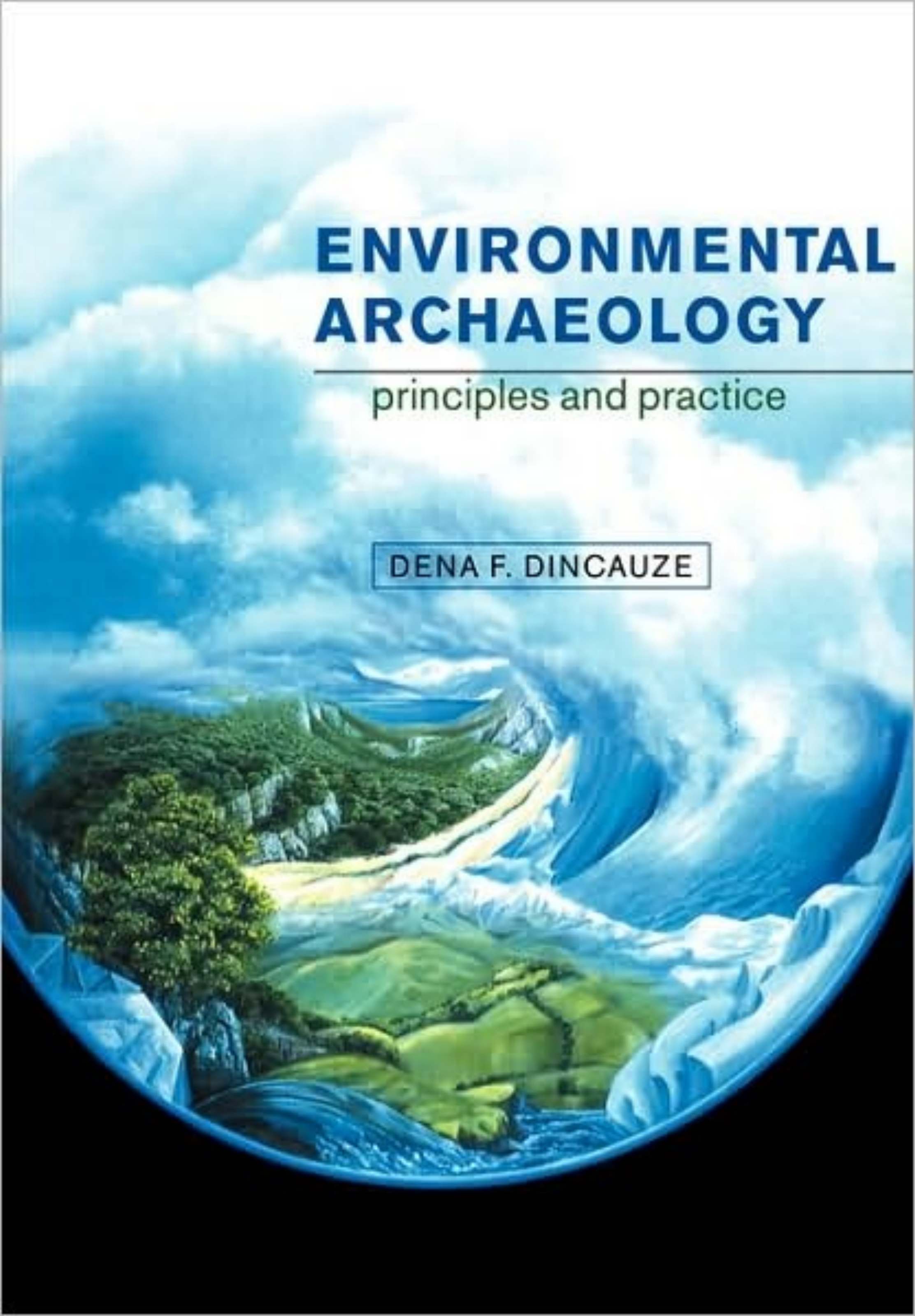
ENVIRONMENTAL ARCHAEOLOGY

principles and practice

DENA F. DINCAUZE



ENVIRONMENTAL ARCHAEOLOGY

Principles and Practice

DENA FERRAN DINCAUZE
University of Massachusetts,
Amherst

CAMBRIDGE
UNIVERSITY PRESS

2

CONCEPTS FOR PALEOENVIRONMENTAL RECONSTRUCTION

> If the theories behind borrowed concepts are not clearly understood and taken into account in the application of the concepts, not only will the results of the concept application be suspect, but misunderstandings may arise between practitioners of the science from which the concept was borrowed and the concept-borrowing archaeologists.
>
> CREMEENS AND HART 1995: 16

The study of the human past requires knowledge of the solar system as well as of the home planet and its geophysical and biological systems, of which we are inextricably a part. The eternal fascination of archaeological research is that it challenges all our creativity, discipline, and enthusiasms; scarcely any knowledge is irrelevant to it. That is especially true of environmental archaeology – the study of paleoenvironments as human habitats. Habitats pose problems and opportunities for resident organisms of whatever size and complexity; humans are not excepted from this imposition. If we are to understand the behaviors of human beings in their unique cultural contexts, we must be able to define and examine crucial aspects of their habitats. Human environments, originally restricted to sub-Saharan Africa, now include the entire world and parts of space – so, the study of human ecology, which is at the core of environmental archaeology (Butzer 1982), is necessarily comprehensive and resolutely dynamic. Not surprisingly, it is still very immature and experimental.

The means for defining and interpreting elements of human environments, both past and present, are expanding. Archaeologists, especially environmental archaeologists, employ techniques and concepts developed in the disciplines of anthropology, biology, ecology, zoology, botany, geology, oceanography, climatology, and pedology (soils), among others. Of course, no one can be expert in all these subjects; both compromise and consultation are required. Under such conditions, it is fair to say, the borrowings have not always been efficient, or even effective.

Donald Hardesty (1980: 161) warned of "the hazards of crossing disciplinary boundaries on search and seizure missions." Borrowed techniques and methods should not be isolated from the concepts and theories with which they were developed. When people fail to respect that relationship, they create unnecessary difficulties. Too often, such failure has resulted in misapplication of methods, oversimplification of interpretation, and error. This work aims to minimize those hazards by introducing the fundamental concepts, theory, and vocabularies of the disciplines most often borrowed from, to help archaeology students cross the borders into the domains of other disciplines prior to specializing in any of them. Further specialization in one or more non-archaeological disciplines is strongly encouraged.

Research in the scientific mode is a search for insight and minimization of bias and error. There are many styles of scientific research. The physical sciences are experimental and quantitative. The natural and historical sciences are more descriptive and qualitative, less suited to controlled experiments. The former enjoy particular prestige for the **precision** of their methods and results. The latter stumble along dealing as well as possible with the complexities of the world and the intricacies of human perceptions, motives, and interpretations. The physical sciences are equally subject to the messy limitations of human cognition and conceptualization, but they deal with phenomena more radically distinct from common human experience. It can be argued that, in dealing forthrightly with complexity, the historical disciplines are more realistic than the conventional physical sciences.

Archaeology shares methodological constraints with all the historical sciences, from astronomy to paleontology. Among these are incomplete and discontinuous data sets that yield few representative samples and are inaccessible to experimentation, poor control of time for measuring rates, reliance on analog arguments, and no direct access to causation. Archaeology shares some of the strengths of historical studies as well, such as their long-term perspectives on change and historical processes and their respect for context and contingency.

Within the historical disciplines, the mode of discourse and the tone of reporting in the natural sciences may be less qualified, less hesitant, than that of the social sciences. That is tradition and style; it is not a reflection of a tighter grasp on "truth."

"[F]rustration with inherently imprecise data too often gives rise, in ecology as in the social sciences, to self-conscious 'schools' of thought, and to equally self-conscious obsessions with one or another 'scientific method' " (May and Seger 1986, *American Scientist* 74: 260).

Paleoenvironmental studies integrate the historical sciences in a manner similar to the way archaeology integrates the social sciences, by focusing on past worlds and ways to understand them. Science is no less scientific when it deals with variation rather than permanence, with change rather than continuity or stasis. The historical sciences, unlike chemistry and physics, deal with the dynamics of *particular* conditions in the past rather than with universal, static, laws. This emphasis on **contingencies** – unique, historical configurations of phenomena – is a powerful heuristic and seems to be the way the historical sciences do, in fact, proceed (Gould 1986). The acceptance of contingency is a far cry from the dogma of historical determinism. The importance of the historical sciences is that the past affects the present because past events and processes have constrained the range of options open to events and processes today.

There are further constraints in the historical disciplines. Scientific theory is often said to be quantitative and lawlike, and to aim for universal, rather than contingent, results. The study of quantitative relationships requires representative samples, which are rarely available and hard to recognize in historical data. Qualitative attributes, used in analogical and inferential reasoning, are more accessible in historical disciplines. Furthermore, several of the historical disciplines are based on concepts which have proven intransigent to quantitative definition or measurement. For example, anthropology has "culture," archaeology has "site," climatology has "climate," ecology has "system," and biology has "community." These concepts have been fundamental to the successes of their disciplines, yet are not amenable to the basic scientific activity of *measurement*, probably because they are artificially isolated from context. They appear to exist outside of time and space. If time and space are "added" in the sense that temporal and spatial limits are defined for each of the concepts, they become operational, but only within the constraints of the definitions. This state of affairs is a quality of these disciplines, not simply an attribute of their supposed immaturity. Efforts to impose quantification at the expense of qualification are futile diversions from the proper tasks of the historical disciplines.

Without apology, then, quantification will not be emphasized in this work, but will be discussed and applied when appropriate in various sections of the text. Scientific concepts that require quantification and assume representative samples are not a major feature of this work.

PALEOENVIRONMENTAL RECONSTRUCTION IN ARCHAEOLOGY

> We seem perilously close to that characteristic failing of interdisciplinary study – an enterprise which often seems to merit definition as the process by which the unknowns of one's own subject matter are multiplied by the uncertainties of some other science.
>
> SAHLINS 1972: 51

The discipline of archaeology has recently been through a period of methodological self-consciousness during which it tried to achieve scientific standards comparable to those of the quantitative physical sciences, while keeping its social science credentials. Historical context was devalued in the search for eternal verities ("laws" of human behavior).

Bailey (1983: 172) has noted that, in the same period, most of the natural sciences that archaeologists utilize became aware of the historical dimensions of their data; such awareness has been a powerful impetus to theoretical developments. Contextual richness and specificity have now been reinstated in archaeology, in both the natural-historical (Butzer 1982) and sociocultural (Hodder 1986, 1987) senses.

Multidisciplinary research has recently been afflicted by a mood of pessimism or skepticism concerning its potential for success (e.g., Thorson 1990b), especially in paleoenvironmental reconstruction. The problems are numerous and not trivial. They include intellectual as well as economic problems of cooperation, of coordination, of integration, and of differential reward. Investigators in the several disciplines involved in paleoenvironmental research are trained in distinct academic traditions and hired into separate departments and faculties. Consequently, they employ different vocabularies and perspectives on phenomena, perspectives that involve the scales at which they work as well as unexamined habits of mind. It is not unusual for members of an interdisciplinary team to be disconcerted by each others' mind-sets and work habits. The integration of results and the efficiency of cooperative work can be facilitated by a strong, appropriate research design, but the communication of research goals across disciplines can be remarkably difficult. Cooperative research in environmental archaeology appears to be especially difficult because of the incompatibility of scales of problems and data between archaeology, operating within the human dimension, and those of the natural-historical sciences, operating at the levels of species, region, and millennium.

All the disciplines involved in aspects of paleoenvironmental reconstruction have different goals, to which their characteristic scales of observation and data collection are appropriate. For archaeologists, the goal of paleoenvironmental reconstruction is the description of change in the physical and biological contexts of human existence.

The goal has often been only partially achieved because of temporal distance, the need to rely on indirect evidence, and the inherent difficulties of working in a multidisciplinary mode. A better understanding on the part of archaeologists of the basics of the cooperating disciplines can turn difficulties into strengths.

Integrating multidisciplinary research

Science is *learning*, not knowing. The writ of other disciplines is no more holy than that of one's own. All bodies of evidence and interpretation grow and change in time, because the scientific mode of learning involves recursive evaluation of hypotheses and insights, with adaptive change as indicated. Interpretations, especially, remain vulnerable to new data and new thought, and should not be accepted uncritically. There is a special intellectual excitement in working across disciplinary boundaries; all parties to the enterprise must be willing to stretch, grow, and reconsider what they believe they know. The personal rewards can be enormously gratifying.

The differences in scales and data sets in the several sciences can be a source of strength in multidisciplinary work. As definers and integrators of research projects, archaeologists can ease integration with the three *C* goals: *complementarity* of different data sources, *consistency* between data sets, and *congruency* of scale.

Complementarity enlists the strengths of diverse data sets to create interpretations more nearly complete than any single discipline can achieve. As the following chapters will show, the reconstruction of any aspect of ancient environments can utilize evidence from a number of other aspects from the same time and place. Since all data sources are subject to errors of association, representativeness, or interpretation, using diverse data sources reduces the likelihood of error caused by overreliance on any one source.

The goal of consistency requires that the reconstruction of any one aspect of paleoenvironments be compatible with the reconstructions of others. All the evidence, and the resultant interpretations, should agree – not necessarily in detail, but in presenting associations that are not inconsistent with plausibility. To take an extreme example: if pollen study implicates the presence of a lush deciduous forest in a given area and time, evidence from soils that permafrost existed at the same time is clearly inconsistent. As neither data set can be dismissed out of hand, additional research is needed to evaluate the alternatives.

The congruency goal recognizes the need to mediate among data sources at different scales. Paleoenvironmental data from different disciplines reach the archaeologist wrapped in concepts and scales that may not be equivalent. "To the archaeologist looking for an interpretation he can use, one form of biological evi-

Table 2.1[a] Exponential scales in space and time

Spatial scales	Area (km^2)	Spatial units
Mega-	5.1×10^8 km^2	Earth
	$<10^8$ km^2	continents, hemispheres
Macro-	10^4–10^7 km^2	physiographic province, region
Meso-	10^2–10^4 km^2	site catchment, area
	1–10^2 km^2	locality, city, large site
Micro-	<1 km^2	locale, site, house, activity area
Temporal scales	**Duration or frequency (yr)**	**Spans**
Mega-	$>10^6$; >1 ma	more than 1 million years
Macro-	10^4–10^6; 10 ka–1 ma	10,000 to 1 million years
Meso-	10^2–10^4; 0.1 ka–10 ka	centuries to 10,000 years (millennial)
Micro-	$<10^2$; 0.001 ka–0.1 ka	less than one century (decadal)

Note: [a] Each higher unit incorporates and generalizes all those below. Scales in the two dimensions are not closely linked.

dence seems as good as another. But pollen grains are very different from snail shells; they get into the soil in different ways, they survive (in calcareous soils) for different lengths of time, and probably indicate ecological conditions over different spatial dimensions" (Dimbleby 1985: 64). Awareness of incongruence between the regional and local scales, and between short and long time spans, can help the archaeologist interpret each data set in terms of its own scales of time and space, to avoid comparing incomparable entities. Geologists face similar problems:

> At different scale resolution levels, which are mapped out according to our aims and abilities, different problems are identified; different types of explanation are relevant; different levels of generalization are appropriate; different variables are dominant; and different roles of cause and effect are assigned . . . It is also apparent that conclusions derived from studies made at one scale need not necessarily be expected to apply to another.
>
> CHORLEY ET AL. 1984: 12

Table 2.1 presents a scheme of exponential scales in time and space, with four divisions in each dimension. The units shown are heuristic only. They have logically no firm boundaries except when those are specified for particular cases. The four-part scheme simplifies reference to the several hierarchical levels of scale, but the exponents obviously can be subdivided more finely (e.g., Butzer 1982: 23–27). Relevant spatial concepts and methodological considerations vary with the scale under consideration.

In the temporal dimension, methods for measuring time vary with the scale, just as methods for inferring sociocultural phenomena also vary. Note that each higher unit incorporates and generalizes all those below; both detail and diversity are lost as research interest moves up the scale. For each level in each dimension, there are appropriate data categories that differ from those at other levels.

Reasoning and "recursive ignorance" in the paleoenvironmental sciences

Historical studies encounter the obstacle that the phenomena we want to study are rarely directly accessible to research in the present time. Consequently, indirect evidence is interpreted to increase knowledge. Astronomers study light spectra to learn about the chemical compositions of stars; paleontologists study bones to learn about the soft tissues and behavior of extinct animals; climatologists study pollen deposits to learn about atmospheric circulation patterns, and archaeologists study artifacts to learn about human behavior and societies. All of the historical disciplines rely on mathematical and statistical data and reasoning when those are appropriate, and rely on inference and comparisons with observable phenomena as necessary.

> Palaeoenvironmental analysis essentially proceeds by induction. Data from faunal, floral and sedimentological residues in bogs, lakes, river terraces and valley fills are used to infer past environmental conditions such as plant cover and hillslope and fluvial processes. These in turn are used to infer climatic parameters which may then form the basis of archaeological explanations. This may eventually produce a recursive ignorance as successive approximations move backwards and forwards, to and from the original inductive activity and may even prove circular as they are applied to explain the data from which they were derived. Despite the intrinsic shortcomings, which are widely understood, this approach continues to be widely practised in archaeology and geomorphology. Of the several reasons for this the two most important appear to be (i) a dismissal of essential limitations on the grounds that the ends justify the means, which is doubtful, or (ii) the reluctance to develop an alternative strategy or set of strategies to deal with the problem.
>
> THORNES 1983: 326

The following discussion responds preliminarily to Thornes' concerns; the bulk of this volume is devoted to them. If we are to escape from recursive ignorance, we need awareness of our habits of mind and level-headed clarity about our goals, so that we can select methods appropriate to our purposes.

Quantitative reasoning: sampling and probabilities

Archaeologists discovered sampling theory in the late 1960s; palynologists discovered it in the 1970s. Since then, its proper application has revolutionized the two disciplines, restricting or broadening acceptable inferences, depending on the case. In archaeology, inappropriate application of sampling theory and the sophisticated statistical analyses that accompany it has spawned a misleading and difficult literature, which has come under scrutiny and refutation (see, e.g., discussions in Cowgill [1986] and Shennan [1988]).

Formal sampling is a way of estimating what you can know about a population of data from examining a portion of it. The portion examined must be rigorously chosen to eliminate bias, and there must be enough of it to be representative, to a major degree, of the diversity within the population, which, of course, is what you want to know in the first place. Sampling is neither a magic solution to this dilemma nor hocus-pocus. Sampling works best with natural phenomena that are homogeneous at some scale. Unhappily, archaeological deposits are usually anything but homogeneous. George Cowgill has endeavored to elucidate the problems and potentials of sampling in archaeology; his work is the place to begin examining the issues (Cowgill 1970, 1986, 1989).

For paleoenvironmental studies, sampling problems are more prevalent than the literature indicates.

- In field work, the selection and size of samples are often defined casually, whether the subject is areas for investigation, sediments, or range of included materials.
- The descriptive literature is typically uninformative about the criteria for sampling in any given case, and
- the interpretive literature is frequently flawed by failures to consider responsibly the error probabilities inherent in the materials selected for study.

The aspiring student of archaeology or of the geosciences should be sensitized to this aspect of methodology, both in practice and in reviewing the literature. More sophisticated selection and evaluation of samples would go far toward correcting the kinds of error and circular reasoning that Thornes implicates in our "ignorance."

Samples for analysis must be representative of the phenomena under study, and appropriately sized for the analytical methods used. Representativeness of samples varies with the size of the sampling fraction and the distribution ("randomness") of samples collected. The more complex and structured the phenomena under study, the larger must a sample be before it can be considered reliably representative. For example, the diversity of sedimentary deposits within the walls of a single ruined room is likely to be orders of magnitude less than the diversity within the walls of a ruined town. However, if detailed behavioral and ecological interpretations are desired, a single sample from any room would not be sufficiently representative. These matters need to be addressed at the outset; good sampling plans can be modified to respond to circumstances, but selected collections of materials can never achieve the status of representative samples.

The decisions to be taken in designing a sampling program within a stratified deposit are different from those related to horizontal patterning. Good texts on excavation procedures (e.g., Barker 1993; Hester et al. 1997; Stein 1987) discuss the decisions involved in sampling – contiguous or non-contiguous units, small or bulk samples, at interfaces or central to stratigraphic units, and so forth. With any uncertainty, make the samples as discrete as possible, and the intervals small. While samples can always be consolidated later, they cannot be meaningfully subdivided after collection.

Since examination of all sedimentary contexts at any archaeological site is almost impossible, the student must be aware of the constraints on inference that are imposed by the sample available for study. The generality of an interpretive argument necessarily varies with the representativeness of the sample on which it is based. The strengths of sampling theory are that it can help estimate the degree to which the samples taken can be said to represent the diversity of the sampled phenomena. The relevance and appropriateness of samples to the problem(s) under investigation should be evaluated explicitly for each case, and carefully reported with the interpretation (Shennan 1988: Ch. 14).

