The future of archaeological predictive modelling

Verhagen, P.; Kamermans, H.; Leusen, M. van; Kamermans, H; Leusen, M van; Verhagen, Ph

Archaeological Prediction and Risk Management

Leiden University Press

Archaeological Studies Leiden University
is published by Leiden University Press, the Netherlands

Series editors: C.C. Bakels and H. Kamermans

Cover Design: Medy Oberendorff
Layout: Hans Kamermans and Medy Oberendorff
Illustrations: Joanne Porck and Medy Oberendorff

ISBN 978 90 8728 067 3
e-ISBN 978 90 4851 063 4
NUR 682

ARCHAEOLOGICAL STUDIES LEIDEN UNIVERSITY 17

Archaeological Prediction and Risk Management

Alternatives to Current Practice

Edited by

Hans Kamermans
Martijn van Leusen
Philip Verhagen

Contents

2. The future of archaeological predictive modelling[5]

Philip Verhagen[6], Hans Kamermans[7] and Martijn van Leusen[8]

2.1 Introduction

In general, academic archaeologists have been sceptical of, and sometimes even averse to, predictive modelling as practiced in archaeological heritage management (AHM) (see Van Leusen *et al*. 2005). The models produced and used in AHM are not considered sophisticated enough, and many of the methodological and theoretical problems associated with predictive modelling have not been taken aboard in AHM. At the same time, the production and use of predictive models has become a standard procedure in Dutch AHM (Deeben *et al*. 1997; 2002; Deeben 2008), and it has clearly attracted interest in other countries as well.

The main reason for using predictive models in AHM is efficiency. In 'post-Malta' archaeology, the financial, human and technical resources allocated to archaeology have increased enormously. But at the same time, these resources have to be spent both effectively and efficiently. So why not create and use tools that will allow us to do so? Archaeological predictive models will tell us where we have the best chances of encountering archaeology. Searching for archaeology in the high probability areas will 'pay off', as more archaeology will be found there than in low probability zones. It is a matter of priorities: we can not survey everything, and we do not want to spend money and energy on finding nothing. And there is also the political dimension: the general public wants something in return for the taxpayers' money invested in archaeology. It's not much use telling politicians to spend money on research that will not deliver an 'archaeological return'. But how can we be so sure that the low probability zones are really not interesting? And where do we draw the line between interesting and not interesting? These are hard choices indeed for those involved in AHM. Archaeologists who do not have to make these choices have an easy job: they can criticize the current approach to predictive modelling from the sidelines, and do not have to come up with an alternative.

Within the BBO program we have been trying to provide such an alternative to the archaeological community (Kamermans *et al*. 2005). However, at the end of the project, we have to conclude that we have only been partly successful. We have done a fair amount of research, published three books and many papers, made the problems with predictive modelling internationally visible but failed to change the procedures of predictive modelling in the Netherlands. In this paper we venture to offer some explanations for the lack of success of new approaches to predictive modelling in AHM in the Netherlands up to now. And finally, we will try to sketch the future of archaeological predictive modelling, for which we can see three distinct scenarios.

2.2 Problems and solutions

Over the past twenty-five years, archaeological predictive modelling has been debated within the larger context of GIS applications in archaeology, and the processual/post-processual controversy that has dominated the archaeological theoretical debate since the late 1980s (see Van Leusen *et al*. 2005). Predictive modelling is rooted in the processual tradition, with its emphasis on generalization and quantitative 'objective' methods, and its lack of interest in the subjective and individual dimensions of archaeology. In itself, this is not a matter of 'bad' versus 'good' archaeology. Within the context of AHM, generalized maps are necessary tools to reduce the enormous complexity of archaeology to manageable proportions. However, the lack of real interest in using spatial technology and statistical methods in post-processual academic archaeology has certainly slowed down the development of predictive modelling as a scientific method. The feeling that processual approaches no longer offered a real contribution to the advancement of archaeological science has left predictive modelling

[5] A slightly different version of this chapter has been published as Verhagen *et al*. 2008.

[6] ACVU-HBS, Amsterdam, the Netherlands.

[7] Faculty of Archaeology, Leiden University, the Netherlands.

