Krištof Oštir – Žiga Kokalj – Laure Saligny – Florian Tolle – Laure Nuninger
With the collaboration of Françoise Pennors and Klemen Zakšek

Confidence Maps: a Tool to Evaluate Archaeological Data's Relevance in Spatial Analysis

Abstract: Inventory data used in archaeology is often incomplete and heterogeneous. In the framework of the ArchaeDyn program, a method has been proposed to evaluate heterogeneity in archaeological inventories. The purpose of this work is to create a validation tool to interpret the results. This tool is called a "confidence map" and is produced by combining representation and reliability maps. The first step consists of generating representation maps to describe the clustering of archaeological items. The second step is based on reliability maps. Data providers are asked to define and outline the level of reliability of their data. Then the representation and reliability layers are combined using map algebra. The resulting maps allow for the comparison and analysis of data confidence.

Introduction

Inventory data used in archaeology is often incomplete and heterogeneous, making its interpretation, dating and localization a difficult task. In fact it represents a sample of a more complex reality. The analysis of archaeological data using spatial analysis tools requires great caution in the interpretation that is drawn from them. The issue is to avoid the identification of spatial trends that are just a consequence of the degree of archaeological investigation.

In the framework of the ArchaeDyn program, a method has been proposed to evaluate and give spatial insight on the heterogeneity in archaeological inventories. ArchaeDyn combines the efforts of several archaeologists working on various topics, ranging from the diffusion of manufactured objects in pre- and protohistorical times, to the use of land through the study of settlements, parcels and manuring during the antiquity (NUNINGER / TOURNEAUX / FAVORY 2008). A great diversity in analysis scales and studied objects led to different inventory protocols such as systematic field walking, bibliographical studies, museum researches, etc. The variety of available data raises questions on the validity of spatial results based on archaeological materials of a different nature, temporality and spatial extent. The purpose of this preliminary study is to create a control tool that will be used for the interpretation of results while trying to extract the most valuable information to the archaeological interpretation. This tool is expressed spatially through what are called "confidence maps" which is a data layer produced by combining reliability and representation of the data.

Representation Maps

Evidence for data dispersion/location over separate study areas is symbolized with representation maps. They were designed with the aim of being standardised in respect to the theoretical mean of the individual study area (i.e. variations from the average). Therefore they allow the quantification and visualization of spatial heterogeneity in the sampling and the inventory of the different datasets. The number of archaeological items in each pre-defined grid cell is computed and this value is compared to the expected (usually mean) value in the study area, which gives an idea of the over- or under-representation of data.

To begin the analysis grid size has to be defined for each individual study area. The proposed optimal cell size calculation is based on the assumption that archaeological data is approximately evenly distributed, which means that each data object is assigned the same area, defined by the cell. The cell size is therefore "unique" for each study area because it is directly related to the area of investigation and the number of observations and in effect it is an average distance among observations (SÁNCHEZ 2006). In our case we have computed the optimal cell size as cell_size = sqrt(total_area/$N_{observations}$). This empirical method is based on the assumption that if the objects are normally distributed, then a similar area should approximately belong to every object. Therefore, the average area of an object can be computed by dividing the whole area of interest by the number of objects. This average area is square shaped when working with a regular grid, and means that the cell size of the grid can be computed by square rooting the average area. This number is then rounded and represents the optimal resolution. A similar approach is mentioned by Shary (Shary / Sharaya / Mitusov 2002). However, data is rarely evenly distributed. In order to improve the statistical significance we have chosen the first larger grid size, fitting the "standard" resolution system used in ArchaeDyn, i.e. 1, 2.5, 5, 10, 25, 50, 100, 250 km. This produces grids that are both optimal and well populated: that is containing a significant number of points. In order to simplify the process of data transformations and comparison of different datasets further, the common point of origin has been defined for all the grids which means the cell boundaries of different resolutions and study areas overlap at the same coordinates. This means that even different scale phenomena can be processed as imagery in order to combine their information over the same or different areas when it is relevant.

