University of Groningen

Epistemic considerations about uncertainty and model selection in computational archaeology

Peeters, Johannes; Romeijn, Jan-Willem

Published in:
Uncertainty and Sensitivity Analysis in Archaeological Computational Modeling

DOI:
[10.1007/978-3-319-27833-9_3](https://doi.org/10.1007/978-3-319-27833-9_3)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2016

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Peeters, J., & Romeijn, J-W. (2016). Epistemic considerations about uncertainty and model selection in computational archaeology: A case study on exploratory modeling. In M. Brouwer Burg, H. Peeters, & W. Lovis (Eds.), *Uncertainty and Sensitivity Analysis in Archaeological Computational Modeling* (pp. 37-58). (Interdisciplinary Contributions to Archaeology). Springer. https://doi.org/10.1007/978-3-319-27833-9_3

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Chapter 3

Epistemic Considerations About Uncertainty and Model Selection in Computational Archaeology: A Case Study on Exploratory Modeling

Hans Peeters and Jan-Willem Romeijn

3.1 Introduction

Uncertainty is an integral aspect of all scientific models in any field of application: we are mostly unsure which hypothesis is adequate as a means of predicting, representing, or reconstructing the system of interest. In quantitative research, the problem of uncertainty is often tackled by statistical means. Archaeological spatial modeling of human behavior is no exception to this: uncertainty is largely approached from a statistical perspective (e.g., Verhagen 2007). The available data allow us to test and choose among hypotheses, using classical statistical tools, or else to establish a probability assignment over a given range of hypotheses by Bayesian means. It is common to both classical and Bayesian methods that the model precedes any such treatment of uncertainties. Hence, the uncertainty that pertains to the model itself does not normally come into view in these statistical approaches. It is precisely the model uncertainty that is at stake in our paper.

In spatial models of human behavior, the sources of model uncertainty are numerous and compounded. The different input components of the model each come with their own uncertainty (e.g., paleolandscape models, assumptions about past human behavior, biased datasets). Moreover, owing to the complexity of the models, the modeling results become critically sensitive to misspecifications, so mistakes in the modeling assumptions have comparatively large effects on the model output. Finally,

H. Peeters (✉)

Groningen Institute of Archaeology, University of Groningen, Groningen, The Netherlands
e-mail: j.h.m.peeters@rug.nl

J.-W. Romeijn

Department of Philosophy, University of Groningen, Groningen, The Netherlands
e-mail: j.w.romeijn@rug.nl

the data that are used to determine the best fit within a statistical model are often also used to motivate particular modeling choices. In other words, the hypotheses we confront with the data were constructed on the basis of those very data. Apart from the fact that such seemingly double use of the data is subject to methodological criticism (e.g., Worrall (2010), but see Steele and Werndl (2013) for a nuanced response), model specifications and output thus rely heavily on the assumption that the—often scarce—data are somehow representative of the target system. In short, model uncertainty is a serious problem for computational modeling in archaeology.

In this paper, we look into the consequences of model uncertainty in this context. Specifically, we consider a case study of computational modeling for exploratory purposes, and we identify two distinct ways in which uncertainties are dealt with. We then generalize from the case study and provide a broader discussion on model evaluation and construction. In this abstract re-description of the case study we disentangle different notions of uncertainty that computational modelers grapple with, we indicate that robustness analysis is central to our dealings with uncertainty in exploratory use of computational models, and we sketch how such an analysis may lead to modeling improvements. We believe a thorough discussion of the case study will help the appreciation of the general discussion later on in the paper. But before we get to this, let us turn to a general outlook on computational modeling in archaeology and lay out the specifics of our case study.

3.2 Target Systems and Modeling Goals

Before we consider model uncertainty and the tools and methods that may control for it, we clarify some of the goals that model-building archaeologists have (see Kohler and van der Leeuw 2007). This will set the stage for a proper appreciation of the problems and sharpen the focus of our paper.

Developing a computational model of some system involves the definition of starting points, and making choices about which variables to include, which interactions between variables to define, and how to weigh parameters. The definition of starting points is directly connected to the purpose or goal of models: prediction, reconstruction, and exploration. A predictive model outputs a set of expectations that can be tested against the data that define the target system. A reconstructive model offers an abstract structure that resembles some target system on certain salient features. An explorative model, finally, occasions insight into the generative rules that underlie the structure or the workings of the target system, thereby establishing, or at least hypothesizing, certain properties of the system, for example, the band-widths of system variability.

Within this context, a modeling approach has to be chosen. In archaeology, quite a number of them have been practiced (see, e.g., van Leusen and Kamermans 2005) and can be captured under the labels “correlative,” “generalized behavior,” and “system-based.” Correlative approaches primarily investigate statistical relationships between variables (mostly aspects of landscape) and archaeological site occurrence and are frequently used in the context of cultural resource management (CRM). Approaches based on generalized behavior are built on “known” and

	Correlative	Generalized behavior	System-based
Prediction	●	○	
Reconstruction		●	
Exploration		○	●

Fig. 3.1 Generalized relationship between modeling approaches and purpose/goal. *Black dots* represent a certain prevalence of one approach over another

assumed aspects of what people have been doing within a particular socioeconomic setting. Both approaches are mostly using geographical information systems (GIS) as the modeling environment. System-based—including agent-based—approaches primarily focus on the emergence of patterns (of variability) through mathematically defined processes, and rules of interaction (e.g., between individuals or groups/populations) in a more abstract modeling environment. These broad “families” of approaches are not exclusive to one particular purpose of modeling, although some prevalence may be noted (Fig. 3.1).

