Information Science in Archaeological Survey.

Voorrips, A.; Loving, S.H.; Kamermans, H.; Hackens T., Miller U.

Information Science in Archaeological Survey

INTRODUCTION

In this lecture I will discuss some aspects of applied information science in archaeology, using a case study: the Agro Pontino regional archaeological survey, which is carried out by the Albert Egges van Giffen Institute for Pre- and Protohistory of the University of Amsterdam. I will first present some background information on the Agro Pontino; secondly, I will discuss the requirements for data base management systems that can deal with the archaeological variability; thirdly, I will present the construction of a research design and of sampling design; fourthly I will present an example of the implementation of data collecting procedures and of a data base; and fifthly, I will discuss some available software and hardware that can be useful in archaeological fieldwork, based on the experiences in the Agro Pontino.

THE AGRO PONTINO - PRESENT SITUATION AND ARCHAEOLOGICAL POTENTIAL

The Agro Pontino is a low-lying plain, approximately 70 km south of Rome, along the Tyrrhenian coast, and extends inland 15 km to the Monti Lepini and Ausoni, which are part of the Apennines formed during the Mesozoic. The plain is oriented NW-SE, and its boundaries in those directions are the Fiume Astura (south of Nettuno) and the Monte Circeo, a distance of about 60 km. The Monte Circeo is an isolated part of the Apennines.

Running the length of the Agro Pontino between the Monti Lepini-Ausoni and the coastal formations is a graben, a tectonically subsiding area, that is approximately 7 km wide. The depth of subsidence below present day sea level is not known, but is below 500 meters, and the earliest deposits recovered so far are marine sediments dated to the Miocene. Most of the present day surface of the graben lies at approximately sea level and is filled with peaty sediments, the accumulation of which has more or less kept pace with the tectonic subsidence for the last 15,000 years. Graben sediments, however, also include colluvial and alluvial deposits transported from the mountain slopes and the Ameseno Valley, which extends inland, and marine lagoonal sediments that were deposited whenever the sea breached the coastal sand barrier along the Southeast coast, which extends from Monte Circeo and Terracina. At the northernmost end of the graben are travertine formations and the tuffs from various eruptions of the Colli Albani.

The Pontine Marshes proper, which have been periodically reclaimed for cultivation since pre-Roman times, lie in the graben. Reclamation has always involved regulation of the water flowing into the area, through the building of canals and dams, and sometimes channeling water out of the area, using canals, pumps, and locks.

The area running the length of the plain between the graben and the Tyrrhenian coast consists of littoral sediments which have been building seaward during the Quaternary and Holocene. There are four littoral formations that have been identified, which provide a kind of horizontal stratigraphy from the graben to the coast; these are termed, from oldest to youngest, the Latina level, the Minturno level, the Borgo Ermada level, and the Terracina level. Each level consists of beach ridges and associated lagoons behind the ridges. These levels have been identified on the basis of their soil development and current elevations (ca. 25 m, 16 m, 6 m, and 0 m asl, respectively). Estimated dates are 250,000 BP (Tyrrhenian I), 100,000 BP (Tyrrhenian II), 70,000 BP (Tyrrhenian III), and post-glacial. This is a very general picture, and it is quite clear from the soil studies and profiles visible in the field that the extant beach ridges have also been subjected to a number of erosional phases (because of sea transgressions). Furthermore, much of the southern part of the area is covered by aeolian deposits.

Along the Southeast coast of the area, between Monte Circeo and Terracina, only the two youngest littoral formations are evident, the Borgo Ermada level and the Terracina level. The changing relationships among this coastline, the subsidence of the graben, and the changing sea levels is not yet understood.

Along with the most recent reclamation of the area in the 1930's, the area between the graben and the coast, the littoral formations and aeolian deposits, was also brought under intensive cultivation through a sprinkler irrigation system. The salt content of the recent coastal lagoons was raised to help control the mosquito population since malarial infestation has been a problem in this area.

Pebble beds, having pebbles of a size sufficient for stone tool manufacture, are found only along the Southwest coast. Usually these beds occur in the Minturno level deposits, and to a lesser extent in the Borgo Ermada deposits. Virtually all of the flint artifacts encountered on the Agro Pontino and surrounding areas have been made from these beach pebbles. This type of deposit requires a high energy beach to form, and is not being formed along the Agro Pontino coasts today, but may be observed to the north in Anzio. It is not known how these fossil deposits were formed. There are two ideas. One is that they have been more or less deposited in place; that is, a previous calcaric formation has been completely dissolved, leaving only the silicious fill of its cavities (what we now see as the beach pebbles), which then washed upon the beach. The other is that they have been transported along the coastline and deposited in front of a major coastal barrier to coastal sea currents, the Monte Circeo.

The obsidian found in the area comes from the Palmarola Islands, 30 km off the coast.

There are several well-known Paleolithic and Mesolithic sites in the area. On the Monte Circeo are Grotta Guattari, Grotte Fosselone and Riparo Blanc. Grotta Guattari has four discernible Middle Paleolithic layers; because it was naturally sealed by a boulder the faunal remains there were very well preserved. The top layer (the only one that was exposed), a kind of «living floor», contained a human skull in the center of a circle of rocks and was thought to be evidence for ritual cannibalism (although this is now under dispute). Grotte Fosselone is a well-stratified deposit spanning the Middle and Upper Paleolithic. Riparo Blanc is mainly a Mesolithic site with a large mollusc deposit and tools for prying open mollusc shells. On the plain proper, a stratified deposit dating back to before 55,000 BP, appearing to lie slightly above the Borgo Ermada beach ridge, was exposed during the digging of the Canale Mussolini (now called the Canale Acque Alte). Some of the strata were peat deposits. Analyses of the faunal and botanical remains provided the first reconstruction of climatic succession of the area for much of the Würm.

