RELATIONAL STRATEGIES IN STATISTICAL DATA ANALYSIS: MAPPING, INFERENCES, CLUSTERING, TOPOLOGIES, MATCHING

L. Mussio, DIIAR Politecnico di Milano, Italy – luigi.mussio@polimi.it
V. Dante, Italy – valentina.dante@gmail.com

KEYWORDS: mapping, inferences, cluster analysis, topologies, matching.

ABSTRACT:

Modelling reality is representing phenomena: descriptions of events are not always reducible to rigorous mathematical functions, requesting therefore statistical methods to reproduce their effects. Throughout both classical and innovative procedures of statistical analysis, the implementation of software can classify data, improving results with every step of sequence (mapping, inferences, clustering, topologies, matching).
Assuming a critic opinion can only descend by a probabilistic choose, multiple tests formulize hypothesis on nature of data: this method of statistical inference is implemented in *post_analysis.exe* to map a set of raw spot data in 3D, to find stationary points and to identify possible outliers applying a multivariate analysis on results. Then, *cluster_analysis.exe* can group data using two iterative cycles which control clusters with parameters fixed by users.
When geometry is unavailable in data, some spatial references are often provided by relationships of presence, association and proximity between elements of the set: therefore, graphs' theory helps in organizing and managing this information. From data grouped in 3D, *texture_analysis.exe* builds up symbolic matrices and creates a structure for every cluster provided; the aim of optimize memory occupation is then fulfilled by *new_sorting.exe*, which rearrange and dissect graphs.
The problem of matching searches for corresponding characteristics between different descriptions: with a digression in the Psychology's sphere, use of perceptual organization leads the implementation of software. To recognize similarity in comparable graphs, *exact_matching.exe* and *unexact_matching.exe* generate an identification code for every point, classifying nodes to associate structures. Computational charge is clearly different: imperfect matching requires much more calculations.

RÉSUMÉ:

La modélisation de la réalité représente phénomènes: la description d'évents peut n'être pas réductible en fonctions mathématiques, donc requérant l'emploie de méthodes statistiques pour reproduire les effets. Avec procédures soit classiques soit innovatrices, l'implémentation de softwares classifie les informations en améliorant les résultats en chaque étape de la séquence (mapping, inférences, clustering, topologies, correspondances).
Puisqu'une opinion critique peut descendre seulement par une décision probabiliste, tests multiples formulent hypothèses sur la nature des donnés: ce méthode d'inférence est implémenté en *post_analysis.exe* pour griller un set d'informations spot in 3D, chercher les points stationnaires et identifier les faux donnés avec une analyse multi-variée. Puis, *cluster_analysis.exe* synthétise les informations en usant deux cycles itératifs qui contrôlent les clusters avec paramètres fixés par les usagers.
Quand la géométrie des informations n'est pas disponible, on utilise la référence spatiale fournie par relations de présence, association et proximité entre les éléments du set: la théorie des graphs ajute à organiser et gérer les donnés. A partir des informations en 3D regroupés, *texture_analysis.exe* construite les matrices symboliques et créa une structure pour chaque cluster; l'optimisation de mémoire est garantie par *new_sorting.exe*, qui fait la dissection des graphs et les réarrange.
La correspondance recherche les caractéristiques similis entre descriptions différentes: une digression dans la Psychologie permet l'emploie de l'organisation perceptive pour l'implémentation des softwares. Pour reconnaitre analogies entre graphs comparables, *exact_matching.exe* et *unexact_matching.exe* générant un code d'identification pour chaque point et associe les structures: la charge computationnelle est différente, parce que la correspondance imperfective est évidemment plus onéreuse.

1. INTRODUCTION

The introduction of automatic methods in statistical analyses means the possibility of using powerful calculations to get to results: in the geomatic area, elaborations are far more complex consequently to diffusion of computers. Each analysis is processed automatically, in order to have computations that cannot be compromised by intervention of an inexpert user or by lack of attention of a skilled one.
Sometimes, information has a perfect equation to be represented with, but quite always descriptions of events cannot be reduced to rigorous mathematical functions without compromising efficacy of the model. The loss of information caused by use of an incorrect representation can be counterproductive, because data could produce erroneous explanation of events. Vice versa, statistical methods guided by essential parameters should be used to produce a good model: what is capable of explaining events can be an efficient method to forecast predictions.
There are many classical statistical procedures which help organising concepts, but they are incomplete: a quote of information is often lost in elaborations. To avoid loss of available data, which should be processed without compromising power and speed of calculations, the most efficient way to solution is to program new software of statistical elaboration.
Naturally, an automatic model should be obtained working along with simple bases which grow in complexity to represent all characteristic parameters of a phenomenon: in order to build reliable software it is necessary to have good skills in computer programming,. Only perfect acknowledgement of information

