[Article] Information and communication: two special issues for Geomatics

Original Citation:

Availability:
This version is available at: http://porto.polito.it/2294158/ since: December 2009

Publisher:
Politecnico di Torino

Terms of use:
This article is made available under terms and conditions applicable to Open Access Policy Article ("Public - All rights reserved"), as described at http://porto.polito.it/terms_and_conditions.html

Porto, the institutional repository of the Politecnico di Torino, is provided by the University Library and the IT-Services. The aim is to enable open access to all the world. Please share with us how this access benefits you. Your story matters.

(Article begins on next page)
Information and communication: two special issues for Geomatics

Introduction

In a process lasting many centuries, the development of surveying techniques overpasses early agrarian and astronomical knowledge by the rigorous application of Mathematics and Statistics to the treatment of experimental data, and have become formally established as a science with fundamental support from the theoretical assets of Geometry and Cartography.

Thanks to the many achievements in space exploration and the rapid advancements of the Information Science and Technology witnessed in the last decades, the newly available data (observations) have utterly grown both in quality and quantity, making it necessary to investigate new methodologies in data processing and computational statistics.

Moreover, in the last few years, the exchange of information has increased exponentially, as dragged by the impetus of development of media.

In this context, basic Geometry and Mathematical Analysis have been found largely inadequate to build gray models manageable with Statistics, whereas Discrete Mathematics is of help in assembling black models. Recently it became clear how the technological approach of Engineering alone to the handling of Information and Communication is doomed to fail, and experience has been borrowed from other areas of knowledge, such as Human Sciences (Linguistics and Psychology) and Cognitive Science (Artificial Intelligence and Neuro-Sciences). The hope is that Human Sciences will unveil new horizons along the difficult path from language to media, which may be of great help for a better interpretation of information.

The answer of Philosophy

We can see that the same definition of science, coming from the Sophists, or rather from the sages, considering its etymological meaning, who knew how to find an intervention procedure in a fluent society, indicates the methodological support that can be supplied by philosophy for Geomatics.

Espero o kumart pros to epóntero (Eldênas kai kosmikèn pros tò kártatè), to put the logical structure of the observed phenomenon, in order to perform.

The logical patterns are then directed to the composition of the «kosmos», which, obviously, stays opened, because we continue observing it and reshaping it. This deals with defining what exactly is cartography, which is carried out by using information processing instruments and a large amount of news
acquired from satellite-based sensors and, of course, interpretative, selective proceedings should be applicable to them).

Scientists (and those who work in Geomatics), while retaining the dignity of abstract theories, must avoid closed systems, which get stuck in ideology and, moreover, they must claim, for what concerns the cosmological theories, which are addressed to the multiplicity of phenomena, their right to put forward organic proposals appropriate to problems.

It therefore an essential requisite that both the survey and the product of Geomatics, which are clearly directed to an operational use, avoid seeking a proud accumulation of large amounts, losing sight of the operational aim — from the pretexts — and becoming like the Socrates: an unorganized heap, which remains the same, even if adding or subtracting any unit (information).

But if man, following the Sophistic thought, knows how to experience and organize, in order to work (Protagoras: «ιδεῖν τὰ ἐξ ἐμοῦ ἑξ ἐμοῦ» — Plato, Kresimás rítríon káthodos), the use of open source programs, which give an undeniable support to that particular ability «to know how audacious» (to know how to dare), led by the Kantian teaching, to have a large amount of inquiring elements, linked to practical problems, being able to work, with an autonomous organizational capacity, towards achieving projects suitable for the improvement of the human condition.

The answer of Statistics

Data acquisition has, for a long time, required heavy engagement both of manpower and technical support. On the contrary, at present time, large sets of data are easily acquired. However this enormous amount of information contains often data external to the distribution under study (outliers) or correlated data which should be treated with particular care. Indeed to neglect correlation means to consider much more information available than in the reality.

This means to supply estimates with relative errors, one or more times better than the correctly expected ones. Furthermore the same developments show that only innovations, of full rank, give valid contributions, to improve the estimates themselves.