Not even the best sampling plan can overcome the fact that all sediments and materials available from the remote past have been subject to non-random selection before they are scrutinized by researchers. **Taphonomy** is the study of the processes leading to fossilization of biological remains (Chapter 9 and Part V). While obviously an example of qualitative, not quantitative, evaluation, the recognition of taphonomic processes must be considered in any discussion of sampling, as they are major natural sources of non-randomness in deposits. Taphonomic processes include the original deposition of particles on the surface or in water, weathering, disturbance or consumption by living organisms, disaggregation and transportation of remains after death, burial at the site of death or redeposition, chemical and physical modification after burial, and whatever may befall the sedimentary context of burial. The complexities of individual histories of organic remains and their enclosing sediments are almost infinite. Rules of taphonomic inference are being developed by paleontologists and archaeologists to help regularize the observation and evaluation of relevant evidence (e.g., Behrensmayer and Hill 1980; Lyman 1994; Schiffer 1987). Palynologists are beginning to develop a literature on taphonomy, also, which will be noted in Part VI.

Qualitative reasoning: analogy, inference, and causation

Analogical reasoning is fundamental to paleoenvironmental and archaeological research. It is a form of inference by comparison. In historical studies, selected situations contemporary with the observer are assumed to share important characteristics with aspects of the past that are under investigation. Classic **analogy** is a form of logic by which one establishes the equivalence of two things that cannot be directly compared. The argument takes the form: A is like B; B is like C; therefore C is like A. Such logic is usually expanded to the more complex issue of comparing properties of entities. For example, if we observe that A and B have properties *m*, *n*, *o*, and *p*, and that C has properties *m*, *n*, and *o*, we may infer by analogy that C has *p* also. But, unless *p* is logically or essentially entailed by properties *m*, *n*, and *o*, our inference is only weakly grounded. In paleoenvironmental studies, effective comparisons can only be made if there is knowledge or the reasonable assumption that the entities being compared share the properties that we want to study, and that their similarities are crucial to the matter under investigation. When temporal distances between the two entities or situations are great, equivalence must be demonstrated, not assumed, lest we trap ourselves in circular reasoning. Analogies are invaluable for formulating research designs because they help specify what parameters are important for a particular question.

Analogical reasoning is a kind of pattern matching – the similarities of two entities or situations are explicated and evaluated. If the similarities are great, the usual assumption is that there are some causal connections among the shared characters so that they recur for similar reasons. This assumption is not, however, supported by the logic of analogy, which is a weak method for learning about causation.

Analogies (particularly in the form of metaphors) are basic to human speech and probably to human thought (Lakoff and Johnson 1980). However, analogical arguments rarely lead to new knowledge; typically, they show that some phenomena had or have wider distributions than was previously known. Lacking a stronger basis for inference, we must learn to use analogical reasoning responsibly and critically (Kelley and Hanen 1988; Wylie 1985). Analogies help us understand phenomena that are, for any reason, imperfectly observable. The analogical mode of reasoning permeates the historical sciences whenever practitioners not only seek to learn about past states and phenomena but also engage in rigorous observations of phenomena of the present in order to learn more about the past (a method called *actualism*).

In fact, analogical reasoning is so important to the historical sciences that it was elevated to the status of a principle during the nineteenth century. The geologist Charles Lyell is closely identified with the principle of **uniformitarianism**, the assertion that processes in the past were comparable to processes in the present. Taken at face value, this could mean that we are unlikely to learn anything surprising about the past, no matter how distant. Stephen J. Gould and others clarify this conundrum by recognizing that processes in the past were not necessarily equivalent in scale or duration to those of the present, only similar in kind and function. This distinction allows for the existence in the past of continental glaciers larger than any now observable, but requires that the processes that formed and controlled those glaciers were like those we can observe at smaller scales today. This kind of uniformitarianism is a *methodological principle*, not a statement of brute equivalence (Gould 1965; Rymer 1978).

A corollary of uniformitarianism is the use of **proxy** data in the paleoenvironmental sciences. Proxies are phenomena that indirectly relate to the phenomena we want to know about – especially, they are data that can be used to infer aspects of paleoclimates through analogies. For example, pollen evidence indicating the former existence of a spruce forest in an area now populated by temperate deciduous trees can be considered proxy evidence of a former climatic regime cooler than that of the present. The argument works through analogical reasoning because today spruce-dominated forests maintain themselves only in areas of severe winters and brief summers where temperate forest species cannot thrive. Similarly, the discovery of woolly mammoth bones in an area now temperate in climate is taken as evidence for ancient climates colder than those of today. Proxies are observable phenomena used to infer the presence or state of phenomena not directly observable – in these examples, forests and climates.

Some investigators have taken the argument further by recording and measuring attributes of biological systems associated with different climatic regimes today, and then mathematically "transferring" those climatic regimes into the past wherever they observe biological associations similar to those of the present. This extension of analogy has been cogently criticized for involving a great many unstated and untested assumptions about complex ecological relationships in both the past and present (Birks 1986; Hutson 1977; Lowe and Walker 1984: 155–156; Rymer 1978).

Chapter 7 demonstrates how widespread and essential is the use of proxy data in the study of paleoenvironments, and further discusses the methodology. Here it is sufficient to remind ourselves of the complexities of biological organisms and of their interspecific relationships within the various biota of the world. Even very simple organisms are capable of responding to environmental problems in more than one way, and when they occur together in associations of diverse species, the diversity of responses may defy codification. Proxy data for indirect inference about conditions in the past must be chosen carefully on the basis of knowledge as full as possible of their relevance for the issue of concern. Relevance involves such abstract qualities as scale as well as the clarity of the proxy's climatic signal. Both of these qualities are knowable only as they are expressed in contemporary, contingent situations. The application of analogical reasoning cannot eliminate the contingencies in the proxies, and thus the method is subject to the fallacy of transferring the present into the past. The best proxies for paleoclimates would be organisms so well known in all their diversity that the boundaries and mechanisms of their responses to particular aspects of environment could be directly specified and observed in fossil data. At present, we control (or believe we control) the boundaries for some organisms, but the mechanisms for almost none. For example, beetles are considered to be outstanding proxies for ambient temperatures, but only those qualities, as we shall see in later chapters.

As more is learned about the mechanisms of the various environmental systems (Chapter 3), investigators have confronted a new challenge to knowledge of the past: the **no-analog** problem. Environmental circumstances are highly complex and not fully determined by any finite set of known factors (Bradley 1999; Hutson 1977; Webb and Bartlein 1992). Because of the multiplicity of factors shaping any system of organisms, all prehistoric as well as modern systems are unique in some characteristics. No set of environmental circumstances is ever precisely replicated, and so no modern situation can serve as an accurate analog for any multivariate environment in the past. For example, the beetles that are used as indicators of temperatures cannot serve as proxies for rainfall, length of season, or vegetation. Contingency complicates inquiry in ecology and in all historical sciences; it limits the application of analogies of all kinds.

Another powerful limitation on the use of analogical reasoning is the principle of **equifinality**, which is that different sets of antecedent conditions may produce similar results. That is to say, no single set of causes can be readily assigned to each unique situation or circumstance. An example is that of a rising sea level, which may be the result of (1) an increase of water in the oceans, (2) sinking of the land, or (3) reduction in the size of ocean basins. Much more information than the simple observation of rising sea level is needed to distinguish among the possibilities. The principle of equifinality requires that multiple hypotheses be involved in any research (e.g., Haines-Young and Petch 1983).

CAUSATION AND SYSTEMIC RELATIONSHIPS

> The gravest problem in actualistic research is assuming that a given agent is necessary and sufficient cause of an observable attribute when no such relationship has actually been established.
>
> GIFFORD 1981: 394

The search for causes of observed phenomena and situations is deeply ingrained in the methodology of science and in the ways we all think about the world. Whenever some event arouses our interest, we seek its cause, in order to understand, commend, or blame. The historical disciplines have few methods for seeking causation. Conventionally, the observation of positively correlated variables (characters that always occur together or change at the same time) is taken to imply some causal link between them. However, the co-occurrence of two or more phenomena does not imply causation by any one of them, only the likelihood of interdependence and close relationship, possibly through a third variable. Also, pure coincidence may be confused with correlation. Pattern is not cause; arguments from correlation are research problems, not explanations. For the latter, we need to identify and understand the mechanisms in relationships – how things work together or affect each other.

In the middle years of this century, as computers aided the study of complex mathematical relationships, investigators applied ideas about systems and systemic relationships to data in many disciplines. **Systems** are bounded sets composed of entities and their relationships. It is characteristic of systems that a change in one entity will require compensating adjustments in others that are linked to it. We have already met an example of this in the discussion of the Red Queen hypothesis in Chapter 1.

Simple closed systems are assumed to be in a state of equilibrium unless disturbed, and to return after disturbance to conditions close to the original. In the mechanistic language of systems theory, the interactive responses of systems to disturbance, or to any change, are termed **feedback**, and may be either positive or negative. Positive feedback amplifies (increases) deviations from the original state of the system, leading to modified relationships among the parts. Negative feedback dampens (reduces) deviations or innovations to systemic relationships; its effects are to return the disturbed system to something like its original state or configuration. Open systems are those that receive matter, energy, or information from their environments. All complex systems involving living organisms are open and dynamic, with feedback responses initiated from inside as well as outside the system. Rather than tending toward static equilibrium states, feedback in complex systems approximates dynamic balance, with additional change the characteristic result of disturbance.

Systems models represent a major advance in realism from simpler models of linear causation. They emphasize multidimensional relationships among components of a system, and help focus attention on the feedback mechanisms that define the dynamic balances. Systems thinking was introduced into archaeology in the late sixties, most influentially by Lewis Binford (1965), David Clarke (1968), and Kent Flannery (1968, 1972). It rapidly transformed the terminology and metaphors of the discipline (Watson et al. 1984: Ch. 2). Formal systems theory is fundamentally mathematical. In the biological, environmental, and social sciences multivariate relationships are too complex for mathematical modeling even with the latest supercomputers. Nevertheless, systems thinking as a way of dealing with complex relationships that are multicausal has produced significant insights and improvements in understanding in all the sciences that deal in any way with living organisms.

THE EMERGING CHALLENGE TO CAUSAL THINKING

All the sciences of Western industrial society are based on a traditional mode of thought that can be traced back to Aristotle, at least, but was codified and established within the scientific enterprises in the seventeenth, eighteenth, and nineteenth centuries. It is exemplified in the work of René Descartes, Isaac Newton, and Charles Darwin. This mode of thinking, amounting to a world-view, an article of faith in the way the world works, assumes direct linearity in causation and a certain mathematical determinism in systemic relationships. It has served brilliantly as the foundation of the mathematical and experimental sciences that have given humanity unprecedented power over the natural world and the means to shape the destiny of life as we know it.

When linear causation and mathematical determinism are employed with suitable analogies, it is possible to construct predictions for states of affairs in the future, or retrodictions for the past, that carry some scientific authority. A fashionable definition of "explanation" in archaeology equates it with the ability to make predictions that are borne out in the course of events. All of this has been intellectually gratifying; it is "normal" science. Recently, ideas emerging from studies of randomness in physics, ecology, meteorology, and epidemiology are beginning to cast doubt on the universal appropriateness of such deterministic assumptions. These investigations are converging in the "sciences of complexity," one aspect of which is "chaos" studies, a name chosen to attract attention but which unfortunately tends to mislead. In both standard and scientific English, "chaos" until now has meant true randomness, absence of patterning, signals that contain no information, existential "noise." Much of the strength and attractiveness of the new chaos theory, however, is that it asserts the information potential of phenomena that have been considered residual in traditional science.

The new insights are extensions of systems theory that are rapidly superseding some basic concepts about systemic behavior. One of the major interdisciplinary findings of chaos investigations is that complex systems do not exhibit deterministic trajectories in either time or phase space; rather, systems of many different kinds tend to share a "sensitivity to initial conditions" (i.e., to their environments) that makes them liable to veer away from former "paths," to be unpredictable in all but the very smallest scales (Glass and Mackey 1988; Gleick 1987; Kauffman 1995; Schuster 1988; Waldrop 1992). This kind of randomizing system behavior can be replicated by fairly simple mathematical operations provided they are reiterated a great many times, with the results of one calculation feeding into the next. Rather than leading quickly to disorder, some randomness has been shown to be "creative" in ways that reduce disorder, and to be deterministic to the extent that it can be replicated at several different scales (fractals). Snowflakes, achieving an almost infinite diversity of forms based on six-part symmetry, are excellent examples of the creativity of sensitive dependence on initial conditions. The growth of snowflakes from water vapor exemplifies decreasing disorder, each flake developing a unique symmetry as it tumbles in turbulent air currents and encounters impurities in water vapor.

The two insights, that (1) there were states in the past that lacked any modern analogs, and (2) systems exhibit sensitivity to initial conditions, together undermine the authority of both proxies and the linear deterministic thinking that supports their use. One need not, however, abandon confidence in the principle of uniformitarianism. If even deterministic systems may exhibit unpredictable behavior, then contingency (context and history) is extremely important, equifinality cannot be ignored, and effects of scale may be paramount in understanding changes in system states in time and space (Gould 1986).

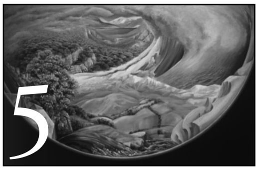
In this age of weather satellites, meteorology's failure to predict even threatening weather more than a few hours or days in advance is shown to reflect realities inherent in the atmosphere, not underfunding of technology. Retrodiction cannot be easier to achieve than prediction. There is no easy, scientific way to reverse time's arrow and observe conditions in the past. We can "know" the past only as the sum of the conditions that have determined the present, and some of those conditions will always be too small or too ephemeral to be observable.

The initial lessons from the new sciences of complexity confirm what philosophers have been saying all along about the limitations of analogical reasoning. The study of present-day phenomena and conditions can provide sound analogies for understanding the structure and mechanisms of systems of many kinds, including the planet's climate and biological cycles. However, analogies cannot be used directly to posit system states in the past or future. Theoretically informed close observation of existing systems exposes the relevant variables of each, and their interactions. This brings the observer to a better understanding of structures and mechanisms, which will lead in turn to enhanced abilities to learn about alternative states. The ability to posit numerous possible alternatives, and to reject those that are not possible permutations of things as we observe them, should bring us closer to the goal of learning about past and future that we thought to achieve by shorter, straighter paths.

Systems that are well understood in the detailed context observable today should permit some prediction, especially at small and medium scales of time and space. This is the case partly because some theoretical alternatives are precluded by immediately precedent situations, thereby reducing the number and complexity of possibilities in the short term.

Whether from the perspective of "normal" science or the new sciences of complexity, the advantages of multidisciplinary studies of complex systems are impressive. The compartmentalized disciplines of modern science have each special strengths for investigating a circumscribed range of phenomena. None can exhaust the complexities of any aspect of the world, but each can specify the likely states of some variables, and the relationships of variables within parts of a given system. Bringing many disciplines to bear on a problem, each contributing data from its special strengths, can significantly reduce the range of alternatives and uncertainties that must be considered in any given case. As complexity theory shows, what is "noise" from one perspective may be crucial information from another. In multidisciplinary investigations, it is very important that all investigators be explicit about the limitations and contradictions in their data, and about the kinds of data not gathered.

The paleoenvironmental sciences offer a vast array of methods of observation applicable to a significant portion of the variables recognized in contemporary environmental sciences. Experimental and descriptive investigations expand the available analogies that enhance understanding of crucial variables and relationships. The study of the past grows in power and insight with the study of the present, but it no longer merely holds a weak mirror up to the present. Informed by theory and experience, archaeologists in collaboration with other paleoenvironmental scientists can learn about a past that was unlike the present, illuminating both the present and the future.



5

INTRODUCTION TO CHRONOMETRY AND CORRELATION

> If there is one issue on which nearly all archaeologists can agree, it is the importance of chronology.
>
> DEAN 1978: 223

Archaeology is necessarily about change, and all change is perceived by looking back from the perspective of one's own peculiar place in space and history. The differences perceived between now and then challenge us to explain them, and we try to do that by using assumptions about the world, time, and process.

For example, consider the story of Genesis as presented in the Judeo-Christian Bible. The Bible incorporates a serious effort to explain change from a legendary Golden Age (Eden) to the world of toil and sorrow most of us experience. In asking "How did the world begin?" we are expressing our assumption that there was a beginning, as we observe with every individual life. The Bible's answer is that God created the world in six days, and then rested. The world that God created was not significantly different from the world we see around us, except that it was Good, and the experienced world is not all Good. If the world in all its diversity and complexity did, indeed, come into being in six solar days, then its existence is proof of a Creation, and a Creator. Many people take comfort in that belief. However, geological and astronomical study has led scientists to posit a slow development of life on the planet, requiring over 2 billion years to shape the planet and its biosphere as we know it today. If this is a reasonably accurate statement of how the modern world was derived, it implicates almost continuous creation, which we call evolution, marked by many beginnings and endings (Gould 1989). This example demonstrates that the amount of time elapsed constrains the interpretation of process; the implied rates of change and duration of development in the two cases *require* very different inferences.

In order to organize and analyze events in time, we require concepts and methods for measuring time – chronometry. The concepts of event, duration, and sequence are essential to chronometry; they are at the heart of historical explanation and of prehistory. Events, or small clusters of events, are dated by various chronometric techniques available to the geosciences, according to the kinds of chronometric evidence available in each case. Two events of different ages define a duration, a span of time that has elapsed between them, and they also implicate a sequence because one is necessarily older than the other. **Durations** (spans between events of known age) are measurable only indirectly; their size depends on the accuracy of the event ages. The measurement of durations is crucial to the study of processes and of life.

Sequence (order of events) is the fundamental concept in chronometry. It can be established by a series of dated events or by various other kinds of sequential relationships among events (see below). Events lack historical significance until they can be ordered into a series by some means, and durations cannot even be defined until events are ordered. Sequence, then, is essential, permitting the ordering of events and estimation of durations. With this much information – events, duration, and sequence – we control relative time. With the addition of ages, however calculated, we achieve chronometry and can think about time in defined, comparable units (e.g., years, centuries).

But what, indeed, is "time" itself? Time is perceived only through the observation of processes that can serve as rough measures of elapsed time. Like language and music, perceived time is linear, experienced only moment by moment, not all of a piece. We speak of time in metaphors, rarely stopping to wonder about the concept itself. Metaphors such as distance ("far back in time") and movement ("time flies") define our sense of time; we are seldom aware that some of the metaphors are mutually contradictory. The concept of the linearity of time is supported by chaos theory and contingency, as well as by radioactive decay, in the sense that precedents cannot be reversed, or contexts of events changed without changing the event. However, because this support is partly tautological it cannot help us evaluate the appropriateness of our metaphors. The tautology is illustrated by the observation that although the concept of time travel is a theoretical possibility in a linear paradigm, it is impossible if time is irreversible (Bailey 1983).

Because we can only think about "time" as it is experienced through some external system of measurement, all expressions of time, past or future, are relative to the rates and ratios of the measurement systems we use. We utilize decay rates, pulse rates, revolution rates, and growth rates, as well as ratios such as the proportions of sand or water in an hourglass, of radioactive to stable isotopes, and of decay products to their parent material. Contrary to some regrettable terminology frequently encountered in introductory textbooks of archaeology (the purported contrast between relative and "absolute" time), *all* time is relative. We have no absolute measurements. Relativity theory has shown that even clocks and calendars are relative cultural concepts (see Bailey 1983). Time exists only as a duration which we can measure against some process or other.

In the historical sciences, the crucial chronological data are sequential relationships and durations. Dealing with events and processes, we ask such questions as:

- Did event A precede or follow event B; that is, is A older or younger than B, or are they synchronous? These are questions about sequence.
- Did process X (or event A) continue longer than process Y (event B)? These are questions about duration.

Given answers, we can then estimate durations and rates of processes and of historical sequences, essential first steps to explanations. Estimates of rates and durations require refined chronologies composed of ordered series (sequences) of small comparable increments of time; none of these relationships requires for its understanding direct measures of calendrical or sidereal time.

MEASURING TIME

The phenomena we use to measure time (**chronometers**) have different basic characteristics that make them useful in that role. Their chronometric characteristics may be *cyclical*, *serial*, or *progressive*.