[8] GIA, Groningen University, the Netherlands.

somewhat 'lost in space' it seems. Which is a pity, because even if we do not want to use predictive modelling in an AHM context, there still is a lot of potential in spatial technologies ('GIS') to develop and test theories of spatial patterning of settlements and human activities.

The criticism of predictive modelling in scientific literature has focused on three main issues: statistics, theory and data. In all three areas, predictive modelling as it stands today is considered by various authors to insufficiently address the complexity of the matter. Statistical methods are used uncritically, often using a limited number of techniques that are not the best suited around. Archaeological theory, especially where it concerns the human and temporal factors in site placement, only plays a marginal role in selecting the variables used for predictive modelling. And archaeological data, which we all know have various degrees of reliability, are used without much source criticism. And while this is all very much true, and many archaeological predictive maps are rather coarse representations of a complex archaeological reality, these criticisms mask a more fundamental question: what is the quality that is *needed* of a predictive model? Because this is precisely why models are made that are not very sophisticated from a scientific point of view: they are considered good enough for the purposes they are made for[9].

Our one fundamental problem with predictive modelling is therefore this issue of quality. No one seems to know what constitutes a 'good' model, and no tools are available and used to make explicit the quality of the models. Within our research project, we have tried to focus on these issues by looking at the potential of new statistical techniques for incorporating uncertainty in the predictions, and by studying the best ways of testing the models (this volume chapters 7 and 8). Our first foray into the uncharted waters of model quality concerned the role of expert judgement in a quantitative framework. When the first criticisms of predictive modelling appeared in the late 1980s, it quickly became clear that a fully 'inductive' approach was in many cases unsatisfactory. The lack of reliable survey data in many areas of the world basically ruled out a rigorous statistical approach, unless vast amounts of money were invested in survey. The pragmatic solution therefore was to stop using statistical methods for developing predictive models, but instead rely on expert judgement, and see if the experts' views were corroborated by the available archaeological data. However, in doing so, a major advantage of statistical methods was thrown overboard: the ability to come up with estimates of for instance site density, in real numbers, and the calculation of confidence intervals around the estimates. Expert judgement models only classify the landscape into zones of low, medium and high probability without specifying the numbers involved. How many archaeological sites can we expect in a high probability zone? And how certain can we be of this estimate with the available data? Statistical methods will provide these numbers, expert judgement will not.

It turns out that Bayesian statistical techniques are very well suited to provide numerical estimates and confidence intervals on the basis of both expert judgement and data (Millard 2005; Verhagen 2006; Finke *et al.* 2008). However, they have not been used in anger in predictive modelling so far. This is probably because of the relative complexity of the calculations involved. There are very few archaeologists in the world that can perform these calculations, even though computing power is now no longer an obstacle. We have however proved that it can be done, and we see Bayesian statistics as a very powerful and useful tool for predictive model building (this volume chapter 9).

We also tested the potential of Dempster-Shafer modelling (this volume chapter 9), which has been suggested as an alternative to 'standard' statistical methods as well (Ejstrud 2003; 2005). While the results of earlier modelling studies indicated that it performed better than most statistical tools including Bayesian analysis it is only partly suitable to include expert judgement. Furthermore, its conceptual basis is rather complex. We will not go into detail here; suffice it to say that Dempster-Shafer modelling is more controversial in statistical science than Bayesian statistics, and it is more difficult to understand.

However, even if tools like Bayesian statistics can build a bridge between expert judgement and quantification, we still need reliable data to have it delivering its potential. The testing issue is therefore of

9 However not all users consider the maps good enough (see this volume chapter 6).

primary importance to predictive modelling. What is probably most problematic in this respect is the lack of attention by archaeologists to the simple statistical principle of data representativity. Nobody seems to wonder if the available archaeological sample represents in any way the original archaeological population. No matter what statistical method is used, this issue needs to be addressed first before attempting to produce a numerical estimate of any kind. And while it is possible to reduce the bias encountered in existing archaeological survey data to an acceptable level, in order to have reliable archaeological predictive models we also need to survey the low probability zones. So here we are facing a real paradox: predictive models are developed to reduce the amount of survey (or even skip it) in low probability zones, yet statistical rigour tells us to do survey there as well.