There are no Neolithic sites excavated from sealed deposits. One Bronze Age deposit was found in the travertines near Cisterna, while there is some evidence for a Bronze Age cemetery in the region of Borgo Ermada. A number of Iron Age period sites have been located near the mouth of the Astura, and the Etruscan site of Satricum immediately north of the survey area contains some earlier deposits. Roman materials, in particular of the Republican period, abound, which is not surprising given that the Via Appia runs down the center of the plain.

Brief overview of the survey activities

The Agro Pontino survey began in 1979 after a report by soil survey crews from the Laboratory for Physical Geography and Soil Science of the University of Amsterdam who were working in the area that they had encountered a large number of archaeological finds while mapping the soils. Two persons (A. Voorrips and S.H. Loving) spent two weeks going to different sections of the area in an attempt to discern how variable the surface materials were. The primary conclusion was that the Agro Pontino offered an excellent opportunity for archaeological research. Therefore, in 1980, a crew of four persons spent three weeks designing and trying out a field survey strategy. In 1982, a crew of about 10 people collected enough material in four weeks to construct a probabilistic sampling design. Based on this design five transects were drawn (see below). The survey of these transects began in 1984 with a crew of 10 people working 4 weeks. In 1986 a crew of 18 people working five weeks completed the transect survey.

Imagine the situation after our first exploration in 1979. Here we had:

a. a reasonably well bounded region;
b. plenty of archaeological materials, from the Middle Palaeolithic up to Roman times;
c. detailed studies about the development and dates of the soils at the present surface;
d. the potential for reconstruction of the palaeo-environment and its changes through time, by means of palynological analyses of both long peat cores from the graben and of shorter ones from the coastal lagoons;
e. a good typological basis for the analysis of the lithic material, thanks to both older and more recent studies of materials found at and around Monte Circeo (e.g. Bietti 1969; Blanc and Segre 1953; Mussi and Zampetti 1978; Taschini 1972, 1979; Zei 1973).

Our first problem was: What kind of information can be pulled out of these (potentially) available data. Before going into this, it is now the time to discuss the relation between data and information and the structure of data base management systems which can cope with that relation.

DATA, INFORMATION, AND DATA BASE MANAGEMENT SYSTEMS

We need to make a differentiation between data and information. Data are the unordered set of observations, abstract notions, and implicit ideas about some aspect of the world-out-there. These data provide information when the observations are ordered, the abstract notions are operationalized, and the implicit ideas are made explicit, such that it becomes possible to build a satisfactory chain-of-reasoning that enables us first of all to construct regularities, that is, to «find» patterns, and, secondly, enables us to explain those patterns in terms of our own outlook on the world and our own scientific interests. In short: information equals processed data.

Information science thus is the science of organizing, presenting, and analyzing data in a manner that leads to an increase of information.

The design of a data base should reflect this attitude. A data base must be constructed such that the data stored within it can be extracted in different ways and combinations to «find» patterns, and to test these patterns for statistical significance. The data base construction has to take place within a framework that has potential relevance to a broad problem domain, a domain that has been defined by making explicit our implicit notions about the archaeological context we want to study. A data base, therefore, is not an uncritical amassing of data into a vaguely structured data bank, under the assumption that different researchers with different research goals will be able to sensibly use those data.

Besides the structured data base for a broad problem domain it is possible to construct specialized archaeological data bases as aids for specific research applications, for instance a file that contains descriptive data on Etruscan mirrors and that is used for classifiable purposes (Moscati 1984, 1986). Such analytical data bases are extremely useful for specialized research, but lack generality.

What are the requirements for a structured data base for a broad problem domain in archaeology? The main consideration is that the data base has to reflect, has to be concordant with, the kind of data that is common in archaeology. Some general characteristics of archaeological data structures are:

1. They are *large*, often involving many observations, each of which is described by a large number of variables.
2. They are *complex*, involving associations and covariations among variables and similarities among observations that may be multidimensional and overlapping. The relationships among observations may be hierarchical or define a plex or network structure (see below). Importantly, the structure of an archaeological data set may vary locally among different sets of observations and variables, rather than be globally uniform. This complexity, in part, relates to the diverse kinds of units that an archaeological data structure may encompass (e.g., geographic regions, sites, proveniences, artifacts) and the diverse scales along which their attributes are measured (nominal, ordinal, interval, or ratio). However, it also relates to the often fluid manner in which even a restricted set of archaeological phenomena may be organized (...).
3. They often involve a high percentage of empty or *zero* cells when fully enumerated as a matrix of observations against variables.
4. They are often used for *multiple purposes* to investigate diverse hypotheses.

5. They are often *open-ended*, i.e., they must be added to continually as work and analysis proceed. Updating a data structure may be two dimensional, involving addition of both new observations and new variables. (Parker *et al.* 1985).