The choice for a particular approach has implications for the selection of model variables and definition of relationships between them. At this point, prior knowledge and conditional factors are fed into the model system, hence introducing constraints (boundary settings) and varying sources and degrees of uncertainty. Clearly, this will be of influence on model returns, but what this actually means for the performance (or quality) of the model is not easy to establish (see, e.g., Kamermans et al. 2009). As yet, the assessment of model uncertainty by means of sensitivity analysis has received limited attention, despite the availability of approaches (Bayesian Theory and Dempster-Shafer Theory) that explicitly incorporate uncertainty as a modeling factor (but see Finke et al. 2008; van Leusen et al. 2009). Moreover, different modeling goals will require different approaches to the uncertainties in the models: a data-oriented statistical technique that assists reliable prediction might lead to models that do badly on the count of reconstruction.

Out of the three modeling goals sketched above, the present paper is focused on the goal of exploration, and for which the role of statistical analysis is not very prominent. And this entails a particular take on the uncertainties at issue. Instead of looking at ways to remedy the uncertainty in the models—weighing parameters and adapting them on the basis of data—our goal is to gauge and control for the uncertainty in the starting points and modeling choices, in the hope that we can use the models for exploration, despite the uncertainties. In particular, we focus on exploration aimed at clarifying the model content and generating hypotheses. As will be seen, this has consequences for the kind of analysis of uncertainty that is appropriate.

3.3 Starting Points and Modeling Choices: A Case Study

We will approach the question of model uncertainty on the basis of an explorative model of postglacial hunter-gatherer landscape use (Peeters 2007). The study area—the Flevoland Polders in the Netherlands—is characterized by a low accessibility and visibility of archaeological phenomena, as these are generally buried under several meters of sediment, and mostly consist of scatters of flint artifacts, and less frequently fragments of (charred) bone and pottery. Excavations do, however, demonstrate that the average preservation of remains is good, thus making their scientific value high. Despite these insights, the study area is basically a black box where it comes to an understanding of postglacial hunter-gatherer behavioral variability.

As indicated, we will consider the use of computational models for the development of hypotheses and more generally for the gradual buildup of a coherent qualitative picture of the target system at hand. The models in Peeters' study were designed to explore the potential of the study area for landscape use by hunter-gatherers after the last glacial, using an approach based on generalized (not agent-based) behavior. Central to the modeling approach was the idea that the area had undergone far reaching environmental changes (in terms of composition and geography) due to structural sea-level rise between 7000 and 4000 BP, and that these changes affected the possibilities of landscape use in a qualitative (what?) and quantitative sense (to what extent?). Hence, the models had to integrate environmental and behavioral parameters, which in combination resulted in a GIS-based assignment of values to individual grid cells.

3.3.1 *Environmental Parameters*

Despite the availability of a large body of geological (bore-hole) data, it was decided not to reconstruct landscape change from these data—for example, because of the lack of chronological control, and issues of spatially variable histories of sedimentation and erosion, as well as sample density—but instead to develop a computer model of landscape change to ascertain a consistent environmental framework. For this purpose, environmental variables had to be selected and parameters set (summarized in Table 3.1), in order to build such a framework. These were fed into an iterative set of “if-then” rules to create a time-series (one century interval) of landscape maps with a spatial resolution of 500×500 m (Fig. 3.2).

3.3.2 *Behavioral Parameters*

Central to the approach is that “cost/benefit” rules—[Characteristic] of/by/to [Constraint] is/are [Qualification] for [Goal]—lead to a “perceived” value of landscape units for any sort of “behavior” (or activity if one likes). In this way,

Table 3.1 Environmental variables, parameters, and parameter settings of the landscape model

Environmental variable	Parameter	Setting
Elevation of the early Holocene surface	Top of Pleistocene surface	Geostatistically interpolated (block kriging; 500×500 m grid-cell setting)
(Ground)water level	(a) Sea-level rise	(a) Reconstructed sea-level at time t_x in 100-year intervals between 7000 and 4000 BP
	(b) Capillary groundwater rise	(b) Fixed % groundwater table increase (difference between water level at time t_x and grid-cell elevation) with a maximum of 1 m
Vegetation	Dominant vegetation zone	Six vegetation zones relative to (ground)water table
		One open water zone
		Boolean categories
Altering surface elevation	<i>Sedimentation</i>	
	(a) Increase of surface elevation through peat accumulation	(a) Fixed % of the difference between surface elevation and (ground)water level at time t_x in wetland vegetation zones
	(b) Increase of surface elevation through clay accumulation	(b) Fixed % of the difference between surface elevation and water level at time t_x in zones where water levels are above surface elevation
	<i>Erosion</i>	
	(c) Decrease of surface elevation during two regionally documented events	(c) Fixed % in zones where water levels are above surface elevation

environmental information was connected to “behavior” through a set of “if-then” decision statements. Based on archaeological data from the study area and generalized knowledge about hunter-gatherer behavior, a range of “behaviors” was defined in connection to the virtual environments. Although the focus was primarily on some aspects of food resource acquisition and dwelling (examples summarized in Table 3.2), any sort of “behavior”—including ritual—could potentially be defined and fed into a similar framework, provided a connection can be made to the dimensions of our virtual world. Similarly, any defined “behavior” can be as simple or complex as one would like it to be. Drawing from the examples provided in Table 3.2, such cost/benefit rules can for instance be formulated as:

[high densities] of [large mammals] are [beneficial] for [hunting]
[proximity] of [open water] is [beneficial] for [traveling]
[presence] of [dense woodland] is [problematic] for [traveling]
[absence] of [open water within 500 m] is [unfavorable] for [dwelling]

Perception values for each grid cell in the spatial model were calculated on the basis of “perception weights” assigned to the parameters included.¹ As such, the maps

¹Weight values range from 0 (bad/low density/costly) to 1 (good/high density/beneficial). For more details on procedures, see Peeters (2007) (available from <http://dare.uva.nl/document/42380>).