technologies allows contributing to credibility of the whole process of statistical elaboration.
As usual in statistical analysis, every data set refers to an object and contains values of statistical variables. Information is elaborated through classical analysis: comparison to parametric models, descriptive analysis, validation controls, everyone of them conducted over complete automatism of calculations.
With enormous quantity of data, automatic procedures are suitable for a computer to manage counts without human intervention: deleting redaction work on data means to cut down time requested by the whole process. To fulfil the goal of complete automation, it is fundamental to forecast and program efficient constraints: only an exhaustive survey predisposes every position in the algorithm.
Nevertheless, speed in elaborations remains a fundamental aim, as quick results of effective software are far more competitive of perfect forecast produced by a slow program.
Therefore, this paper deals with the construction of methods which handle relational concepts to improve reliability of results: through classical analyses to matching, considering useful each spatial information, even though geometry is not provided with data. The inclusion of a weak concept of distance introduces, through a notably complexity of elaborations, availability of interesting results: spatial references of events may provide great accuracy of statistical predictions.
A faithful representation of a data set can be used in order to obtain some realistic methods to forecast future behaviours: the model should be able to show transformations of reality in consequence of a perturbation of the system.
Some brand new software had been programmed to fasten statistical data analysis, in order to have a computer managing and elaborating the whole set of available information. Speed and precision due to use of informatics permit to come to results in very short time, without losing precision or causing errors.

2. STATISTICAL INFERENCE

An opinion on a scientific theme must be critic, as Galileo stated the principles of every acceptable theory on phenomena. Nowadays, almost every decision comes from a choose involving probabilistic arguments: since there is no certainty about exactness of assumptions, only experiments can affirm or refuse hypothesis. Working on a population of observations, the scientific method must refer to available data to find elements proving or rejecting results: robust statistical analysis is the only procedure to distinguish and isolate outliers among a set of correct data.
In some cases, for an inexpert user of statistical methods, it could be very hard to find outliers in a set: if errors are in number comparable to correct data, there could be misunderstandings of the representation. Besides, there are some risky cases in which addressing analysis means obtaining controlled statistics: however, misuse of effective statistical methods and instruments falsifies results. Some errors, finally, depend on the possibility of wrong transcriptions of data or are caused by distraction of the user: this type of incongruous information can be eliminated using a complete automatism.
Then, the use of robust statistics permits to avoid erroneous forecast: since some errors could be contained among data sets, working with statistical inference can localize incorrect information. Employing multiple tests is a strategy in order to control the possibility of using false observations: through an analysis on available information, every step formulizes a hypothesis on nature of data, assuming every property only after a probabilistic evaluation of the situation.
There is nothing magic in a statistical procedure, even if a computer works instead of a human being: automatic software always generates results, working on data (maybe realistic or even false) and running through rigorous, coded hypotheses. The availability of highly detailed models and fast automatic calculation process should not induce the illusion that some software can explain everything, quickly.
Multivariate statistical inference builds and verifies hypothesis for quality problems. A probabilistic choose allows assuming a critic decision on results: through statistical inference strategies, multiple tests permit to forecast parameters for both normal models and distribution free ones.