RESEARCH DESIGN AND SAMPLING DESIGN FOR THE AGRO PONTINO SURVEY

Let us go back to the Agro Pontino survey. The area is well suited for a *regional approach*. Instead of trying to find individual sites for excavation, we look at the region as a *whole*, and try to describe and understand its development through time. This approach leads to the definition of a research design that must be able to cope with the variety and complexity of regionwide collected data, and that will have to serve multiple purposes.

1. Research design

In the Agro Pontino project the soil surveys are basic for our knowledge about geological developments, age of surfaces surveyed, and prehistoric environmental differentiation. The major soil survey of the area (Sevink *et al.* 1984) was completed as the archaeological survey was beginning.

A first problem is to assess the factors that influence visibility during the archaeological survey. The soils information is critical here; but also, information about recent soil transport, the conditions of the field, and other factors that might influence what the surveyor can perceive must be recorded. Using this information we can assess the quality of our data and be able to distinguish between variability caused by environmental or circumstantial factors and that caused by prehistoric behaviours. A statistical procedure for such analysis has been developed (Loving *et al.* 1985) using the data collected up to 1982, and has been improved upon using the 1984 transect data (Verhoeven 1985).

At present one of the research goals is to investigate *changing patterns of prehistoric landuse* in the Agro Pontino using land evaluation techniques in conjunction with archaeological, ethnographical, and historical data. Land evaluation is a technique developed by physical geographers and is widely used by the FAO (Food and Agricultural Organization of the United Nations) in third world countries (Brinkman and Smyth 1973; Beek 1978). H. Kamermans (Kamermans *et al.* 1985) has adapted the land evaluation approach for archaeological purposes. An archaeologist first collects the physical data by translating and combining palaeoecological information to yield a qualitative land classification for different time periods. Then, using ethnographic, archaeological and historical sources, he constructs a series of models of prehistoric socio-economic situations. Then a land suitability assessment is derived by combining the appropriate land classifications with plausible socio-economic models. Finally, the archaeological survey results are examined in order to evaluate the socio-economic models in terms of the land suitability assessment.

The geological and soils data that are needed to reconstruct the palaeoecological situations are available, but not all the palaeobotanical data necessary have been collected and analyzed. One pollen core from the graben, collected in 1981, has been analyzed, however, and some first attempts of palaeoenvironmental reconstructions have been based on this analysis (Eisner *et al.* 1986).

A second major research goal is the investigation of differences between the *mobility patterns* of Middle Palaeolithic and Upper Palaeolithic groups in the Agro Pontino, using the lithic materials found during the survey. The mobility patterns can be considered to reflect aspects of social organization and of subsistence specialization. This research is of interest for the ongoing debate about the differences (biological? cultural? both?) between *Homo sapiens neandertalensis* and *Homo sapiens sapiens*. This part of the research entails, among other things, developing a typology for the debitage and investigation of the pebble variability in the pebble beds along the South-west coast.

A third research goal stems from a general problem in the use of archaeological survey materials: they are frequently difficult to date. The use of «guide fossils» (*fossile directeur*), which tend to be rare in survey collections, is not an optimal approach. Therefore, a new approach is being developed. This approach combines, in a statistical manner, the certainty with which chronological assessments of individual items can be made in order to date the assemblages of which the items are a part. This approach will allow incorporation of less easily dateable materials.

2. *Sampling design*

In order to operationalize the research goals outlined in the previous section, a first step had to be the acquisition of a general overview of the dispersion of archaeological materials in the Agro Pontino. Where it is impossible to do a complete survey, a sampling design needed to be developed.

A main decision, based upon a theoretical and a practical reason, was to take the recent agricultural fields as our sampling units. The theoretical reason was that in order to be able to find (spatial) patterning in the archaeological surface record it was necessary to take a unit that is independent of the density of the archaeological materials. Therefore, restricting data collection to places where artifacts are found would be an inadequate approach, given the research goals. As for the practical reason, in most literature on archaeological sampling some kind of quadrat system is imposed on the landscape, and then some of the quadrats are used for data collection. This may be a good solution for areas without clear landmarks. The Agro Pontino, however, is divided into a multitude of, on average, rather small agricultural fields. It was deemed much easier to use these existing fields, which can be traced back on air photographs and topographical maps without many problems. Correction for size differences between fields can be done during the analysis and reporting phase of the research.

We can think of the sampling design for the Agro Pontino survey as a step-wise design, meaning that the results of one phase of the design are taken into consideration when making selections for the next phase. Presently there are three phases to the design:

1. A non-random sample of elements (the recent agricultural fields) that provided estimators of the variance used for determining the sample size required to make inferences about parameters within a given bound on the error of estimation. This sample consists of the information collected from fields in the Agro Pontino that were selected non-randomly for survey from 1979 to 1982.
2. Using these estimates, a systematic non-aligned transect sample was designed to select (a) a sufficient sample size for making probability statements about the statistical populations of archaeological materials in the Agro Pontino as a whole, (b) a sample that spatially «covers» the NE-SW length of the Agro Pontino plain and is thus theoretically capable of detecting NE-SW variability in the populations, and (c) a sufficient sample size from three environmental strata (the coastal formations, the aeolian area and the graben) to make probability statements about the archaeological record in relation to soil parent materials. This is explained in more detail in Appendix I.
3. Using the results of the transect survey and analyzing them in the context of land evaluation research — which is the basis for the investigation of man-land relationships through time in the area — a «purposive» survey will be made to fill in «gaps» in the data collected.