Our approach has been to re-assess the value of using statistical methods in predictive modelling (this volume chapters 7 and 8). We are convinced that this re-assessment is necessary, and think that it can offer a valuable contribution to AHM as well. If we can base the models on sophisticated statistical methods and reliable data, then we can really start using predictive models as archaeological and/or economic risk management tools.

In the end, we have not been able to get this message across to the AHM community in the Netherlands yet. While we have not done an exhaustive survey among our colleagues, we think that the following reasons may be responsible for it:

1. the innovations suggested are too complex. While it is sometimes said that statistics are not very difficult, but only very subtle, in practice most archaeologists are not working with it on a daily basis. Some even have difficulty grasping the most fundamental principles of quantitative methods. This makes it hard to get the message across, as it does not really help when we have to bridge a large gap in knowledge between the statistical experts and the people who have to use the end results of statistical models.
2. shifting from the current expert judgement approach to a more sophisticated statistical approach is too expensive. Improving the models in the way we suggest does not *replace* anything in the current way of dealing with predictive modelling, it only *adds* to it. So, on top of the things we already do, like gathering and digitising all the available information, and interviewing the experts, we now also need to have a statistical expert doing the modelling, a data analysis programme to detect and reduce survey bias, and perhaps even a test survey as well.
3. it is irrelevant. While academic researchers may be bothered with the quality of the models, most end users are not. They trust the experts. Especially those responsible for political decision making will not care, as they only need clear lines drawn on a map, that will tell them where to do survey and where not. If the archaeologists are happy with it, then they are as well.
4. and this ties in to our last explanation: archaeologists may have reason to be afraid for more transparent methods that will give insight into the uncertainties of predictive models to non-archaeologists. When anyone can judge model quality, they will lose their position of power in dealing with politicians and developers.

2.3 Recommendations

For the moment we have not been able to convince our AHM colleagues of the need to improve predictive modelling along the lines suggested. We certainly did not have enough time and money to fully develop these new approaches into practical working solutions, but at least we have showed some promising avenues to follow. From our research, a number of recommendations result for each of the themes originally identified in the baseline report (van Leusen *et al*. 2005).

Concerning the quality and quantity of archaeological input data we recommend:

- to use sophisticated statistical methods (Bayesian statistics, Dempster-Shafer modelling) to integrate expert judgement and archaeological data in one quantitative framework
- to improve the way of registering archaeological fieldwork data, including the methods of data gathering, and the location of areas where no archaeological remains were found
- to analyse archaeological fieldwork data for the presence of systematic biases that are either the consequence of data collection strategies, or of geological conditions, and use methods for correcting these biases to come up with more representative data sets

Concerning the relevance of environmental input data we recommend:

- to use new relevant data sources like LIDAR elevation models, geological borehole databases, historical maps and remote sensing images
- to explicitly describe and judge the quality and reliability of environmental input data

Concerning the need to incorporate social and cultural input data we recommend:

- to take a critical look at modern archaeological theory, and see where the social and cultural factors involved in site location can be translated into spatial predictive models

Concerning the lack of temporal and/or spatial resolution we recommend:

- to always produce ‘diachronic’ predictive models, in which both environmental and cultural dynamics over the long term are represented
- to adapt the scale of the mapping to the intended use; maps that are only meant to provide a baseline for spatial planning zoning can be less detailed than maps that should also provide detailed advice on the type of archaeological research that should be done in those zones

Concerning the use of spatial statistics we recommend:

- to apply appropriate statistical methods for modelling (like Dempster-Shafer modelling) and testing (like resampling), in order to prevent creating spurious correlations and over-optimistic views of model quality

Concerning the testing of predictive models we recommend:

- to set up long-term field testing programs, focusing on the areas where uncertainties are highest
- to use appropriate statistical methods to judge the quality of predictive models