2.1 *post_analysis.exe*

Software *post_analysis.exe* realizes and manages preliminary mapping processes on data, adding results to a statistical inference analysis. Then, it reduces the quantity of available information, deleting outliers and classifying data through classical tests and non-parametric ones.
Software realizes and manages a preliminary mapping on raw spot 3D data: it identifies maximums, minimums, saddles and hyper-saddles searching for stationary points among data, using a moving window to control regions of information.
The program finds outliers too: it reduces the quantity of available information cutting off errors between data, through a work on robust statistical procedures.
Besides, *post_analysis.exe* classifies information applying a multivariate inference analysis on results, using tests. The procedure scheduled to validate information is a step – by – step method, assuming only one hypothesis a time: software controls with a probabilistic evaluation every conclusion before accepting it, to grant correctness of the whole process.
The program controls data normality using the Pearson et al. test, with asymmetry and kurtosis coefficient obtained from preliminary calculations. Then, the software studies independence between data: it uses Lawley test to operate on covariance structure (expressed by correlation coefficients of normal information) or Kolmogorov test to have a non – parametric control on a distribution. Bartlett test is a classical way to assure equality of sample variances of normal and independent data: after that, Fisher test or Welch test on variance analysis allow controlling equality of sample averages.
A bibliographical research on Welch's work revealed a hidden innovation in his ideas: multiple test of Welch can help explaining phenomena in the traditional theory of classical tests, without using the distribution free ones.
Vice versa, if data on hand do not respect condition of normality, analyses are to be conducted with non – parametric tests: in this case, distribution – free controls verify equality between dispersion values and central values. With independent information, the software uses Kruskal – Wallis rank test, otherwise it employs the generalization by Friedman of Thompson sign test.
Structure of output can be used to launch further procedures, so that an automatic system could work on a database with correct formats and efficient dimension.

3. CLUSTER ANALYSIS

Working with direct geospatial observations, digital images or maps from numeric cartography requires operating the transformation of enormous, incoherent data sets in compact ones. The only possibility to achieve this goal is to select significant information through deletion of blank pixels from

database: then, compactness comes from identification of zones with same approximation.
Besides, since accidental objects refer to data, this information is necessary to employ a deterministic method, such as interpolation or Fourier's analysis. Otherwise, the function of autocorrelation between variables is required to work with stochastic models, such as collocation method or analysis on fractal theory. Both applications work on object classification and they are useful to realize solid models or to produce orthoimages of an area.
Finally, it could be interesting to divide regions of data in homogeneous zones, in order to reduce data number or to underline presence of singularities.
Pursuing solving of each stated application, it is useful to employ cluster analysis techniques: gathering statistical procedures and computational algorithms is the way to classify data sets and to build groups of homogeneous information.
Cluster analysis allows the resolution of problems in exploration, compactness of information, classification or interpretation, since the very fundamental aim of this theory is maximizing intern compactness of identified clusters and exasperating existing differences between them.
Employing cluster analysis to geodetic and geomatic problems allows a preliminary exploration of data, while it allows producing more compact sets and classifying the whole information system. When interpretation of data is uncertain, situation is not suitable to choose neither a functional model nor a stochastic one: in these cases, cluster analysis addresses decisions, without requiring too strict initial hypotheses. Once recognised some homogeneous data sets, it is possible to launch sophisticated analysis.

3.1 *cluster_analysis.exe*

The innovation in programming new software employing cluster analysis techniques is implementing relational strategies in classical structure.
Spatial reference may contain geometry associated to available information. Otherwise, weak references provide a topology and every concept of distance is uncertain. No measures are available and it is necessary to find other variables to have an entity able to establish differences between observations. In this case, algorithms of region growing and line following in 3D can manage processing spatial information, producing a program able to compose and separate groups.
Region growing and line following play an important role in formation of clusters. Indeed classical strategies, performed by norms (euclidean, Manhattan or other), fail when data sets are ill conditioned: for instance, taking into account concave, non – stellar or multi – connected bodies.
Software *cluster_analysis.exe* implements the very general method to classify data sets: it analyses an incoherent set of spot measures and manages formation, differentiation and union of homogeneous groups, using two iterative cycles to control clusters with parameters fixed by users.
The process runs through the division of the whole universe of observations in sets with same dimensions or highest likelihood in information. Then, software proceeds with calculation of statistical parameters characterizing every set and definition of cluster points. The execution of a proper cluster analysis on every group, at this time, can help control significance of each cluster point: eventually, it could be necessary to modify a set. Finally, division of data sets follows indications provided by results of cluster analysis.
An example shows some aggregated data, sorted by software to preliminary results obtained from executing the first iterative cycle (first step), towards more satisfying divisions (second step). In this case, the program runs twice to reach good representation of population: naturally, complex cases require getting over some more steps to obtain interesting results, but it is important to underline that the procedure finally converges.