DATA COLLECTING AND DATA BASE

The research design and the sampling design discussed in the previous paragraphs show all the characteristics of an archaeological data structure as defined by Parker *et al.* (1985); *large; complex; on many occasions zero cells; to be used for multiple purposes; open-ended.* To handle such data, well-planned collection procedures and a well-planned data base structure are necessary.

Appendix II gives an impression of the kind of data that are collected in the Agro Pontino survey. The various record numbers refer to the *levels* in the data base that has been constructed for the survey data. This data base is located at the main frame computer of the University of Amsterdam and uses the *Scientific Information Retrieval* (SIR) data base management package. The data base has a mixed hierarchical - and network structure.

A hierarchical structure implies that only simple relationships between the various parts of the data can exist: groups of more specific data-items are each related to a single «parent»-item. For example, an item describing the properties of a single agricultural field can have one or more «children», each of which contains the data specific to different visits that were made to the field. The relationships between the different levels of records are defined by means of keys. The key for a field is its field number; the key for a visit to that field is the field number plus the visit number. There is only one way to reach a record at one of the lower levels; down from the top, via (great) great-grandparents, grandparents, and parents.

A hierarchical data base is *parsimonious*: data are stored only once, at the level where they belong. Access to the data is normally fast because of the restricted number of routes through the data base.

A hierarchical data base cannot, however, reflect the *complexity* of archaeological data where items often have multiple relationships to higher levels of observation. One way to cope with the complexity is to add a network or plex structure to the data base, by means of which multiple relationships can be handled. The Agro Pontino SIR data base is an example of such a structure.

Cases «own» records containing data for the case, and the *case number* and the *record keys* are what allow the user to locate the information desired. Since the observation unit of the survey is the agricultural field, the agricultural field is the logical choice for the definition of a case. But we also have an interpretive unit called a site that we might want to use as a case as well. Some of the information we collect is pertinent only to the field (such as field visibility conditions), whereas some other information is pertinent only to the site (such as the areal extent of Roman artifacts), but some is pertinent to both (such as the individual artifacts). We want to avoid entering the same information twice by being able to aggregate all fields interpreted to belong to a site and being able to aggregate all sites interpreted to belong to one or more fields. To make this possible:

a. two types of cases have been defined, a site case and a field case, each of which owns a separate set of records and has its own hierarchy;

b. «link» records are put in each hierarchy so that the user can collect information from the other hierarchy.

At present our site hierarchy is undeveloped because our analysis of the materials has not proceeded to the point where we really need much information about sites. A diagram of the existing structure gives an idea of how the data base is interlinked.

SITE CASE

Record 1 (site record)
keys: site case number

Record 2 (link record)
keys: site case number
field case number
visit number
find number

Records 16-23
(not yet defined)

FIELD CASE

Record 3 (field record)
keys: field case number

Record 4 (visit record)
keys: field case number
visit number

Record 5 (link record)
keys: field case n°.
site case n°

Record 6 (find record)
keys: field case number
visit number
find number

Record 7 (artifact rec)
keys: field case number
visit number
find number
artifact number

Records 8-15
Analysis records by
material type

Software and hardware in the field

Micro-computers become cheaper and more powerful every month. They now are useful tools for archaeologists, not only in the office but also in field situations. When data can be stored immediately after having been collected, using a data base system that can handle the complexities of archaeological data, all kinds of checks can be made; preliminary reports and catalogues can be produced, etc. The possibility of immediate checking for errors and consistency is extremely important: for every day a data-error stays unnoticed, the time to correct it seems to double.

In order to use computers for field work, two questions need to be answered:

a. What kind of computer to select?

b. What kind of data base package to use?

I will try to answer these two questions from my own experiences with micro-computers in various field situations.

The actual *brand* of computer is almost unimportant as long as it meets the following requirements:

1. it must run under a well-known and well-supported operating system, like MS-DOS or CP/M;
2. it must have a memory of at least 256 Kbytes, but preferably a bigger one;
3. it must have at least two 360 Kbytes floppy disk drives, but preferably a 10-20 Mbyte hard disk;
4. it has to be steady, reliable, and preferably portable, or at least luggable;
5. it must be a brand that can be repaired in the country or area where the field work is located.

The second question is more difficult to answer. There are many professional data base packages for modern micro-computers, but these have been written almost without exception with business applications in mind, which tends to make them less useful for archaeological purposes. A powerful package that can be used for the building of an appropriate archaeological data base structure is dBase III. This package, however, is less a data base management system than a programming language. It takes months to years to design and implement a data base for archaeological purposes that does more than a bare minimum of tasks.

The best solution I have found until now is a package named MINARK. MINARK is a data base management system, written by the Australian archaeologist and computer scientist Ian Johnson. It is presently available for micro-computers which run under PC-DOS, MS-DOS, CP/M80 or CP/M86. The difference with a package like dBASE III is that MINARK is not a *programming language* for the development of one's own applications, but a well-structured data base management system, already tailored for archaeological data and the archaeologist's needs. This means, among other things that as soon as a data base structure has been defined the user has at his or her disposal all the functions that serve to maintain a database, like data entry, data correction, browsing through the database, data retrieval, possibilities for data selection, standard output formats, and formats for graphical output.