- Cyclical chronometers are essentially planetary and celestial phenomena. For millennia, people have observed cycles of day and night and of the seasons, as well as the apparent movement of stars and planets, and have kept records of their changes and reappearances. Each cycle has a characteristic duration, and few of them are in phase with each other. Astronomers today track the movements of stars and galaxies and measure their cycles against atomic time. We also track and time subatomic cycles, and use those to create chronometers of unparalleled precision. Cyclical time is best for measuring durations.
- Serial chronometers depend upon the construction or knowledge of a series of unrelated or discrete events. Thus we relate some event to another, and thereby give it an approximate age and chronological equivalence or priority in relation to

others. References to events in such contexts as "in the Kennedy administration," "in gravel lens C," "in January," "before Joe was born," are references to unconnected serial events that nevertheless represent relative time. The series must be recorded or memorized to establish its sequence, and thereby its broader continuing usefulness.

- Progressive chronometers measure time as rates or ratios of phenomena subject to regular change. The changes may or may not have finite limits. Examples of finite progressions that are good chronometers include biological aging and radioactive decay. Radioactive accumulations, chemical changes, and stylistic changes are less finite but essentially linear phenomena that provide useful measures of time.

These many different systems of time measurement are not closely compatible. Think of the differences between the regular lengths of Moon months (counted as cycles of Moon phases as viewed from the Earth) in comparison to the calendar "months" which are unequal divisions of a solar year, compensating for the fact that 12 lunar cycles are completed in less than one solar year. Similar disharmonies between the rates or intervals of any two systems of time measurement are expectable. Even two systems based on annual increments, one biological (tree growth rings) and one geological (varve sequences), may fail to produce precisely comparable chronologies, because of inherent irregularities. Radiocarbon years are well known to be not always equal to solar years (see below); the rate of **obsidian** hydration changes with temperature and therefore the amount of hydration varies per solar year, and so on with many other methods. In order to compare any two or more of the large number of chronometers available, we must convert, correlate, or **calibrate** the incremental or ratio measures of age or duration to a standard; the standard of choice is calendar years, an approximation to sidereal time.

Sidereal time ("star" time) is a convenient and relatively precise convention with considerable heuristic value. As occupants of the third planet revolving around our star, the Sun, our bodies and minds have become accustomed to it. Star time provides a way of defining and measuring increments of time that are independent of almost every other phenomenon we measure chronologically, so that every rate or process that interests us can be expressed in units of star time (day, season, year), and thereby be compared directly with any other rate or process. Because of the centrality which such time expressions have gained in our thinking about time, measures of time that are precisely translatable into sidereal years are elevated in our esteem. Archaeologists' habitual preference for "dates" over "ages" seems to derive from our socialization to sidereal years as enshrined in calendars. This cultural preference is incompatible with the imprecision inherent in most archaeological chronometers.

It is crucial to bear in mind that there are constraints upon the **accuracy** and precision of all chronometric methods. Accuracy is a statement about the degree to which a measurement approximates an abstract, absolute standard – its "truth." Precision is a statement about the fineness of resolution of a measurement, or its replicability in repeated measurements (R. E. Taylor 1987: 106). A measure can be accurate and not precise, as in "Anne was born in 1940" (the correct year), as well as precise and not accurate, as in "Anne was born at 1:34 A.M. on July 1, 1943." While the ideal is an age statement that is both accurate and precise, in both archaeology and the geosciences one must balance the ideal against the possible. An awareness of the different potentials for accuracy and precision in various measurements of time is essential to good problem definition and clear thinking. In archaeological time-measurement we can settle for accuracy without precision when we are trying to arrange things in sequence. Precision is necessary when the degree of contemporaneity, or the fine-scale ordering of events, is at issue, as is the case in detailed studies of cultural processes. Precision without accuracy is always to be avoided.

Although archaeologists borrow most of their techniques of time measurement from the geosciences, the scale of time resolution desirable for most archaeological applications is much finer than what satisfies geologists and other Earth scientists. *Geological time* is typically measured in millennia or larger intervals, so that "± a million years" still makes sense to geologists, whereas in archaeology it is difficult to use such an interval even to think about the physical evolution of hominids. Archaeologists should try to achieve time intervals at scales that have some meaning in human societies; for *cultural time*, even a radiocarbon century is a bit coarse. Given the uncertainties and approximations of archaeological deposits and available chronometric methods, time intervals of less than a century can rarely be perceived. Therefore, problem formulations that require the *social time* of decades rarely fit archaeological data sets at all well. Archaeological time, in the context of the geosciences, requires high-frequency **resolution** (the potential for precision; the property of being reducible to equivalent units); that in turn entails pushing the available technologies to their limits in rigorous and imaginative ways (Table 5.1).

CALIBRATION

Calibration is the expression of one kind of measurement in terms of another, to establish equivalences. For example, measurements based on rates and ratios are expressed relative to various other processes independent of them; the preferred calibration standard is sidereal time. All chronometric methods have inherent uncertainties, and some inherit uncertainties from others with the calibrations. The uncertainties and ambiguities of chronometric methods are such that the best results

Table 5.1 Exponential scales in time with chronometric methods of appropriate precision

Temporal scales	Duration or frequency (a = yr)		Chronometric resolution
Mega-	$>10^6$	(>1 ma)	K/Ar; paleomagnetism
Macro-	10^4–10^6	(10 ka–1 ma)	^{14}C; TL; ESR; K/Ar; fission tracks; U-series; paleomagnetism; obsidian hydration
Meso-	10^2–10^4	(0.1 ka–10 ka)	^{14}C; dendrochronology; TL; ESR; U-series; obsidian hydration; fission tracks; varves; archaeomagnetism
Micro-	$<10^2$	(0.001 ka–0.1 ka $<$ a century)	calendars; dendrochronology; ice layers; archaeomagnetism; (radiocarbon indeterminacy)

will be obtained when it is possible to use more than one technique to estimate the age of a particular event.

Each technique brings its own particular uncertainties and limitations, and its own calibration problems, but in combination there is strength. Even if chronometric results in a suite fail to correlate precisely, they make possible a rough estimate of the error factors involved in the comparisons (Aitken 1990; Betancourt 1987; Browman 1981; Tooley 1981). Archaeologists typically have preferred simple, single answers about age to multiple estimates that force evaluation of error sources, even when the first choice conceals large errors. We consider calibration further in the sections on particular methods.

The accuracy of clocks and calendars is achieved by definition; units are not precisely coterminous with any natural process, but are cultural conventions. We are reminded of the non-equivalence of the Hebrew, Chinese, and Roman-derived calendrical systems by colorful New Year's celebrations in different months. The elaborate and precise Mayan calendar, based on several astronomical cycles, "floated" unlinked to the conventions of Western calendars for many frustrating decades of research. Calendars themselves require calibration. The familiar annual calendar used in the Western world must be adjusted every four years by the addition of an extra day to keep it synchronous with the Earth's positions in its orbit around the Sun.

Cross-dating and correlation

The most desirable archaeological data are those that permit the recognition of discrete events, rather than palimpsests of the remains of many commingled events.

The goal in cross-dating and correlation is to relate an event in one context to an event in another. Jeffrey Dean defined a typology of events that is helpful in thinking about dating in archaeology (Dean 1978: 226–228). He calls those events of ultimate archaeological interest "target events"; these are the cultural or behavioral events whose age we wish to know. The event that provides the information about age is the "dated event," such as the death of a tree for tree-ring or radiocarbon dating. Fundamentally important to the dating enterprise is the conceptual distinction between the target and dated events; they are rarely identical. Only these two of Dean's four-class scheme will be used here.

Environmental reconstruction requires accurate correlations and cross-dating in order to compare events recorded in different universes of data. Again, relative age (sequence) is the critical information needed to establish chronological equivalence or priority. When events are separated in space, and therefore cannot be directly compared in terms of a single related event, cross-dating is required. Comparisons across space are facilitated by chronometers that are easily calibrated, such as radiometric techniques, tree rings, volcanic ash falls, or cultural associations (all of which represent relatively small spans of geological time). Calendars are important only as devices for comparison and calibration – for relating past events to our present sense of time.

The correlation of archaeological or geological deposits from one exposure to another can be accomplished by several different means. The best is continuous exposure, with sections and plans connecting two or more areas, but this is rare in archaeological excavation even at the site scale. Depositional units can be correlated by comparing the sedimentary structure, texture, chemistry, and the included fossils and cultural materials between the individual units (Part V). Because sediments have lateral extent, all of the above characteristics can change within a sedimentary unit that nevertheless retains integrity, thereby introducing a major complication. The geological concept of **facies** is helpful here – change of depositional environment within the same depositional episode. Note that adjacent archaeological profiles separated by baulks technically require correlation or cross-dating; they cannot automatically be considered equivalent while they are discontinuously exposed.

Artifactual cross-dating has a long and honorable history in archaeological chronologies, although its pitfalls are only now being fully realized (see the case study on pp. 132–135). Fundamental to the method is the demonstrable and repeated observation that styles of artifacts change through time. Objects of roughly the same age are most alike; the longer the time span between the production of two items, the less similar they are likely to be. On the assumption (rarely demonstrated and equally rarely tested) that items are reliably incorporated into archaeological deposits shortly after the time of their manufacture, archaeologists infer the synchroneity (coevality) of deposits that contain similar artifacts. The development of chronometric methods independent of artifacts has exposed the uncertainties, and sometimes the circular logic, on which long-distance correlations were assumed and cross-dates asserted (e.g., Hardy and Renfrew 1990; A. C. Renfrew 1973).

Similarly, the assumption that similar pollen assemblages can be used to establish the contemporaneity of stratigraphic units separated by space carries its own pitfalls, but has proven to be a valuable foundation for working hypotheses. The more complex and unique the pollen assemblages compared, the stronger the basis for the inference of equivalent age, with the additional caveat that the strength of the inference is negatively correlated with the spatial distances involved in the cross-dating.

Faunal materials offer special possibilities for cross-dating, and carry their own particular sources of error. The distance mobility of animals is related to their size, their environmental tolerances, and their means of locomotion; their usefulness for cross-dating cannot be separated from those characteristics. The presence or absence of faunal species in archaeological deposits rarely can provide chronological resolution within a millennium. Faunal cross-dating is most useful in the broad expanses of geological time, when millennia matter little, or in those instances of introduced species when the time of introduction can be reliably known. The case of the sixteenth-century Spanish introduction of horses to the Americas is an example of such an event; the introduction of the European rat to the New World, or to the British Isles by Romans, are other potentially datable events that can cross-date archaeological deposits.

Volcanic ash (**tephra**) may be carried by wind or water for great distances from its sources in volcanic vents; indeed, when ash is injected high into the atmosphere, it may travel around the globe several times before settling to Earth. Individual ash sources, and sometimes the products of individual explosions, are identifiable by their chemical or physical signatures. Their appearance in stratigraphic contexts permits cross-dating by "tephrochronology." Ash fall events are datable by means of radiocarbon, dendrochronology, thermoluminescence, potassium–argon, obsidian hydration, or fission-track dating (see below). When dated ash falls can be identified at distant points, they carry direct chronometric information and provide cross-dates at the event scale, sometimes less than a year.

Archaeologists and geoscientists have available a wide range of methods for determining age. Choice among them must depend upon the materials available and the problems addressed. The following brief summaries touch upon several kinds of chronometers that measure rates and durations in past time. All but the first group, seasonal clocks and biological rhythms, give coarse resolutions for historical problems at the human scale.

SEASONAL CLOCKS AND BIOLOGICAL RHYTHMS

Time measurement methods derive ultimately from the natural cycles of the planet – the alternation of day and night and of the seasons. Organisms of the biosphere are sensitive to all these normal dynamic conditions of life. When biological responses to these cycles accumulate without interruption, they can be interpreted to measure duration and sequence at time scales familiar to us as fellow organisms of the planet.

Dendrochronology

Dendrochronology is the most precise chronometric method available to archaeology (Baillie 1995; Dean 1978), providing dates closely correlated to sidereal years, sometimes even seasonal fractions of years. The method is based upon recording the proportional widths of annual growth rings in climatically sensitive trees and then matching pattern sequences from tree to tree back in time from a known year. Chronosequences are created by matching "signature" sets of rings between trees of different ages, to extend sequences back into the past (Fig. 5.1).

Trees grow outward in annual increments, adding a sheath around the entire tree just beneath the bark. In temperate zone trees, "early wood" composed of large cells is emplaced by rapid spring growth. The "late wood" near the end of the growing season has thicker and tougher cell walls. Trees are dormant during the cold or dry season because growth is prevented by the fall of leaves or other mechanisms such as insufficient water or low temperatures. The following year's early wood normally presents a visible contrast to the cells from the end of the previous growth season.

Growth stressors

Ring widths vary according to (1) the growth conditions of the environment and (2) the age and (3) size of the tree. In years with abundant water and sunlight rings are relatively wide. Rings are narrow on young shoots, wider afterward, and narrower again on large mature trees, because of the changing ratio of leaves to wood mass, and therefore food available for growth. Ring widths also narrow with height, as distance from the roots increases. Trees growing in conditions where light and moisture vary little from year to year develop rings closely similar in width, called "complacent" rings. In the American Southwest, where dendrochronology has long been an

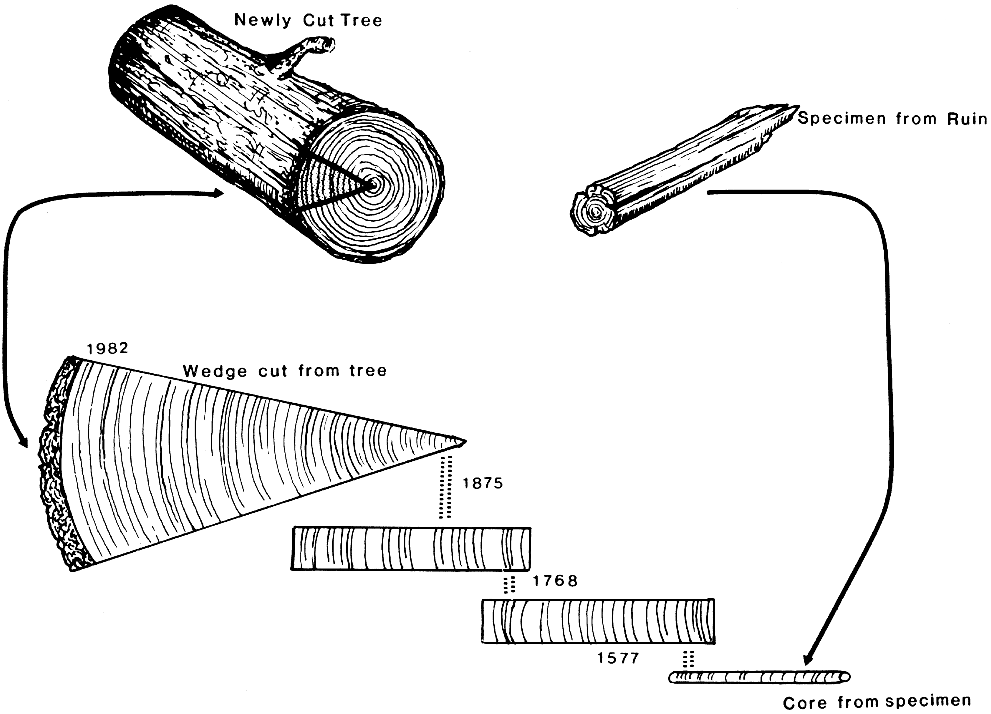


Figure 5.1 Schematic of chronosequence building in dendrochronology. See text for explanations. (Reproduced from Cordell 1984: Fig. 2.4 with permission of the illustrator, Dr. Charles M. Carrillo.)

important component of archaeological research, trees receive abundant sunlight; they are sensitive to (stressed by) varying rainfall. Elsewhere, inadequate sunlight can be a significant stressor, especially at high latitudes, where it is expressed as seasonal differences in insolation and temperature. The construction of dendrochronological sequences requires trees that respond to stressors at the regional scale, so that matches can be made over distance (Grissino-Mayer 1995). Very local stressors, such as earthquakes, slope changes related to landslip, and insect defoliation have subregional effects that must be interpreted to permit correct reading of ring sequences. Insect defoliation can halt growth in midseason, giving a narrow ring or a split ring if growth resumes later in the summer.

Tree death, the termination of growth of an individual tree, is the main dated event for dendrochronology. It is best marked by wood with bark adhering, secondarily by the **sapwood** directly beneath the bark. The relationship of tree death to any archaeological target event must be independently established (Baillie 1995; Dean 1978).

Methods

Tree-ring records are collected as wedges, slabs or cores, taken perpendicular to the bark. The surface to be read is planed or trimmed with a razor before counting. Ring-width plotting is the preferred method in arid zones (Fig. 5.1). Thin-sections are prepared for **X-ray densitometry**, a promising new technique in cool, wet environments where ring growth is typically complacent and the wood is waterlogged. Densitometry measures the varying density of cellulose in rings.

Counting is not a simple, straightforward matter. Rings may fail to form, or be doubled, under some conditions of stress. Wetland species are prone to ring complacency. We have noted that ring width varies over the life of the tree, independent of the regional variance, so that ring widths cannot be directly compared from tree to tree. Ring-matchers must disregard the actual widths of rings in favor of recording their relative widths and patterned clusters; a variety of techniques have been used, with electronic ones now dominating. The heart of the method is pattern-recognition; the actual physiological mechanisms creating the contrasting widths are complex and poorly understood. Short-term regional climatic variation is the basis for the patterned signatures that permit matching.

Ring matches are done not on the wood itself, but usually by scanning computer. Ring records are normalized to screen out life-cycle variation in order to create data of regional relevance (Baillie 1995; Parker et al. 1984: 216–217). Long-term regional trends are of interest mainly to climatologists (see Chapter 8); they may be statistically isolated in order to facilitate the construction of supraregional chronologies.

Regional chronologies

The time spans applicable for dendrochronology are controlled by regional preservation conditions, of which there are many kinds: roofed buildings; arid climates; saturated environments (bogs, lakes and rivers, ocean, anaerobic alluvium); charred wood; chilled wood (e.g., Arctic, high elevations). The ideal situation for establishing dendrochronologies is the recovery of multiple specimens of species sensitive to regional climatic variation, as exemplified in the arid Southwest and in the bogs of Western Europe and the British Isles.

Regional chronologies have been established in many parts of the temperate zones (Baillie and Brown 1988; Becker 1993). The first and for years the longest was the bristlecone pine sequence from the White Mountains of southern California, over 8000 years long. The oak chronologies of Ireland and Central Europe have now surpassed it, with over 11,000 years recognized in some areas (Becker 1993). The famous archaeological chronologies of the southwestern United States consist of many small regional sequences of varying lengths, few longer than a thousand years.

The southwestern American sequences are compiled from piñon pine and Douglas fir timbers from ruins, and bristlecone pine from mountain slopes. Red cedar is valued in the eastern United States because of its preservation and ubiquity, but it is relatively complacent (Stahle and Wolfman 1985). West European sequences use mainly oak, which is also preferred for sequences compiled from old furniture, building timbers, and painted wooden panels.

Applications

Dendrochronology is a useful tool for architectural history and art history, where its use is fairly straightforward since the specimens are usually closely related to the target event. Archaeological applications are often more complex and demanding. The target events are typically construction dates and use-spans of structures; for these applications, original timbers and repair timbers are sought, whose dated events are times of death (cutting). In dry and arctic climates, where dead trees may be useful construction timbers and fuel over many decades, problems proliferate. As for radiocarbon dates, the contextual integrity and cultural relevance of the sampled rings must be demonstrated on site (Baillie 1995; Dean 1978; Parker et al. 1984), and include sapwood to maximize precision and association.

Varves

The chronometric advantages of both stratigraphy (see below) and tree rings are combined in **varves**, annual increments of bottom sediments in certain lakes of the arctic and temperate zones. Annual laminations of lake-bed sediment, with typically a lighter-colored increment deposited in spring and summer and a darker layer in the winter, are usually thoroughly mixed by bottom-dwelling (benthic) invertebrate animals and by vertical circulation of water. However, lakes deep enough for the bottoms to be below the seasonal turnovers, and with oxygen deficits severe enough to inhibit benthic fauna, retain the laminations. Paired laminae may be formed in different ways, each of which produces a variant of the summer:winter contrast in sediments that receive a spring–summer influx of coarse material such as diatoms, calcareous or iron-rich sediment, and mineral sediments. The dark winter layers are formed by fine organic detritus deposited from suspension while the lake is sealed under ice. These seasonal rhythmites, or varves, retain the associations of all the material settling to the bottom of the lake, and offer annual resolution within the deposits, rather as tree rings do. Series as long as 13,000 years have been compiled (Stuiver et al. 1986).

Varves are utilized for fine-resolution chronology by palynologists and limnologists. Under ideal conditions, varve widths can vary as a climate signal with wet or warm summers giving thicker varves, permitting correlation of sequences from lake to lake. In this way, varve sequences have provided significant chronologies for Late Glacial events in Europe and North America.