Fig. 3.2 Time-series of paleogeographical reconstructions based on the Flevoland environmental model with clay accumulation set at 20 % (see Table 3.1). Grid cells measure 500×500 m (from Peeters 2005, Fig. 4, p. 156)

Table 3.2 Example “behaviors” and their model parameters

Behavior/activity	Parameter
Large mammal hunting	Large mammal encounter probability
	Overland traveling possibility
	Overwater traveling possibility
	(Overland and overwater traveling determine “traveling weight”)
Dwelling	Overland traveling possibility
	Overwater traveling possibility
	Soil moisture
	Presence of water within 500 m

produced for each time slice represent “perception surfaces” (cf. Whitley 2000) instead of predictions of site location (Fig. 3.3). Where and when certain types of behavior actually occurred cannot be predicted, as this depends on decisions that were made on the basis of many factors at the “ethnographic scale” (e.g., occurrence of game, hunter’s experience, perceived gain and needs). These factors were in constant fluctuation at a temporal and spatial scale that is unattainable in the specific model environment outlined above (already our 100-year time slices easily include three generations of hunter-gatherers). Indeed, the problem of temporal resolution provides a major factor of uncertainty with regard to the understanding of the archaeological record (Bailey 2007; Holdaway and Wandschneider 2008), and one that can only be approached through computational modeling as a means to build a link between “ethnographic” and “archaeological” time.

3.4 What to do with Uncertainty?

The above broad outline of starting points and modeling choices makes clear that the modeling work in this case study involves a myriad of sources of uncertainty, each of which having their own problems. The environmental part introduces simplifications of landscape dynamics, whereas the behavioral part brings in biased assumptions about how prehistoric hunter-gatherers may have “perceived” some possibilities of landscape use. On the other hand we have to bear in mind the goal of the modeling work at hand. In this case, the modeling primarily serves an explorative purpose: a heuristic device that helps to ask questions and interrogate the archaeological record, and in addition—through inclusion of an erosion map—helps to sort out at a regional scale which areas are likely to bear the best preserved archaeological resources. To some extent this is a predictive use of the model in question, but the most prominent use of the models remains that they invite hypotheses about the nature of sites potentially present in some geographical space rather than directly supporting, for example, CRM decision making.

Fig. 3.3 Example of the large mammal hunting model for the 6500 BP time slice (from Peeters 2005, Fig. 8, p. 160). The environmental model for this time slice is shown in Fig. 3.2 (top left). *Top left*: cost/benefit surface for overwater traveling (dark = good; light = bad); *top right*: cost/benefit surface for overland traveling (dark = good; light = bad); *bottom left*: large mammal encounter probability (dark = high; light = low); and *bottom right*: perception surface for large mammal hunting (dark = good; light = bad)

In view of the many difficulties to evaluate or “ground-truth” the outcome of these models—for example, due to the very problems of detecting and assessing sites in the study area altogether—it seems to us that posterior analysis of model uncertainty may add little. The data are so sparse that they will hardly help to reduce overall uncertainty. Of course specific parts of the model may be improved

by relying on data. The landscape modeling part, for instance, could be evaluated against bore-hole data, as the model was not based on reconstructions from these very data. The “virtual stratigraphy” of grid cells emerging from the modeled landscape dynamics appeared to fit the actual sequences recorded in bore-hole columns rather well, which provides at least some confidence that there exists concordance between the computational and the empirical model. In fact, this aspect lends itself for statistical calibration, an approach common to modeling in the geosciences (Brouwer Burg et al., [Chap. 1](#)). The behavioral part of the model is another matter of course, and one that does not lend itself very easily for calibration on the empirical facts. Across the board, we encounter many uncertainties that do not seem amenable to a standard statistical treatment.

This, then, brings us to the issues that are central here. How problematic are the uncertainties when it comes to the exploratory use of models? Can we control for uncertainties at the front end of model building, in order to safeguard the quality of the model output? And if so, what would be the best way to do this? Specifically, we are asking how the uncertainties in the computational models just described impact on the role of the models as catalysts of theorizing and hypothesis formation. The idea is that potentially adverse effects of the uncertainties can be controlled for by employing specific statistical tools: we can perform a sensitivity analysis, otherwise known as a robustness analysis, of the models and hence determine the reliability of the conclusions we draw from them. To flesh this out, we will first return to the case study and see some exploratory use of the computational model at work. In the light of this we identify two kinds of uncertainties, one within statistical models and one about such models. After this we will discuss the ways in which uncertainty may be controlled for in the case at hand, and which may be employed more broadly in computational archaeology. In the final part of the paper, we will then suggest how these techniques, once suitably developed, lead to better models in computational archaeology.

3.5 Back to the Model

In our case study, the model of hunter-gatherer landscape use is further explored in terms of interactions between environmental and behavioral parameters. More specifically, we ask how sensitive the model outcomes are to changes in environment–behavior interactions. In order to hold grip on the effects of parameter settings and model outcome, we will look at one single factor: the increase of surface elevation through clay accumulation.

In this model clay accumulation only affects landscape zones in which the surface is below groundwater level (reed-sedge, reed-rush, and open water). In the initial model, clay accumulation was set at 20 % relative to water depth. When varying the accumulation rate between 0 and 100 % in steps of 10 % ([Fig. 3.4](#)), it can be noted that no clear changes occur between 0 and 30 %. However, from 40 % onwards, major fluctuations in the relative importance of open water and the reed-rush zone

Fig. 3.4 Relative importance (percentage on the *vertical axis*) of vegetation zones for different rates of clay accumulation in 100-year time lags (age BP on the *horizontal axis*). The *lower right* graph plots all values obtained for those vegetation zones which are affected by clay accumulation, with a clay accumulation setting between 0 and 100 % in 10 % lags

occur. Open water and reed-rush are in competition, and eventually reed-rush becomes dominant over open water. Between 80 and 90 % of clay accumulation, reed-rush and reed-sedge get in competition, while the relative importance of open water is further reduced.