It is possible to perform checks on the validity of the entered data by means of defining ranges in which the data values must be. Data records can be edited and retrieved in the order they have been entered, but also in the order of indexes that are defined by the user. Data entry is not only possible by means of the keyboard but also files that are made with for instance a word processor or contain data collected from peripherals like a digitizer can be directly read into the data base.

Output is provided in several standard formats. One can select:

1. only the record identifications, all variables, selected variables;
2. creation of data files for statistical packages like SPSS, SAS and SYSTAT;
3. a number of statistical functions, among others frequencies, percentages, totals, averages, standard deviations, minimum, maximum, and cross-tabulations;
4. the layout of reports and list can also be completely defined by the user.

The possible applications are manyfold. Because of its general setup, MINARK can handle databases of different types. For instance:

1. A museum catalogue of new acquisitions;
2. A database of all the sites, registered by an archaeological institution, or all the sites in a certain area;
3. A database of all the sites, excavated by a certain person or institution;
4. A database of all the excavation data from a certain site;
5. An index on the photographic files of a museum;
6. Field data for an anthropological/ethnographical study;
7. An annotated bibliography.

It is of course possible that there are even better systems for archaeological field (and office) data bases, but I have not found one yet.

H. Voorrips, S.H. Loving, H. Kamermans
Albert Egges van Giffen
Inst. v. Prae- en Protohist.
University of Amsterdam
Singel 453
NL-1012 WP AMSTERDAM, The Netherlands

References

Beek, K.J., 1978, *Land Evaluation for Agricultural Development*, Wageningen: (*ILRI Publication*, 23).

Bietti, A., 1969, *Due stazioni di superficie del Paleolitico Superiore nella Pianura Pontina*, in *Bulletino di Paletnologia Italiana*, 20 (78), p. 7-39.

Blanc, A.C. and Segre, A.G., 1953, *Excursion au Mont Circé*, in *IVe Congrès International de l'Association Internationale pour l'étude du Quaternaire*, Rome.

Brinkman, R. and Smyth A.J. (*eds.*), 1973, *Land Evaluation for Rural Purposes*. Wageningen, (*ILRI Publication*, 17).

EISNER, W., KAMERMANS, H., WYMSTRA, T.J., 1986, *The Agro Pontino survey: Results from a first pollen core*, in BIETTI-SESTIERI, A.M., (*ed.*), *Convegno di Prospettive Storico-Antropologiche in Archeologia Preistoria*. In press.

KAMERMANS, H., LOVING, S., VOORRIPS, A., 1985, *Changing patterns of prehistoric landuse in the Agro Pontino*, in C. MALONE and S. STODDART (*eds.*), *Papers in Italian Archaeology. Part I: The Human Landscape*. Oxford, (BAR *International Series*), p. 53-68.

LOVING, S., VOORRIPS, A., KAMERMANS, H., 1985, *Old finds in new fields: First results of the Agro Pontino survey*, in *Bulletino di Paletnologia Italiana*. In press.

MOSCATI, P., 1984, *Richerche matematico-statistiche sugli specchi etruschi*, in *Contributi del Centro Linceo Interdisciplinare di Scienze Matematiche e loro Applicazioni N. 66*. Roma: (*Accademia Nazionale dei Lincei*).

MOSCATI, P., 1986, *Analisi statistische multivariate sugli specchi etruschi*, in *Contributi del Centro Linceo Interdisciplinare di Scienze Matematiche e loro Applicazioni N. 74*. Roma: (*Accademia Nazionale dei Lincei*).

MUSSI, M., ZAMPETTI, D., 1978, *Siti preistorici di superficie nel territorio di S. Felice Circeo (Prov. di Latina)*, in *Quaternaria*, 20, p. 41-48.

PARKER, S., LIMP, W.F., FARLEY, J.A., 1985, *Archaeological data base management as conceptual process*, in C. CARR (*ed.*), *For Concordance in Archaeological Analysis. Part III: Data Organization and Data Base Management*. Kansas City: (*Westport Publishers Inc.*), p. 91-113.

SEVINK, J., REMMELZWAAL, A., SPAARGAREN, O.C., 1984, *The Soils of Southern Lazio and Adjacent Campania*. Amsterdam: (*Laboratorium voor fysisch-geografisch en Bodemkundig, Publicatie*, 138).

TASCHINI, M., 1972, *Sur le Paléolithique de la plaine Pontine (Latium)*, in *Quaternaria*, 16, p. 203-223.

TASCHINI, M., 1979, *L'industrie lithique de Grotta Guattari au Mont Circé (Latium): Definition culturelle, typologique et chronologique Pontinien*, in *Quaternaria*, 21, p. 179-247.

VERHOEVEN, A., 1985, *The Agro Pontino survey project: Results from the 1984 field season*, in *Internal report, Albert Egges van Giffen Instituut voor Prae- en Protohistorie, University of Amsterdam*.

ZEI, M., 1973, *Un giacimento epigravettiano nell'Agro Pontino a Modella di Sabaudia (Latina)*, in *Rivista di Scienze Preistoriche*, 28, p. 107-132.

APPENDIX I: DETAILS OF THE SAMPLING DESIGN

The total area of the Agro Pontino is 877 km^2 with 678 km^2 deemed accessible for survey (after excluding urban and rural development areas, the Parco Nazionale, roads, etc.). In this design the sample element is the same as the observation unit, the agricultural field. On the basis of the 374 elements observed in the first phase, the sampling frame has been estimated 74,800 elements; that is, there are approximately 74,800 fields in the Agro Pontino.