Varved sediments, formed in the centers of deep lakes, rarely include archaeological materials, but nevertheless can be very informative about human activities in the watersheds. A notable example is Crawford Lake, Ontario, where maize pollen recognized in a core led to the discovery of significant late-prehistoric Iroquoian occupations on the windward side of the lake, and permitted the close dating of the maize, of woodland clearance episodes, and the integration of the pollen-core data with human activities (McAndrews and Boyko-Diskonow 1989). Studies of a varved lake in Finland have revealed a detailed picture and chronology of several episodes of land clearance and farming (cited in Saarnisto 1986). Varve studies produce high-resolution data on such environmental changes as fire frequency, disturbance by land-clearing, and eutrophication (oxygen depletion by phosphate enrichment) of lakes.

Ice accumulation banding

Analogous to varves in the manner of accumulation and study is the annual banding in glacier ice, observed at high latitudes and altitudes (Baumgartner et al. 1989). Deep cores drawn from the Greenland and Antarctic ice sheets, and from mountain glaciers in Peru and China, show annual banding defined by seasonal differences in accumulation on the snow fields. As the fallen snow turns to ice, the bands retain sufficient integrity to be traced to great depths in the glacier. Atmospheric dust and other contaminants permit some radiocarbon dates for calibration, with sufficient success that now, in heavily studied cores, reliance is placed not on radiocarbon but on counts of annual bands, which are defined by several different criteria (Hammer 1989). Such real-year counts make the ice-core chronologies potentially synchronous with **astronomical time**.

Volcanic dust layers and other unique event records support extension of ice dates to special cases far removed from the glaciers. The principal phenomena dated are changes in atmospheric circulation and temperatures, which reflect climate relevant to large areas (e.g., Mayewski et al. 1994; Thompson et al. 1990). Precision to single years is claimed in special circumstances, increasingly supported by multiple criteria isolating annual accumulations and by cross-checking adjacent cores (Thompson 1991).

STRATIFICATION AND STRATIGRAPHY

> The only truth is stratification.
>
> D. F. DINCAUZE

In archaeology, as in most of the field sciences, the basic method for establishing and measuring chronological relationships is **stratigraphy**. Stratigraphic relationships, properly handled, are the least ambiguous and arguably the most accurate of all time measurements available to archaeologists – they represent sequence, irreversible order. In addition, they are capable of relating to each other highly disparate kinds of data – in fact, the entire range of data that can occur within sedimentary matrices. Thus, mineral sediments, soils, pollen and plant macrofossils, micro- and macrofaunal remains, artifacts, features, and human remains can be related chronologically to each other to the extent that they co-occur and can be observed in the same stratified sediments. Co-occurrence of this sort is the firmest basis on which to establish sequence and synchroneity of data, and therefore of the events represented by those data. For this reason, "stratigraphic analysis . . . is the starting point for all palæoenvironmental reconstructions" (Tooley 1981: 47). It is crucial, therefore, that some fundamental concepts of stratigraphic analysis be established clearly right at the beginning.

Stratification is the record of past events, processes, and states preserved as phenomena and relationships in sediments. Stratigraphy is a record of the interpretation of stratification by a stratigrapher. These distinctions, although often ignored by archaeologists, are crucial, because no effort to understand can succeed if data are confused with interpretation. Stratigraphic plans and profiles, which are recording devices, encode the interpretations given to observations made in the field; they cannot be considered or interpreted as if they objectively represent field phenomena. As recording devices, they represent what observers were aware of seeing, and what they thought about their observations, including judgments about relevance, priority, scale, order, and diversity. Photographs record what the photographer aimed at, with technological limits on phenomena recorded; color and texture are not reliably recorded on film. Stratigraphic records reduce three-dimensional phenomena to two-dimensional pictures or to the linearity of language. By their nature such records must be selective, and therefore they represent an incomplete record of the complexity of field situations.

There is no reason to be apologetic or defensive about the fact of an interpretational screen intervening between phenomena and record; what is crucial to field science is an awareness of this screen, and a constant, conscious effort to make it as transparent as possible. To this end, archaeologists, like other field scientists, need to become self-conscious about their roles as stratigraphers – selectors and recorders of data. The simple acceptance of the principle that younger sediments overlie older ones does little justice to the complexities of either field situations or the translation of observation to interpretation within human minds (Ager 1993; Barker 1993; Stein 1987). Geologists have developed international standards for stratigraphic conventions and terminology (nomenclature), ratified by practicing scientists and observed by all who wish to communicate clearly. The terminology is changed and developed with need. Archaeologists should be aware of these standards, should use them when appropriate, and should refrain from arbitrary reinvention of stratigraphic terms and concepts (Farrand 1984; Gasche and Tunca 1983).

However, a basic distinction in stratigraphical practice must be made between geological and cultural stratification. The difference derives from the agents of transport and deposition, as well as from the environment of deposition. Discriminating between geological and cultural deposits is not always a simple matter, but it is crucial to archaeological interpretations (Stein 1987). In many sites, the two kinds of deposition and transformation processes may have alternated through time but, clearly, they must be distinguished if the sequence of events at a site is to be understood. Stratified archaeological sites normally involve both kinds of deposition, as, for instance, when human occupational debris is buried by river flood deposits, sand dunes, hillside slumping, or cave deposits. Outside of urban contexts, it is rare to find deposits that are entirely products of human transport and deposition. Cultural materials are more typically embedded intricately within natural deposits, and their interpretation entails understanding the relationship between the geological and cultural events that combined to create the stratification.

Stratigraphy

Stratigraphic units include *deposits* that are the result of discrete depositional events and processes, *erosional or constructional interruptions* (unconformities, pits, and walls) and the *interfaces* they define, and the lateral *gradations* that indicate a change in depositional environments over space ("facies" to geologists). Each of these units must be understood in three dimensions and recorded in such a way that the dimensional relationships can be understood by someone who did not see them *in situ* (Harris 1989). Because sedimentation is typically episodic, stratification is likely to be incomplete in the sense of not recording an unbroken sequence of events. Episodes of non-deposition and erosion separate episodes of deposition. Any assumption of steady rates of deposition must be tested, since rare and large-scale events are important in compiling the sedimentary record.

A deposit is an irreducible component of stratification, the result of a discrete event or episode of accumulation. The discreteness of a deposit must be observed, evaluated, and demonstrated *in the field*; it cannot be unambiguously reconsidered after excavation. Interfaces, including the interfaces that bound archaeological features whether destructional (pits) or constructional (walls), are equally important units of stratigraphy. The relationships of interfaces to all their contiguous deposits must also be clarified in the field, and recorded explicitly. The process of excavation, ideally, isolates every deposit and interface descriptively. There are many correct ways to do this, depending upon the nature of the deposits and the goal of the excavator, but there are an even greater number of incorrect ways to excavate and record archaeological phenomena. Progress has been made in rationalizing techniques of excavation and recording (Barker 1993; Harris 1989); everyone going into the field should be familiar with the most recent discussions of methods, in order not to repeat old errors. The conventional square, flat-floored archaeological excavation units usually include parts of more than one deposit. Unless the deposits are carefully discriminated in the field and recorded in standardized formats, there is the real danger of conflation of deposits and excavation units, resulting in confusion regarding the formation of deposits and their chronological relationships. Vigilance is required to clarify relationships in the field at every opportunity. Interpretation should be nearly continuous, in order that conflation be recognized and corrected as soon as possible.

Of course, neither depositional units nor excavation units can be assumed to be coterminous with units of cultural activity at archaeological sites, or to represent event-scale phenomena. Interfaces may sometimes have that quality, but deposits may occur over units of time varying from the instantaneous to the millennial. When a deposit or interface represents an episode of construction (walls, floors, etc.) or destruction (e.g., pit digging), the relationship between cultural event and depositional event may be very close. However, with natural deposits, or mixtures of cultural and natural deposits, the relationship must be teased out for each case.

The contents of individual deposits may vary greatly, even when the sedimentary materials themselves vary little. Deposits may include redeposited materials unrelated in origin to events or circumstances contemporary with the deposit itself. Cultural materials may be displaced across depositional boundaries by postdepositional processes, moving up or down across sedimentary units. Thus, the content and the matrix must be separately analyzed and their congruence evaluated (Barker 1993; Gasche and Tunca 1983; Harris 1989; Villa and Courtin 1983). Ideally, this is done without logical circularity so that the interpretations can be complementary rather than dependent.

Sampling in stratigraphy

If stratigraphic relationships are to yield their store of chronological information, they must be properly recorded; a crucial aspect of recording is sampling (Chapter 2). Samples are isolates taken from stratigraphic contexts in order that they may be minutely scrutinized and subjected to special analytical techniques. Sampling strategies must be integral parts of the research plan from the beginning. Everyone with responsibility for the field work and laboratory study (that includes everyone involved in the project) must understand what samples are needed, why they are needed, and how they are to be selected and handled. Materials removed from the field lose their representational value unless the purpose of sampling is reflected in the choice and handling (labeling, packaging, and storing) of samples.

Samples must be *adequate* for the intended analytical purposes (in size, target, and frequency), *discrete* (representing isolated sedimentary units), and carefully and fully *documented.* It follows that consultation between laboratory technicians and the field personnel is the best basis for field practice in sampling. The adequacy of a sample depends upon its being representative of the phenomena of interest, a relationship that can only be determined in the field (Fig. 5.2). It must be large enough to support the analytical techniques to which it will be subjected, and there must be a sufficient number of samples to assure comparability among the various site contexts.

Samples taken for any purpose are best when selected from within discrete stratigraphic units; the analytical resolution of a sample is compromised whenever stratigraphic boundaries are crossed. This concept will be encountered in various guises in the substantive chapters that follow (the exceptions for micromorphology are discussed in Part V). Discreteness has important implications for the kinds of interpretations that can be made from sample studies, and the problems may be very subtle. An example from pollen zone chronology exemplifies some of the complexities: "When comparing pollen zone boundaries that have been dated radiometrically, there is the problem that sample thicknesses [from cores] have varied, thereby increasing or decreasing the age range of material dated" (Tooley 1981: 17). Samplers selecting materials for dating must be alert to identify the range of contaminants that may affect the intended dating techniques, and for indications of stratigraphic disturbances that may have displaced materials. The discreteness and integrity of any sampling site is best considered a working hypothesis to be scrupulously tested both in the field and in the laboratory.

In interdisciplinary studies, field sampling requires close collaboration between specialists in the several disciplines, in order that samples be relatable across disciplines. Specialists in the different disciplines should be in the field together in order

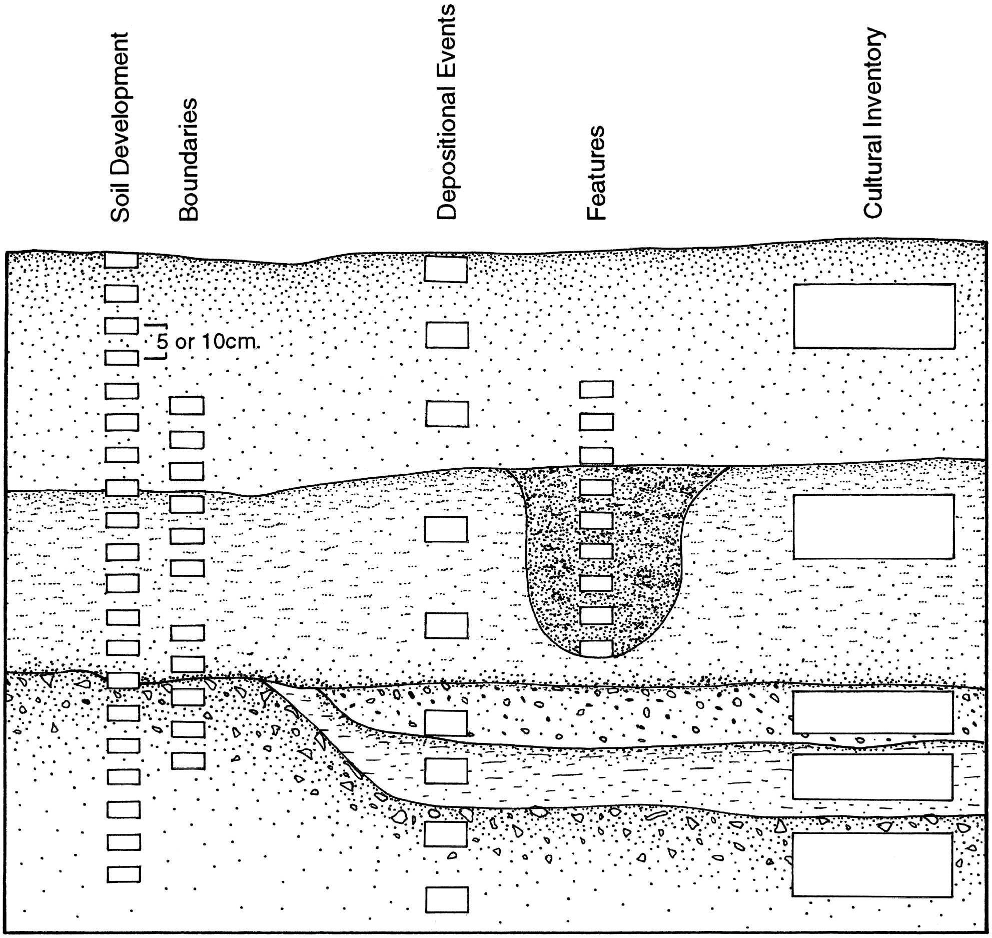


Figure 5.2 Sampling locations on an archaeological section. Five different sampling strategies are exemplified: collecting data for soil development, depositional boundaries, depositional events, constructional features, and cultural inventories. Note that each strategy requires special sample sizes and locations. The rectangles are proportional to the sample sizes needed for different purposes; the smallest samples require ca. 200 g. The soil development sample column on the left represents the case of a single deep soil formation with multiple cultural surfaces. The buried soils shown on the several surfaces each require its own sample series at smaller scale. (After Stein 1985: Fig. 1.)

to take samples closely and significantly related to each other (splits; adjacents), and to relate the different samples to the microstratigraphic controls (Fig. 5.2).

CHRONOMETRY BASED ON DIAGENETIC CHANGES

Diagenesis, the suite of chemical and physical processes by which organic and inorganic materials become rock, provides a number of ways by which relative and chronometric ages can be estimated. The rates of the several processes are closely dependent upon immediate environmental conditions and long-term climatic states, and so all must be calibrated if they are to yield information about time.

Organic and inorganic changes during diagenesis

The mineral and organic fractions of bones and shells (discussed further in Part VII) make possible a range of chemical dating techniques that can establish the relative ages of bone and shell within deposits, and sometimes between them. Such relative dating techniques are based on exchanges between buried organic tissues and minerals and various constituents of the sedimentary matrices surrounding them.

Bones buried in the earth are either leached and disintegrated into their constituent molecules and elements, or they become fossilized by mineral exchanges with groundwater. The rates by which these processes occur vary according to the density of bone, the duration of burial, the availability of groundwater, the temperature regime, and the chemistry of the soils and groundwater. Bones in the same place, buried for comparable lengths of time, should have roughly comparable rates of mineral exchange or disintegration. This principle is used to establish relative ages of bone materials in cases where other methods are not applicable, or where controversy surrounding finds has justified or required multiple avenues of research.

Fluorine and uranium, widespread in natural sediments and soils, are carried by groundwater and exchanged for some of the original constituents of the bone mineral hydroxyapatite. Their concentration can be measured by a variety of chemical and physical means and compared with that of other bones in the same deposits. Strong contrasts between bones in apparent association within a deposit imply that some of the material was added at different times, subsequent to deposition of the matrix. The infamous Piltdown Man hoax was exposed in this way (Weiner 1955).

The loss of organic components of bone, essentially the collagen, can be measured roughly by the remaining constituent nitrogen. Highly variable from one context to another, and from one tissue to another, such loss is a progressive and irreversible process, the development of which can be compared to other material to approximate relative ages.

These chemical methods are at their best in assessing relative contemporaneity, exposing mixtures of materials, and sorting out difficult stratigraphic situations. In such applications they are comparable to aminostratigraphy and obsidian hydration techniques (below). However, they are less suitable for calibration, since rates of change vary with the original material, age, temperature, groundwater, and other highly variable conditions of the soils. They cannot, therefore, provide independent rate-based estimates of calendar time.

Amino-acid racemization dating

The proteins that comprise living organisms are themselves made up of chains of amino acids, of which there are many kinds. Amino-acid molecules are built around a central carbon atom, with other atoms or atom groups attached to it at various angles in three dimensions. Most amino-acid molecules are capable of asymmetry: they can occur as mirror images ("stereoisomers"), with left ("L") and right ("D") versions that can be distinguished in polarized light. In living organisms the L form predominates; over time after burial the molecules change to the D form by a process called **racemization**. Thus the ratio of D to L forms ("isomers") increases with time, tending toward equality, and can be measured by gas or liquid chromatography. More complicated amino acids with larger molecules undergo a related change, producing converted forms that differ in some physical properties from the originals; these can also be measured as ratios that increase in time (Rutter et al. 1985; Wehmiller 1984). However, the changes are not simply dependent on time: temperature strongly affects racemization rates, and other complications are involved (Aitken 1990: 204–214; chapters in Hare et al. [1980]).

Because time and temperature are major factors in these chemical changes, isomeric ratios and racemization rates are utilized as measures either of time since death of an organism, or of the average temperature conditions affecting the buried remains. For these purposes, it is fortunate that different amino acids have different racemization rates. The amino acids most used are aspartic acid, leucine, and isoleucine; their different conversion rates result in different potential dating spans. Furthermore, the several rates may be used as internal checks against each other where rates overlap. The potential for amino-acid racemization dating is best in the meso- and macro-ranges, limited by contamination or the attainment of D:L equilibrium (Fig. 5.3). Recent work in shorter time spans shows significant promise at decadal scales in special conditions (Hare et al. 1997: 281–286). The analytical method requires only a very small (but

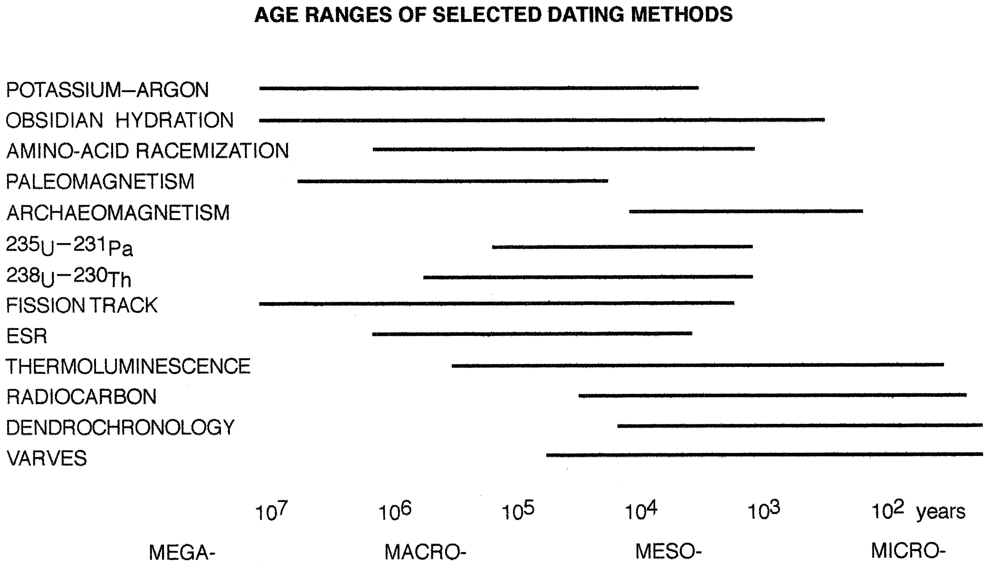


Figure 5.3 Age ranges of selected dating methods. These ranges are subject to change with technical development. The spans shown are approximately those of Aitken (1990).

representative) sample of organic material, ca. 5 mg, depending upon the material employed. Laboratory comparisons have demonstrated a reasonably high level of replicability (precision) in the analyses involved to establish D:L ratios (McCoy 1987).

The application of amino-acid racemization to the dating of organic materials such as bone and shell is a secondary application of an analytical method that is mainly used to establish paleotemperatures. The method is quite successful in estimating the average temperature conditions for a buried sample of organic material, once the age of that material can be determined independently. For dating applications, however, departures from the simplifying assumptions of the method introduce serious difficulties.

The racemization rate is in fact a product not only of the average temperature but of temperature extremes; the process may be slowed almost to nothing by very low temperatures, and it is accelerated by high temperatures. A difference of only ± 2°C in the temperature estimate can reduce the accuracy of an age estimate by as much as 50%. Consequently, estimating a temperature average in order to solve for time does not give good results. It is now recognized that the depth of burial is also an important factor, since it involves heat sensitivity; burial depth, of course, can change over time due to disturbance, deposition, or erosion.