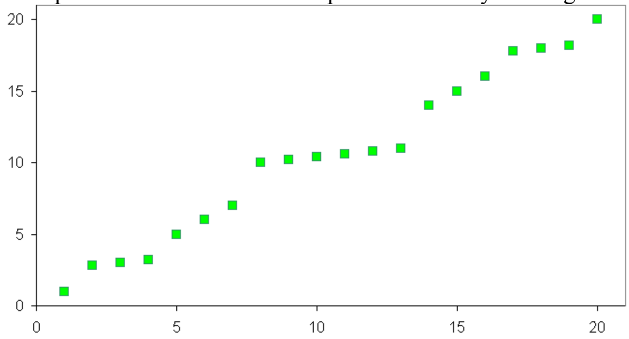

Figure 1. Cluster analysis: available data

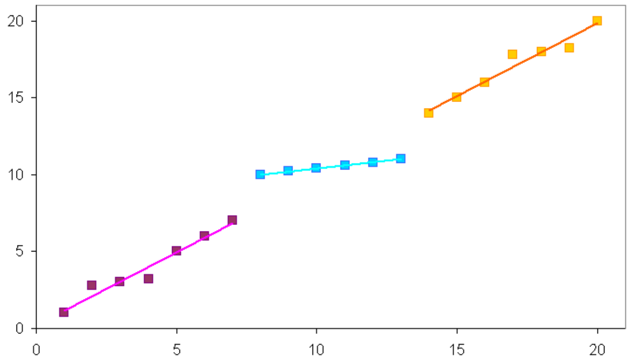

Figure 2. Cluster analysis: results at first step

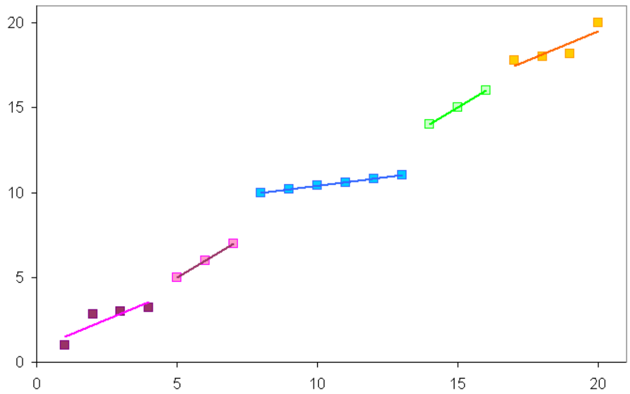

Figure 3. Cluster analysis: results of second step

4. GRAPH'S THEORY

In statistical analysis, it is frequently asked to get over the calculation of minimum squares compensations: they require building and solving a linear normal system of sparse matrices, as topological structure of problems they refer to. Nevertheless, in order to find a solution, it is convenient to operate on a normal matrix with clustered non-null elements. This grouping may improve, for instance modifying order of unknowns with specific intervention on available information.
In this case, geometry is unavailable, but relationships of presence, association and proximity between elements of the set provide some spatial references to information. This type of data produces a geometrical structure into available measures: nevertheless, for an insufficient clustering generating too many

groups, it is convenient to employ another graph structure to reduce number of information.
Besides, large availability of huge quantity of data requires employ of efficient methods to reduce memory occupation. Information on hand has great dimensions: because weight of calculations is proportional to number of stored data, optimizing lines means get faster to results.
Therefore, in order to increase simplicity and speed of elaborations, it is useful to try transforming the database: to improve managing possibilities, the new built system defines observations employing statistics of existing clusters. Creating a structure allows visualizing efficiently singularity between available information, underlining topologies attributed to data. In fact, graphs' theory helps organizing and managing observations, since these mathematical statistical structures are naturally apt to representation of problems, like choosing a path or connecting points through connecting nets.
To succeed in the aim of improving elaborations' speed through selection of remarkable data, two programs are born: one of them manages the most efficient sorting of available information, while the other adds an algorithm of dissection to widespread graphs. Each one builds up a file .txt to report achieved results: every step produces some reduction in memory occupation.