Given a standard Gauss-Markov least squares problem, the normal matrix, coming from a linear functional model and the classical stochastic model, where all observations are assumed to be un-correlated, has the form:

\[ C = A^T P A \]

being \( A \) the design matrix and \( P \) the weight matrix.

On the contrary, if the "observations" are assumed to be "correlated", the normal matrix becomes:

\[ D = A^T (P^{-1/2}(R - I)P^{-1/2} + P^{-1/2}P^{-1/2})A \]

where the coefficients \( r \) of the matrix \( R \) represent the correlation among the observations. Consequently the inverse matrix of the second normal matrix \( D^{-1} \) is different, with respect to the inverse matrix of the first normal matrix \( C^{-1} \):

\[ D^{-1} = A^T (P^{-1/2}(R - I)P^{-1/2} + P^{-1/2}P^{-1/2})A^{-1} \]

and its main diagonal elements are larger than the corresponding elements of the first one.

Furthermore given a standard Gauss-Markov least squares problem, the normal matrix, coming from a linear functional model, containing also a priori un-correlated information on parameters, and the classical stochastic model, has the form:
AMBIENTE E SVILUPPO SOSTENIBILE

\[ C = A^TPA + (I - Q)^{1/2} \]

where \( Q \) is the weight matrix of the pseudo-observations on the parameters.

On the contrary, if the "a priori information on parameters" are "correlated", the normal matrix becomes:

\[ D = A^TPA + (I - Q)^{1/2}(R - I)(I - Q)^{1/2} \]

where the coefficients \( r \) of the matrix \( R \) represent the correlation among the a priori information on the parameters. Consequently the inverse matrix of the second normal matrix \( D^{-1} \) is different, with respect to the inverse matrix of the first normal matrix \( C^{-1} \).

\[ D^{-1} = (A^TPA + (I - Q)^{1/2})(R - I)(I - Q)^{1/2} \]

\[ = (A^TPA + (I - Q)^{1/2})(R - I)(I - Q)^{1/2} \]

\[ = C^{-1} + C^{-1}(R - I)(I - Q)^{1/2} \]

\[ = C^{-1} + C^{-1}(I - Q)^{1/2}(R - I)(I - Q)^{1/2} \]

\[ = C^{-1} + C^{-1}(I - Q)^{1/2}(R - I)(I - Q)^{1/2} \]

\[ = C^{-1} + C^{-1}(I - Q)^{1/2}(R - I)(I - Q)^{1/2} \]

\[ = C^{-1} + C^{-1}(I - Q)^{1/2}(R - I)(I - Q)^{1/2} \]

\[ = C^{-1} + C^{-1}(I - Q)^{1/2}(R - I)(I - Q)^{1/2} \]

and its main diagonal elements are larger than the corresponding elements of the first one.

Notice that the main diagonal elements of both the inverse matrices are vice versa elements of the corresponding elements of the inverse matrix without any a priori information.

In the previous expressions, since \((R-I)\) is not positive defined (being null the trace of the matrix), we must consider different cases. Indeed correlation may also play in favour.

Notice that correlation may also play in favour of variance reduction:

- in the mean with negative correlation;
- in the difference, with positive correlation, with positive correlation (as the functional model has itself a minus sign).

On the other side, since one generally computes the mean of nearby data a possible correlation would obviously be positive.

For the computation of differences, a simple procedure is dealing with uncorrelated data (in space and/or time): positively correlated data, as it is widely known, are very close, even almost the same, and their differences are close to zero, that means uninteresting.

Multi-collinearity is the case when independent variables are inter-correlated: then confidence intervals for predicted values are very ample.

So, ill-conditioning may ensue, thus making estimated values very sensitive to random variability of samples.

Gauss-Markov theorem is the basis of the usage of LS linear regression. This theorem states that, in the field of linear unbiased estimators, least square estimator has the minimum variance: obviously, this statement does not prevent that other estimators (biased or not-linear) may be found, which have a lesser variance.

We may be tempted to choose different estimators, in case we can't accept defaults of Least square: specially the lack of robustness and the hard prediction.

As to the first question, robust techniques are available; for the second one, often predictions are at hand when one reaches a balance between bias and variance.