Representation classes were defined to stand for no data, normal, over and extreme representation *(see Fig. 1)*. It was found that these types of classes correspond to the nature of archaeological data, whose frequency is typically exponentially distributed and hardly ever normal. If it were the case then classes would be under, normal, and over represented. The approach is different from the previous work done by the group (Nuninger / Tourneaux / Favory 2008). Some unresolved issues that remain are the automatic or semiautomatic selection of thresholds for classes and the no-data phenomenon.

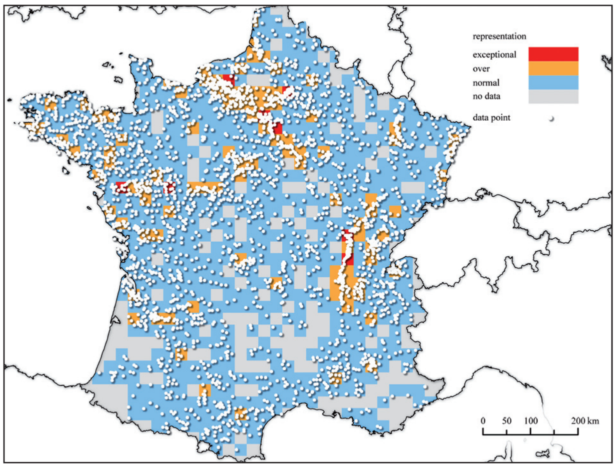

Fig. 1. A representation map of dated archaeological bronze objects in France (map: Z. Kokalj, data: F. Pennors).

Even though the process was designed with the aim of being non subjective and based purely on statistics, a uniform automatic statistical division of classes based on average proved to be unreasonable. This was due to the extreme data heterogeneity that included different distributions, differences in absolute values, no data phenomenon, and the use of integer values. According to our tests, the classification process has to be done (semi)manually and individually for every dataset with the help of statistical and mathematical tools. The usual procedure is based on histogram analysis and its modification using a logarithmic function, and defining the natural breaks in the data. The latter are especially difficult to define if absolute frequencies (representations) are low. This implies the importance of selecting the optimal grid size.

The problem of handling no data values has not been solved satisfactorily, but rather bypassed. The statistics can be significantly altered with the inclusion of cells with no data values in the calculation. The argument for including such values is the fact that the space is continuous and areas cannot be left out, however in cases where data is highly concentrated this can lead to dramatic decrease of the average and as a result even the areas with only one object can be classified as over represented. Increasing the cell size by one "standard" step and manual delimitation of classes avoided this problem because with the latter, the interpreter can manually classify such areas as normally represented and then the initial number of no data cells is effectively decreased anyhow. A problem which arises is the further concentration of extreme values and the resulting reduction of "contrast", but if this is not the primary concern it is well supplemented by improved overall legibility and accuracy of the final map.

Reliability Maps

Reliability maps express the settings (and limitations) of inventory exploration (i.e. how the archaeological sources were explored) in terms of common indicators such as survey level – sampling, visibility level, the quality of references etc., about a specific dataset. A reliability map gives information on the intensity of research and exploration (reliability of the inventory), and is not primarily concerned with the quality of the data's location. This