Most of these recommendations will not come as a surprise, even though we place a stronger emphasis on statistical rigour than is currently usual. Improvements in the quality of the environmental input data and the temporal and spatial resolution of predictive modelling are already clearly observed over the past few years. Even the inclusion of social and cultural factors in predictive models is now gradually becoming a feasible option, and can easily fit into academic research projects. In contrast, the rigorous testing of predictive models is still far from becoming a standard procedure. The reasons for this are on the one hand found in the highly scattered nature of ‘post-Malta’ archaeological prospection that is only done in high-probability areas were survey is enforced. Field testing, while sometimes part of a predictive modelling project, is usually only done to check the modelling assumptions: are *e.g.* the geomorphological units mapped correctly, and where are the highly disturbed zones? Systematic collection and analysis of survey results is a time-consuming task, and when it is done, it is primarily used to produce up-to-date catalogues of archaeological finds, rather than for purposes of model development and testing – which in many cases would imply collecting data for larger areas than the region modelled. On the other hand we observe a lack of interest in the issue of model quality by authorities from the national to the local level. This may partly be attributed to the fact that archaeology is not such a big economic risk after all; developers and authorities do not expect a direct benefit from investing in risk management methods. The *scientific* risk involved however should worry all archaeologists.

There is no way to prevent the development of a highly biased archaeological record if we do not take the issue of predictive model quality and testing seriously. This is primarily a task for the archaeological community itself, including companies, universities and the national archaeological service. And yes, it will need money – which implies defining different priorities and making different choices when applying for funding and when spending budgets. The current procedures in Dutch archaeological heritage management however preclude prioritising this kind of research, as they are purely focused on selecting and eventually excavating individual archaeological sites. The recently published National Archaeological Research Agenda will probably only reinforce the focus on localized interpretive archaeological research, rather than on methodological improvement and synthetic research.

2.4 Conclusions

What will be the future of predictive modelling in AHM in the Netherlands? We think that there are three scenarios that may be followed in the next ten years or so. The first one is a scenario of ‘business as usual’. There are some points that speak in favour of this option. First of all, predictive maps are fully accepted by the Dutch archaeological community and are relatively well embedded in planning procedures. Secondly, as far as is known, the use of the predictive maps that are currently around has not led to archaeological disasters, even though some grumbling is occasionally heard about the quality of the maps. Rather, it is the other way around. Predictive maps, in spite of there theoretical and methodological shortcomings, are more than effective in protecting the archaeological heritage: they over-protect. This of course is an unwanted situation. In practice, municipal authorities commissioning predictive maps for their own territory do this with the explicit aim of reducing the area where preliminary research is needed. We might be heading to a future where commercial predictive modelling will have as its highest aim the reduction of the zones of high archaeological probability – without having the tools to judge whether this reduction is supported by the archaeological data.

Cautious archaeologists would therefore certainly prefer the second possible scenario: this is, to stop using predictive models, and do a full survey of all the threatened areas. Obviously there are many countries in the world that can do archaeological heritage management without predictive maps. Even in the United States, full survey is sometimes seen as a feasible alternative. In its favour speaks the reassuring thought that all archaeological remains present will be detected, and if necessary protected or excavated. However, this scenario is a political impossibility in the Netherlands, for the reasons mentioned above: politicians want less money to be spent on archaeology, not more. And even in countries where full survey is supposedly done, as in France or the United Kingdom, preliminary selection by means of desk-based assessment plays an important role in deciding where to do what kind of archaeological research. The question then is: is the Dutch method of selection by means of predictive maps better than doing the desk-based assessments that are common practice in France and the United Kingdom?

There is no way we can escape doing selections in archaeological heritage management. However, we need adequate tools on which to base these selections. Which brings us to the third scenario: the further development of predictive models into true risk assessment tools? There are, we feel, at least three supporting arguments for moving towards quantitative mapping of model uncertainty. First of all, there is the question of model quality and testing we already discussed. At the moment, expert judgment is among the inputs used to determine whether a zone is placed into high, medium of low probability, and uncertainties regarding this classification are never specified. However, expert judgment can never serve as an independent criterion of model quality. For independent model testing, we need data sets based on representative samples of the archaeological site types predicted. Secondly, the absence of estimates of the uncertainties in predictive models may lead to ‘writing off’ zones of low probability that are in fact zones where little archaeological research has been done. By including uncertainty measures in the models, it may be possible to break through the vicious circle of self-fulfilling prophecies that is created by doing ever more surveys in zones of high probability. And thirdly, the use of true statistical estimates and confidence intervals brings with it the perspective of making risk assessments in euros, rather than in relative qualifications of site density. Predictive modelling then may provide a first assessment of the bandwidth of the archaeological costs of a development plan.

