Doerr, Martin and Apostolis Sarris (eds) 2003. The Digital Heritage of Archaeology. CAA2002. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 30th CAA Conference, Heraklion, Crete, April 2002.

Spatial Analysis Utilities: a quantitative tool for studies on archaeological distribution

Giancarlo Macchi Jánica

Department of Archaeology and History of Arts, University of Siena
Via Roma 56, 53100 Siena (SI), Italy
macchi@unisi.it

Abstract. The study of the distributive characteristics of settlement patterns is one of the most important topics in the reconstruction of historical landscapes. Unfortunately, advanced knowledge of GIS techniques, as well as quantitative techniques and methods are required in order to achieve positive results in this area. Furthermore, the application of spatial analysis in the archaeological research requires a large and expensive investment of software and hardware to accomplish such tasks. In order to address all these issues, this year we started the development of an analytical tool for the study and analysis of spatial distributions. This software rather than a GIS application shall be conceived as a general utility for data processing. In fact, the main aim of Spatial Analysis Utilities is to accomplish in a quick and easy manner the entire quantitative process. The philosophy of the software is to provide conclusive statistical data with just some mouse clicks to non advanced GIS users.
Keywords. GIS, Spatial Analysis, Software development, Quantitative geography

1 Introduction

Every phenomenon studied by the archaeologist took place inside a spatial–temporal dimension (Wegener 2000). Therefore, his science largely involves studying and understanding evidence that can be synthesized in distribution schemes. Usually distribution maps of artefacts inside archaeological sites, or maps of settlements on the landscape.

As a result archaeological research has always demanded a careful and meticulous approach to distributions in space. And it is for this reason that archaeologist had always seek better methods and instruments to achieve this goal in the most comprehensive and objective manner (Griffith 1999). Early, spatial analysis became an answer to distribution concerns. This geographic discipline provided, as Hodder and Orton confirmed in their book, exceptional means and techniques to deal with the complexity and chaos inside distribution systems. However, a major improvement to a quantitative approach to spatial distributions in archaeology was the introduction of Geographic Information Systems. These applications allowed the archaeologist, for the first time, to deal with large amounts of spatial data and with its analysis in a reasonable amount of time. Actually GIS may be used as a management tool that provides means for data collection and presentation as well as an analytical tool.

There is no doubt that GIS programs are the most sophisticated and powerful solution that the market can offer to the archaeologist. Nevertheless such programs –and particularly those able to perform advanced spatial analyses– are complex tools. Usually this kind of software solution requires long periods of time to reach a satisfactory level of expertise and confidence.

Also, it must be considered that GIS have high costs. When it comes to «spatial analysis extensions and modules», the expenses for a GIS solution may increase remarkably. And even in the case the archaeologist decides to afford the expenses and undertake the long learning curve, soon he will discover that usually GIS yields results in a rough form. Which means that results must be re-imported and reanalyzed in a spreadsheet or statistical software; fact that certainly will worsen the two issues indicated before. The question would be: Does every archaeologist have to accept these obstacles in the case he wish or decide to perform spatial analysis and quantitative studies on his «point distribution maps»?

2 A tool for spatial analyses

The aim of this paper is to illustrate the characteristics of Spatial Analysis Utilities or S.A.U.; a stand–alone application developed in the Laboratory of Information Technology applied to the Medieval Archaeology of the Department of archaeology at the Siena University.

S.A.U. is an application for the study and quantitative approach to archaeological distributions. Beside the scientific requirements and objectives that motivated this project, S.A.U. was developed trying to solve the three issues described above:

1. high cost expenses
2. long learning curve
3. rough GIS yields

The main objective of S.A.U. is to offer the researcher a flexible tool capable of carrying out the whole of an analysis in just a few seconds. This software allow the characteristics and nature of the spatial structure of distribution maps to be discovered by the archaeologist with the use of just a few simple commands, and without any knowledge of the field of GIS or even statistical geography. S.A.U. has a simplified user interface which permits its use even for computer beginners. Its purpose is to provide the archaeologist a set of tools for the measure and analysis of spatial distributions offering a large number of quantitative indices and indicators about the spatial structure of the pattern.

