# Quantifying the State of the Art

Christian Mayer
Federal Commission on Historical Monuments,
Vienna, Austria

Abstract

Databases of archaeological sites provide a huge amount of information on the archaeological landscape. This information does not only refer to the location of a site but also to exactness of dating, number of reports and their temporal distribution thus providing an important source for assessing the state of the art in an archaeological perspective.
To analyse this information, each site is scored according to formal aspects (location, dating etc.) as prescribed by the data model of the database to which the information is stored and weightened against the maximum possible score of a site.
The spatial distribution of the scores is then analysed by means of Geostatistics, spatial autocorrelation and Geographically Weightened Regression (GWR) to provide an overview over the knowledge of the archaeological landscape under investigation.
The paper will present a study of about 25,000 sites from about 9,000 geographical units in Austria where information is taken from the National Archaeological Survey conducted by the Federal Commission on Historical Monuments, Vienna. Aim of the study is to explain the local differences in knowledge of Austria's archaeological landscape such determining regions of poor archaeological provision.

To quantify the state of the art is an important task, since this is the only way to point out deficits in the general set-up of a science. The usual approach to obtain such quantification is to rate publications by rank of the journals in which they are published. Certainly, that approach will give a general picture of the quality of scientific work done so far. But as far as archaeology is concerned, a spatial concept is of interest, that provides an idea how well an archaeological landscape is known. In this paper an alternative approach will be put forward that uses a database of an archaeological survey, actually its data model, to develop a measure for the state of the art of an archaeological landscape.

The database used for this purpose is the National Archaeological Survey (NAS) run by the Austrian Federal Commission on Historical Monuments. Its data model (MAYER 1996, 2002) discerns between topographical places and sites, where the latter contains - among others - information on dating, location, cultural affiliation and class of site (like settlement, burial place, deposit and stray finds) grouped by class of site and dating.

| score | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| location | unknown | cadastrial unit | field name | coordinates | parcel of land |
| dating | unknown | period 1 (prehistoric,..) | period 2 (Iron Age) | period 3 (Early Iron Age) | phase (Ha C,..) |
| cultural affiliation | unknown | known | | | |
| class of site | unknown | contex unknown | settlement, burial place | specification of the latter | |

Table 1 Scores

Basically, an entry into the database sets off with a record of any archaeological item, either find object or locality and this information is stored according to its available evidence. The data model formalises this evidence into a number of elements such giving an idea how well an archaeological item is known. In fact the terminology by which the evidence is transformed into a record is hierarchically stratified. The more evidence is available, the higher is the level in the hierarchy of the term applied. This provides us with a concept of exactness for each of the 4 major elements mentioned above (Tab.1): Assigning a score to each level of hierarchy in the terminology gives a measure of the amount of knowledge about each element. The sum of these scores serves as a measure of the amount of knowledge about the item as a whole. Since more than one site may be found at the same topographical place, the scores of each site are summed and the sum is divided by the number of sites at a place.

As an illustration of some methodological aspects of the approach introduced here, scores of prehistoric sites will be used that stem from the parts of Austria already covered by the NAS. The study area comprises 44,613.58 km² that is 53.18% of the Austrian territory. 5,640 out of 10,082 mapable places are considered. A full account of the results obtained from the analysis of all data will be given by a mapping project and be published in the Fundberichte aus Österreich.

The distribution of the sum of scores is of major importance for the methodological aspects of this approach to quantify the state of art. Firstly, the scores used here are discrete, meaning that only 13 distinct values can be attributed. Of course, a finer terminological grid would produce a greater variety in the values, therefore the distribution of scores is a distribution in blocked form. Secondly, this distribution gives an idea how we can rate the state of the art as good or bad: It is clear that a better state of the art than found in the working area would have a higher average score. If the state of the art is homogenous through the working area, the standard deviation would be small. Further, if the average of scores gets close to the maximum obtainable score, the distribution may be skewed to the right. Of course, life would be easier if the scores would be normally distributed which they are not. Therefore, we cannot use the three first moments of a normal distribution (average, standard deviation and skewness) but at least we can replace them by their robust counterparts.