For the second phase, a systematic unaligned transect sample was chosen as the most efficient design for selecting elements because it allowed the sample to retain the approximate proportions of the environmental zones present in the Agro Pontino as a whole, and, at the same time, provided a way to cross-cut the major axis of environmental variation on the plain. It was decided to make spatial strata, hereafter referred to as *blocks* to avoid confusion, of equal size across the length (NW-SE) of the Agro Pontino and to randomly select at least one transect within each block. The transects were to be drawn NE-SW from the mountains to the Southwest coast, thus crossing the colluvium, the graben fill, the Latina level lagoonal deposits, aeolian sands, and finally the series of coastal beach ridge-lagoon deposits. The position of the blocks would reflect the NW-SE differences in the extension of the aeolian sands and the beach ridge-lagoon complex.

In order to decide how many blocks and transects to draw, it was necessary to know how many fields would be crossed by a single transect and how large the sample should be. By counting the number of fields along several lines from the coast to the mountains, it was ascertained than an unobstructed transect (i.e., one that did not cross an urban area, the park, etc.) could be expected to yield about 150 fields. The sample size required is not so easily estimated because it varies according to the parameter queried and to the size of the error of estimation one is will to accept (*not* according to an «acceptable» proportion, such as 10 % of the area, as is frequently supposed). Thus, four general questions about the archaeological record were asked and, using the first phase sample results as estimators of variance, the sample sizes required for a simple random sample to provide answers at the 95 % confidence level were calculated.

The formula used in questions 1, 3 and 4 below is:

$$n = \frac{Npq}{(N - 1)\, D + pq}$$

where: n = sample size required;
N = number of elements in the sampling frame;
P = proportion of interest;
q = 1 - p;
D = $B^2/4$, where B is the bound on the error of estimation.

It should be noted that B is in the same unit terms as the estimators p and q. That is, B = .05 means an absolute error of ± 0.05.

Question 1: What proportion of fields in the Agro Pontino can be expected to contain archaeological remains?
The first phase sample of 343 surveyed fields showed that 75,8 % of the fields contained artifacts (p = .758) and 24,2 % did not (q = .242). A randomly selected sample of 293 observations would be sufficient for an inferential statement about the proportion of fields containing artifacts with an error of .05 (± 5 %).

Question 2: Of the fields containing archaeological remains, what is the mean density of the remains?
Since this question concerns absolute counts rather than proportions, the formula used is slightly different:

$$n = \frac{No^2}{(N - 1)\, D + o^2}$$

where: o^2 = the population variance, estimated by the sample variance; all other symbols are the same as above.
Density was calculated on the first phase sample by using the area of the field actually covered rather than by using the total field area and increasing the number of finds in a field. Because not all fields were surveyed systematically and had recorded coverage information, density calculations could only be made for 213 elements of the first

phase sample of 343. This smaller sample did not differ significantly from the larger sample (Kolmogorov-Smirnov two-sample test, $\alpha = .99$) in either area or number of finds. Density ranged from .000168 artifacts/m^2 (or 1-2 per hectare) to .084/m^2 (840/ha) with a variance of .0001. With a bound on the error of estimation of $\pm .001$ (± 10/ha), n = 398. Assuming that there is 75 % chance of finding a field with artifacts, approximately 560 fields should be surveyed to achieve a .95 probability of securing a sufficient number of fields that contain one or more artifacts.

Question 3: What proportion of fields contain materials of various time periods?

In the first phase sample, 281 fields have materials that have been dated in a preliminary way. A number of these fields have materials from more than one time period. We decided to set B at one-tenth of the proportions of interest as estimated from the first phase sample, with a maximum of 0.05. Thus, the error of estimate would never be more than 10 % for the proportion of interest (p) found in the first phase sample.

Period	Present (p)	Absent (q)	B	Number of observations required
Middle Paleolithic	.376 (129)*	.624 (214)	.0376	658
Upper Paleolithic/Meso	.402 (138)	.598 (205)	.0402	591
Neolithic/Bronze Age	.394 (135)	.606 (208)	.394	623

* Absolute counts in parentheses

Following a more precise dating of materials and finer chronological divisions, it is probable that certain time periods will be represented in even lower proportions. If so, we will most likely have to accept a wider bound on the error of estimation since making additional observations will not be possible.

Question 4: What proportion of fields contain materials of a density exceeding 20 finds per hectare?

The frequency distribution of find densities shows that there are a very large number of fields with a low density of finds and a very small number of fields with a high density of finds. A greater density of finds is desirable for a number of reasons, such as a higher certainty for dating, detecting patterns in the relationship between finds and environmental variables, etc. The «standard» of 20/ha is the mode of the frequency distribution (which has a mean of 71.5, a Sd of 111.3, and a median of 36.7). Using the density sample of 213 fields (see question 2 above) from the first phase collection, 60.6 % of the fields had artifact densities equal to or greater than .002/m^2. The sample size required with a bound on the error of estimation of .05 is 380 observations. Again, assuming a 75 % chance of locating fields with finds, a .95 probability of securing a sufficient number of fields to determine the proportion of fields with greater find density can be obtained with approximately 675 observations.