One of the motivations of our commitment on this project was the great number of archaeological and historical databases that our department has produced over the last decade. For example, among the projects developed in our department there is the spatial study of the castles (see figure 1) and the ecclesiastical sites in Medieval Tuscany. While the first database is composed by more than two thousand records, the second consists of more than four thousand ecclesiastical sites. In each record all the chronological indications found in the historical documentation and sources were collected in a very accurate manner. Therefore, we were able to deal with a complete and precise sequence of the medieval settlement in central Italy. In the same time, this completeness implies an extraordinary amount of possible and necessary statistical and spatial analyses to reach a reconstruction of spatial patterns and settlement models. For this reason it was necessary to set up a specific strategy in order to accomplish the full range of spatial studies and to explore the spatial structure of the full chronological sequence of sets obtained from these two distribution databases.

Before a general description of the program, it must be said that apart from the scientific and analytical features of S.A.U., this application represents a helpful means and an effective channel for educational purposes. So far, the application has been used as an introductory tool for «quantitative geography lessons» in our university.

2.1 A general description of the program

S.A.U. is a multiplatform application. In fact this program works on PC's, Mac's and even on the new OSX, the new Apple's operating system. It was developed with Realbasic (from RealSoftware), an Object–Basic like language. The reason we choose to develop the application in this environment, instead of Visual Basic, was simply the possibility to compile the code also for Macintosh Operating Systems.

At the moment S.A.U. is composed by 5 modules: barycentre (or mean centre), nearest neighbour, quadrat analysis, proximity analysis and *Thiessen* polygons. A random distribution generator that allows the user to carry out evaluations between real and random distributions is also included in the software.

The S.A.U. user interface is composed of a main bar that contains the most commonly used tools in the analysis process, a display window for the graphical presentation of the distribution and results, a command panel or inspector that changes in association with the current analysis module, a graph window for the presentation of histograms and a log window for an alphanumeric reading of results (Figure 6).

Once an analysis module has been chosen, a control panel window guides the whole process and lets the user set some of the processing and visualization options. Each cycle starts with the importation of data and finishes with the results being exported in a bitmap or vector format. S.A.U. supports file exports in ArcView shape, dxf or a wide range of bitmap formats. For some modules an ASCII raster export is also supported.

In S.A.U. each analysis session begins with importing a data file which describes a distribution through a series of X (eastings) and Y (northings) coordinates. The data file must present the data as coordinates in a Cartesian plane. S.A.U. does not have the capacity to analyze data expressed in degrees and minutes; we expect to include this option in future version of the program. Basically S.A.U. imports distribution data in text format. This import strategy was preferred in order to avoid unnecessary obstacles for unskilled or inexpert users. In fact, this kind of file may be even compiled manually in a simple text editor. The following is an example of an importable text file:

```
 1691097    4854002
1726264     4774682
1635999     4865535
1650097     4823260
1604350     4892851
1612813     4889125
1614331     4887575
...         ...
 ...        ...
```

S.A.U. processes the imported file without reference to the measurement units in which the co-ordinates are recorded. Among the options for some modules there is the possibility to process additional attributes for each data point. For example, in the calculation of the barycentre, it is possible to indicate the weight or size of each point in order to carry out a weighted calculation rather than one in which all points are equally weighted. In this case, it is necessary to include the value or attribute in a third column.

3 The modules

In order to understand the full potential (and meaning) of this application, it is necessary to consider and ponder its features in terms of the simplicity required in order to obtain significant results. To be precise, the growth and development of GIS application offers today a wide range of potential modules for spatial analysis purposes. In general a user can always accomplish different spatial analysis tasks with the support of a GIS package. The difference relies on how complicated the procedure is, and consequently how much energy (and time) the user has to spent in order to accomplish it.

3.1 The barycentre

The first example is related to the calculation of the barycentre or central mean in the study of the evolution of possession of the different castles lords and rulers (like bishops, aristocratic families or communal government) in medieval Tuscany. The aim of this analysis was to measure or examine the differences or similarities in the evolution of the possession and behaviour of these social groups. For example, how wide was the progressive movement of the barycentre of possession for each family through the time line. A preliminary assumption was in fact that the aristocratic families would have a more accented change of the central mean of their possession that the ecclesiastical authorities.

Even if the calculation of the central mean (Lloyd 1972) of a distribution set may appear as a very easy procedure or method, S.A.U. became very important in terms of simplicity and usefulness. It made it possible to determine more than six-hundred barycentres and export them directly in an ArcView format in less than two days (see figure 7). One of the most important characteristics of S.A.U. is its straightness. Once the file is imported and the analysis is set up the user can obtain his results almost immediately.