The bias-variance trade-off (or "bias-variance dilemma") is a very important issue in data modelling. Ignoring it, is a frequent cause of model failure, and although it has a deep theoretical rooting, it can be explained in simple terms:

- Models with too few parameters are inaccurate because of a large bias (not enough flexibility),
- Models with too many parameters are inaccurate because of a large variance (too much sensitivity to the sample).

Identifying the best model requires identifying the proper "model complexity" (number of parameters).

What is really important is that, on the average, the estimate \( \delta \) be close to the true value \( \delta \). So we will end to favour estimators such that the mean-square error be as low as possible, whether is biased or not:

\[ E[\delta - \delta]^2 = \min. \]

(Such an estimator is called a minimum mean-square-error estimator).

Mean square error can be split into variance, bias and noise:

\[ E[(\delta - \delta)^2] = E[(\delta - \delta)^2] + 2E[(\delta - \delta)(\delta - \delta)^2] + E[(\delta - \delta)^2] \]

\[ = E[(\delta - \delta)^2] + 2E[\delta - \delta]E[(\delta - \delta)^2] + E[(\delta - \delta)^2] \]

\[ = E[(\delta - \delta)^2] + 2E[\delta - \delta]E[(\delta - \delta)^2] + E[(\delta - \delta)^2] \]

As the problem of outliers, we have had for some decades the well tested robust methods of estimation.

The problem of correlation is faced with some warning indexes (condition number, Variance inflation factors, ... and with other approaches (ridge regression, multiple regression, principal components, ...).

Information, communication and language

Without unnecessary details, we would like to show the state of art of Language Theory.

The period from the 1940s through the end of 1950s saw two
The probabilistic paradigm placed a role in the development of speech recognition (use of the Hidden Markov Models by Jelinek and Rahmim, 1988) and promoted some interplay with fields of statistics and electronics, obtaining some successes in computational Linguistics.

The view that takes language as a tool of transmission of meaning is based on the codes model: this model relies upon the mathematical theory of Information of Shannon: the thought is coded by the speaker into an array of sounds, which in turn the listener will decode in order to extract the right meaning, to be shared with the speaker.

This means primacy of thought above Language.

Not everybody agree with this: some scholars think that this not comply with a number of examples from non-literal language:

- Chomsky, like Descartes, says that Language marks the difference between mankind and other animals. Indeed, human language is a rational system, both independent and self-related as to the other cognitive attitudes: firstly, independent from general intellect.

Chomsky states that language and grammatical ability are almost the same thing. In 1957 he introduced a hierarchy of classes of languages, based upon the complexity of the underlying grammar (Chomsky, 1957):

- Unrestricted grammars (0-type grammars)
- Context-sensitive grammars (1-type grammars)
- Context-free grammars (2-type grammars)
- Finite state or regular grammars (3-type grammars).

The most general class of grammar is type 0, which has no restrictions on the form that rewriting rules can take; for the other grammar types, the form of the rewriting rules is increasingly restricted and the generated languages are correspondingly simple. The simplest languages, generated by 2 type and 3 type grammars, are unable to describe human languages. It has become clear that the problem of handling computationally the most general formal language is a highly challenging task. Context-free grammars are not able to represent every natural language, so TAG (Tree Adjoining Grammars) were devised and gave place to "Mildly sensitive grammars", which are suited to express human languages.

House, Chomsky and Fitch (1995) consider a faculty of language at large (FLB, Faculty of Language Broad) and a faculty of language in the narrow sense (FLN, Faculty of Language Narrow). The first (FLB) is made up with FLN and two more elaboration systems (at least); the second (FLN) refers simply to the computational linguistic system, taken as itself.

The Minimalist Program (MP) is the present view of Chomsky, and shows a remarkable change from his previous theory (transformational grammar). This is summarized in the following.

Human language is the optimum solution to the problem of expressing thought through sound. Linguistic expression is formed under three conditions:

- computational limits;
- interface between computational limits and sensory-motorial system;
- interface between computational system and conceptual system.

Principles of universal grammar are a consequence of combined effect of the former three limits, and this is new.

At any step of derivation of linguistic expression, phonetical and semantical compatibility must be granted.