Thus, sample requirements for an acceptable estimate on the parameters queried ranged from 293 to 675 observations. Five transects were therefore deemed the smallest sample feasible. Accordingly, the area was partitioned NW-SE into 5 blocks, each approximately 12 km wide, and a transect was randomly selected from each of the blocks.

Appendix II: The Agro Pontino Survey Recording Forms

AGRO PONTINO PROJECT

Record 3, part 1 Name - Nome

Code	Variable	Coding scheme
— — — — — —	CASENUM field number numero di campo	(See field manual, Section VII, C, 1)
—	NUMVIS	Number of visits
— — — — — —	XCOOR	X coordinate in regional grid
— — — — — —	YCOOR	Y coordinate in regional grid
— — — — — — — —	AREA	Area of field in square meters
— — — — — — — —	PROFILE	Area of profile in square meters
— — — —	JSMAP	(See field manual, Section VII, E, 3)
—	SLCL2 field slope class declivio di campo	

Record 3, part 1 (cont.)

Code	Label	Coding scheme
— —	SLAS2 field slope aspect orientamento di declivio	11 - no slope, non applicable - no declivio, no applicabile [0] 01 - North - nord 01 - Northeast - nord-est 02 - east - est 03 - Southeast - sud-est 04 - South - sud 05 - Southwest - sud-ouest 06 - West - ouest 08 - Northwest - nord-ouest 09 - Multiaspect ridge 10 - Multiaspect basin 99 - Missing, not recorded - dati assenti

Code	Variable	Coding scheme
— —	SAMP sampling transect number	10 - Not a transect field [0] 01 - First transect field 02 - Second transect field 03 - Third transect field 04 - Fourth transect field 05 - Fifth transect field
—	STATUS survey status recognizione compiuta	1 - Field surveyed - campo recognito 2 - Field and profiled surveyed - campo e profile recognito 3 - Profile surveyed - profile recognito 4 - Unsurveyed, other information, non recognizione, altra informazione 5 - Unsurveyed field and/or profile - non recognizione 6 - Survey impossible - recognizione impossibile
—	FINDS1 finds reperti	1 - Finds present - reperti presenti 2 - Finds absent - reperti assenti 8 - Not applicable, non survey, non applicabile, non recognizione
—	PEBBLES1 who pebbles ciottoli intatti	1 - Whole pebbles natural in soil matrix intatti ciottoli presente in suolo 2 - Whole pebbles not natural in soil matrix intatti ciottoli assente in suolo 8 - Not applicable, no survey, non applicabile, non recognizione
— — — —	LOCVAR1 to LOCVAR2 physiographic associations relazione fra campo e topografia locale	10 - No (other) physiographic associations - niente [0] 01 - Strategic spot - buona visibilità 02 - Drainage channel - vicino a corso d'acqua 03 - Coast - vicino a costa 04 - Marsh - vicino a palude 05 - Lakeshore 06 - Natural spring
—	LOCVAR3 vegetational zone zona di vegetazione	1 - Zone A (aeolian - eolico) 2 - Zone B (Latina lagoon) 3 - Zone C (peaty graben) 4 - Zone D (colluvium) 5 - Zone F (beach ridge-lagoon complex)

AGRO PONTINO PROJECT

Record 3, part 2 (soil variables)

Field number - numero di campo name - nome date - data

Code	Variable	Coding scheme	
— —	SEDIMENT	01 - alluvium - alluvionale	06 - littoral ridge - litorale
	soil sediment	02 - colluvio-alluvium	07 - lagoonal - laguna
	sedimento di	03 - colluvium - suoli da declivio	08 - tuff - tufo
			09 - aeolian on lagoonal
	suolo	04 - travertine	
—	TEXTURE	1 - coarse - di grano grosso	
	soil texture	2 - medium - di grano medio	
		3 - fine - di grano fino	
— —	DRAINAGE	10 - very poorly drained - palude [0]	
	soil drainage	01 - poorly drained - non permeabile, acque stagnante	
	conditions	02 - imperfectly drained - più meno	
	quant è permeabile	03 - moderately well drained	
	il suolo	04 - well drained	
		05 - somewhat excessively drained - permeabile	
		06 - excessively drained - molto permeabile	
— —	GEOAGE	00 - older than Latina level	07 - Aeolian 2
	age of soil	01 - Latina level	08 - Terracina level
	età relativa	02 - Minturno level	09 - Alluvium 1
	di superficie	03 - Older gravelly ridge; aeolian 1	10 - Aeolian 3; alluvium 2
			11 - Pre-Roman/Roman
	di suoli	04 - Colluvium 1	12 - Recent
		05 - Borgo Ermada level	
— — —	SOILTYPE	3 - eutric histosol	49 - solidic planosol
	FAO soil type	4 - lithosol	69 - haplic phaeozem
	tipo di suoli	5 - pellic vertisol	91 - gleyic luvisol
		6 - chromic vertisol	92 - albic luvisol
		8 - calcaric vertisol	93 - vertic luvisol
		10 - eutric fluvisol	96 - chromic luvisol
		17 - mollic gleysol	97 - orthic luvisol
		19 - calcaric gleysol	99 - gleyic luvisol
		29 - cambic arenosol	100 - vertic cambisol
		31 - calcaric regosol	101 - calcic cambisol
		33 - eutric regosol	105 - chromic cambisol
		35 - rendzima	106 - eutric cambisol

other soiltype - altro tipo di suolo

secondary soil types present: tipo di suoli subordinati

Profile section drawn? yes ——— no ———
Designe schizzo di profile?