	Level 1 (reliable)	**Level 2 (fairly reliable)**	**Level 3 (not reliable)**
WG2 (Gandini et al. 2008) and WG1-manuring (Poirier / Tolle 2008)	1) areas where systematic field walking with spacing of 10 m maximum has been completed, and 2) where there are optimal visibility conditions (ploughing or vineyard or lavender).	1) areas where systematic field walking with spacing of more than 10 m has been completed, or 2) where systematic field walking has been carried out but there is only partial visibility of the ground (wildland, fallow, meadow, woods)	1) areas where only partial or no field walking has been performed and/or 2) there is very poor visibility due to land use and/or areas where significant taphonomic problems are assumed (sedimentary covering or erosion).
WG1-field systems (Georges-Leroy / Tolle / Nouvel 2008)	1) areas where systematic field walking (under forest condition) has been completed and 2) where there are optimal visibility conditions, 3) with a good precision in recording features < or = 10 meters	1) areas where punctual field walking has been completed or 2) where there is poor visibility (high density of vegetation...) and/or 3) imprecise records of features (error > 10 meters)	1) areas where very punctual or ancient field walking has been completed
WG3-Bronze objects (*Fig. 2.* and Gauthier 2008).	1) areas where the author of the database paid a special attention. 2) where field walking and excavation have been completed with a relatively high density of research/field walking (due to preventive archaeology, dredging) on the study area. 3) where data is easily accessible (straight access to raw data, no access limitation to the stored data – archaeological services, museum, private collection) and with many publications.	1) areas where the author of the database paid a special attention and/or 2) where field walking and excavation have been completed with a relatively medium to high density of research/field walking on the study areas but with less sufficiency and/or 3) where data are easily accessible (straight access to raw data, no access limitation to the stored data – archaeological services, museum, private collection) but with few publications only.	1) areas where the author of the database paid a good to fairly good attention and/or 2) where only partial or no field walking/excavations have been performed with almost no archaeologists working on the study area or without sufficiency and/or 3) where data are less accessible (no or partial access to raw data, limited access to the stored data data – archaeological services, museum, private collection) and with few publications only.

Tab. 1. Reliability rules (examples) defined by the workgroups of the ArchaeDyn's project (Nuninger / Tourneaux / Favory 2008).

means it also can be interpreted as a correlation between intensity of research and actually identified sites or archaeological evidence. In our case a reliability map covers the entire study area and distinguishes three reliability levels: reliable, fairly reliable and not reliable. It has been defined by the providers of individual datasets and has been mostly drawn by hand according to a predefined set of rules. The rules were defined by each workgroup and by each archaeological team. Indeed, these rules are depending on the kind of investigation. Nonetheless, each set of rules is written in accordance to the three predefined degrees which then allow comparisons to be made. The definition of reliability levels is adjusted according to the nature of data. For example, instead of field walking, data availability in museums or publications can be considered *(Tab. 1)*. The identification of individual levels is based on an empirical method as its foundation is the knowledge of the data quality, and is therefore inherently biased. It is also highly

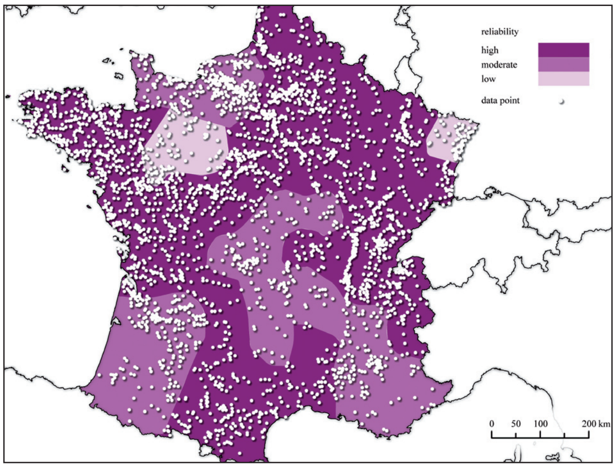

Fig. 2. A reliability map of dated archaeological bronze objects in France (map: Z. Kokalj, reliability zones and data: F. Pennors).