Chomsky says: "...although many aspects of FLB are shared with other vertebrates, the core recursive aspect of FLN currently appears to
lack any analog in animal communication and possibly other domains as well... FLN contains a wide variety of cognitive and perceptual mechanism shared with other species, but only those mechanisms underlying FLN, particularly its capacity for discrete infinity are uniquely human."

So, the scholars who agree with Chomsky, think that human language has two distinctive proprieties:

- signs may be used even out of context;
- signs may be built up.

For instance, special "dances" of bees are out of context, but are not built up.

This is the opinion of Chomsky, opposed to the one of Jackendoff and Pinker (2003), who had rejected the recursivity as a main character of human language, also stating that Chomsky had opposed to the evolutionary Darwin's thought.

Nowadays, scholars mainly agree that:

- notwithstanding the fact that the form of communicative signal is linear, syntax is hierarchical and organization of words;
- all human grammars appear to come down from an universal scheme (that is the old dream of a Mendelev's table for human languages);
- present techniques of neuroimagery cooperate to prove these assumptions: actually, it may be found that the structure of human adult brain can catch the differences between rules of human syntax and rules which contrast with them (however, early in 1861, Pierre Bocca had suggested that the seat of human language stays in a well defined area of brain).

A. Turing has proved that a so-called universal Turing machine is able to develop any computational sequence, when it is based upon a real procedure. So, some think that it is possible to generalize these concepts to the behavior of human brain.

However, other scholars in the field of neuro-sciences have stressed the point that in the development of brain faculties, a number of casual events is involved, and this would not agree so plainly with the similitude brain-universal machine.

The Nobel prize winner (1972) G. Edelman says that no matter how wonderful the technological progress may appear, the roots of scientific imagination are not quite different from the origins of Poetry, Art in general (or even Ethics), as that a clear separation between Sciences and Human Disciplines becomes irrelevant and unnecessary.

**Checking quality**

The field of Data Processing meets the realm of Human Sciences, by sharing expertise and learning from one another. This is particularly relevant in Geomatics, where images, maps and 3D graphics are subjected not only to be modeled and computed, but also recognized and understood. In this frame, Linguistics (phrase – structure grammar, pattern recognition, parsing), Communication techniques, Psychology (e.g. Gestalt theory) and Philosophy of Science (a bridge between Epistemology and History of Science and Technique) contribute to form cognitive tools.

On the other hand, the survey and mapping disciplines, particularly Photogrammetry, Remote sensing and GIS, can give very important contribution to many human activities, but they can also be negatively used to support, in very sophisticated ways, the escalation of armies, wars and destruction, being able to produce devastating effects both in intensity and extension.

The development of science and technology must take into account human life, while positively playing a fundamental role in the real world. A peaceful use of mature and innovative technologies towards the achievement of concrete social, economic and cultural benefits for all the people must be encouraged; this would increase to a high-level of standard the quality of life of the human gender.

Cultural tools proper of human Sciences, must be used in order to face:

- The world of space-referenced GIS (time varying or not);
- The wide ensemble of images (from digital image processing to 3D models);
- The widening structures of information (such as Internet and other).

We may derive from Human Sciences the well-known diamond scheme, and use it for circulation of information:

- Human society is the sum of economical, political, social and cultural needs;
- Messenger makers, are the authors of the mass and popular culture;
- Receivers are common people and, as reciprocal exchanges, the same messengers;
- Cultural Object is the commun-shared meaning, enclosed in a form.

Information inside the Informative Systems (spatially referenced or not), is fully compatible with the proposed sketch; actually, information systems may be taken as a part of universal media.

In post-modern civilization,
a special issue is development of networks: this is the natural evolution of XIX-century railways and of XX-century highways. So, Internet's WWW is not simply a "software", but also a cultural active element.

A correct information passes through a differently conceived Cartography: see the Peters' atlas (Arno Peters, 1980), edited in the framework of Willy Brandt's report about North-South.

It is an equivalent projection, specifically an equal-area one, which shows in the simplest way the real proportions of different areas in the planet, other than usual maps, which generally stress the relative weight of different areas (as to population, resources, economic development...).