AGRO PONTINO PROJECT

Record 4 Name - Nome

Code	Variable	Coding scheme
— — — — — — —	CASENUM field number numero di campo	(See field manual, Section VII, C, 1)
—	VISITNUM visit number numero di visita	
— — — — — — —	DATE data	day - month - year giorno - mese - anno
[— — —]	BOOK	day book number

Code	Variable	Coding scheme
— —	TYPE survey type tipo di recognizione	1 - field systematic - recognizione systematica di campo 2 - field unsystematic - recognizione non systematica di campo 3 - profile systematic - recognizione systematica di profile 4 - profile unsystematic - recognizione non systematica di profile 5 - (1) and (3) 6 - (1) and (4) 7 - (2) and (3) 8 - (2) and (4) 9 - redeposited dirt survey - recognizione di materiale redepositato 10 - soil description only - solo descrizione del suolo 11 - inquiry only - solo questione 12 - other - altro
— — —	PERCENTF	——— No, times field crossed - numero delle striscie di campo If rectangular field - per rettangolare campo: ——— lengthwise or ——— widthwise crossing longitudinalmente o latitudine orientamento delle striscie If irregular field - per campo irregulare: Draw a sketch of the field to show how the field was crossed Disegna uno schizzo del campo e mostra orientamento delle striscie

Code	Variable	Coding scheme
— — —	PERCENTP	Percentage of profile coverage - quanto % copertura di profile
—	TEMP temperature temperatura	(indicate how it feels - come el sentita) 1 - hot - troppo calda 4 - other - altro 2 - warm, comfortable - 9 - missing data - calda dati assenti 3 - cool - fresca
—	CLOUDS cloud cover tempo	1 - clear - sereno 2 - partly cloudy - parziale nuvoloso 3 - scattered showers - qualche rovescio 4 - rain - pioggia 5 - other - altro 9 - missing data
— —	TIME	Field surveyed from ——— to ——— Ora di recognitzione di campo/profile
—	PLOWING plowing conditions condizione superficiale di suoli per l'aratro	1 - plowed, large peds - grande blocchi di suoli 2 - plowed, roughly hoed - media blocchi di suoli 3 - plowed, finely hoed - fino blocchi di suoli 4 - rolled - spianato 5 - abandoned, harvested - mietuto 6 - pasture - pastura, non arato 9 - missing data
—	DUST condizione dust conditions superficiale di suoli per pioggia o irrigazione	1 - recent rain, irrigation (uitgeregend) recente pioggia 2 - disturbance since last rain (e.g., plowing), but not dusty - scompiglio dopo pioggia 3 - dusty - polveroso 9 - missing data - dati assenti
—	VEGET vegetational cover quanto vegetazione	1 - non or very sparse - niente o sparsa 2 - less than 50 % of the surface - meno di 50 % 3 - more than 50 % of the surface - più di 50 % 4 - vineyard - vigneto 9 - missing data - dati assenti
—	EROSION local soil movement movimento locale di suoli	1 - no apparent erosion - niente 2 - generalized slope wash - dilavamento 3 - erosion channels present - canali di erosione

Code	Variable	Coding scheme
		4 - (2) and (3) 5 - indeterminate - indeterminato 6 - field levelled 7 - field terraced 9 - missing data - dati assenti
—	SOILADO soil added addizione di suoli	1 - no soil added - niente 2 - infer aeolian cover (from analysis) - eolico secondo analisi 3 - levee - argine di fiume o di canali 4 - canal dredge, report - da canali secondo relazione verbale (data dai padroni) 5 - canal dredge, infer (from analysis) - da canali secondo analisi 6 - soil brought in from other place - da altro campo 7 - indeterminate - indeterminato 9 - missing, not reported - dati assenti
—	SOILMIN soil removed come rimozione di suoli	1 - no soil removal - non rimozione 2 - excavation, visible - rimozione visibile 3 - excavation, report - rimozione secondo relazione verbale (data dai padroni) 4 - excavation, infer (from analysis) - rimozione secondo analisi 5 - indeterminate - indeterminato 9 - missing, not reported - dati assenti
—	FINDS2 archaeological finds reperti	1 - present - presente 2 - absent - assente 3 - present, not all collected - presente, non predibile 8 - non applicable - no applicabile
—	PEBBLES2 whole pebles intatti ciottoli	1 - whole pebbles natural in soil - intatti ciottoli presenti nel suolo 2 - no whole pebbles in soil matrix - intatti ciottoli assenti nel suolo 8 - non applicable - no applicabile 9 - missing data, not reported - dati assenti

Record 6

Code	Variable	Coding scheme
— — —	RECFIND	estimated number of uncollected finds numero di reperti non predibile

Record 25

Code	Variable	Coding scheme

Tiles present? tegola presente? —— yes —— no

Record 7

Code	Variable	Coding scheme

Profile finds - reperti in profile

Artifact number Numero di reperti	ZCOORD cm	HORASSOC soil horizon association
______	______	______
______	______	______
______	______	______
______	______	______
______	______	______
______	______	______
______	______	______
______	______	______
______	______	______