As archaeologists we would expect in future releases of this utility an automatic importing–processing routine that allows series of progressive files to be analyzed sequentially. Or for instance an algorithm that allows the user to import a general distribution file related to chronological attributes that can be processed automatically by setting up a temporal class size.

But the ability to provide an instant export of the cartographic results is just one of the qualities of S.A.U. In fact, this application presents also powerful tools' for the analysis of the distribution structure just, like dispersion indices or frequency values. This is the case of the next two modules: the nearest neighbour and the quadrat analysis.

3.2 The nearest neighbour

The nearest neighbour is a module that allows the study of the attributes and qualities of the dispersion index quickly and easily. It was mainly conceived as a tool for calculating the spatial relation inside a distribution system. One of the most important features of S.A.U. is that right after the conclusion of an analysis the user can obtain valuable statistical information like histograms without any difficulty. In this specific case, the results or statistical histograms are readily available not only for the nearest neighbour but also for the second and third nearest neighbour.

In addition to a graphical representation between each point and its nearest neighbour this module also calculates the dispersion index (R) by dividing the observed mean (r_o) by the nearest neighbour expected mean (r_e). The more closely the settlements are clustered together the closer to zero this value will be. If the points are located randomly this ratio will present a value close to one (1). A dispersed regular pattern will reach instead the limit of 2.15 in the case the pattern is organised on a perfect hexagonal grid (Silk 1979: 106-110).

$$r_e = \frac{1}{2\sqrt{d}} \quad \textbf{(1)}$$

Where d is the density: number of point/area.

In order to calculate the nearest neighbour expected mean, it is necessary to determine the area of the study region (see formula 1). The boundary definition file should be imported in dxf format. The import algorithm proceeds by acquiring the first polygon described in the dxf file. Once imported, the user can calculate the area of the study region and then with the number of points calculate the expected nearest neighbour distance. After that, the user can read the relation between the observed mean distance and the expected mean distance in the log window.

The nearest neighbour module also allows the user to import a second distribution file to measure the nearest neighbour distance between two different systems. S.A.U. made it possible for example to determine without any difficulty the chronological evolution of the spatial relation among the castles and parishes, or castles and cities.

The following chart presents the evolution of the mean distance between castles and parishes in the middle ages. It is very interesting to observe how the curve describes instead of an expected reduction, due to an increasing number of castles, a progressive growth of the mean distance among the two settlement systems.

3.3 Quadrat analysis

The third tool of S.A.U. is the quadrat analysis module. Its cartographic outcome is a density grid that represents a general comparative frequency index of the structure of the distribution map. Once the distribution file is imported the only action requested of the user is to set-up the cell size and the grid position. This is done in an interactive manner through the display window. After that, S.A.U. counts the point and assign this value as an attribute to each cell of the grid. Then the user can export the cartographic results in a shape or dxf format. But the quadrat analysis module may offer more than cartographic representation of frequency indices. Statistics applied to the quadrat analysis can generate essential information about dispersion in distributional data. In fact the relation between the standard deviation and the mean can indicate or distinguish the nature among a clustered, random or organized settlement pattern. Zero for a regular patterns, one for stochastic, and greater than one for clustered one (Silk 1979: 98-100).

It is unusual that a distribution map in archaeology, like an excavation, topographic or regional map, matches the rectangular shape of a grid. For this reason it is essential to exclude those cells outside the study region or area. The user may therefore import a boundary mask in order to exclude grid–cells allowing S.A.U. to calculate more accurate results. Inside this module the log window also indicates the most adequate cell size based on the extension of the grid and on the number of points. This size may be used to obtain within this module the best possible statistical result.

In the following chart we can observe the comparison between the dispersion indexes of castles based on the quadrat and the nearest neighbour analyses. The entire work (48 calculations) was finished in just an hour of work. The possibility to calculate these two dispersion complementary indices for the full chronological spectrum in S.A.U. allowed a more comprehensive and insightful perception of the whole settlement process. For example, in this chart we may observe a major change in the whole system around the year 1170.

Currently S.A.U. doesn't allow the execution of statistical or significance tests for the nearest neighbour or the quadrat analysis. We expect to implement these options among the tools available for this two analysis modules in future versions in order to offer more accurate statistical readings to the user.