4.1 *texture_analysis.exe*

Software *texture_analysis.exe* manages stored information, merging double indications and deleting incomplete items.
From 3D data with an existing grouping, it builds up symbolic matrices and creates a structure for every cluster provided. Then, software manages a transformation of graphs, to cut off memory occupation due to stored information: minimization of structure dimensions runs through an algorithm of dissection. It traditionally works on level order, but innovation consists in trying to employ a weak concept of distance: obviously, when available information does not provide geometry, there cannot be such parameters. Therefore, efforts leaded to introduction of topological dissection of graphs, working on spatial references without unique measures of distances.
Software reaches the most reliable results working on parameters obtained running some other program (such as *cluster_analysis.exe*): data grouped using topological information can be easily transformed in symbolic matrices, to produce graphs at last.

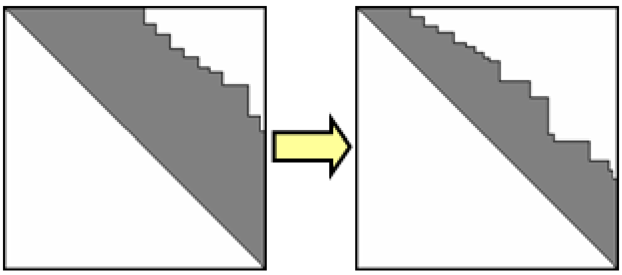

Figure 4. Reduction of memory occupation – *texture_analysis.exe*

4.2 *new_sorting.exe*

The aim of optimizing memory occupation, obtained working on both dissection and sorting of information on hand, is then fulfilled by other software, *new_sorting.exe*, which rearranges and dissects graphs. In fact, it has been implemented with all functions provided in *texture_analysis.exe*, but this software can also rearrange the set, using a simpler representation of the same data. Same time employ of both steps produces greater reduction in memory occupation.
In particular, software *new_sorting.exe* manages a structure for information in 3D, allowing both reorder and dissection. In particular, it is possible to choose type of execution for dissection: it can follow geometric constraints, as well as only topological criteria, on decision of the user, who also sets calibration parameters values. With data provided with topological constraints, software can execute as usual algorithms of reorder and dissection, working on available information in the same way employed for classical analysis.
Software algorithm works at its best on results produced by running other program *cluster_analysis.exe*, assigning a new sorting to points on graphs after executing dissection.

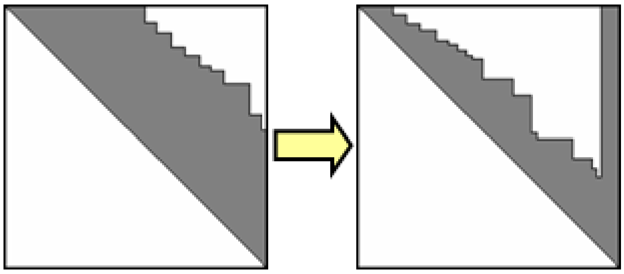

Figure 5. Reduction of memory occupation – *new_sorting.exe*

5. RELATIONAL MATCHING

Going further, to future connections with Computer Graphics, classical statistical analyses are no longer useful: automatic elaboration of real time collected data requires a brand innovative approach to more specific search in information.
Digital photogrammetry has been correlated to Computer Vision, because they both refer to automation of phenomena related to vision through filter methods: in particular, matching represents a complex problem in many fields, from photogrammetry to Computer Vision techniques to artificial intelligence. In fact, matching solves correspondences between structures, searching for correspondent items in data descriptions: typically, it searches for homologous points in models and images. Then, automatic object identification is solved developing a system recognising homologous characteristics between images.
To simplify the computational system, the best technique uses relational descriptions of data and employs principles of perceptual organisation to evaluate abstraction level and type of available information. Identification of an object on a scene derives from a research for correspondences between relational descriptions of images and models: every candidate in image resembling the model is considered, but solutions are accepted only after geometric validation.
Relational matching allows working without preliminary data on images, facing on the converse greater computational complexity deriving from dimensions of research space. Construction of relational descriptions funded on perceptual organisation principles is based on studies on human vision system: implementing the capacity of identifying complex structure in a scene is working on characteristics of continuity, regularity, repetition and symmetry, which are rarely accidental and then can be assumed as correct.
Software can reproduce relational matching to search for correspondences between series of data, working recursively on perceptual organisation principles. This theory was born in the Psychology’s sphere, from the frame of German Gestalt movement, at the beginning of XX century: first studies on