Experimental work has indicated additional factors affecting results. Microorganisms introduce D-amino acids into buried material, and soil bacteria can influence racemization rates in collagen. Groundwater conditions (very difficult to establish for antiquity) will affect the material by leaching, and contribute to the definition of the temperature ranges. Rates vary with the type of organic tissue involved (bone, tooth, mollusk shell, eggshell) and differ among genera and species as well. That means that any calibration attempt must employ closely comparable material. Additionally, the molecules themselves convert at different rates depending upon their positions in the protein chains during diagenesis. Molecules bound on or within amino-acid chains convert at a faster rate than unbound molecules.

The problems may yet be overcome by careful comparative and experimental research. Geologists are achieving good results using racemization rates of mollusk shells to establish *relative* chronologies in order to check or cross-date stratigraphic sequences. Aminostratigraphy might supplant the more traditional chemical relative-dating methods using diagenetic changes in organic materials (see below). Aminostratigraphic applications could be useful in sorting and correlating discrete archaeological deposits, once the difficulties are overcome. For instance, amino-acid racemization analyses have reduced some of the uncertainties in correlating deposits within limited space, as among the several shell-rich archaeological middens at Klasies River Mouth in South Africa (refs. in Hare et al. 1997: 285). There, spatial separation of the middens precluded the establishment of relative ages by stratigraphic criteria.

Archaeological deposits and materials bring special problems as well as opportunities to racemization dating. The most favorable geological matrices are those that were rapidly buried and subsequently undisturbed. Archaeological matrices, in contrast, begin as superficial terrestrial deposits, subject to heating, mixing, complex chemical environments, and contamination. The method's promise for archaeological applications lies in the optimal time range, within 10^4–10^6 years, extending into the Middle Pleistocene beyond the limits of radiocarbon. If accuracy can be improved, racemization analyses might solve some problems of studying human development in the Middle Paleolithic. Regrettably, however, research applications on hominid specimens have been unsatisfactory. Bone has proven to be a poor material for racemization analysis because it is more subject to leaching and contamination than are more suitable carbonaceous materials such as various shells, both molluskan and avian. For instance, ostrich eggshell has unique chemical and physical properties that appear to make it an unusually appropriate material for amino-acid racemization dating (Brooks et al. 1990). Good results have been achieved with ostrich shells in Africa, leading to expectations for successful applications of the method to eggs of other bird species.

Obsidian hydration

Obsidian is a volcanic glass formed when molten silicate rock is chilled quickly upon exposure to air. Because it is not crystalline, obsidian is capable of taking water into its mass through a process of diffusion (adsorption), which is dependent on temperature. Water increases the bulk of the material, causing mechanical strains throughout the zone of diffusion, which progressively widens with time. A sharp diffusion boundary characteristically separates the saturated from the unsaturated zone. Rock surfaces of different ages thus have rinds of different thicknesses, varying with the age of the surface's exposure by fracture. The relative thicknesses of hydration rinds provide a measure of time elapsed since the exposure of the surface. At some point the saturated zone will spall off the mass of the rock, limiting the method's application in geological time.

In thin-sections cut perpendicular to the surface of the rock, the width of the diffusion zone can be clearly seen and measured under polarized light microscopy because the strains change the refractive index of the glass. Analysts measure the thickness in microns (μm: 1 millionth of a meter, 1 one-thousandth of a millimeter). The measurement technique itself is relatively quick, inexpensive, and precise, making this method useful for a number of purely archaeological applications, most especially for relative dating within sites with long occupation spans or complicated stratigraphy (Aitken 1990: 214–218). Where applicable, the method is superb for establishing sequences of artifacts or deposits.

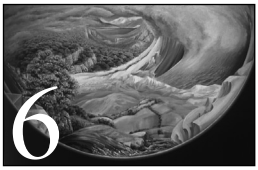
The thickness of the hydration rinds can be used for chronometric dating of events (the exposure of a fresh surface) after calibration with stratigraphically associated objects of known age, so that a site-specific hydration rate can be calculated. Calibration has been variously based on the Egyptian calendar, radiocarbon, potassium–argon dating of the source lava flow, and dendrochronology. A successful calibration requires that (1) the association between the obsidian and the calibration standard is correctly known; (2) there is a suitably large range of rind widths within the collection, so that differences are significant; (3) the specimens are of known chemical composition; and (4) there is a suitably large number of samples (60–500) to avoid small-sample bias. Rates calculated from calibration necessarily incorporate whatever errors are characteristic of the calibration standard; the rates can be no more precise or accurate than the technique providing the calibration (Friedman et al. 1997).

As hinted above, the hydration rate is not simply a factor of time. It varies significantly with temperature, humidity, and the physical and chemical properties of different glasses (Ridings 1996). Therefore, no rate can be transferred from one glass to another, nor can a rate be extended over a very large area. Furthermore, the fact that the rate will vary with temperature, over the full range to which the glass has been exposed, introduces a strong unknown into the equations which solve for time, because paleotemperatures are notoriously difficult to determine and will vary additionally with depth of burial. This problem is comparable to the difficulties inherent in dating by amino-acid racemization. Research continues to refine a method for determining rates of diffusion directly, using samples in controlled laboratory experiments (Friedman et al. 1997; Mazer et al. 1991; Stevenson et al. 1996). The fundamental research on establishing diffusion rates has been overwhelmingly empirical; the rates are not yet understood as processes in contexts, so that research so far has not fully clarified the theoretical relationships between temperature, humidity, time, and composition of the natural glasses.

Where hydration rates have been established for glasses of known composition, the optical measurement technique provides inexpensive relative and calibrated ages for specimen series and deposits. Optimally appropriate for use within the range of Holocene temperatures, hydration measurement may be extended into Pleistocene ages in the subtropics. The method has proven itself in obsidian-rich areas such as Mesoamerica, and has important potential for East Africa, the Near East, and the Pacific rim.

SUMMARY

All dating methods are relative to some measure of time. As long as we choose to privilege astronomical time, dating based on rates or cycles must be calibrated to that master chronometer. Methods that yield what are sometimes called "absolute" time are those that can be closely calibrated to astronomical calendars. In the future, if atomic time supplants astronomical cycles, all methods will be recalibrated, their relative status thereby affirmed. The choice of a dating method will vary with opportunity and the scale and nature of the events to be dated. Whatever choices are made, they must be supported by appropriately precise sampling in the field. "I would strongly argue . . . that very precise relative dating with only a rough idea of absolute time is usually of much greater importance than absolute dating per se" (Wolfman 1990b: 344).



6

MEASURING TIME WITH ISOTOPES AND MAGNETISM

> Archaeological entities, processes and explanations are bound by metaphysical concepts of time and space.
>
> CLARKE 1973: 13

Even metaphysical time is measured by means of geophysical processes. Ages calculated from measurements of processes such as radioactivity and magnetic-field variations have gained such prominence in archaeology that they threaten to eclipse the more fundamental stratigraphic method. Their claims to accuracy, however, have proven unreliable. It is essential that archaeologists understand the weaknesses as well as the strengths of these esoteric chronometric methods. The application of sound, careful stratigraphic methods of observation and recording in the field can help control for the grosser errors of radiometric and magnetic dating methods by calling attention to discrepancies that require special attention and interpretation.

CHRONOMETRY BASED ON RADIOACTIVE DECAY

Elemental atoms may have one or more unstable isotopic forms with different atomic weights, subject to loss of **alpha** (α) or **beta** (β) **particles** by spontaneous emission. A radioactive **isotope** has a characteristic **half-life**, the time during which half of all the radioactivity will be spent. The rate at which various materials emit particles, therefore, can be used to estimate the passage of time from a defined beginning point. Counting apparatus counts particle emissions; over a short span of time average emission rates can be recalculated as portions of half-lives. The emission of beta particles by radioactive carbon, and of alpha particles by uranium and its radioactive "daughter" products in decay series, are the basis of several chronometric methods that have redefined the reach and potential of the historical geosciences and archaeology. Particle counting methods are being replaced in many laboratories by mass spectrometry, to measure directly the mass of various isotopes in a sample. Direct measurements are considerably more precise than the particle counts for estimating the amount of radioactive isotopes, and thus are able to support more accurate age estimates.

Radiocarbon dating

We begin with radiocarbon, "this fortuitous isotope" (Butzer 1971: 30), because it is the premier archaeological method and because many other methods depend for their applicability upon calibration with radiocarbon time. The discovery by Willard F. Libby that the decay of the radioactive isotope of carbon can be used to measure the passage of time revolutionized the historical sciences and made possible paleoenvironmental studies as they are currently practiced. The importance stems not only from the relatively high accuracy and precision of ages calculated from radiocarbon decay, but principally from the fact that carbon is the element fundamental to all life forms on this planet, and is therefore practically ubiquitous. What is measured is the time since the cessation of metabolism in an organism incorporating radioactive carbon. The relevance of such a time measurement is left to the insight and ingenuity of the archaeologist to demonstrate.

Carbon occurs in the form of two stable isotopes (^{12}C and ^{13}C) and one radioactive isotope (^{14}C, radiocarbon), by far the rarest. Radiocarbon originates in the upper atmosphere when neutrons bombard nitrogen-14 and form carbon-14 + hydrogen ($^{14}N + n \Rightarrow {}^{14}C + 1H$). The radioactive atoms combine with oxygen to produce radioactive carbon dioxide, which is distributed by atmospheric turbulence and is then incorporated into the hydrosphere and biosphere (Fig. 6.1). While subaerial organisms live, the proportion of radiocarbon in their bodies remains close to that of the atmosphere. In the case of marine organisms the reference is to the upper levels of ocean waters. When organisms die and no longer metabolize new carbon, the finite amount of radioactive carbon in their tissues begins to diminish without replacement. Radiocarbon has a half-life close to 5730 years, which means that half the radioactive atoms disintegrate in that span, each producing a nitrogen atom and a beta particle ($^{14}C \Rightarrow {}^{14}N + \beta$). The time since the death of an organism can, therefore, be calculated from the concentration of the radioactive isotope in the material today. Because of the small amount of radiocarbon in the universe, and the length of its half-life, precise measurement of residual amounts is difficult in matter

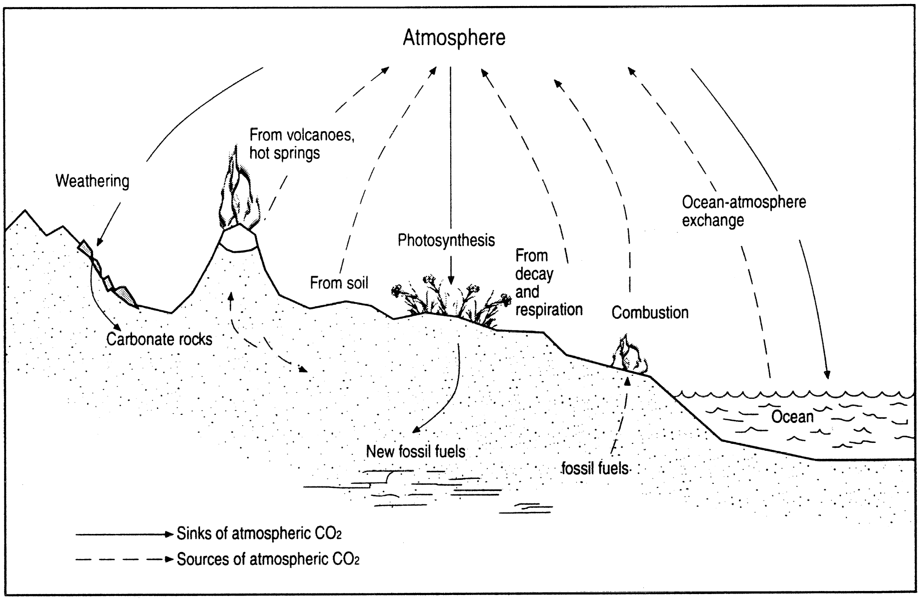


Figure 6.1 The carbon cycle, showing sources and sinks of CO_2. (After Oliver and Hidore 1984: Fig. 15.6.)

more than 30,000 to 40,000 years old; that is the effective range of the method in its conventional "beta-counting" mode. Accelerator mass spectrometry is expected ultimately to extend the range of measurement, but that improvement is elusive. Methodological refinements may permit extension to ca. 70,000 B.P. (Long and Kalin 1992).

Method

As indicated above, there are two ways to determine the amount of radioactive carbon in organic matter: count the emissions of the beta (β) particles, or measure the amount of the isotope directly. The β-counting methods were the first developed, and are conventional. The organic sample is cleaned and reduced to a purer form of carbon, in gaseous or liquid forms. Proportional gas counting and liquid scintillation are the main approaches for laboratories operating today. The purified carbon is placed in a counter and left for specified periods of time, during which the β emissions are counted by the apparatus. The counts are then averaged, and the average converted statistically into an estimate of age based on the half-life properties of the isotope (see R. E. Taylor [1987] or Aitken [1990] for technical details).

The age of the sample thus estimated is always accompanied by a figure expressed as a standard deviation (± some quantity), which is the statistical deviation of the

β emissions for the series of measurements taken, usually including an estimate of apparatus error as well. In two out of three cases (68%), the true age of a sample should lie within the bounds of one standard deviation (one sigma or σ); the probability increases to 95% likelihood within two standard deviations. Certainty, however, is never achieved; some age calculations will be wrong simply because of the randomness of the β emissions. *Which* age estimates suffer from this inescapable uncertainty is, of course, indeterminable from the method itself.

The new accelerator mass spectrometry method (AMS or TAMS [tandem accelerator mass spectrometry]) directly counts the atoms of the several radioactive isotopes in a sample, thus avoiding the problem of random decays. This method requires much smaller samples and theoretically produces more precise results. However, AMS age estimates are also burdened by uncertainties related to machine instabilities that affect the precision and accuracy of the ages calculated (Gove 1992). The choice of method for the archaeologist depends upon the size and age of the samples, cost considerations, and laboratory interest in the problems being addressed.

Laboratories use a conventional approximation of the true half-life called the "Libby half-life" (5568 ± 30), which is shorter than the "Cambridge" half-life now universally acknowledged as more accurate (5730 ± 40). The Libby half-life is used so that all published ages since the beginning of the method are directly comparable. When accuracy is more important than comparability, the conventional figures can be converted to reflect the more accurate half-life by multiplying the age by 1.029. The *age estimates* calculated by the technicians, with their accompanying error ranges, are normalized by reference to the year A.D. 1950, a convention established by international agreement to calibrate all laboratory results to a single reference year ("before present"), and to express the utilization of a standard measure of ^{14}C concentration approximating the time before nuclear weapons testing changed the composition of the atmosphere ("before physics"). This resulting figure, compounded of a series of estimates, probabilities, and conventions, is the age "B.P." (or BP; "b.p." in English convention) of the sample, clearly a major departure from anything like an "absolute" age. A further simple calculation yields an approximation to calendar years: subtract 1950 to convert B.P. ages to B.C./A.D./B.C.E. dates. However, international radiocarbon reporting standards now discourage this latter shortcut to calendrical years.

Among investigators involved in paleoenvironmental studies, only the archaeologists tend to use ages converted to dates, and to talk in terms of calendar years B.C., A.D., B.C.E., and so forth. The reasons for this are perfectly sensible from the perspective of researchers who have access to calendrical precision for part of their time span of interest, and who wish to have their data in superficially comparable form. However, as we have seen, radiocarbon ages are properly expressed as a range of time

($\pm 1\sigma$) – a "span" not a "spot" (Orme 1982: 10). Even when expressed as calendar years, these spans do not command calendrical precision, and to pretend that they do is to confuse matters significantly. The special case of dendrochronologically "corrected" radiocarbon ages, and the calibration techniques related to them, are discussed below. In general, when working with interdisciplinary teams and combining archaeological and Earth-science data and results, one can avoid confusion and assure comparability by quoting radiocarbon assay results in terms of the direct expression of ages and durations, and remembering that these are counts of radiocarbon years, which are not always equivalent in length to calender years.

Complexities and uncertainties

The method is based on some key simplifying assumptions, three of which are classic.

1 The production rate of ^{14}C is constant,
2 All ^{14}C produced is rapidly and evenly distributed around the world, and
3 All organisms take up ^{14}C in the proportions in which it occurs in the atmosphere.

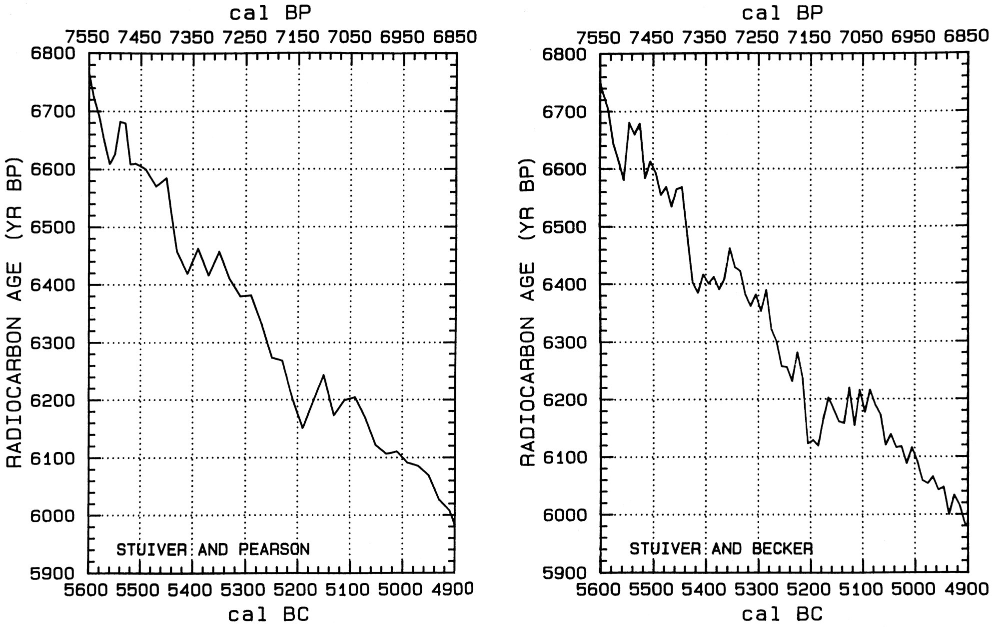
A fourth simplifying assumption has been recognized recently:

4 All laboratory results are comparable.

As it happens, none of these assumptions is actually true; reality, as with most of the world, is more complicated. The violation of the methodological assumptions (1–3) results in inherent, "systemic," uncertainties, uncertainties that are inseparable from the method. Fortunately, all four sources of imprecision are now fairly well understood. Even though their effects cannot be eliminated from the laboratory results, they can be allowed for and sometimes corrected by calibration.

The production rate of radiocarbon is not constant because fluctuations in the Earth's magnetic field permit different amounts of cosmic rays to enter the atmosphere at different times. Furthermore, the solar wind that carries the cosmic rays itself varies (secular variation). Variation in solar activity as measured by sunspots affects radiocarbon production such that quiet-Sun periods are periods of radiocarbon maxima, yielding spuriously young ages on the 210-year cycle of sunspot minima (Pecker and Runcorn 1990). The solar variation is not precisely predictable in its influence on radiocarbon ages. Its variability, in combination with the fluctuations in the Earth's magnetic field, limits the degree of resolution and accuracy that can be achieved within the method.

The secular variation resulting from fluctuations in the magnetic field and in solar activity is partly controlled for by enlisting dendrochronology in the service of



(a) (b)

Figure 6.2 Two high-precision calibration curves for the period 5900–6800 B.P. in radiocarbon years. Curve (a) of Stuiver and Pearson (1993) was compiled by averaging 20-year (bidecadal) series of tree rings. Curve (b) of Stuiver and Becker (1993) was compiled from 10-year series. The calibration is read from the ^{14}C age at the left across to curve intersection, then up for cal B.P. or down for cal B.C. ages. Multiple intersections flag uncertainties. Note that curve (b) has more detail but less accuracy. "[B]idecadal curves should be used for most purposes" (Stuiver and Becker 1993: 35). (Reproduced from Stuiver and Pearson [1993: Fig. 1K] and Stuiver and Becker [1993: Fig. 20] with permission of *Radiocarbon*, © 1993 by the Arizona Board of Regents on behalf of the University of Arizona.)

radiocarbon dating. By testing tree rings of known age in high-precision laboratory apparatus, physicists calculate, decade by decade, the departure of radiocarbon ages from the "true" annual age of the sample. Calibration charts have thus been created, showing the spans of calendar date equivalents for radiocarbon ages (Fig. 6.2). Studies with old wood, particularly the tree-ring series that extend back more than 10,000 years, show that secular variation is of significant magnitude and is partly random: it cannot be predicted closely. Recent work in the northern hemisphere has produced calibration charts of refined precision for the last few thousand years as well as a computer program (CALIB 3.0.3) readily available to researchers (Stuiver 1993). Checking radiocarbon assays against dendrochronology revealed a disconcertingly large residue of error in ages previously published (Baillie 1990). Calibration of greater ages with other dating techniques and materials, including subtropical corals, glacial ice, and annually laminated sediments, discloses even wider separation in deep time (e.g., Bard et al. 1990; Kitagawa and van der Plicht 1998). Archaeologists should develop an attitude of informed skepticism about the precision of radiocarbon ages beyond dendrochronological calibration, and work to identify and avoid erroneous estimates (Taylor et al. 1996).