3.4 Proximity

The purpose of the fourth module is to calculate a general proximity index through the production of a cartographic surface that describes or synthesizes the general structure of the points.

The method is very simple: the measurement of the degree of proximity in the whole system, by counting for each point the number of neighbours within a certain distance set by the user. Once the user has entered the value of the searching radius in the control panel the file can be processed. Once the analysis is completed, S.A.U. offers the possibility to accomplish an inverse distance weight interpolation to expose or render clusters of areas of concentration. This interpolation can be masked with the same import procedure described above. After the interpolation is completed, the user can also modify the number of shades of the surface in the display window. S.A.U. allows the exporting the surface in a RASTER GRID format.

3.5 Thiessen

The last module calculates the Thiessen polygons for each point of the distribution file. However, the innovative feature is the fact that S.A.U. can offer instant histograms of the number of sides, area or perimeter of the polygons of the processed points. Those polygons, which by their spatial characteristics constitute exceptions or differ remarkably from the others, are considered external and are excluded from the calculation of these statistics. The exclusion condition or requisite is the inclusion of all vertices of the polygon inside the convex hull.

This module also calculates and allows the export of the *Delauney* network Cartographic layer that can be of great usefulness during network analyses inside a GIS environment. This module also allows the visualisation and the export of the convex hull.

3.6 The random distribution generator

The random generator works simply by importing a dxf that de-scribes the study region and by entering the number of desired points. The user can export the points in a text-format file, import them in S.A.U. and process them with any of the previous five analysis modules.

4 Conclusions

The list of future improvements and new tools for S.A.U. is very long. The major upgrade may be a variety module, that calculates a cartographic surface of the variation or transformation of an attribute in a nominal scale. However, we feel that there is yet too much to do for each one of the existing modules. Maybe future versions will include some sort of map-algebra like ability. Perhaps it will become a separate application.

Certainly this program may appear to the advanced user as an unsophisticated tool; indeed S.A.U. is just six month old. We are convinced that S.A.U. even in its initial stage represents a potential solution for the study and analysis of distributions in archaeology. With your comments, critics and recommendation we will be able to improve its features in order to provide a better and common tool to the archaeological research. S.A.U. is free. You can download the current version at the following web address:

http://archeologiamedievale.unisi.it/sau/sau.html/

http://medievalarchaeology.unisi.it/sau/sau.html/

http://192.167.112.135/sau/sau.html

References

EBDON D., 1985, *Statistics in geography,* Blackwell Basil

GRIFFITH D., 1999, *A casebook for statistical data analysis: a compilation of analyses of different thematic data sets*, Oxford University Press.

HODDER I., ORTON C., 1976, *Spatial analysis in archaeology*, Cambridge University Press

LLOYD P.E., 1972, *Location in space – a theoretical approach to economic geography,* Harper & Row

SILK, J., 1979, Statistical concepts in Geography, George Allen & Unwin.

WEGENER M., 2000, *Spatial Models and GIS*, in Fortheringham A. S., Wegener M., *Spatial Models and GIS*, GISDATA 7, Taylor & Francis