In view of the deterministic definition of the landscape zones in Boolean categories, these trends are to be expected: with the decrease of water depth due to increasing clay accumulation, vegetation types that favor more shallow water will gain importance. Nonetheless, there are some surprising patterns to note. The differences in the relative importance of open water and reed-rush between 30 and 40 % clay accumulation become quite substantial but were expected to be gradual. However, when increasing clay accumulation with only 1 %, it appears that a rather abrupt “turn-over” occurs from 37 to 38 %, initiating increasing fluctuations in the course of time (Fig. 3.5). A somewhat comparable, yet less marked shift occurs from 84/85 to 86 %, when the relative importance of reed-rush starts to fluctuate strongly compared to the rather stable situation at lower clay accumulation rates (Fig. 3.5). So apparently, even a simple deterministic linear model like this can produce unforeseen outcomes, suggesting some degree of nonlinear behavior.

The question here is: how do such changes to the relative importance of inundated landscape zones affect modeled human behaviors that are somehow connected to these zones. To explore this further, we will look at the “large mammal hunting” behavior summarized in Table 3.2. The calculation of perception values for large mammal hunting involved the following transformations:

$$\begin{aligned} TW_c &= {}^{avg}LT_{\sum c1...c9} + {}^{avg}WT_{\sum c1...c9} \\ HP_c &= TW_c * LM_c \end{aligned} \quad (3.1)$$

where TW_c is the traveling possibility weight of the target grid cell; ${}^{avg}LT_{\sum c1...c9}$ the average overland traveling possibility weight of the target grid cell, including neighboring grid cells; ${}^{avg}WT_{\sum c1...c9}$ the average overwater traveling possibility weight of the target grid cell, including neighboring grid cells; HP_c the hunting perception weight of the target grid cell; and LM_c the animal encounter possibility of the target grid cell.

The only parameter influenced by the effects of clay accumulation on the landscape is “overwater traveling possibility,” set (on a scale from 0 to 1) at 0.25 for the reed-sedge zone, 0.5 for the reed-rush zone, and 1 for open water (non-inundated zones clearly have 0 overwater traveling possibility). With the decrease of open water in the advantage of reed-rush and eventually reed-sedge, overall overwater traveling possibilities are reduced. However, as overwater (WT) and overland traveling (LT) are complementary in the calculation of traveling weights in the original model, the overall effects are buffered by TW_c and LM_c , none of which are affected by clay accumulation. Hence, HP_c is indifferent in connection to varying clay accumulation settings. This, then, leads us to a theoretical reconsideration of the front-end modeling choices.

Fig. 3.5 Relative importance (percentage on the vertical axis) of vegetation zones that are affected by clay accumulation in 100-year time lags (age BP on the horizontal axis). The graphs show the abrupt transition from a relatively stable situation up to 37 % of clay accumulation, towards an unstable situation between 38 and 50 % of clay accumulation. This unstable phase is followed by a stable phase between 60 and 84/85 % of clay accumulation, after which new fluctuation sets in

Fig. 3.5 (continued)

We have seen that the simple, deterministic landscape model composed of Boolean categories is sensitive to—occasionally minor—changes in an environmental parameter. In this example, clay accumulation influences the relative importance of open water, reed-rush vegetation, and reed-sedge vegetation. The modeled hunting perception that involves the relation between traveling possibilities and encounter possibility of large mammals is, however, not influenced by these landscape

changes. In other words, the modeled behavior is insensitive to changes in the environmental parameter “clay accumulation.” One cause of this indifferent model outcome may lay in the strict Boolean definition of vegetation categories (Table 3.1) relative to (ground) water level, which creates hard divisions in weight values of behaviors connected to these zones. Using fuzzy categories with overlapping boundaries will certainly provide better proxies of “natural” variability (cf. Arnot and Fisher 2007). Another cause may lie in the applied buffering distance of neighboring grid cells to calculate TW_c . This distance was set at one grid cell, which implies that TW_c only changes if the average possibility values of WT and LT for the nine grid cells taken into account had been affected.

Here we get to the key issue for the computational model: the influence of an environment changing at the local level (i.e., the target and neighboring grid cells) on the potential behavior within the target grid cell is only felt if changes occur within the neighboring grid cells. Now, the modeled behavior in our example of large mammal hunting is particularly bound to non-inundated vegetation zones, which are normally bordered by a reed-sedge/rush zone as soon as (ground) water levels reach the land surface. The replacement of open water by reed-rush in the landscape model is in fact occurring at some distance (more than one grid cell) from non-inundated land. This implies that the average possibility values of WT are not affected in the context of the large mammal hunting model. Consequently, WT will only have a noticeable effect on HP_c , if TW_c is considered over larger distances (more than one grid cell).

If the calculation of HP_c indeed should include grid-cell values over larger distances, we also have to conclude that a “remote evaluation” or grid-based evaluation of target grid cells is maybe not such a good way to proceed. Our large mammal hunters were not flying over the landscape, nor were they parachuted onto a grid cell to evaluate its hunting potential: they moved through the landscape. And when traveling overwater, they had to find their way through reed vegetation, which varied in spatial extent and density. This is probably what will have influenced decisions of where to go, to find routes, and see where good possibilities—not necessarily the best or most optimal—for a successful hunt would be. In other words, our hunters, and their targets, are (decision-making) agents.