The opposite approach is also useful, precisely when one wants to stress the differences in situation or usage of various areas, neighboring or not.

In this case, one seeks to keep the shape recognizable, thus zooming the site, as compared to other areas.

In a special case, the pro-capita consumption of energy, Holland would appear much larger than India or Mexico.

Peters's map would have emphasized the gap between rich and undeveloped countries: actually, by representing the true sizes of continents one can appreciate the real vastness of the southern part of the dry land, which is the less developed. It may be seen at first sight the presence of problems of peaceful coexistence of the poor and the rich in the world.

Another case could be the map of population, of income, general or pro-capita, or any social question (Figures 3-6).

These maps are not realistic, being not equivalent-area maps, but they are the more representative insofar as they show areas which are directly proportional to the characteristics one wants to analyze.
Fig. 5. Thematic map of global consumptions. 
Carta tematica dei consumi globali.

Fig. 6 (a,b). Thematic map of families with over 200, below 20 dollars daily income. 
Carta tematica del reddito familiare superiore a 200, inferiore a 10 dollari al giorno.

An example: the orthodox monasteries and churches of Kosovo

Kosovo is a relative small area in Balkans (some 10,000 km² and 200,000 inhabitants).

The area has a mild, humid climate, with some Mediterranean influent; this means in principle an adequate agricultural support, a sufficient forest coverage, and generally favorable conditions for human development, also helped by a considerable mineral wealth (mainly coal, zinc, silver, lignites, ...).

In the area of Kosovo-Metohia have been present for many centuries populations of different nationalities and religions: Serbs, Albanians, Turks, Croats, Roma, ... (Christian-Orthodox, Christian-Catholic, Muslim, ...).

In the centuries of transition from Byzantine to Slavonic and finally Turkish dominion, Kosovo has seen a peculiar development in religious architecture, which deserves the general definition of serb-Byzantine art, mainly in XI and XII centuries. We deal with an important part of medieval culture, in the special area of Balkans, where Byzantine culture interferes with local interpreters of Slavonic and Greek schools, not neglecting local influence from the great western architectural and pictorial tradition, mainly through Dalmatia.

This important heritage has suffered heavy damage, even destruction, from war events and general disorders, in the years from 1998-2004.

We consider it a civil duty to save the memory of this past and to keep what has remained, in the area of information archiving, such goal can be best achieved by using architectural GIS.

For a long time, storage and archiving methods have been used in order to preserve and restore cultural heritages. Although new powerful methods are available today, the attitude of mankind towards conservation in general has not improved, so this heritage is in real danger.

The first stage of our work has been to put together cartographical data, collect historical and social information for the area and implement the list of religious and architectural monuments.

A survey of the spot has proved quite difficult, so our data have been checked through internet links, general and specialized literature and personal communication.

Agosto 2009
At first, we have taken into account some forty monasteries, churches and other relevant constructions, all considered important from either a religious, artistic or civil point of view.

For economic reasons, but also to grant a wider access to potentially interested people, we have often used Open Source data; also, the conceptual database has been shaped starting from easily accessible sources and tools.

At present, some forty geo-referenced spots are available, which are linked to a database including historical-architectural information and data about past damage, and some pictures (Fig. 10).

In order to achieve a 3D GIS, we need to have more surveying data, pictures, photos, plans, etc. GIS works as an archive and data access: so we shall have a WEB component, inside the system, in order to publish the collected information and in the future to help editing data.

Acknowledgement

We are grateful to Chiara P Lotz for her valuable support in philosophical themes and Gordana Paskovic for her precious knowledge of Kosovo problems.
Fig. 10. Monasteries and churches in the data base.
Monasteri e chiese della base di dati.

References

AMBIENTE E SVILUPPO SOSTENIBILE


Jugoslavia: gli affreschi medioevali, UNESCO, 1965
ESRI Data & Maps — ArcGIS Europe Data Set, 2006
UNEP, Grid Arendal, Maps and Graphics Library, http://maps.grida.no/
Penn State University Library, Digital Chart of the World, http://www.maproom.psu.edu/dcw/

14

Agosto 2009