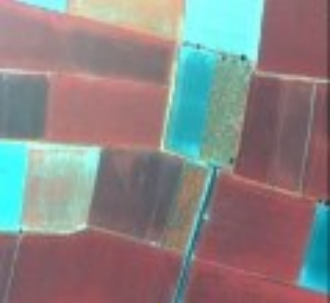
A-posteriori analysis

- A-posteriori analyses carried out with the aim to optimize procedures of data processing, to obtain best result in terms of accuracy of final estimates.
- A-posteriori approach:
 - does not try to include the source of disturb/error into the final model
 - tries to simulate/highlight problems to measure impact and minimize effects
- Veregin (1994): simulation modelling is an attractive alternative when little is known about error propagation mechanisms and can be used to:
 - simulate the effects of GIS operations on the data
 - quantitatively assess the impact on the final model
 - Where possible, to optimize the GIS procedure.



Sensitivity analysis

- Cancellieri *et al.* (1993) adopted sensitivity analysis approach for measuring influence of different variables and effects of GIS operations on statistical yield forecasting model
- Saltelli, *et al.* (2012) distinguish between uncertainty and sensitivity analysis:
 - uncertainty analysis is responsible for analysing the propagation, into results of models, of the uncertainty embedded in some variables. The uncertainty analysis answers the question "how reliable / uncertain this model is?"
 - sensitivity analysis is concerned with measuring the strength of the impact / relationship between variables and model. That is: "what is the impact of "each" factor on the variability of the final outputs?"



Ruolo della statistica

- Ruolo fondamentale della statistica
- Analisi dei dati
- Individuazione di relazioni
- Stima su punti non osservati
- Stima per piccole aree
- Sviluppo di questionari
- Campionamento statistico - campionamento spaziale
- Combinazione ed integrazione di dati non comparabili
- Trasformazioni ecc.
- Combinazione di diversi piani informativi in un GIS
- Valutazione dell'impatto delle trasformazioni dei piani informative in un GIS
- Valutazione dell'errore dei modelli

Bibliografia

- Arbia, G. & Petrarca, F.: Effects of Scale in Spatial Interaction Models. Journal of Geographical Systems (2013)
- Benedetti R.: Developing more efficient and accurate methods for using remote sensing, Second meeting of the Scientific Advisory Committee of the Global strategy to Improve Agricultural and Rural Statistics, FAO Headquarters, 29 - 30 January (2014)
- Cancellieri M.C.: Carfagna E., Narciso G., Ragni P.: Aspects of Sensitivity Analysis of a Spectro-Agro-Meteorological Yield Forecasting Model, EARSeL (European Association of Remote Sensing Laboratories) Advances in Remote Sensing, vol. 2, n. 2 - VI, , pp. 124-132 (1993)
- Carfagna E.: Multiple Frame Sample Surveys: Advantages, Disadvantages and Requirements, in International Statistical Institute, Proceedings, Invited papers, International Association of Survey Statisticians (IASS) Topics, Seoul August 22-29, pp. 253-270 (2001)
- Carfagna E.: Using satellite imagery and geo-referencing technology for building a master sampling frame, the 59th World Statistical Congress, Invited Paper Session (IPS110) Developing a master sampling frame for integrated survey, Hong Kong, 25-30 August (2013) <http://www.statistics.gov.hk/wsc/IPS110-P1-S.pdf>
- Carfagna, E., Carfagna, A.: Alternative sampling frames and administrative data; which is the best data source for agricultural statistics?, in R. Benedetti, M. Bee, R. Espa & F. Piersimoni, eds. Agricultural Survey Methods. Chichester, UK, Wiley (2010)
- Carfagna, E., Carfagna, A.: Appropriate Survey Methods for Different Country Profiles - Key Challenges, Gaps and Remaining Methodological Issues, ICAS VI, technical session IDCB 7 - Estimates, forecasts, expert opinion and assessment - their role in the official statistics, Rio de Janeiro, 23-25 October (2013)
<http://eventos.ibge.gov.br/en/icas2013/programa-da-conferencia/programa-detalhado>