This negative conclusion might lead us to believe that our case study misses the mark, but we don’t think so. The exercise demonstrates strengths (the landscape model) and weaknesses (parameter connections) in the computational model, and brings certain sensitivities of parameter settings, or lack thereof, into focus. We believe that our illustration of exploratory model use is in fact rather informative on how model uncertainties are often dealt with, and on what problems we run into when doing so. We show that exploring model outcomes under variation of front-end modeling choices—sensitivity analysis—offers a means to reconsider the theoretical basis that, in this case, defines environment–behavior interactions. Notably, the usual format of a sensitivity analysis is that variations over model input lead to particular, often qualitative patterns in the outcome variables, so that these patterns can be concluded from the model despite the uncertainties over input. Our case study is unusual in that, for us, the lack of response in the outcome variables invites a theoretical

advance; we will return to this below. At this point, we simply note that this is a theoretical advance and hence a fruitful exploratory use of the model nonetheless.

Returning to the above case, we have to conclude that HP_c in this particular model setup is insensitive to WT. The inclusion of WT is not “wrong” per se, but in this case it is ineffective in the posteriorly defined model settings. At the same time we feel that a leap forward can be made when a spatial modeling environment, like the one presented here, is combined with an agent-based approach. In doing so, the abstract “environment” in which digital agents usually operate in the context of ABM approaches (but see Danielisová et al. 2015; Janssen and Hill 2014) is replaced by a modeled environment that can, however, be validated (even statistically) on the basis of empirical data. In this way, the modeling environment and approach permit to build exploratory frameworks that help to increase our understanding of the archaeological record in dealing with issues of time perspectivism.

3.6 Statistical Uncertainty

With our case study firmly in place, we now turn to a more systematic discussion of the kinds of uncertainties involved, and the methods used to control for them. We discuss statistical uncertainty in this section and model uncertainty in the next one, focusing mostly on sensitivity or robustness analysis in both. We end by briefly considering, in the penultimate section, if uncertainties can be resolved by invoking theoretical criteria for models. It will be seen that the two sections following this one are more important for the paper as a whole but we nevertheless believe that some insight into ordinary statistical uncertainty is needed first: we hope that it reminds the reader of standard dealings with uncertainty and it introduces the idea of robustness analysis in a, more or less, familiar context.

As we said, the dominant response to uncertainties in scientific modeling is to deploy statistics. Statistical techniques can be used whenever the hypotheses that we entertain express expectations about empirical facts that are cast in terms of probabilities rather than certain facts. Rather than predicting the presence of a settlement with certainty, a hypothesis might for instance determine that a settlement here is more probable than there or more specifically that the chances of finding a settlement are only 20 %, and so on. Standard statistical methods allow us to confront such hypotheses with data, for example, records of excavations that have or have not laid bare settlements, and subsequently lead to a choice among the available hypotheses.

We can illuminate an important distinction within statistics by focusing on the kind of choice the statistical tools allow us to make. Classical statistical methods offer categorical choices among available hypotheses. We test a hypothesis, possibly against an alternative, and then decide to go along with it or not. Or else we estimate a parameter and so choose the hypothesis with the parameter value that makes the data come out maximally probable. By contrast, Bayesian statistics outputs a probability distribution over the hypotheses under consideration, as an

expression of our opinion about the hypotheses. The choice among hypotheses is not categorical but by degree, for example, we might end up assigning a probability of 90 % to the hypothesis that the chance of finding a settlement in a particular area is 20 %. Bayesian statistics thus regulates how data impact on our opinions over hypotheses.

In order for Bayesian statistics to output such verdicts concerning hypotheses, a modeler needs to input a so-called prior probability at the outset: a probability distribution over the hypothesis that expresses her initial opinions about them. This aspect of Bayesian statistics is often lamented, because it introduces a subjective starting point into the statistical analysis. It is not always clear what can motivate the probability over hypotheses. However, the input component might also be considered an entry point for opinions that researchers already have about the subject matter, for example, insights based on experience with the subject matter that may be brought to bear on the model. In evaluating archaeological hypotheses, the elicitation of expert judgments can be a welcome addition to the empirical data, which is often scarce and contested. Moreover the formal treatment of opinions over hypotheses, as offered by Bayesian statistics, may help to streamline the theoretical debate and evaluate arguments in it. In other words, the Bayesian methods may have an edge over classical ones.

We can easily make this concrete by reference to the case study. In the landscape model, an important role was played by the parameter that accounts for clay accumulation. Of course the accumulation rate will vary over place and time, but the modeler will typically have an idea of what range of values will be appropriate, and might even be able to provide a probability distribution over rates that expresses her expectations. A statistical analysis of the landscape model using bore-hole data may benefit greatly from the prior opinions of an expert, and the Bayesian framework offers ways to accommodate these in the analysis.

Recall that this paper focuses on the exploratory use of models, for which the role of statistical analyses is not very prominent. However, it turns out that there are other uses of probability assignments over model parameters, uses that do not involve data but that fit very well with the goal of model exploration. Returning to the case study, we saw that the clay accumulation rate influenced what the model predicts about the landscape. In certain regions of the parameter the dependence of the vegetation on accumulation rate is critical: a small rate change might result in substantial changes in vegetation. In other regions, the vegetation patterns remained more or less stable. Assuming that this is not an artifact of the model's use of binary variables, the uncertain expert opinion may thereby play a crucial role. If, according to the expert, the range of probable values includes such a critical region, then the model predictions vary wildly. But this is not the case if the expert excludes such critical regions from an established range of values. The probability assignment over parameter values, delivered by the expert, thus determines whether or not the model predicts unstable vegetation patterns.

The suggestion here is that the use of probability assignments over model parameters, as an expression of expert opinion, may provide robust qualitative conclusions, and thereby support the exploratory use of models. By dealing with the uncertainty

concerning the model parameters in a particular way, we manage to bring out what the model tells us about the target system, despite the statistical uncertainties that surround the model. To draw robust conclusions of that sort, we rely on the model as an adequate representation of the target system, and we assume that the uncertainty about model parameters is adequately captured by the expert opinions, that is, by the probability assignments. Accordingly, a model may be improved by reformulating it in such a way that experts have a better grip on the uncertainties.