human vision explained simple criteria, basis of human vision process. In fact, it would be controlled by six basic principles of proximity, continuity, homogeneity, closure, similarity of shape and common path, which would rule perceptual organisation.
Hermann von Helmholtz stated visual perception is an unconscious inference, because vision derives from probable interpretation of incomplete data: recognising constituent elements of an object leads to association with the whole object though integration of characteristics. In fact, it seems impossible to obtain perfect correspondence between relational descriptions of data: the most probable possible combination has to be selected as correct. Actually, dimensions of problems prevent evaluation of all solutions: number and completeness of topological and geometrical relations limit wideness of space in which to search for possible correspondences.
Therefore, since matching searches for corresponding characteristics between different descriptions, use of perceptual organisation leads the implementation of software.
Criteria would constitute a system able to distinguish between two sets of data, to recognise similarities and to solve differences: generating an identification code for every point, they can classify nodes and then associate structures. Precisely, it was necessary to program two software, in order to optimize requests on memory occupation: since an imperfect matching requires much more calculations, computational charge of programs is clearly different.

5.1 *exact_matching.exe*

Software *exact_matching.exe* identifies every correspondence between identical databases or structures with same number of nodes and same configuration. In particular, it associates homologous points on every set, assuming a unique correspondence does exist between data.
Indeed, locally perfect matching has been implemented through representation of connections on graphs: software executes automatically the association between points with same code and compatible geometry to available information. Local exactness indicates the possibility of recognising perfect correspondences between structures identified by relational descriptions, associating one – to – one elements on graphs.
After clustering data, structuring them in graphs and optimising memory occupation, software searches for a classification of points coherent to topological characteristics: on both structures, every point is associated with a code. It has been necessary to impose a limitation to maximum number of confluent relations in a node, to allow data memorisation in double precision variables.
Comparison between data sets finally allows finding a correspondence, searching node – by – node for identical codes on graphs. Every ambiguity in point codes generates uncertainty in matching results: if there are not points with the same descriptions in a graph, matching is perfect and correspondences are definitive, because associate all nodes one – to – one. Otherwise, software builds a second code for ambiguous points and tries to associate them. If ambiguity still persists, a third description is produced to try association: over these attempts, ambiguities are reported on output because identical codes indicate indistinguishable points.
An example comparison between two identical structures (case A and B) illustrates work executed by software, since every point in correctly recognised and associated to assigned names: results show superimposition of graphs.