Properly calibrated dates typically have larger ranges of uncertainty than do uncalibrated dates or ages, as is also the case with properly calculated averaged ages. Furthermore, some precision is lost at the start to averaging multiple ages, because the age span represented in individual organic samples from archaeological contexts is typically considerably greater than the spans included in the tree-ring samples used to compile the calibration charts. Calibration corrections of radiocarbon ages should be used when it is necessary to compare events dated by ^{14}C to events dated by tree rings or calendar years, and whenever *rates* of change or processes are an issue. They can also be used to judge the integrity of "clusters" or "gaps" in radiocarbon series, by indicating major variational episodes in the secular trends (Fig. 6.2). Because of the fundamental differences between raw radiocarbon ages, calibrated ages, and calendrical dates, conventions for labeling them have been defined. The international standard for reporting ages is that conventional ages (as defined above) are reported as "years B.P. where 0 B.P. is the year 1950. Dendrochronologically calibrated ages are to be reported as cal A.D. or cal B.C., or, if required, cal BP" (Mook 1986). The international standard is a very useful reference tool for reporters and interpreters of radiocarbon ages; adherence to its several rules confers scientific credibility on the user.

The assumption of rapid and even distribution of CO_2 is violated because the Earth's deep oceans serve as reservoirs of dissolved carbon dioxide – 93% of all carbon in the exchange system (Aitken 1990: 61). The "reservoir effect" delays the redistribution of ^{14}C to the biosphere, and also regularly releases older stored ^{14}C into both the biosphere and atmosphere. Marine organisms, consequently, are impoverished in radiocarbon relative to most terrestrial plants, and tend to give radiocarbon ages greater than their actual age, reflecting not their contemporary atmospheres but the distortions of the marine reservoir amplified by the food chain, and with additional complications in special circumstances (e.g., Kennett et al. 1997; Little 1993). The tissues of modern North Atlantic seals have been radiocarbon-dated to as much as 400 years old. Pacific deep water is even older than that (Boyle 1990; Shackleton et al. 1988). Awareness of the exaggerated reservoir effect in the southern hemisphere, which has more marine than terrestrial surface, has inspired recent collaborative research into inter-hemispheric calibration.

The tissues of some organisms diverge from the atmospheric ratios of carbon isotopes because of metabolic fractionation: the organisms selectively metabolize carbon isotopes by their atomic weights, normally discriminating against the heavy ^{14}C. Plants that discriminate less than most against the heavy isotope ^{14}C (i.e., take in more) produce tissues with isotopic ages somewhat younger than those expected. Maize (Indian corn) is a notorious example of this, but it is only one of a number of plants, mostly tropical and semi-tropical grasses, which convert carbon dioxide to food through a special ("C_4") metabolic pathway that utilizes heavy carbon efficiently. Plants, the primary users of carbon dioxide, are the fractionating agents. Animals that consume fractionated plant foods accumulate the unrepresentative carbon signal, which can build up to significant levels over their lifetimes. The presence of fractionated isotopic ratios can be recognized by comparing the ratios of the stable isotopes, ^{12}C and ^{13}C; correction factors for the ^{14}C discrimination can then be calculated (Browman 1981; R. E. Taylor 1987). The ratio of the ^{13}C isotope in a radiocarbon sample, expressed as $\delta^{13}C$, is essential to evaluation of the age calculation. The standard is woody vegetation of the temperate zone, with a $\delta^{13}C$ value ca. $-25‰$ ["per mil," parts per thousand]. Any departure of $\delta^{13}C$ values from the typical -25 ‰ is normalized as part of the calculation of conventional ages.

In order to evaluate the precision and accuracy of radiocarbon age measurements, one needs to know the sources of difficulties, and to be able to estimate their effects upon given sample ages. R. E. Taylor (1987) classifies the uncertainties and error sources in radiocarbon dating under four "factors" – sample provenience factors, sample composition factors, experimental factors, and systemic factors. Systemic factors and some sample composition factors (e.g., fractionation) are discussed above.

In archaeology, a significant source of dating error involves sample provenience factors: discrepancies between the real age of the organic material assayed and the age of the event of interest to the archaeologist. Dean's "event" concepts (introduced in Chapter 5) are useful for thinking about this. The target event is the event that the archaeologist wants to date; the reference event is the link between problem and method that permits dating of the target event. The dated event is the actual event whose age is measured. For example, in a case where the target event is the time during which a pithouse was occupied, the discovery of a hearth with charcoal in the middle of the floor may offer a reference event, to the degree that the use of the hearth is related to the occupancy of the house (i.e., the hearth was not intrusive). However, the dated event is the time of death of the tree; this may or may not closely approximate the time of the hearth fire. Clearly, rigorous stratigraphical observations in the field are essential for estimating the relationships between reference, dated, and target events; upon their integrity everything else depends.

Among the important sample composition factors are issues of sample contamination; these are likely to be especially prominent in archaeological sites because of the complex chemistry and depositional history of such areas. Fractionation problems aside, the introduction of organic matter of different age from that of the main body of the sample will complicate any assessment of the sample's age. Such contamination typically occurs after death and burial, and then normally involves the adulteration of the sample by humates or carbonates carried in groundwater, which may be deposited in the pores and interstices of the sample. These may be older or younger than the sample, while the intrusion of roots or the introduction of animal waste into buried samples can bring in younger materials. Contamination in the form of older or, more typically, younger material may also be introduced to a sample during collection, handling, and storage. Standard well-publicized handling procedures can minimize such error (Aitken 1990; R. E. Taylor 1987). Samples of organic material submitted for dating (Table 6.1) are cleaned to remove extraneous materials before the carbon is purified, as part of the laboratory processing.

The contribution of experimental factors to radiocarbon age uncertainties is under investigation. Experimental factors include those involved with counting random events, as expressed in the standard deviation that accompanies each age figure. Also included, but not often expressed, are the errors introduced by background or machine contamination in the laboratories, and machine error – those unpredictables that intrude into any elaborate and precise measurement technology. These errors are usually quite small, and therefore ignored, but can cause problems. They account for some of the differences, sometimes systematic, in the age calculations by different laboratories from "equivalent" samples. Among experimental factors, the statistical nature of radiocarbon measurements cannot be overemphasized, and must never be lost sight of. As discussed above, the probability of the true age of a sample lying within the interval expressed by one standard deviation is 68%; of two standard deviations, 95%. However, "the probability that the actual age of the sample is exactly equal to the value cited as the 'age' approaches zero" (R. E. Taylor 1987: 123). Radiocarbon measurements express *intervals*, not moments in time.

Laboratory pretreatment processes vary with the field and sample conditions, and can reduce the available size of the sample significantly. Chemical pretreatments vary with the sample material and suspected contaminants. The conventional treatments include washing in a strong base, NaOH (sodium hydroxide), for the removal of humic and tannic acids, and an acid bath in HCl (hydrochloric acid) for the removal of carbonates and certain unstable organic compounds (Aitken 1990; R. E. Taylor 1987). Since the pretreatment process chosen may affect the results, clients of radiocarbon laboratories should consult on any possible complications. It should be

Table 6.1 Typical sample sizes for radiocarbon dating

Counting mode	Sample preparation	Sample size in mg C[a]	Count time in hours
β decay	Standard gas/LS	250–5000	24–72
	High-precision gas/LS	10,000–20,000	72–168
AMS/TAMS	Solid	2–5	1–2

Notes:
[a] mg C = milligrams of carbon.
LS = liquid scintillation.
Source: Adapted from R. E. Taylor 1987: 96, Table 4.1; see original for more details.

obvious that careful scrutiny of the sample in context at the time of collection, and full reporting of the context to the laboratory are both essential to decisions about appropriate pretreatment, and thus to the success of the assay.

In order to monitor and minimize such errors, radiocarbon laboratories check their results by separate runs and by inter-laboratory comparisons. Archaeologists should be aware that inter-laboratory differences are real. Even within a given laboratory, age estimates may vary with different pretreatment regimes and other factors still unknown. Tests have shown that laboratory variance may exceed the expressed error ranges, and that inter-laboratory differences may be greater still (Scott et al. 1998). The program of inter-laboratory testing is proving its advantages for defining, measuring, and controlling the discrepancies.

There is a large literature on problems of archaeological sampling for radiocarbon assays, amounting to a major body of cautionary tales and sound advice. Anyone anticipating the collection of samples for radiocarbon dating should become familiar with the literature, and should take special problems and questions to the radiocarbon laboratory prior to field sampling (see Aitken [1990: Ch. 4] and R. E. Taylor [1987: 107–112] for examples of sampling errors and things that go bump in the night). "[L]ittle reliance should be placed on an individual ^{14}C 'date' to provide an estimate of age for a given object, structure, feature or stratigraphic unit" (R. E. Taylor 1987: 105). Multiple dated samples from the same or closely related stratigraphic units, or dates on different fractions of the same sample, can help expose anomalies that indicate discrepancies. Anomalies are themselves information, implicating the need for further evaluation of the data in hand, or for more data. (See Kra [1986] for guidance in recording and submitting radiocarbon samples.)

Radiocarbon dating approximations create a number of subtle problems for the historical sciences, especially those, like archaeology, that depend on the ordering, correlation, comparison, and interpretation of cultural-scale events (Buck et al. 1994). The secular variation in radiocarbon production means that the lengths of radiocarbon years vary in time; especially in time ranges beyond the ninth millennium B.P., the length of radiocarbon centuries is significantly shorter than the length of sidereal centuries. An eloquent warning about the non-equivalence of uncalibrated ^{14}C years and the astronomical calendar (Braziunas 1994) might usefully be celebrated by archaeologists as the "Braziunas caveat," and heeded whenever comparisons are undertaken between the radiocarbon time scale and paleoclimatic chronologies. Periods of enriched cosmic ray bombardment result in radiocarbon "plateaus," spans of time within which fine chronological distinctions cannot be made. One such plateau complicates the isolation of chronology and event sequences during the Younger Dryas climatic reversal, coincident with the significant postglacial human population expansions into high middle latitudes in North America and Europe (Edwards et al. 1993; Hughen et al. 1998). Furthermore, because event sequences and chronologies are the foundations for inferences about processes, expect that further precision in dating methods will require reconsideration of some established models of historical and evolutionary processes.

With so many sources of possible error impinging upon radiocarbon measurements of elapsed time, it is strongly recommended that multiple interpretive hypotheses be entertained in efforts to understand any anomalies that are recognized. If the anomalies cannot be evaluated so that their sources can be identified, the responsible investigator will publish the results with a discussion of a range of possible explanations, to provide future investigators with clues for solving the problems.

Citation conventions

Archaeologists have habitually published their radiocarbon results in a variety of formats, some quite idiosyncratic. Furthermore, within the English-speaking world, archaeological conventions for distinguishing dendrochronologically calibrated ages from others, and calendar dates from ages, have diverged on both sides of the Atlantic. Here, assuming that the reader is familiar with at least the local practice, I call attention again to international conventions agreed by practitioners of the radiocarbon craft (Mook 1986). Archaeologists may benefit from the discussions addressed to them by Bowman (1990: 42) and R. E. Taylor (1987: 4–6). Most laboratories report in this "conventional" form whether the age is ultimately calibrated or not. Archaeologists should follow the format to enhance the clarity and usefulness of their published results. In brief, the published report of a "conventional" radiocarbon result should provide the following information:

- ages calculated with the Libby half-life of 5568 years
- the ages normalized to $\delta^{13}C = -25$ ‰
- the results in uncalibrated years B.P., with B.P. zero being A.D. 1950
- an error estimate of $\pm 1\sigma$
- the laboratory number.

Any calibration, correction of the half-life, or expansion of the error term to 2σ should be separated from the provision of this essential information. Also, interpretation will be facilitated if the $\delta^{13}C$ value of each sample is also provided. In other words, it should always be possible to distinguish the primary data on the radiocarbon age calculation of a given sample from later manipulations undertaken to compare ages to dates. A table with all the essential values is an ideal accompaniment to a discussion of ages. In every case, it is essential to realize that "The probability that a certain cal[ibrated] age is the actual sample age may be quite variable within the cal age range" (Stuiver and Becker 1993: 390). No matter how thoroughly manipulated, no radiocarbon age expression carries a guarantee of accuracy.

Prior to 1995, summaries of radiocarbon results regularly appeared in the journal *Radiocarbon*, which remains the best source for current and innovative research on methods and applications. Research results in the form of laboratory date lists are available in many different media now: on-line, on disk, in print. Consult radiocarbon laboratories in the area of interest; they are listed with addresses in *Radiocarbon.*

Uranium-series dating

The radioactive isotopes of uranium, ^{238}U and ^{235}U, decay through a series of "daughter" products to end after very long periods of time as stable isotopes of lead. The daughter products are many in each series, and all of them have their own particular half-lives. Some daughters may be separated from the "parent" materials by escaping the matrix as gases or precipitates, to continue on their own decay routes in other contexts. Because of the large number of different isotopes and elements produced during these decay series, there are various methods for measuring time elapsed since the series activity began in a particular material, and they cover different ranges of time, depending upon the critical half-lives. Some methods measure the decay of daughter products as ratios to the parent, others measure the accumulation of radioactive isotopes in materials since a given initiating event – the deposition, burial, or formation of the material.

U-series methods are applied to dating organic or inorganic carbonates, and have been successful in dating geological contexts and events. For archaeology, the most successful applications have been dating archaeological materials enclosed in **speleothems** (carbonaceous cave deposits). Methods frequently applied to date Quaternary geological phenomena and deep ocean cores vary in precision and accuracy among themselves (Aitken 1990: Ch. 5; Ivanovich and Harmon 1992).

The methods that have proven most useful in Quaternary paleoenvironmental studies measure the decay of ^{238}U through ^{234}U to ^{230}Th (a thorium isotope called ionium), and ^{235}U to ^{231}Pa (protactinium), or measure the ratios of ^{234}U to ^{238}U. In all cases, the crucial assumption is that the carbonate material whose age is being measured has constituted a closed system since the event of interest; that is, that no uranium or its major daughter products have entered or left the material in the time span being estimated. This assumption has been supported by research models and by results in coral reefs on raised beaches (taken out of the water) and in tufas and speleothems that crystallize after being deposited from groundwater. "Open system" involvement with groundwater contaminants typically complicates attempts to date carbonates such as caliche, bone, teeth, and mollusk shells.

Another problem with the methods is establishing the amounts of uranium or of the target daughter products that characterized the material at the time from which we wish to measure. In the case of uranium isotopes themselves, in sea water, the assumption of a constant ratio between them seems tenable, and their different rates of decay provide a measure of the time elapsed since they were taken up by marine organisms such as corals. In terrestrial systems, it is difficult to determine the amount of uranium available, for instance, in groundwater depositing calcites or aragonites in carbonates, bones, or shells.

The introduction of mass spectrometric methods for the measurement of isotope ratios established a new level of precision and usefulness for uranium series dating. A notable improvement in measurement of isotope ratios is TIMS (thermal ionization mass spectrometry), which permits greater precision and speed in U-series analyses (Wintle 1996: 134). Research published in 1990 measuring U/Th ratios in Barbados corals appears to have achieved precision and accuracy at least equal to that of radiocarbon, without the problems of radiocarbon's secular variation. The coral U-series ages match well their radiocarbon ages within the range of dendro-calibration. Beyond that, they diverge from the radiocarbon ages, but U-ages may be more accurate than ^{14}C for time before 9000 years ago (Bard et al. 1990). If the U/Th ages are as accurate as appears, the radiocarbon discrepancy may be as great as 3500 years by 20,000 years ago. If the accuracy of U/Th dating is indeed good enough to make it a calibration standard for ^{14}C, geochronology will be fundamentally improved worldwide and nearly all archaeological models of change rates in deep antiquity will require reconsideration.

Archaeologists must be aware of the highly experimental nature of these methods within the time spans and precisions that make them useful to archaeology. Referring to the most recent literature, mainly in geological publications, and consulting with investigators directly involved in development and application of the methods, are strongly advised.

Potassium–argon dating

A radiometric method based on the decay of radioactive potassium (^{40}K) has become indispensable to geologists. With a very long half-life, the decay of ^{40}K to its daughter product, ^{40}Ar, is capable of expressing the ages of certain rocks essentially from the first consolidation of the geosphere (Aitken 1990: Ch. 5; Faure 1986).

Potassium, which has one unstable and two stable isotopes, is a constituent of some of the most common minerals in igneous rocks. The decay of ^{40}K produces an isotope of the rare gas, argon, which is trapped in the crystal lattices. In rocks cooling from magma, gases are freed, and the newly solid rock is theoretically devoid of argon, thus setting a radiometric clock that will measure the time since cooling (compare the emptying of electron traps to set the thermoluminescence "clock," below). Measuring the amount of potassium in a rock sample, and then the amount of argon-40, permits calculating the age of the rock. The assumptions of the method, that (1) no residual argon remained after cooling, and (2) the system has been closed since, neither receiving nor giving up argon, are in fact violated often. Weathering or exposure of the rock may liberate argon or permit atmospheric argon to penetrate it. Various methodological refinements have reduced the errors introduced by the violation of the assumptions, but error ranges of a million years or so adhere to geological age determinations.

In the historical geosciences, K/Ar has helped build a basic chronology for the ice ages, and for the seafloor spreading that measures continental drift. Attempts to correlate geosphere events with climatic changes inferred from other kinds of evidence, including orbital parameters, are similarly beset with scale problems because of the large error ranges of this method.

The K/Ar dating of Cenozoic and Quaternary sites of hominoid and hominid finds has provided a rough chronology for the development of human beings and their earliest experiments in material culture, most notably in the Rift valleys of East Africa. It has made possible some comparability between sites widely separated in space, and some approximation to developmental rates. However, it is yet undetermined how much of the growing complexity seen in early human development may be attributed to the large imprecisions, and consequent inaccuracy, of this dating method.

Recent improvements in measurement technology have overcome some of the problems besetting classic K/Ar dating. Single crystal laser fusion (SCLF) has made feasible direct comparison of argon isotopes ^{40}Ar and ^{39}Ar, thereby increasing precision and circumventing some problems of sample contamination (Wintle 1996: 129). The precision so achieved has brought Ar/Ar ages into the time span of radiocarbon, offering the potential for lengthy calibration series.

CHRONOMETRY BASED ON RADIATION DAMAGE

Within the geosphere, natural radioactivity occurs in the form of alpha and beta particles and gamma rays, emitted by isotopic decay. This radiation, in combination with the weaker cosmic radiation, damages other elements, creating free electrons by ionizing radiation. The free electrons may recombine with nearby atoms or become trapped in gaps in crystal lattices. These trapped electrons may be freed by heating, in which case they emit measurable light (thermoluminescence). Alternatively, the density of trapped electrons can be measured by exciting them with microwaves and measuring the intensity of the resonance signal in a magnetic field (magnetic resonance). On the assumption that the intensity of the light or resonance signal of the electrons bears a linear relationship to the accumulated dose of ionizing radiation, these phenomena can be used to measure time elapsed since the electron trapping began. Further basic assumptions include (1) the expectation that the modern intensity of ionizing radiation at the collection site has been fairly stable over the time being measured, and (2) that the collecting material has neither lost nor gained electrons by any other means.

Groundwater is a major complication. Since water is more effective than other materials at absorbing radiation, its presence attenuates the radiation received by the sample, thereby reducing the dose, and consequently the apparent age. Water may also leach away or redeposit radioactive materials, thus changing the dose over time by violating the first assumption (stability). The difficulty in estimating the groundwater history of a site makes this complication particularly challenging for applications of radiation-accumulation dating methods.

Thermoluminescence dating

Electrons trapped in crystal defects are freed as light (thermoluminescence; TL) when minerals are heated above a critical point (ca. 300–450° C). If, at some archaeologically significant time in the past, the material was heated enough to free all the trapped electrons and “reset” the TL clock, then the time elapsed since that heating event can be estimated. The intensity of light measured correlates with the duration of electron accumulation. Baked clay is an excellent material for this method; a range of other archaeological and sedimentary materials can also be dated. The developmental history of the TL method has been well summarized by M. J. Aitken (1985, 1990) and A. Wintle (1996).

The method is self-calibrating: field and laboratory tests estimate the level of radiation dose a sample received at its burial point and the sensitivity of the particular sample material to receiving, storing, and releasing the free electrons. With samples up to 3000 years old, the method compares favorably in accuracy with radiocarbon dating; it is less accurate with progressively older samples. However, as it is applicable back to over half a million years, it is well worth development despite the complications and limitations. With improvements in U/Th dating, opportunities will expand for calibrating ancient TL ages, to check on the accuracy and precision of the latter method.