To illustrate, we return to the case study. The Flevoland model introduced earlier seems to be rather unfit for a direct evaluation of the hypotheses in the model, by computing their fit with the data or the posterior probabilities. Although a statistical approach may suit the environmental dimension of the framework in view of the vast body of data available, the behavioral dimension causes problems, as the archaeological data interpretable to a specific level of behavior are particularly few. Of course, if one would accept a strong environmental dependency of hunter-gatherer behavior, one could argue that a reduction of uncertainty in the environmental dimension implies the same in the behavioral dimension. However, despite the subsistence-focused examples provided, we do not support such purely deterministic lines of reasoning. Not only do hunter-gatherers also take decisions on the basis of cosmologically inscribed factors (Descola 1999; Lavrillier 2011; Nadasdy 2007), there is furthermore a “sociohistorical” aspect to the use of landscapes based on, for instance, acquired information through sharing of knowledge and movement along paths (Aporta 2009; Lovis and Donahue 2011; Mlekuž 2014).

The aspects of behavioral complexity may, however, offer an opening to employ robustness analysis. Candidate models of various context of behavior, that include diverse ranges of parameters—bear in mind that theoretically any factor one would like can be included—can be analyzed to identify invariances. As invariant features of modeled behavior are less sensitive to variations in the starting points, such features can be expected to leave an archaeological echo, in contrast to features that return high degrees of variance. In this way it may become possible to identify patterns in model outputs that can be “tested” against archaeological data that are more of a qualitative than a quantitative nature, as in the case of our example area. Robustness analysis, then, may provide insight into the sensitivity of model parameters to differences in starting conditions, and—without the neglect of uncertainty—give way to the definition of models that return “perception surfaces” that, in a way, get closer to the “active landscape” as it was used by hunter-gatherers than the “neutral landscapes” in Peeters (2007) exploratory playground.

3.7 Model Uncertainty and Exploratory Modeling

One presupposition is central to both classical and Bayesian statistics and is highly relevant to our present concerns: all statistical approaches to uncertainty require that we choose a set of hypotheses, or theoretical possibilities, over which the experts, and perhaps the data, then produce a verdict. In statistics, this range of hypotheses is often called a model. The basis for a model is typically a set of causal relations,

perhaps a mechanism, or some other structure, in which salient quantities and their qualitative relations are determined, without fully specifying the parameter values that are associated with them. The foregoing exposition on landscape use offers a good example. For our purpose, it is important to notice that the model serves as a presupposition for the statistical treatment of uncertainties: we cannot express any statistical uncertainty if we do not, at the outset, come up with a set of hypotheses or parameter values.

What to do when we are uncertain about the quantities and relations that need to be included in the model? The usual application of statistics is not appropriate when the uncertainty pertains to the very conceptual structure that is used to control for uncertainties, so the uncertainty that pertains to the statistical model itself. Looking at the case study, we see that exploration of the model led to a criticism of the way in which environmental and behavioral models were connected, or more precisely, the way in which behavioral responses to the environment were conceptualized. The uncertainty here is fundamental. It does not concern the value of some parameter or other, it rather concerns the way in which agents and their relations to the environment are conceived within the model. In other words, it concerns the entire model setup. And it is to this kind of uncertainty that we now turn. Although there are at least as many approaches to the issue of model uncertainty as there are uses of models, our present goal is quite specific: we want to control for the detrimental impact that model uncertainty has on the use of models for exploratory purposes. We consider three approaches: statistical model selection, robustness analysis, and in the next section the use of measures of informativeness and surprise.

A first and rather natural response is to partially remove the uncertainty by fitting the models to data. In other words, we convert model uncertainty into statistical uncertainty and repeat the statistical procedures on the level of models. We are then in the business of statistical model selection: the models are taken as hypotheses, and evaluated according to their respective fit with the data, or according to other data-related quality criteria. Several classical approaches to adjudicating between models are on offer, all based on their own formalization of model quality. And there are also model selection tools along Bayesian lines. For one, we may express our opinions concerning the candidate models by assigning probabilities to them, and then compute the so-called posterior model probabilities, on the basis of the data and the prior probabilities. The development of such model selection tools for the archaeological context provides an interesting avenue to explore.

It seems clear, however, that statistical model selection cannot be the only answer to the issue of model uncertainty, certainly not when the use of models for exploratory purposes is at stake. Dealing with model uncertainty in the afore-mentioned manner will not align with the exploratory purposes of modeling for at least three reasons. First, the application of such selection tools requires the availability of ample data, whereas exploratory modeling often happens when little data are available. Moreover, off-the-shelf model selection tools are based on idealizations that are typically not met in the archaeological context, while advanced tools that may be applicable in this context are less well established and hence prone to technical and interpretive problems. Finally, and most importantly, if we select a single model

in response to model uncertainty, or average over a number of models, we seem to cover up something that may in fact be highly informative, namely that the models have certain qualitative features in common. Some features of the models might be constant, despite the uncertainty over them.

We submit that, in the case at hand, we can better deal with model uncertainty by focusing on these commonalities among the models, rather than to opt for one of them to the exclusion of others, or to average over the models. The idea is, in other words, that we deal with model uncertainty by selecting those results and insights that are invariant, or robust, under a wide range of models. This is a reiteration of the robustness analysis explained in the preceding section. Statistical and computational methods can be employed to identify such invariances, namely by supporting a systematic search of the parameter spaces of several models and charting the range of predictions that the models then generate. Such an exercise presents a view of the theoretical possibilities that the models under consideration offer and connects these to the potentially observable consequences of the different modeling assumptions. In short, an exercise like that gives the modeler a feel for the system she is modeling.