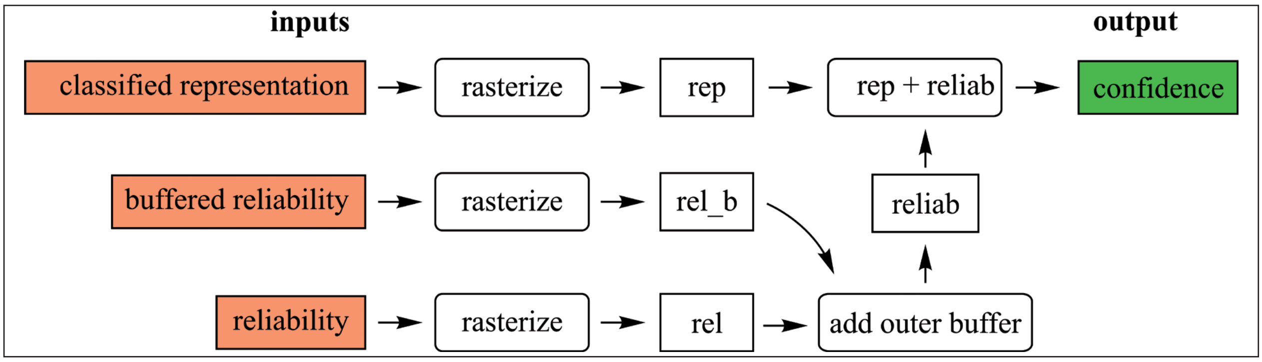

Fig. 3. Confidence map processing model.

dependent on the phase of studies and this directly connected to the state of the studied database. The ArchaeDyn's databases are, from now on, fixed at the present state of the investigation in order to provide analysis. New discoveries or new development of the database will be used by the end of the project during validation and for final interpretations.

Confidence Maps

Confidence maps act as a tool to evaluate the relevance of archaeological data in spatial analysis. They give an impression about the confidence and faith that a user can have about the final results based on the input data. The representation and reliability layers are combined using map algebra to produce confidence maps. The logic behind this lies in joining two spaces: location-based density (representation) and intensity of inventory (reliability). Results allow for the comparison and analysis of data confidence and thereby the evaluation of the interpretation and spatial modelling with respect to trustworthiness. They also give information about the correlation between data representation and reliability. The map can be used to eliminate "spurious" zones for space-time analysis over the long-term according to the comparison of each study area along with its chronology and the interpretation key of the representation map.

The proposed process is essentially based on simple algebraic operations and "binary" logic. The confidence was coded into two digit numbers, with one digit reserved for representation and the other for reliability. To technically enable the addition, the representation map has to have "denary" classes, 10, 20, 30, and 40, being either an extreme representation, over representation, normal representation or no data, respectively, the reliability map was given values of 1, 2, and 3, ranging from high to moderate to low reliability. Another technical issue is an accurate rasterization of the reliability map. Normal rasterization omits border areas with less than half cell occupancy. Consequently a 3/4 cell size buffered layer with preserved attributes has to be created and rasterized. Its outer buffer is then added to the rasterized reliability and the result combined with the representation map. An ArcGIS tool was designed to speed up and enable batch processing.

The ensuing confidence map is in effect an overlay of both maps *(see Fig. 3)*. By inspecting the map one can immediately find areas of different representation but also areas with low data reliability. The strongly coloured areas are more reliable than the light coloured areas. Both can and should be included in the analyses with a degree of caution. The proposed process can also be applied to analyse and compare other spatial phenomena, and tests are underway to evaluate the process for effectiveness in representing temporal changes.

Some difficult to manage issues still remain

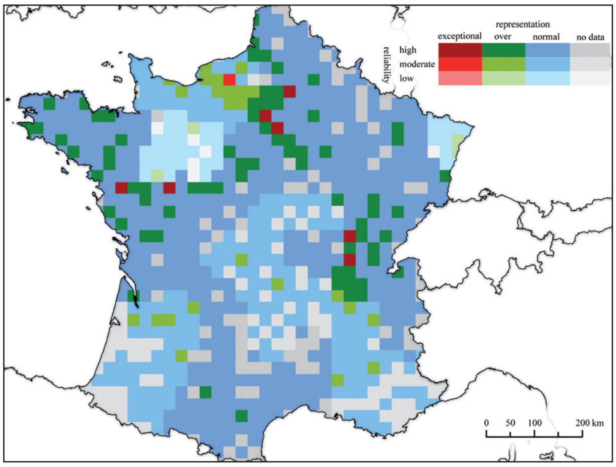

Fig. 4. A confidence map of dated archaeological bronze objects in France (map: Z. Kokalj, data: F. Pennors).

in this approach. Questions, such as how to discretize representation maps and how to interpret areas with no data will need to be addressed in the future.