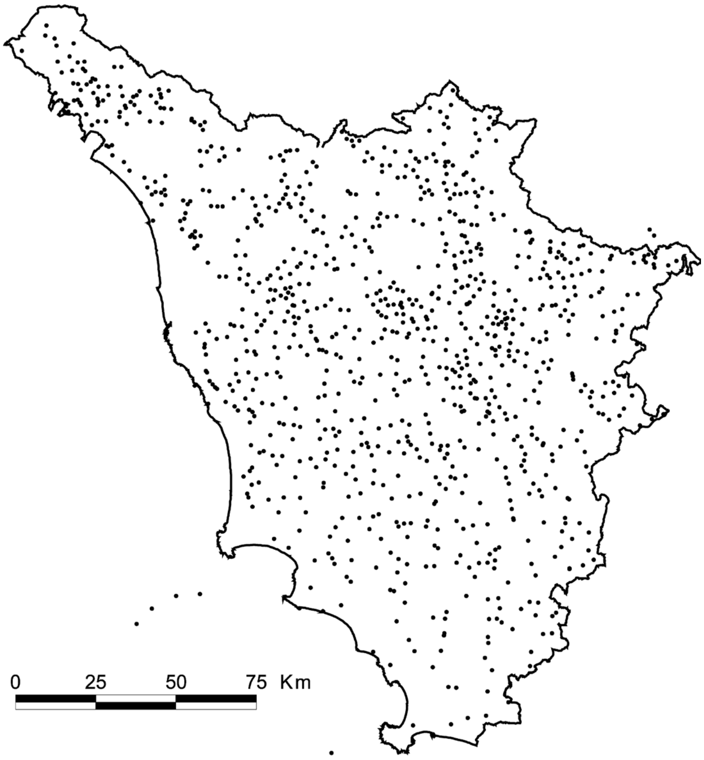

Figure 1. *The castles database developed at the Department of Archaeology at the University of Siena is composed by more than 2000 records.*

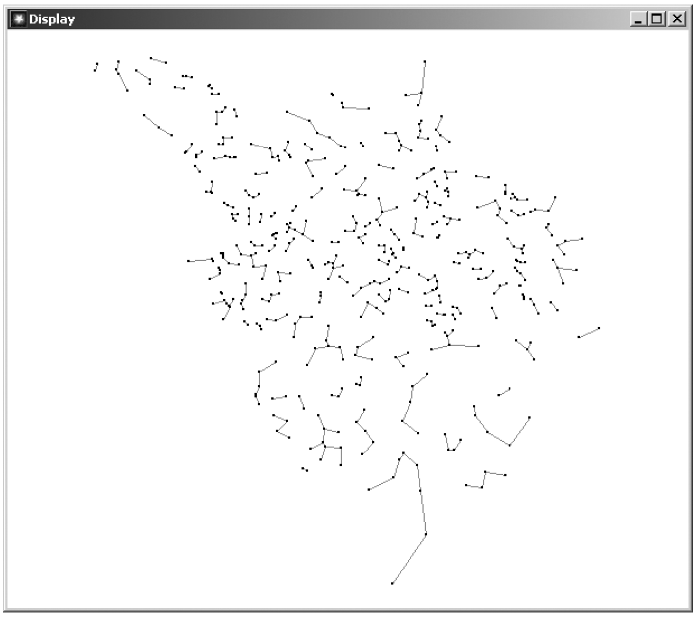

Figure 2. *Example of the cartographic outcome of the nearest neighbour module in S.A.U.*

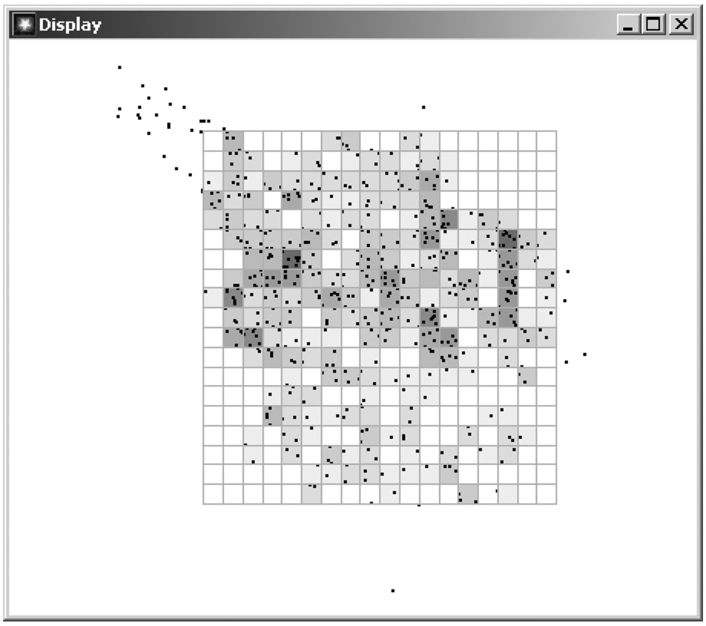

Figure 3. *In S.A.U. the user can place the grid interactively through the display window.*