Several methods are used to prepare and read sample luminescence, depending upon the characteristics of the sample and site (Aitken 1985, 1990; Wintle 1996). Basically, a prepared sample is heated gradually, and the glow emitted by the sample is graphed as a function of heat and intensity of light (sample glow curve). In combination with the information gained by the several self-calibration tests, the glow curve can be interpreted as an expression of received dose (“equivalent dose”), which permits calculation of time elapsed since the last heating of the sample.

The method incorporates two critical assumptions: (1) there is a linear relationship between the dose and the TL emissions, and (2) there has been no loss (fading) of accumulated TL. Both of these assumptions are violated by reality, and special tests for complications are required. Radiation sources may be both internal (in the sample material) and external (in the matrix), along with a minor contribution from cosmic radiation. The dose originating from the matrix can be measured on samples in the laboratory or directly in the field with a dosimeter when conditions require, providing there is close coordination between the excavators and the laboratory technicians. Variations in the nature of samples and matrix necessitate choices among several preparation and reading methods, and place special responsibilities on the sample collector (Aitken 1985, 1990).

The method was first developed for archaeological use in dating ceramics, where the mineral additives are the preferred target. In addition, TL ages since firing can be calculated for burned flint, burned rock, and slags. For different reasons, the time since deposition of crystalline calcite and aeolian sediments is also measurable by TL. Calcites, essentially radiation-free at deposition, receive radiation from external sources. Consequently, the age of speleothems can be calculated to help date archaeological materials in caves. Wind-blown mineral grains, on their way to becoming aeolian sediments, are bleached by exposure to sunlight, which removes most of the trapped electrons from the grains. Accumulation of electrons resumes after deposition and burial, incorporating a fresh TL signal representing the time since burial. TL dating of loesses, first developed in the USSR, has been widely applied in Quaternary studies, both terrestrial and marine (Wintle and Huntley 1982). The technicalities of all these age determinations are considerable; there is no substitute for consultation with a specialist while field investigations are in the planning stage. "TL dating is no routine work" (Wagner et al. 1983: 39).

The age of the sample or samples (they are best run in sets from a single context) is expressed as an average of the samples with two error figures. The first, usually the smaller of the two, expresses the statistical error in the calculation of the average age of the samples in the set. It is used to compare TL ages within a site or between similar contexts. The second, usually larger, uncertainty factor expresses the predicted systematic error, accounting for some of the field complications known to affect the sample. It is appropriately used when comparing TL ages with calibrated radiocarbon ages or between sites. Both error expressions are given as one standard deviation; both increase with the age of the samples, averaging 5–10% of the age (Aitken 1985: 31).

Used with sensitivity to its particular complications, TL can provide important information to several of the sciences involved in paleoenvironmental reconstruction. The range of TL, far beyond that of radiocarbon (to ca. 10^6 years), and its relative precision in comparison with other long-range dating techniques such as potassium–argon, promise an important future in geochronology.

Optical luminescence dating

A related approach to dating is showing promise for unburnt sediments, since it relies directly upon luminescence stimulated by light (Aitken 1990: 175–177; Wintle 1993, 1996: 132). This method, especially suitable for aeolian sediments because it quickly measures sunlight bleaching, is finding favor among Quaternary geologists for such applications, and for archaeological sites on or under aeolian sediments (Wintle et al. 1994).

Electron spin resonance dating

Electrons freed from paired bonds by particle bombardment may be trapped in crystal lattices of many materials, as summarized above for TL. Trapped electrons measured by electron spin resonance extend this accumulation clock to organic matter and calcite minerals. So far, it has worked better on teeth than on bone, and is applicable to mollusk shells, corals, foraminifera, speleothems, and **travertine** (massive carbonate rock deposited from freshwater). Contamination during diagenesis by radioactive minerals carried in groundwater is a serious complication, especially for bones and porous (poorly crystallized) carbonates.

The radioactivity of the sample material and of the sample matrix must be measured to determine the dose rate; the matrix can be measured in the field by a buried dosimeter. The efficiency of a sample for trapping electron charges (and positively charged "holes") in the lattice is measured, so that the environmental dose rate can be corrected to yield an accumulated dose rate specific for the sample. The accumulated dose is assumed to bear a direct relationship to time, although it has also proved to vary with ambient temperatures; the lower the temperatures, the slower the accumulation and thus the longer the time span that can be measured prior to saturation.

Measurement involves mounting a sample in a magnetic field of known force and subjecting it to microwaves to excite the trapped electrons. The intensity of the response by unpaired trapped electrons correlates directly with their number (a function of dose); consequently, it can be converted into a time measurement. The accumulated dose divided by the annual dose (established from rate and sensitivity) gives a measure of the time since crystal growth, or since subsequent significant heating that resets the chronometer.

The ESR method, still experimental and of unproven accuracy (Rink et al. 1996; Wintle 1996), has been used on archaeological teeth, shells, and calcite minerals with apparent success despite a wide range of uncertainties (Aitken 1990; Grün and Stringer 1991). Since it gives especially good results with hydroxyapatite in tooth enamel, it is being intensively developed for paleoanthropology because it is applicable to time beyond the reach of radiocarbon (Aitken et al. 1993: Jones et al. 1994). Nevertheless, experience is revealing an increasing number of complicating variables in the range of Middle and Lower Paleolithic sites (Grün 1997).

Fission-track dating

The massive nucleus of the heavy isotope of uranium, ^{238}U, is subject to spontaneous fission, releasing great amounts of energy and sending fragments outward at high velocity. These fragments damage other materials in their paths, leaving scar-like fission tracks in glassy mineral matter. On the assumption that the rate of decay is a constant, then the accumulation of tracks in a mineral is a function of time and the amount of uranium in the material. The applications of this dating method within archaeological time spans depend upon the presence of appropriate new material (uranium-rich minerals or glasses in a volcanic tuff) or some thermal episode related to the archaeological target event that would "set the clock" by raising the temperature of an appropriate material enough to eliminate older tracks by annealing (melting), so that damage since the annealing event could be counted (Aitken 1990: 132–136).

Materials suitable to this method include a number of glassy minerals (zircon, apatite, mica, etc.) and obsidian. Given such a material in primary association with an archaeological or natural event of interest, the analyst prepares the material by grinding a fresh surface and etching it, to emphasize the damaged zones surrounding the tracks. Sample preparation varies with the material, in recognition of different hardnesses and compositions. After preparation, the tracks are counted under a microscope. The uranium content of the material must be determined for each sample before time can be calculated (Wagner and van den Haute 1992).

The method is dependent upon some simplifying assumptions, which are not demonstrably accurate. The principal assumptions involve the uranium and the tracks themselves. It is necessary to assume (1) that the uranium decay rate is constant (although several different figures are in use for the constant), and (2) that the uranium is uniformly distributed within the material, which is sometimes demonstrably not the case. The analyst must also assume that all the tracks formed since the event of interest are still visible, that is, that none has healed naturally, and that the density can be adequately sampled by the polish-and-etch method. In fact, only some tracks intercepted by the polished surface can be seen on it; there are critical angles of interception that render some tracks unobservable. The damage is calibrated against damage inflicted on the sample in a nuclear reactor.

All of these uncertainties make the error in fission-track counting difficult to estimate; thus, both precision and accuracy can be in doubt. The method has wide application in geology, especially to volcanic rocks. Once enthusiastically embraced by archaeologists, the method is superseded in suitable materials by TL and ESR.

CHRONOMETRY BASED ON COSMOGENIC NUCLIDES

Cosmic rays, whose high-altitude bombardments make radiocarbon out of nitrogen atoms, also produce rare isotopes of many other elements. Some of these, especially ^{10}Be, ^{26}Al, and ^{36}Cl, with half-lives much longer than radiocarbon, have been used in experimental dating of geological and site surfaces. The concept is that the accumulation of these rare isotopes, or their ratios, can be used to estimate the time since a surface was first exposed (e.g., Bard 1997; Beck 1994; Plummer et al. 1997). The method shows promise, but the many complicating variables need more study (Tuniz et al. 1998). Watch for developments.

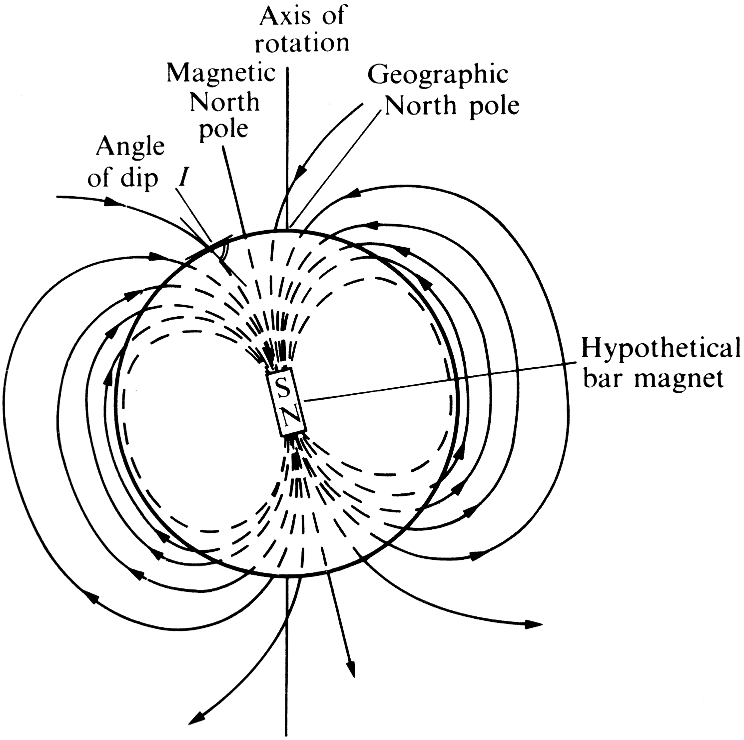


Figure 6.3 Earth's magnetic field, represented as a bar magnet. A small magnetized needle would point along the lines of force. The close lines show relative field strength or intensity. (Reproduced from Aitken 1990: Fig. 9.2, with permission of Addison, Wesley, Longman.)

CHRONOMETRY BASED ON THE EARTH'S MAGNETIC FIELD

> [A]rchaeologists are beginning to accept the fact that effective solutions of many of the most crucial problems in the discipline will require their greater participation in a wide range of natural-science applications.
>
> WOLFMAN 1990B: 354

The Earth's **magnetic field** is created by electrical currents originating in the geodynamo, the fluid dynamics in the molten outer part of the core. Imagine the field as the effect of a bar magnet running through the center of the Earth, at a slight angle to its axis (Fig. 6.3). The force field produced is roughly parallel to the surface at the equator, roughly at right angles to the surface near the poles, and at angles varying with latitude between. The angle of force against the surface of the Earth is the **inclination**; it is that which makes compass needles dip toward the Earth's surface if free to move in that direction. The poleward orientation of the imagined "magnet" defines the **declination** of the field, that is, its variance from true north. A third variable property of the magnetic field is its **intensity**, or strength.

The position of the magnetic pole, and therefore the declination, wanders almost constantly, although at varying rates. On short time scales it drifts around the direction of true (axial) north. Moreover, the period for changes in inclination is different from that of the declination, and these intervals are different from changes in the intensity. Therefore, all these phenomena can be used as dating methods if their changes through time are known (Aitken 1990: Ch. 9).

Additionally, over long but variable time intervals, polarity reverses direction, exchanging north for south. Polarity has reversed several times within the last few million years (Jacobs 1994). Wandering of the magnetic pole and reversals of the field are expressed in the "frozen" magnetism of rocks and sediments, especially volcanic rocks, which are dated by K/Ar and Ar/Ar ratios.

Paleomagnetic and archeomagnetic dating

The choice between these terms is often dictated by no more than the research specialty of the user: "paleomagnetism" is the term favored in the Earth sciences, "archaeomagnetism" in archaeology. However, there is a useful distinction to be made. Paleomagnetic dating as practiced in oceanography and geology is based on the sequence of polarity reversals in the Earth's magnetic field, which provide a reference chronology going back many millions of years. Archaeomagnetic dating is based on short-term changes in the orientation and intensity of the field in recent millennia, expressed at regional scales.

Archaeomagnetic dating is normally undertaken for archaeological materials within the Late Holocene, utilizing the secular variation of the Earth's magnetic declination, inclination, and intensity. Records of former positions of the declination and inclination, and of the intensity of the field, are held in minerals that are susceptible to magnetic force (the iron oxides magnetite and hematite dominate). Small magnetic "domains" within the mineral grains tend to align with the Earth's field when they are mobilized sufficiently to do so. This mobilization occurs when minerals cool out of molten rock, or cool after being heated up to ca. 700°C by fire or lightning, or when sediments settle freely through water and are then compressed, or are strongly weathered chemically.

Thermoremanent magnetism (TRM) is the signal of the Earth's field at the time the heated materials cooled; it remains fixed in its original orientation while the Earth's field changes around it, and is measured with appropriate instruments. The

TRM orientation in materials not directly dated is then compared to dated curves, and a date obtained by calibration. Similarly, detrital remanent magnetism (DRM) can be measured in fine sediments undisturbed since they settled out of water (Verosub 1988), and chemical remanent magnetism (CRM) from strongly weathered sediments. The latter (CRM), less stable than either TRM or DRM, is considered a contaminant.

Once research has established the meandering path of the magnetic pole (declination) and the vagaries of the inclination, as measured from points on the Earth's surface within an area of ca. 1000 km, and has dated reference points along the paths, then other measurements of these characteristics can be fitted to the data to provide approximate ages (Fig. 6.4). Dating accuracy varies in time and space; on dependably calibrated and well-dated curves, the accuracy can be within 25–50 years for declination and intensity.

The orientation of a sample in its original position is crucial information. Therefore, the best archaeological samples are baked earth under old fireplaces or ceramic kilns or the unmoved contents of kilns. Portable pieces of baked clay (ceramics) cannot give the positional data, but may sometimes be dated on the basis of intensity variation when curves for that have been established previously.

The critical assumptions for archaeomagnetic dating are: (1) the remanent magnetism reflects the field at cool-down or settling, and (2) in the case of TRM, there has been only one critical heating (that is, the sample must be relevant to the archaeological event of interest). As with other methods, exceptions to these assumptions complicate the applications. Remanent magnetism may decay in time, and it can be reoriented when sediments undergo chemical change (but that is usually obvious). Subsequent heating, e.g., by lightning or other fires, or resuspension in water, can reorient the particles. Errors may occur in sampling, by the selection of unevenly heated samples or of weathered samples, or by movement of the sample before or during recording or by simple recording errors. There are, as may be expected, additional complications in measuring the RM of samples and in creating the curves (Aitken 1990; Batt 1997; Eighmy and Mitchell 1994; Sternberg 1990; Wolfman 1984, 1990b), as well as a growing awareness of lability in the magnetic field.

Information on the varying locations of the virtual geomagnetic pole and other aspects of the secular variation of the field obtained from archaeologically dated samples are of interest for their own sakes to geophysicists (Verosub and Mehringer 1984). Variations of intensity inform about changes in the effective magnetic shielding from solar radiation, and thus, indirectly, about some aspects of climate and ^{14}C production that may enhance understanding of the linking of these systems (e.g., Stuiver et al. 1991).

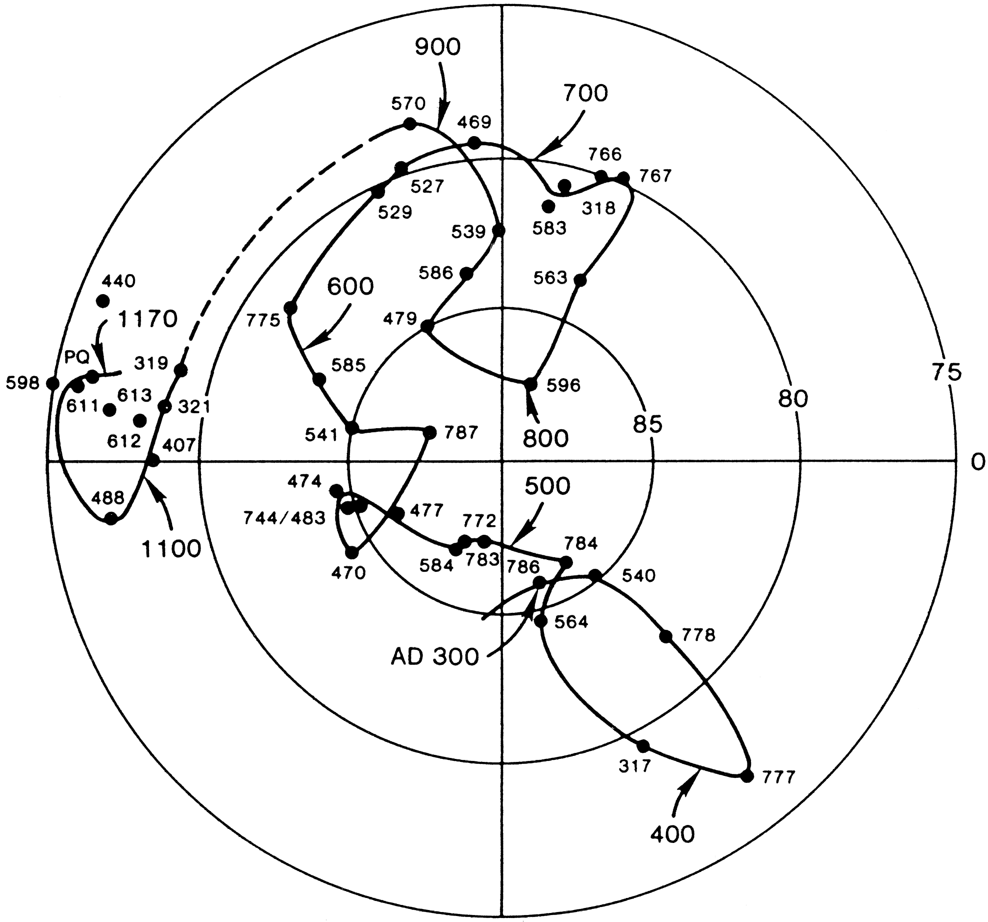


Figure 6.4 Wandering of the Earth's magnetic pole, as graphed from Mesoamerica for the period A.D. 300–1170. The curve is drawn on a polar projection above 75° latitude. Each point represents a "virtual geomagnetic pole" (VGP) calculated from declination and inclination in TRM at a particular sampling site. (Reproduced from Wolfman 1990a: Fig. 15.4B, with permission of the University of Arizona Press.)

Paleomagnetic dating based on polarity reversals provides a coarse chronology with a span in the millions of years. It is the basis for some of the dating of deep ocean cores, which are calibrated to potassium–argon-dated terrestrial volcanic materials that retain a TRM signal. Major, long-lasting polarity reversals are called **chrons**; briefer reversals are subchrons and excursions (Fig. 6.5). By matching the sequence and number of reversals recorded in the sediments of deep cores, investigators establish a chronology on materials otherwise difficult to date. Of course, with phenomena at such large scales, precision is impossible, and skepticism about accuracy may

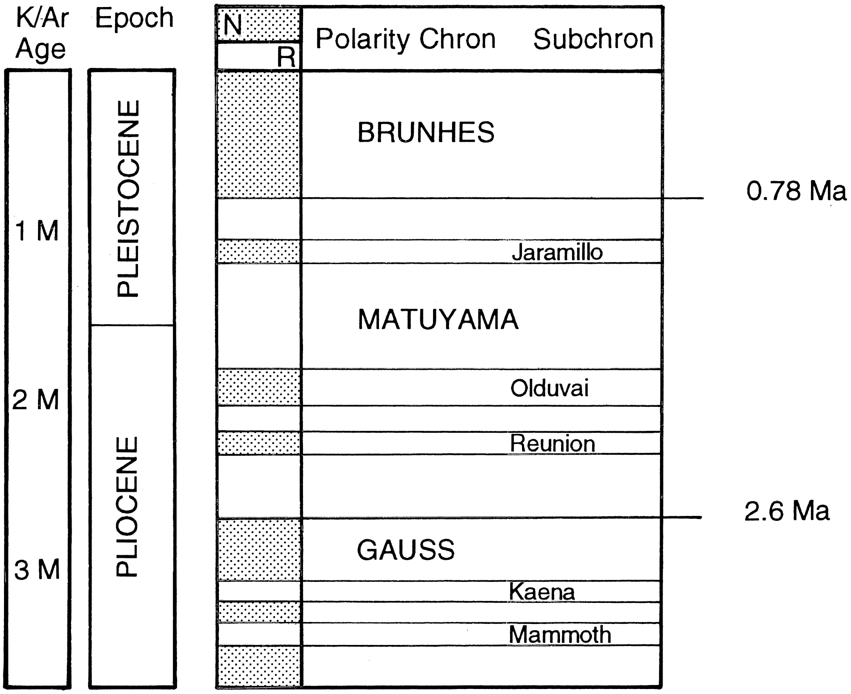


Figure 6.5 The most recent three paleomagnetic chrons (Brunhes, Matuyama, Gauss), with subchrons and ages. Shaded areas are Normal polarity (N), others Reversed (R). Polarity reversals are the basis for dating and correlating sedimentary and eruptive events worldwide.

often be well founded (Bradley 1999: 95–96; Merrill and McFadden 1990). Nevertheless, for glacial and interglacial sequences far removed in time, and for early hominid evolution, such dates may be the best currently available.