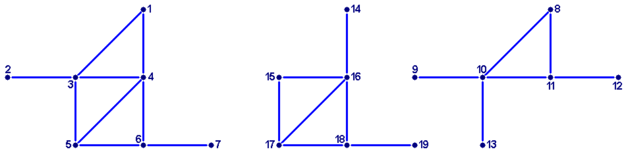

Figure 6. Exact matching: case A

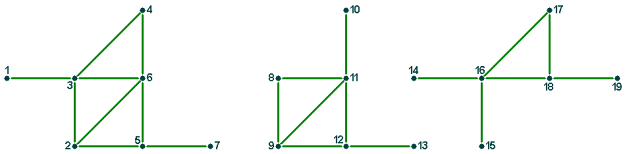

Figure 7. Exact matching: case B

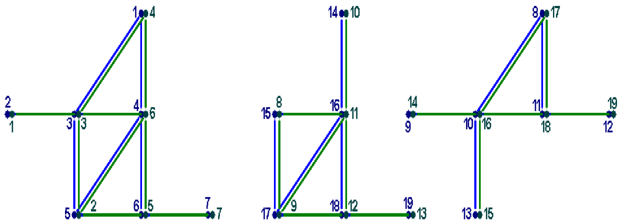

Figure 8. Exact matching: results

5.2 *inexact_matching.exe*

When databases are different in configuration and number of nodes, employ of more complex software is required.
Working to superior complexity, *inexact_matching.exe* makes no hypothesis on provided information: an imperfect system of matching researches analogies between graphs, controlling all structures and memorising only correct correspondences.
Finally, using this software implies much more calculations to get to results, but also a wider field of applicability.
This software provides best results with combined use of *cluster_analysis.exe*, *texture_analysis.exe* and *new_sorting.exe*: after clustering, construction of graphs and optimisation of data sets, comparison is executed one – to – one between nodes: matching cannot be exact, then the software executes some changes chosen by the user, who indicates correct points.
To solve ambiguity in point codes software implements an additional control on a supplementary parameter, related to graphs' density.
This software is only semi-automatic, because human intervention in this case reduces error percentage: the user is required to address software in the correct direction, indicating exact correspondences between the found ones. This procedure allows the software using topological information to search for other associations, because from first set of correspondences some points could be selected as invariant: emanation points, supposed to be rare because they lead the most relations, are selected establishing some correct associations of couples.
Then, software analyses the structures to verify correspondences testing their compatibility with the structure, through as many tree paths as settled emanation points.
In particular, as associated couples cannot be modified, then it is important to set only correct correspondences to address software: analyses of assigned structures verifies associations through their compatibility to geometry or topology of graphs.
Presence of correspondences between nodes is identified through verify of points order control: *inexact_matching.exe* can set the value limit of high connections. Employing this parameter can address the software, privileging a preliminary

control of high grade nodes before executing a complete verify on the whole graph.
An example comparison between two different structures (case A and B) illustrates work executed by software, since every point in correctly recognised and associated to assigned names: results show superimposition of graphs, with evidence of complexity successfully managed by *inexact_matching.exe*.