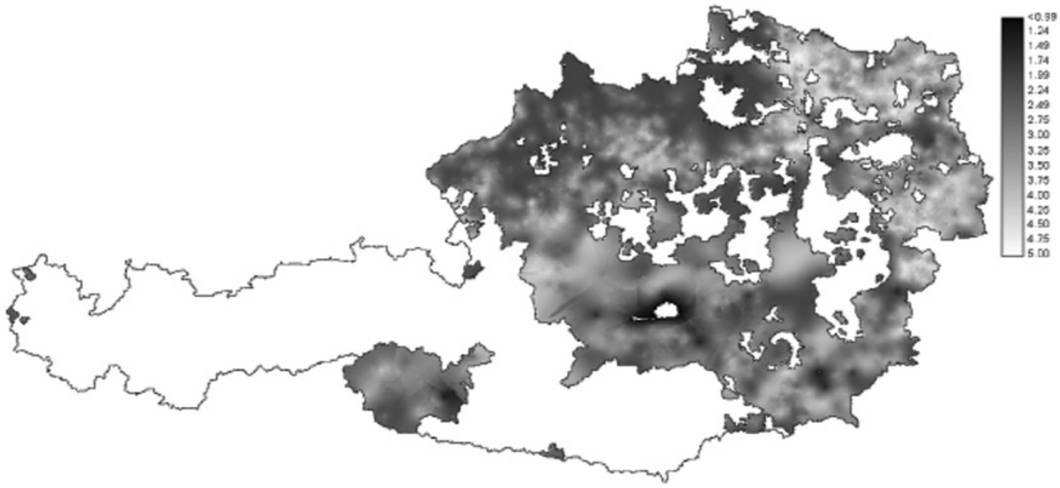


Figure 1 State of the Art in Austrian Prehistoric Archaeology

By virtue of the problem, we are dealing with a spatial phenomenon. Therefore, the overall distribution of scores calculated from the working area does not describe the state art in a region satisfactorily. Applying krigeing (Cressie 1993:29) as a method of spatial analysis to the data, we obtain a spatial picture of the general situation. But this picture is as informative as deceptive. In fact the cross validation coefficient of the underlying model is only about 0.45. Although a highly significant value, the determination coefficient is only about 20%. Of course, one wonders where the rest of the variation has gone. Looking at the variogram we see that the variation within the data sets off on an already high value at low distances and the spatial dependency is measurable only to rather small distances (about 1.8 km). This means, that good knowledge of one site does not necessarily guarantee a good knowledge of a site nearby. What about exactness of location, class of site, dating and cultural affiliation ? Exactness of location as well as class of site showed simply no spatial dependency. This is because the scores of these two attributes of the sites have very low standard deviations, the exactness of location due to the excellent archive work done with the archaeological survey, class of site because most of the sites are known only by surface finds. The map shown in Figure1 shows thespatial distribution of the sum of cultural affiliation and exactness of dating. The other variables are ignored for being of less importance. The correlation between estimated and actual data is only about 0.5 for reasons discussed already. Clearly "good knowledge" is quite rare.

Of course, one is interested to know, by what effects the scores are influenced. For instance, one would like to know, whether the vicinity of a museum influences the state of art in its vicinity. Typically, this question is transformed into a regression model by using the scores as the dependent variable and the distance to the nearest museum as the regressors. To illustrate this, the location of about 370 museums were collected and the distances to each mapable site calculated. To provide a better picture, the museums were classified to their body of responsibility being run by a private person, a local authority or a local society as the first group, the federal states as second group and the Republic of Austria as the third group. A classical regression model using these variables give a general correlation of only 0.125. Again this result is deceptive since classical regression does not consider the local dependency between the variables. As an alternative, geographically weighted regression has been applied (Fotheringham et al. 2002). Applying this method to the data, we obtain a correlation coefficient of 0.61. Variable selection detects the federal museums to be of little interest in this respect, since they are all situated at Vienna. Leaving them out we obtain a correlation coefficient of 0.60, museums run by the federal states and local museums produce a correlation coefficient of 0.57 and 0.58 respectively. Looking at these figures a second time the results seem paradox: One has to expect, that the closer a site is to a museum, the higher the knowledge of the site would be. In that case, the correlation coefficients would be negative, which they are not. In a map of the local correlation coefficients there are areas with very large negative values that are indeed close to museums. But again, this is not a general phenomenon since there are quite a lot of museums that have obviously no influence on the knowledge of their surrounding archaeological landscape. Further, one should not mix high negative correlation coefficients with good knowledge of the region. When the scores are multiplied with the sign of the correlation coefficient, it becomes obvious, how little the presence of a museum influences the knowledge of the surrounding archaeological landscape. Obviously, whether a museum influences the state of art in its vicinity depends on the museum's activity but not simply on its existence.

Undoubtedly, the quantification of the state of art in archaeology by the proposed approach has produced very interesting insight into Austrian archaeology, although results presented are not necessary good news.

## References

CRESSIE, N.A.C, 1993. Statistical for Spatial Data. Wiley.

FOTHERINGHAM, A.St., BRUNSDON, Ch. and CHARLTON, M., 2002. Geographically weighted regression, Wiley.

MAYER, Chr., 1996. Fundstellenbezogene Daten in der Abteilung für Bodendenkmale des Bundesdekmalamts, Fundberichte aus Österreich 35:321.

MAYER, Chr., 2002. Some aspects of SMR management in Austria. In Mapping the future of the past 37.