References

Deeben, J., D. Hallewas, J. Kolen, and R. Wiemer 1997. Beyond the crystal ball: predictive modelling as a tool in archaeological heritage management and occupation history. In W. Willems, H. Kars, and D. Hallewas (eds), *Archaeological Heritage Management in the Netherlands. Fifty Years State Service for Archaeological Investigations*, 76-118. Amersfoort: Rijksdienst voor het Oudheidkundig Bodemonderzoek

Deeben, J.H.C., D.P. Hallewas and Th.J. Maarlevelt 2002. Predictive modelling in archaeological heritage management of the Netherlands: the indicative map of archaeological values (2nd generation). *Berichten ROB* 45, 9-56. Amersfoort: Rijksdienst voor het Oudheidkundig Bodemonderzoek

Deeben, J.H.C. (ed.) 2008. *De Indicatieve Kaart van Archeologische Waarden, derde generatie*. Rapportage Archeologische Monumentenzorg 155. Amersfoort: RACM

Ejstrud, B. 2003. Indicative Models in Landscape Management: Testing the Methods. In J. Kunow and J. Müller (eds), *Landschaftsarchäologie und geographische Informationssysteme: Prognosekarten, Besiedlungsdynamik und prähistorische Raumordnungen. The Archaeology of Landscapes and Geographic Information Systems: Predictive Maps, Settlement Dynamics and Space and Territory in Prehistory*, Forschungen zur Archäologie im Land Brandenburg 8. Archäoprognose Brandenburg I, 119-134. Wünsdorf: Brandenburgisches Landesamt für Denkmalpflege und Archäologisches Landesmuseum

Ejstrud, B. 2005. Taphonomic Models: Using Dempster-Shafer theory to assess the quality of archaeological data and indicative models. In M. van Leusen and H. Kamermans (eds), *Predictive Modelling for Archaeological Heritage Management: A research agenda*. Nederlandse Archeologische Rapporten 29, 183-194. Amersfoort: Rijksdienst voor het Oudheidkundig Bodemonderzoek

Finke, P.A., E. Meylemans and J. Van de Wauw 2008. Mapping the Possible Occurrence of Archaeological Sites by Bayesian Inference. *Journal of Archaeological Science* 35, 2786-2796

Kamermans, H., J. Deeben, D. Hallewas, P. Zoetbrood, M. van Leusen and Ph. Verhagen 2005. Project proposal. In M. van Leusen and H. Kamermans (eds), *Predictive Modelling for Archaeological Heritage Management: A research agenda*. Nederlandse Archeologische Rapporten 29, 13-23. Amersfoort: Rijksdienst voor het Oudheidkundig Bodemonderzoek

Leusen, P.M. van, J. Deeben, D. Hallewas, P. Zoetbrood, H. Kamermans and Ph. Verhagen 2005. A Baseline for Predictive Modelling in the Netherlands. In M. van Leusen and H. Kamermans (eds), *Predictive Modelling for Archaeological Heritage Management: A research agenda*. Nederlandse Archeologische Rapporten 29, 25-92. Amersfoort: Rijksdienst voor het Oudheidkundig Bodemonderzoek

Millard, A. 2005. What Can Bayesian Statistics Do For Predictive Modelling? In M. van Leusen and H. Kamermans (eds), *Predictive Modelling for Archaeological Heritage Management: A research agenda*. Nederlandse Archeologische Rapporten 29, 169-182. Amersfoort: Rijksdienst voor het Oudheidkundig Bodemonderzoek

Verhagen, Ph. 2006. Quantifying the Qualified: the Use of Multi-Criteria Methods and Bayesian Statistics for the Development of Archaeological Predictive Models. In M. Mehrer and K. Wescott (eds), *GIS and Archaeological Predictive Modeling*, 191-216. Boca Raton: CRC Press

Verhagen, Ph, H. Kamermans and M. van Leusen 2008. The future of archaeological predictive modelling. *Proceedings of Symposium The Protection and Development of the Dutch Archaeological Historical Landscape: The European Dimension, 20-23 May 2008, Lunteren*