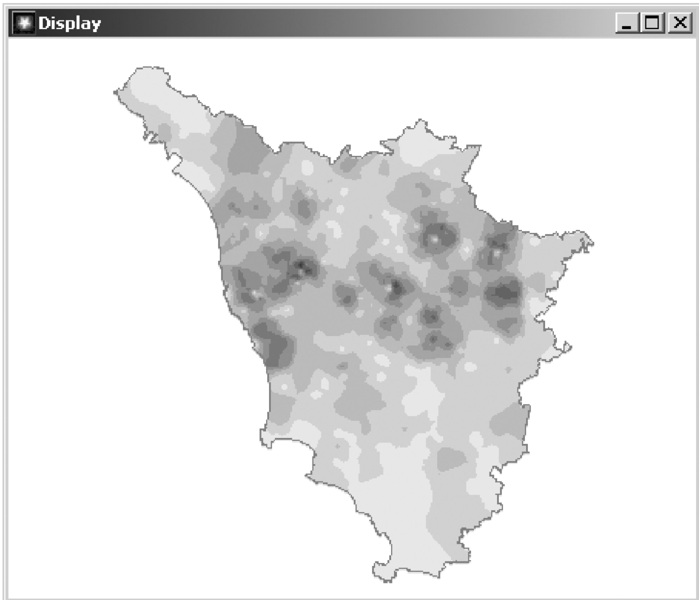

Figure 4.*The user can make a IDW interpolation with the results of the proximity module.*

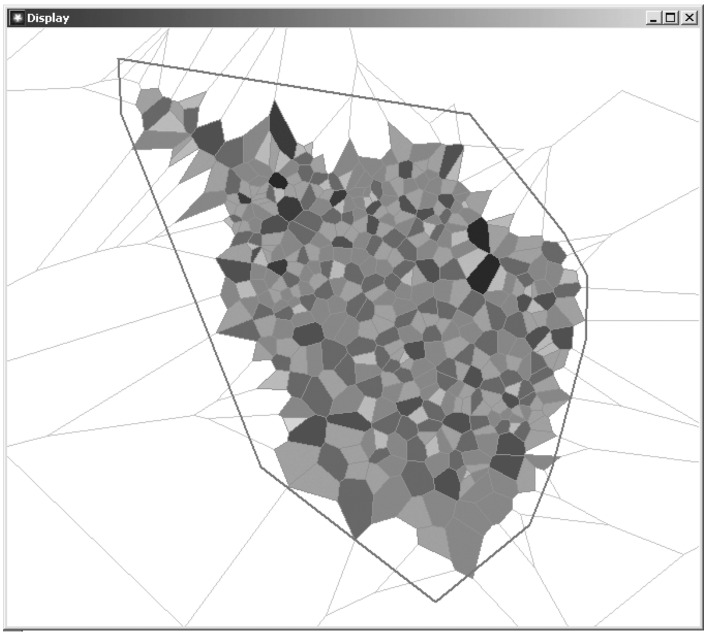

Figure 5. *In this image the colours are used to represent the number of sides of the polygons.*