It may be viewed with some suspicion that the modeler does not have full command over the model she has built and so needs simulations and statistics to trace the role of the assumptions that have gone into it. Surely, it may be thought, the role of those assumptions should be in plain sight! But the practice of computational modeling is just not like that. Often models are highly complex, involving a multitude of parameters that are densely interlinked. It is generally not visible how such models relate to the empirical facts that they are supposed to account for. In fact modelers may well be surprised by the stringency or flexibility that a model offers in this respect. Investigating the spread of predictions relative to variations in the starting points, as done in robustness analysis, is a natural way of exploring the nature and contents of the models under scrutiny.

Looking again at our case study, we can quickly see that the analysis does not fit the mold of a robustness analysis. So what approach to model uncertainty is taken in the case study then? Recall that the uncertainty at stake is one about the entire model setup, and in particular about the connections between the environmental and the behavioral model. How should agents be conceptualized in the first place, and how do we relate them to their environment? The insight obtained from the exploratory use of the model was that agents and environment are independent where they should not be; the appropriate coupling of the environmental and behavioral model is missing. The behavior of the modeled agents appeared to be robust under substantial environmental variation, flying in the face of widely shared ideas about the relations between agents and environment. In other words, in our case study a faulty robustness casts doubt on modeling assumptions and so invited substantial revisions of the model.

From a distance, the inferential pattern that emerges is similar to the robustness analysis sketched before. Researchers will normally harbor intuitions about relations that should manifest between the variables that characterize their system of interest, in this case: a causal relation between vegetation and hunting opportunities.

By simulating the impact of variations in one such variable on other variables, researchers can check if their model adequately captures their intuitions. So where robustness analysis helps to find salient invariances, the analyses here help to check systematic covariation.

3.8 Theoretical Criteria

This last point connects to a rather speculative aspect of the view we have developed so far. Researchers check models for their robust properties but they also check them against numerous items of background knowledge, often tacitly. Good computational modeling offers researchers a grip on these background checks and allows them to make the checks explicit. So when it comes to the quality of models for the purpose of exploratory use, researchers will consider highly theoretical aspects of the models. They might ask what model will generate interesting hypotheses, what model presents surprising theoretical possibilities, or what model will be most informative. In the case study at hand, the model was judged to be defective because it failed to show dependencies that are expected on the basis of background knowledge of the system under scrutiny, that is, theoretical background knowledge of hunting behavior.

Speaking more generally, in the face of uncertainty over models researchers might select their favorite model not on the basis of how well it fits with available data but rather on these highly theoretical aspects. A model that passes specific checks against background knowledge is preferable, even if there are no empirical data that can be used to make those checks empirical. Or, more theoretical still, a model might be preferred because of the checks that it occasions, the insights that it might deliver, because of surprising predictions it might offer, or because of the opportunity it gives to formulate testable hypotheses, all of this quite independently of the data. Such theoretical aspects of models are very hard to formalize and quantify. But we are optimistic that some progress can be made in this direction, primarily by adapting and refining extant model selection tools. The informativeness of a model is naturally related to the specificity of a model, and so is the surprisingness, although in a different manner. And in turn, model selection tools provide a handle on the specificity (cf. Romeijn et al. 2012). We believe that a conceptual clarification of these theoretical criteria for models will be beneficial to a wide range of sciences, including archaeology, and we think that a formalization and quantification will contribute to this clarification.

Leaving aside these speculations on formalizing the theoretical virtues of models, we hope that the foregoing has made clear that model uncertainty is a serious methodological concern, and one that cannot be tackled by standard statistical means. We have shown that statistical model selection will typically not provide resolution, whereas robustness analysis and model comparison on the basis of theoretical criteria may present fruitful, certainly in the context of the exploratory use of models.

3.9 Conclusion

We hope to have made clear that model uncertainty needs proper attention but also that it is a methodologically complex problem that cannot be dealt with in a standard and straightforward fashion. The case study presented here makes very clear how important it is to critically (re)consider the relationships to be built in models from a theoretical perspective. It occurs to us that, generally speaking, models of human behavior and that of hunter-gatherer behavior in particular, are difficult—if not impossible—to calibrate through validation on empirical facts. Although Bayesian approaches could be very useful to explicitly deal with model uncertainty through assignment of prior probabilities to parameter settings and candidate models, the frequent lack of unambiguous data obstructs computation of posterior model probabilities. However, in an exploratory context of modeling purposes, we think it seems better to deal with sources of uncertainty at the front end of model building, and apply techniques such as robustness analysis and work on the development of systematic selection tools that rely on theoretical criteria, as sketched above. Such tools cannot aim at the selection of “a best” model but will help to identify families of models which return invariant or—conversely—highly variant outcomes, hence providing a basis to choose among those models which seem to offer the best possibilities for hypothesis testing. And it is exactly this possibility that will help to improve the acceptance of computational modeling as a useful tool for archaeology, a position that is not generally shared among archaeologists. With a critical approach to model uncertainty and model selection from an epistemic perspective, we believe that the research program of computational modeling in archaeology is engaged in a process of continuous self-improvement. This self-correcting character is reason for optimism about the viability of models in archaeology, which is much in line with the more general views expounded in Henrickson and McKelvey (2002) with regard to agent-based modeling in the social sciences.