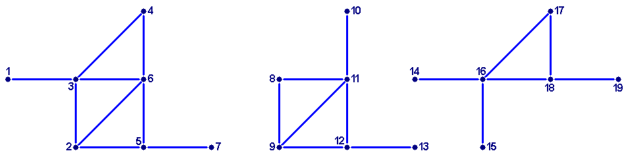

Figure 9. Inexact matching: case A

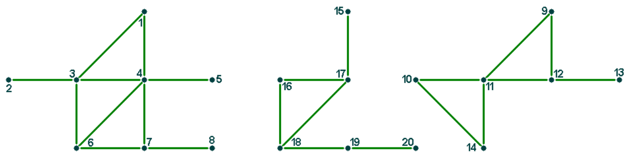

Figure 10. Inexact matching: case B

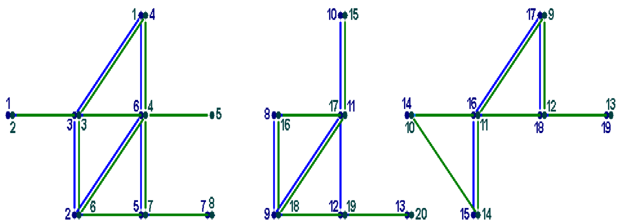

Figure 11. Inexact matching: results

6. CONCLUSIONS

Reality is far too complex to have its perfect representation with mathematical models: its best approximation is in probability of processes and events, according to statistics.
This paper runs through some operative problems typical of geomatic data management: applied mathematics allows solving these difficulties, dealing with huge dimensions of databases and complete automation of procedures. Relational strategies applied to statistics rule analyses from mapping, through inference, to complex attempts of minimal organisation: grouping, sorting, dissections and matching.
Nevertheless, mathematical rigour cannot grant even the existence of linear, unique, normal solutions. In fact, only clean models underwent to outliers deletion can be considered normal and could be represented by linear models and determine unique estimations. At the contrary, the whole reality rarely offers multiple solutions, without reference to known models.
This paper explains work methods on some classical statistical procedures, changed by introducing new items, in order to obtain better results. Relational strategies can lead implementation of software, which can classify data throughout the employment of both classical and innovative procedures of statistical analysis: providing an improvement of results with every step of the sequence means generate more realistic representation of events or more reliable forecast, or even both.
As the subtitle suggests, this paper moves from mapping, to inferences and clustering techniques, towards topologies and matching problems.
Software *post_analysis.exe* produces grid of correct information, executing the cut off of eventual anomalous observations in the meanwhile. Therefore, through employ of robust procedures on mapped data, it realises a multivariate statistical analysis. It verifies data normality with the test of Pearson et al., working on asymmetry and kurtosis coefficients. It controls data independence employing test of Lawley with normal distributions and otherwise test of Kolmogorov. The first one operates on covariance structure, expressed by correlation coefficients, while absence of normality requires verifying complete independence between variables. Test of Bartlett on variance components controls variances equality, in case of both normality and independence: then, test of Fisher or test of Welch on variance analysis can give information on averages equality.
Therefore, employing non – parametric tests is necessary: for independent data, equality between dispersion values and central ones is verified by test of Kruskal – Wallis; otherwise, it is provided using the test of Friedman.
In particular, reference can only be done to generic central values or dispersion ones, since absence of normality inhibits classical statistics (averages and variances).
Then, an analysis on spatial reference of data can be done, to state the presence of geometrical information or, at least, topological structure: with weak spatial reference, only topological information is provided and without having any geometry, the concept of distance becomes uncertain. Brand new software has been implemented: combined execution of *cluster_analysis.exe*, *texture_analysis.exe* and *new_sorting.exe* allows managing grouping and structuring of spot data in graphs, eventually submitted to reorder and dissection.
Working on these relational strategies, it sounds realistic to believe that almost every classical statistic problem can be faced with this innovative approach: considering similarities between digital photogrammetry and computer vision, studies on human perception are interesting in order to make computer see. Throughout examination of Gestalt theory, it seems a resemblance does exist between principles of human vision and implementation of a software allowing computer matching: therefore, a program considering six principles of perception could give good results.
Then, if it is necessary to compare databases, it could be useful to proceed with specific matching operations: interoperability and normalization of information generate high probability of intersections between graphs' structures. The simplest matching operation between identical databases requires *exact_matching.exe*: every single data set is clustered, structured in graphs and optimized before assigning a topological classification to nodes. Indeed, the two data sets are compared punctually to find a correspondence, while the report underlines possible ambiguities in node codes. If there are any, uncertainty cannot be contemplated in matching: association node – to – node concludes without errors.
If databases are different, *inexact_matching.exe* is required: after clustering, structuring in graphs and optimization, nodes are associated one – by - one. Nevertheless, since matching cannot be exact, from the first found correspondence software proceeds with some swaps. Then, assuming some rare points as emanation points, software uses the two structures to verify matching compatible with sides of graphs.
Finally, this paper addresses the reader towards introduction to innovation. As Psychology can help define principles of human behaviour, computer graphics may be concerned by some ideas born in studies about vision: various arguments can meet, solving a complex problem involving more topics.
This work on research of minimal organisations is a contribution on heavy path to Utopia. Development comes through unconventional relationships between concepts, to

define a composed frame in which every aspect becomes a part of the whole.
Perhaps this goal sounds like Utopia, a place that never exists. Or maybe Utopia is simply where everything has its own dimension, confused with the whole it composes: as German scientists stated, visual perception is an unconscious inference on probable interpretation of incomplete data. Only recognising constituent elements of an object can lead to integrate single characteristics to the whole object.

SELECTED BIBLIOGRAPHY

References from Journals:

Shapiro, R. G., Haralick, L. M., 1987. Relational matching. *Applied optics*, vol. 26, Washington DC, USA.

Vosselmann, G., 1991. Relational matching. *Lecture notes in computer vision*, n. 628, Berlin, Germany.

References from Books:

Arnheim, R., 1962. *Arte e percezione visiva*. Feltrinelli, Milan, Italy.

Kanitza, G., 1980. *Grammatica del vedere*. Il Mulino, Bologna, Italy.

Köhler, W., 1961. *La psicologia della Gestalt*. Feltrinelli, Milan, Italy.

Lowe, O. G., 1985. *Perceptual organization and visual recognition*. Kluwer, Boston, Massachusetts, USA.

Metzger, W., 1971. *I fondamenti della psicologia della Gestalt*. Giunti Barbera, Florence, Italy.

Minsky, M., 1990. *La società della mente*. Adelphi, Milan, Italy.

Sachs, L., 2000. *Applied statistics*. Springer Verlag, Berlin, Germany.

References from Other Literature:

Dante, V., 2006. *Strategie relazionali nell'analisi statistica di dati. Mapping, inferenza, cluster analysis, topologie, matching*. Thesis, Politecnico di Milano, Milan, Italy.

Marr, D., Poggio, T., 1979. A computational theory of human stereo vision. In: *Proceedings of Royal Society London*, series B, vol. 204, London, UK.

Nardinocchi, C., 1996. *Strategie di matching relazionale in fotogrammetria digitale*. PhD thesis, Politecnico di Torino, Turin, Italy.

Minsky, M., 1988. *Perceptrons – An introduction to computational geometry*. Cambridge University Press, Cambridge, United Kingdom.

Posch, S., 1998. *Perceptual grouping*. University of Bielefeld, Bielefeld, Germany.