Conclusions

To represent the level of trust of the spatial analysis and modelling results we have defined a tool called confidence maps. Confidence maps provide the user a spatial impression about the representation and the reliability of the input data at the same time, which gives us the opportunity to then detect "artefacts" in the data. The same methodology has been defined for different scales and for different observed phenomena. Despite the fact that the data used can be very dissimilar the interpretation of confidence maps is the same. This is a welcome innovation especially when considering the extent of the ArchaeDyn project.

There are still some problems that remain to be solved. Confidence maps are not suitable for all databases. They better suit databases containing "noise" and perform better with large amount of statistically well represented data. We have also found a rather strong scale dependence of the results. Different tests have shown that the tool does perform better with small scale (big area), a large quantity of points (often it will be studies of objects and not sites or settlements), and a low positional accuracy (studies about the diffusion of material, circulation of artefacts).

The confidence maps methodology is still in development and in the future we intend to improve the individual processing steps and overcome the mentioned limitations.

Acknowledgments

Archaeological data used in the study was obtained in the frame of ArchaeDyn project. Part of the work has been performed within the ModeLTER (European Laboratory for Modelling of Landscapes and Territories over the Long Term), institute founded by ZRC SAZU and CNRS.

References

GANDINI ET AL. 2008
C. GANDINI / F. BERTONCELLO / E. GAUTHIER / L. NUNINGER / F. TRÉMENT, Hierarchical Typology and Settlement Patterns Modeling at Interregional Scale. In: Present volume, 278.

GAUTHIER 2008
E. GAUTHIER, Consumption and Circulation of Prehistoric Products in Europe: Characterization of Spatial Evolutions Using Map Algebra. In: Present volume, 375.

GEORGES-LEROY / TOLLE / NOUVEL 2008
M. GEORGES-LEROY / F. TOLLE / P. NOUVEL, Analysis of the Intensity of Agrarian Exploitation by Spatial Analysis of Ancient Field Systems Preserved by the Forest Cover. In: Present volume, 281.

NUNINGER / TOURNEUX / FAVORY 2008
L. NUNINGER / F. P. TOURNEUX / F. FAVORY, From Archaeomedes to ArchaeDyn. In: Present volume, 278.

POIRIER / TOLLE 2008
N. POIRIER / F. TOLLE, Measurements of Diachronic Stability of Agrarian Exploitation. In: Present volume, 281.

SÁNCHEZ 2006
J. SÁNCHEZ, The Accuracy of Grid Digital Elevation Models Linearly Constructed from Scattered Sample Data. International Journal of Geographical Information Science 20:2, 2006, 169–192.

SHARY / SHARAYA / MITUSOV 2002
P. SHARY / L. SHARAYA / A. MITUSOV, Fundamental quantitative methods of land surface analysis. Geoderma 107, 2002, 1–32.

Krištof Oštir

Scientific Research Centre
ZRC SAZUModeLTER
Gosposka ul. 13
1000 Ljubljana, Slovenia
kristof@zrc-sazu.si

Žiga Kokalj

Scientific Research Centre
ZRC SAZUModeLTER
Gosposka ul. 13
1000 Ljubljana, Slovenia
zkokalj@zrc-sazu.si

Laure Saligny

Pôle Géomatique et Cartographie
Maison des Sciences de l'Homme de Dijon UMS 2739,
Dijon, France

Florian Tolle

University of Franche-Comté,
Laboratory of Chrono-Ecology UMR 6565
16 Route de Gray
25030 Besançon, France
florian.tolle@univ-fcomte.fr

Laure Nuninger

CNRS, Laboratory of Chrono-Ecology UMR 6565
16 Route de Gray
25030 Besançon, France
laure.nuninger@univ-fcomte.fr

Françoise Pennors

Protohistoire égéenne-UMR 7041 (ArScAn)
Maison R. Ginouvès,
21 allée de l'Université,
92023 Nanterre, France

Klemen Zakšek

Scientific Research Centre
ZRC SAZUModeLTER
Gosposka ul. 13
1000 Ljubljana, Slovenia
klemen.zaksek@zrc-sazu.si