References

- Aporta, C. (2009). The trail as home: Inuit and their pan-Arctic network of routes. *Human Ecology*, 37, 131–146.
- Arnot, C., & Fisher, P. (2007). Mapping the ecotone with fuzzy sets. In A. Morris & S. Kokhan (Eds.), *Geographic uncertainty in environmental security* (pp. 19–32). Dordrecht, The Netherlands: Springer.
- Bailey, G. (2007). Time perspectives, palimpsests and the archaeology of time. *Journal of Anthropological Archaeology*, 26, 198–223.
- Danielisová, A., Olšehočová, K., Cimler, R., & Machálek, T. (2015). Understanding the Iron Age economy: Sustainability of agricultural practices under stable population growth. In G. Wurzer, K. Kowarik, & H. Reschreiter (Eds.), *Agent-based modeling and simulation in archaeology* (pp. 183–216). New York: Oxford University Press.
- Descola, P. (1999). Des proies bienveillantes. Le traitement du gibier dans la chasse amazonienne. In F. Héritier (Ed.), *De la violence* (Vol. II, pp. 19–44). Paris: Odile Jacob.
- Finke, P. A., Meylemans, E., & van de Wauw, J. (2008). Mapping the possible occurrence of archaeological sites by Bayesian inference. *Journal of Archaeological Science*, 35–10, 2786–2796.

- Henrickson, L., & McKelvey, B. (2002). Foundations of “new” social science: Institutional legitimacy from philosophy, complexity science, postmodernism, and agent-based modeling. *Proceedings of the National Academy of Sciences*, 99 (Suppl.3), 7288–7295.
- Holdaway, S., & Wandschneider, L. (2008). Time in archaeology: An introduction. In S. Holdaway & L. Wandschneider (Eds.), *Time in archaeology: Time perspectivism revisited* (pp. 1–12). Salt Lake City, UT: University of Utah Press.
- Janssen, M. A., & Hill, K. (2014). Benefits of grouping and cooperative hunting among Ache hunter-gatherers: Insights from an agent-based foraging model. *Human Ecology*, 42, 823–835.
- Kamermans, H., van Leusen, M., & Verhagen, P. (Eds.). (2009). *Archaeological prediction and risk management. Alternatives to current practice*. Leiden, The Netherlands: Leiden University Press.
- Kohler, T. A., & van der Leeuw, S. E. (2007). Introduction. Historical socio-natural systems and models. In T. A. Kohler & S. E. van der Leeuw (Eds.), *The model-based archaeology of socio-natural systems* (pp. 1–12). Santa Fe, New Mexico: School for Advanced Research Press.
- Lavriellier, A. (2011). The creation and persistence of cultural landscapes among the Siberian Evenkis: Two conceptions of ‘sacred’ space. In P. Jordan (Ed.), *Landscape and culture in northern Eurasia* (pp. 215–231). Walnut Creek, CA: Left Coast Press.
- Lovis, W. A., & Donahue, R. E. (2011). Space, information, and knowledge: Ethnoscatterography and North American boreal forest hunter-gatherers. In R. Whallon, W. A. Lovis, & R. K. Hitchcock (Eds.), *Information and its role in hunter-gatherer bands* (pp. 59–84). Los Angeles, CA: The Costen Institute of Archaeology Press of UCLA.
- Mlekuž, D. (2014). Everything flows: Computational approaches to fluid landscapes. In G. Earle, T. Sly, A. Chrysanthi, P. Murrieta-Flores, C. Papadopoulos, I. Romanowska, & D. Wheatley, D. (Eds.), *Archaeology in the digital era, volume II: e-Papers from the 40th annual conference of computer applications and quantitative methods in archaeology (CAA)*, Southampton, 26–29 March 2012 (pp. 839–845). Amsterdam, The Netherlands. Retrieved from <http://dare.uva.nl/aup/en/record/500958>.
- Nadasdy, P. (2007). The gift in the animal: The ontology of hunting and human–animal sociality. *American Ethnologist*, 34(1), 25–43.
- Peeters, H. (2005). The forager’s pendulum: Mesolithic–Neolithic landscape dynamics, land-use variability and the spatiotemporal resolution of predictive models in archaeological heritage management. In H. Kamermans & M. van Leusen (Eds.), *Predictive modeling for archaeological heritage management: A research agenda* (pp. 149–168). Amersfoort (Nederlandse Archeologische Rapporten 29).
- Peeters, J. H. M. (2007). *Hoge Vaart-A27 in context: Towards a model of Mesolithic-Neolithic land use dynamics as a framework for archaeological heritage management*. Amersfoort.
- Romeijn, J. W., van de Schoot, R., & Hoijsink, H. (2012). One size does not fit all: Derivation of a prior-adapted BIC. In D. Dieks, W. Gonzales, S. Hartmann, F. Stadler, T. Uebel, & M. Weber (Eds.), *Probabilities, laws, and structures* (pp. 1–28). Berlin, Germany: Springer.
- Steele, K., & Werndl, C. (2013). Climate models, calibration, and confirmation. *British Journal for the Philosophy of Science*, 64(3), 609–635.
- van Leusen, P. M. & Kamermans, H. (Eds.). (2005). *Predictive modeling for archaeological heritage management: A research agenda*. Amersfoort.
- van Leusen, P. M., Millard, A. R., & Ducke, B. (2009). Dealing with uncertainty in archaeological prediction. In H. Kamermans, P. M. van Leusen, & Ph. Verhagen (Eds.), *Archaeological prediction and risk management. Alternatives to current practice* (pp. 123–160). Leiden.
- Verhagen, P. (2007). *Case studies in archaeological predictive modeling* (p. 14). Leiden, The Netherlands: Archaeological Studies Leiden University.
- Whitley, T. G. (2000). *Dynamical systems modeling in archaeology: A GIS evaluation of site selection processes in the Greater Yellowstone region*. Unpublished Ph.D. Dissertation, Pittsburgh.
- Worrall, J. (2010). Error, tests, and theory confirmation. In D. G. Mayo & A. Spanos (Eds.), *Error and inference: Recent exchanges on experimental reasoning, reliability, and the objectivity and rationality of science* (pp. 125–154). Cambridge, England: Cambridge University Press.