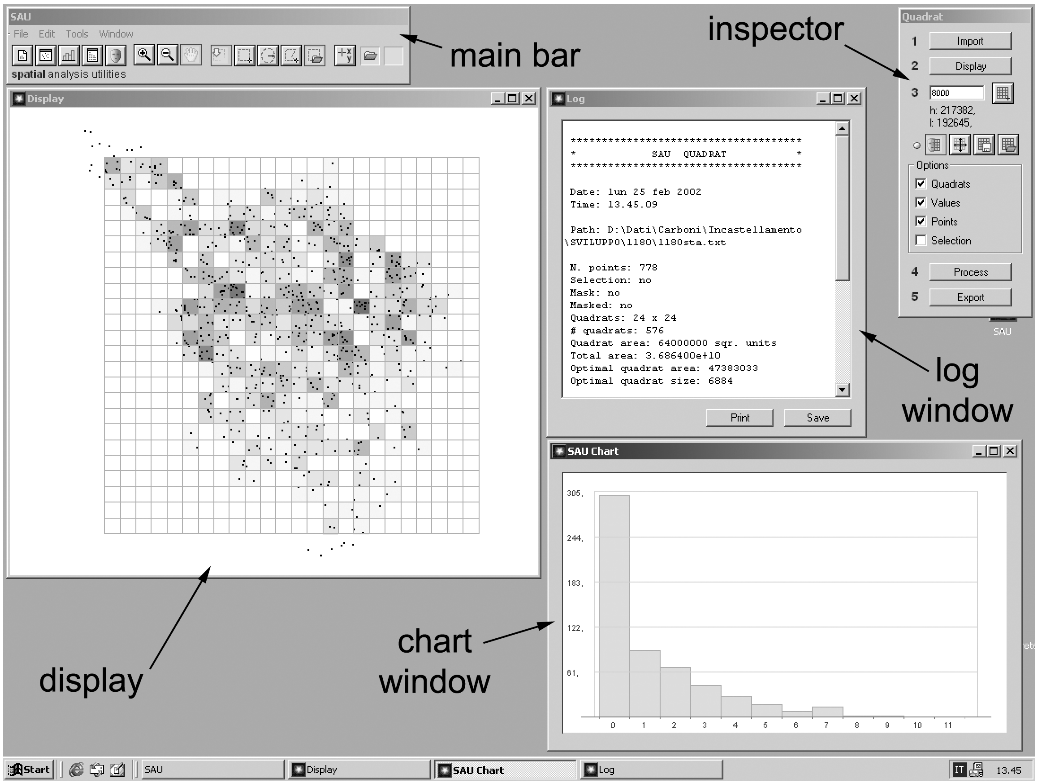

Figure 6. *S.A.U.'s Graphic User interface*

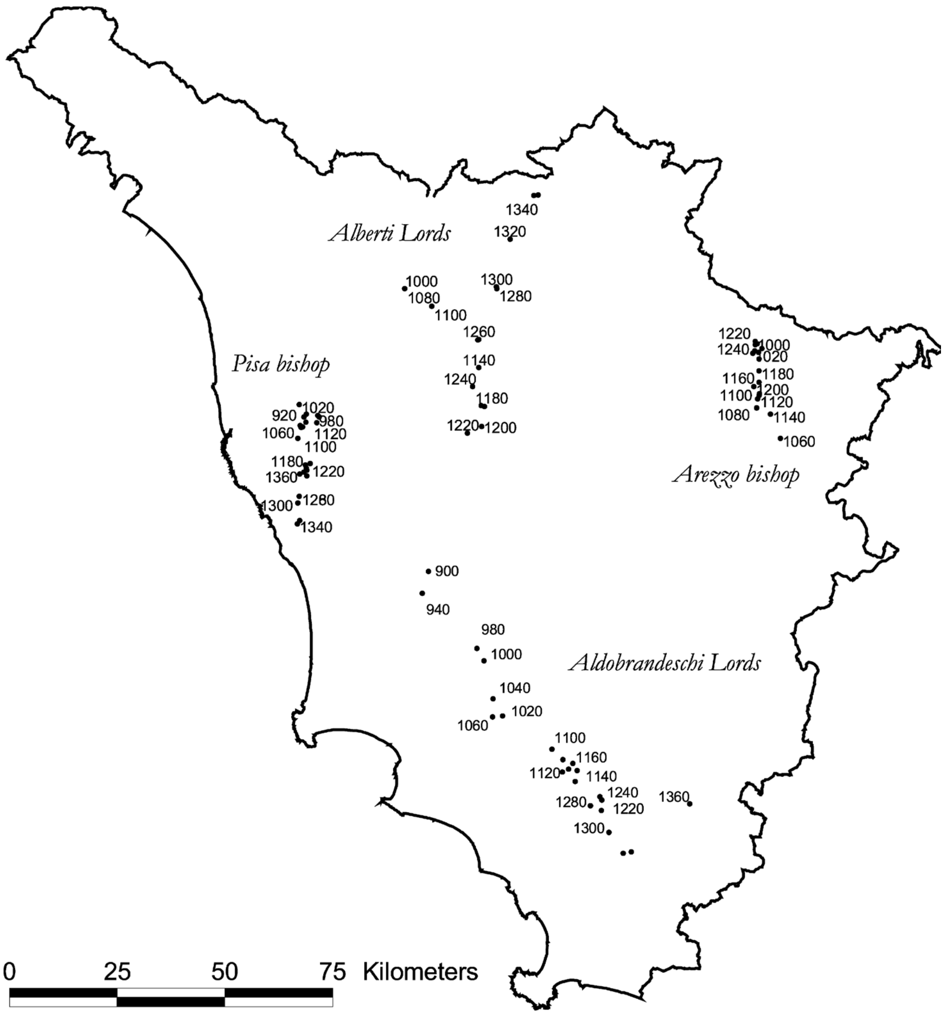

Figure 7. *The chronological evolution of the barycentre.*