SUMMARY

Archaeologists and geoscientists can choose among an expanding array of methods for measuring the passage of time (Fig. 5.3). These methods vary in their applicability to various materials and situations, useful spans, precision and accuracy, and expense. An informative example of synergistic ^{14}C applications in interdisciplinary contexts is discussed by Bartlein et al. (1995).

All methods have limitations and complexities, including some still undefined. Archaeologists should be familiar with the basics of the methods currently most useful, as summarized above, but need to understand that each method makes highly technical demands upon its practitioners, so that no one can be expert in more than a few. Once aware of the range of materials and situations that can be analyzed in terms of expired time, archaeologists must assume responsibility for consulting with expert practitioners prior to collecting critical samples.

All chronometric methods, even calendars and stratigraphy, are subject to improvement. Users should expect that refinements will render any textbook or handbook presentation obsolete at some point. New methods appear frequently as the Earth sciences receive better funding to study aspects of planetary ecology and climate change, among which knowledge of rates and processes of change are recognized as crucial for effective management. Archaeologists aware of the dynamism in geochronology can make informed use of it.

THE ELUSIVENESS OF TIME

Decades ago, B.P. (before 1950), dating Bronze Age sites in the eastern Aegean seemed direct and reliable, given artifactual cross-ties with the astronomically calibrated Egyptian king lists that served as calendars. For instance, the association of Late Minoan IA ceramics in a Greek tomb with Egyptian scarabs provided a *terminus ante quem* age estimate for the ceramic style at ca. 1500 B.C., a nice round, memorable date. Ceramics of LMIA style were buried when the Thera volcano erupted on Santorini island, north of Crete, and 1500 B.C. was proposed as the date of the event (Marinatos 1939). The scale of the eruption, devastating to the town of Akrotiri, led archaeologists to try to relate it to catastrophic fires and building destruction elsewhere in the Aegean. When radiocarbon dating became available, scholars wanted to use that method to refine the dating of the eruption and test its synchroneity with destructive events nearby. The result of those efforts, and applications of additional dating methods, has been a vast, expanding, contentious literature that remains inconclusive. Why?

In the decade of the 1970s, charred organic samples from the Akrotiri excavations were sent to the radiocarbon laboratory at the University of Pennsylvania, a respected research facility. The immediate results supported the traditional age, but tree-ring calibration produced dates implying an age greater than 1600 B.C. (Fishman et al. 1977). Efforts to explain the results focused at first on contamination by ancient carbon in volcanic gases venting nearby (Weinstein and Michael 1978).

As these issues were being pondered and contested (Cadogan 1978), other dating methods were applied. Archaeomagnetic studies on burned buildings in Crete and the tephra on Santorini showed diverse orientations: the destructive events on the two islands could not be coincidental in time (Downey and Tarling 1984; Liritzis 1985; Sparks 1985; Tarling 1978). Voices from western North America chimed in with a hypothesis for dating by teleconnection through climate proxy. Studies of tree rings in the California mountains disclosed notable frost rings at 1628/1627/1626 B.C., which were interpreted as evidence for a major volcanic eruption in the northern hemisphere. Those were correlated to Thera on the basis of some of the contested radiocarbon dates (LaMarche and Hirschboeck 1984). Although immediately assailed from the perspective of artifact cross-dating (Warren 1984), the claims soon attracted supporting evidence from tree-ring sequences in Ireland and Germany (Baillie and Munro 1988).

Geological stratigraphers joined the discussion, reporting on the distribution and relative dating of deposits of tephra from Thera as observed in marine cores and on land eastward into Anatolia (Turkey), which provided estimates of the size of the eruption and implied the possibility of more refined stratigraphic dating (Stanley and Sheng 1986; Sullivan 1988; Watkins et al. 1978). From the Dye research station on Greenland, ice-core studies produced another set of interpretations, building on the climate proxy argument begun by the tree-ring researchers (Hammer et al. 1987). These authors championed 1645 B.C. as the date for a major volcanic eruption with a strong sulfur signal observed in ice layers, retracting a later age they had proposed in 1980. By the late 1980s the radiocarbon debate had heated up again as more organic samples were distributed to several different laboratories. The conflicting results of the cross-laboratory tests pleased no one (Aitken et al. 1988; Betancourt 1987; Betancourt and Michael 1987; Manning 1988; Warren 1988).

There matters stood as the Third International Congress on Thera and the Aegean World convened in 1989 to consider the evidence. Conferees reported on radiocarbon analyses of sample sets carefully selected to avoid some of the problems raised earlier, which nevertheless produced similarly scattered results tending toward ages in the seventeenth century B.C. rather than the sixteenth (Friedrich et al. 1990; Housely et al. 1990; Hubberten et al. 1990; Nelson et al. 1990). Neither AMS dating of single seeds, the involvement of several laboratories, nor calibration of multiple dates clarified the situation. Champions rose to urge compromise on demands for precision, and to consider the ice-core and tree-ring ages to date the same volcanic event, which might be the Thera eruption. Their mediating efforts were properly rejected as compromising accuracy as well. Other voices argued against using the ice-core and tree-ring dates because the climate proxy argument could not be tested; volcanoes other than Thera might have produced sulfur veils and chilly climates in the centuries at issue (Pyle 1990).

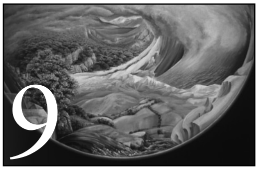
The archaeologists considering the dating problems were incompletely familiar with geological advances in understanding the eruptions. At the conference stratigraphers called attention to evidence for several chemically and petrologically distinct tephra bodies on Santorini, and noted that tephra deposits elsewhere differentially matched more than one of them. Geologists demonstrated that the Thera volcano had erupted at least twice, with the possibility that the resultant ash deposits could be confused with one another (Hardy et al. 1990). The sequence of events as well as their synchroneity – eruptions, earthquakes, fires in buildings, destruction of buildings, sources of organic samples – was therefore exposed as an unresolved issue given the stratigraphic, archaeomagnetic, and radiocarbon complexities and contradictions (Housely et al. 1990). As the conference ended, consensus remained elusive despite significant gains in clarification of the arguments advanced (Manning 1990).

Since the Third Congress, new evidence and argument from high northern latitudes has called into question much of the debates that took place there. Investigators of paleoenvironments in Iceland published a strong caveat against confusing coincidences with correlations or correlations with causes, highlighting the unrealistic efforts of archaeologists to achieve calendrical precision through methods incapable of yielding it (Buckland et al. 1997). Tracing sulfur deposits in the GISP 2 core in Greenland, Zielinski and Germani (1998a) supplemented the seventeenth-century B.C. sulfur signal by extracting tephra sherds from the same ice layer and analyzing them chemically. They found no match for any Thera tephra, nor for those of four other eruptions close in time. Their responsible conclusion – the results neither support nor refute the relevance of the age of the sulfur signal for the Thera eruption – has survived one immediate challenge (1998b). Furthermore, research in Anatolia on tree rings and ash deposits has not yet clarified the situation, despite an attractive, but ultimately incomplete, argument supporting the early date from tree rings (Kuniholm et al. 1996). Inconclusiveness is properly a stimulus for research, not an embarrassment.

Why is this so important? Why have so many excellent investigations been directed to this enigma? The entire east-Mediterranean Bronze Age chronology rides on the results, since the validity of the traditional chronology based on links with Egypt is now strongly challenged. If LMIA is earlier than 1500 B.C., the entire archaeological scenario for the Bronze Age must be extensively revised and lengthened, with implications for connections in all directions.

Why has resolution been so elusive? When otherwise reliable methods fail to provide information, one must evaluate the assumptions that guided the choice and application of the methods. Even after the Third Congress eliminated some old assumptions, such as the reliability of the 1500 B.C. cross-dated age, other assumptions appear to be obstacles to resolution. The paleomagnetic and stratigraphic studies highlighted uncertainties in the sequence of events and assumptions of synchroneity; more rigorous stratigraphical analyses or new excavations on both

Santorini and Crete are needed. How many volcanic events, and of what kinds, affected the region? What was the magnitude of the eruptions, and can they be positively identified in Greenland ice cores? Do boundaries between archaeological periods relate in any way to external, natural events? Were the Cretan fires lit by human or natural agencies, and what was the sequence? How do stored seeds relate to the volcanic events at Akrotiri? Essential to any dating effort is a clear idea of what is to be dated, accompanied by rigorous selection of the appropriate methods and samples. An awareness of unforeseen complexities in human affairs is also necessary.



9

LANDFORMS

> Geology . . . is a historical science concerned with past configuration of the Earth, dealing with successions of unique strictly unrepeatable events through time.
>
> HALLAM 1981: 11

For as long as the Earth has had an atmosphere, its surface has been continually shaped by air, water, and ice disaggregating, transporting, and depositing mineral matter. Human lives are lived on surfaces, but archaeological surfaces are not necessarily those on which the archaeologist walks. Depending on their age and situation, ancient surfaces have been buried or lost to erosion.

Conceptual reconstruction of past landforms and surfaces is an essential aspect of modern archaeology because the spatial context of a site is crucial to its interpretation, and to understanding its relationships to other sites. Space and landscapes define the resources available to any human group, and landform changes through time are related variously to changes in other elements of the environment – climate, hydrography, and biota (mainly vegetative). The mechanisms by which landforms, as elements of the geosphere, are shaped by processes originating in the hydrosphere, cryosphere, atmosphere, and biosphere are imperfectly known although believed to be determinable. Because of these interdependencies, landform reconstruction informs about the past states of variables in all five spheres. However, as with all complex dynamical systems, our ability to predict future states or understand past states and conditions is limited by the element of chance influencing combinations of mechanisms in several scales of space and time.

The "reconstruction" of ancient landforms is not done with earth-moving equipment; it is, rather, a conceptual exercise based on the study of remnants available for observation in the present. The results are perspective drawings, maps, or descriptions of landforms whose past shapes, sizes, and positions in space are approximated.

Unlike past climates, the reconstruction of which is dependent on proxy evidence, landforms can leave within the geosphere some accessible evidence of their former states, evidence that can be found by applying the methods and insights of geomorphology. For example, field prospection and recording of tangible remains provide evidence of past states of the surface of the Earth. **Sediments** preserve some records of past landforms, accompanied by evidence about former states of the atmosphere, hydrosphere, cryosphere, and biosphere. All these kinds of data are, in turn, used to reconstruct former landscapes with their climates, hydrography, and biota.

As geological data have earned a more central place in archaeological interpretation, a new specialty has emerged within archaeology, that of the geoarchaeologist. Self-identified geoarchaeologists have tended to be much more interested in sediments than in landforms (e.g., Waters 1992). When geoarchaeological study is taken beyond the immediate site locale and into the realm of past landforms, the assistance of an archaeological geologist is typically required – a geologist with a predilection or tolerance for problems at archaeological scales. The two combined terms indicate that in each case field workers are crossing disciplinary boundaries and dealing in part with concepts, terminology, problems, and scales that are exotic to their primary training (Butzer 1982; Rapp 1975; Stein 1993; Thorson 1990b). Archaeologists typically gain more from the expertise of a geomorphologist working with them than the geomorphologist expects to gain from the archaeologist, because the small-scale resolution and historical particularity which the archaeologist seeks may be less familiar or even useless for the research geomorphologist. Archaeologists should cultivate assiduously those geomorphologists whose interest in rates and processes incline them to work on Holocene phenomena, and who therefore can potentially benefit from the fine spatial and chronological resolution intrinsic to much archaeological field research.

Collaboration can be successful only with a great deal of discussion, cooperation, mutual goodwill, and intense communication between the investigators. "[A]rchaeologists and geologists, to work effectively together, must be aware of each other's values and paradigms, and of the strengths and limits of their respective data sets" (Thorson 1990b: 33). As part of their own preparation for field research, archaeologists should know at least something about how geologists and geomorphologists think about the Earth and approach its study. Control of basic geomorphological concepts and vocabulary by archaeologists will facilitate communication and enhance cooperative research.

INTRODUCTION TO GEOMORPHOLOGY

Geomorphology, the study of landforms, has developed from being mainly a qualitative and descriptive discipline to being a science based on a systems paradigm and quantitative methods. Like other sciences, it is moving beyond mechanistic systems models toward the complexities of process–response models for morphogenesis. Concepts of scale and process currently dominate in explanations and research designs, succeeding emphases on form and age. Advances in instrumentation have made possible the study of a larger number of variables in both the field and laboratory. Landscape evolution involves complex combinations of processes working at different scales of space and time; the major landforms take shape over time durations in the 10^6–10^7-year scales. Observation, on the other hand, rarely can be undertaken over more than 10 years and in limited areas; such prolonged or repeated studies are still exceptional. Theory in geomorphology, therefore, currently lags observation and mathematical modeling; complex four-dimensional problems (the fourth dimension being time) are not yielding readily to explanatory models built in two and three dimensions. The assumptions and premises underlying geomorphological studies, once implicit, are being explicated, challenged, modified, and replaced (Schumm 1991; Thorn 1988).

Much research in geomorphology has been undertaken in order to describe and understand paleoclimates by using landforms as proxies – the literature of glaciology and Quaternary geomorphology especially is dominated by such concerns. The link to climate history, particularly strong in Europe (e.g., Summerfield 1991: Table 1.4; Tricart and Cailleux 1972), originated with the discovery of terrestrial ice ages through interpretation of landforms on the basis of alpine glacial analogies. More recently, appreciation of the number and complexity of variables involved in landform processes, and the realization that cycling between stable and unstable states can occur at different scales and involve time lags, has tempered expectations of reading the details of climatic changes from landforms. To the extent that landforms can serve as climate proxies, they must be used as such in combination and complementation with proxies from different data domains, and with comparative data sets at a range of scales (Bull 1991). Research directed toward the measurement, explication, and interpretation of land modification processes has exposed the scope of our ignorance about the critical variables and the measurement of rates. Investigators have turned to laboratory and computer models in hopes of identifying the critical variables and vectors in systems at different scales. These developments make modern geomorphology more valuable to archaeologists because of its enhanced explanatory powers and finer chronological resolutions, at the same time that it becomes less accessible because of its increased technicality (Chorley et al. 1984; Goudie 1981; Thorn 1988; Summerfield 1991).

For archaeologists, ancient landforms are of interest as the locales and geographical contexts of sites – the home spaces and habitats of human communities. At large scales, landforms define the physiography and other elements of the environment in which human communities exist, and in many respects the climates and resources available to them. At intermediate scales, landforms constrain communication and travel, and to various degrees determine both the state and condition of the biotic resources that are available to humans. At small scales, sedimentary landforms may comprise the matrix or physical context of sites, and thus must be understood for chronological control. Human modification or construction of landforms is often evident at small and medium scales, and may require interpretation. Description and interpretation of small-scale landforms in terms of formation and deformation processes become crucial for understanding the location, integrity, and natural history of a site. Understanding landforms at all scales aids in correlating a site location with other coeval surfaces across space, and thereby interpreting the spatial component of human behavior.

SCALES OF LANDFORM ANALYSES

In conformity with the discussion of scales presented in previous chapters, we can consider landforms within a nested series of four scales (Table 9.1). At the mega-scale, global or hemispherical landforms lack immediate relevance for archaeologists, although they are important, even prominent, in paleoclimatology. Mega-scale landforms of continental size, involving areas of 10^7–10^8 km^2, are relevant as background to human evolution and behavior, and to large-scale climatic phenomena. Macro-scale phenomena include physiographical provinces involving areas ranging between 10^4 and 10^7 km^2; they have direct relevance to human territoriality, resource exploitation, and communication and are perceptible components of habitat. The typical archaeological scales of region and locale, with areas as large as 10^2–10^4 km^2, belong in the meso-scale division, leaving the micro-scale for very local and site-scale phenomena.

Landform analysis and reconstruction can be undertaken at any of these scales, each having its particular degree of archaeological relevance. The processes and variables that define landforms vary with the scale under consideration (Table 9.2). Continental-scale landforms reflect mainly crustal structure; province-scale landforms reflect mainly the intensity and duration of erosional planation and tectonism. At smaller meso-scales, the diversity and number of relevant processes and

Table 9.1 Landform scales

Spatial scales	Area (km^2)	Illustrative landforms
Mega-	global: 5.1×10^8 continental: $<10^8$	(geoid) continents; ocean basins
Macro-	physiographic province: 10^4–10^7	mountain ranges; continental glaciers; major drainage basins
Meso-	regional: 10^2–10^4	sand seas and loess sheets; river basins; volcanoes; karst terrains; fault zones
	locality: 1–10^2	small volcanoes and lava flows; river floodplains and terraces; minor drainage basins; dune fields; glacial valleys; mesas; arroyos
Micro-	local: <1	river channel features; glacial kames and minor moraines; periglacial features; beach ridges; buttes

variables become very large, while at the scale of microrelief, definable short-term processes dominate.

The analyst's ability to isolate the relevant variables, and consequently the completeness and accuracy of the resulting analysis or reconstruction of landforms, is likely to be best at the largest and smallest scales. This rule holds for two basic reasons: (1) at the extremes of scale, fewer variables control the processes, and (2) the variables are likely to be comparable in scale themselves (clearly either large or small) and therefore to present similar measurement or sampling problems. We can illustrate this best by looking at the extremes. Phenomena at the continental scale reflect mainly atmospheric and structural states of long duration and low frequencies. The controlling variables are, therefore, few in number and among the most stable and regular processes involved in landform development. At micro-scales, on the other hand, landforms are very closely responsive to the small-scale and high-frequency processes that shape them in the short term; the identification of relevant variables is least complicated because they can generally be observed directly. Small-scale landforms deserve close attention by archaeologists because they are likely to be unstable and therefore to reflect conditions of the present rather than of the past.

For archaeologists, these observations bring both cheer and disappointment. Geomorphological analyses can be very helpful for the particularistic kinds of problems posed at micro-scales, to define and interpret landforms supporting and

Table 9.2 Landform processes at four scales

Spatial scales	Area (km^2)	Selective morphogenetic processes
Mega-	global: 5.1×10^8	plate tectonics
	continental: $<10^8$	crustal structure; tectonism
Macro-	physiographic province: 10^4–10^7	orogeny; tectonism and faulting; continental glaciation; hydrography; erosion; climate
Meso-	regional: 10^2–10^4	erosion; tectonism and isostasy; hydrography; volcanism; aeolian deposition; river-basin development; chemical solution; glaciation and deglaciation; regional climates
	locality: 1–10^2	volcanism; fluvial processes (floodplains and channels); gullying; fans and deltas; aeolian deposition (dunes); valley glaciation; glaciofluvial processes; slope processes; mass wasting; seismicity; isostasy; local climates
Micro-	local: <1	small-scale fluvial processes; glaciofluvial processes; periglacial processes; volcanism; beach processes; solifluction and gelifluction; seismicity; microclimates

Note: The duration and age factors vary by scale; the processes expressed in large-scale landforms influence form over longer time spans than do those affecting the smaller scales. The processes effective at each descending scale are to a greater or lesser degree dependent upon those at larger scales, which set parameters. No priorities are implied by the order within lists.

enclosing archaeological sites. When attention moves to the meso-scale, the habitats of human communities, description will be more easily and less equivocally attained than will interpretation. It is specifically at the meso-scales, where human activities are diverse and adaptationally crucial, that the number and complexity of environmental variables affecting landforms is also greatest, reducing seriously the analytical resolution obtainable with the methods and concepts currently available to geomorphologists (Stein 1993). This is not a brief for avoiding the issues; the need for landform analysis to support archaeological investigations at meso-scales is very great, and neither technical nor theoretical difficulties pose insurmountable obstacles. At these scales, human influences on landforms are often recognizable, while some landforms, particularly glacial and glaciofluvial features, may be mistaken for human artifacts. It is especially at the challenging meso-scales where the relatively fine-grained resolution of archaeological chronologies can be of most help to the field geomorphologist, and where archaeologists, therefore, can help to repay their debt to their colleagues, potentially advancing both disciplines in the process.

Archaeologists excavate sediments and erosional disconformities that must be interpreted in terms of the processes that formed them. Were the critical changes at regional or local scales? Were they triggered by climate, land use, or other factors? What spans of time and what ages were involved in the cycles? How did the changes affect human societies? How did human responses in turn affect erosive cycles? How have archaeological sites been preserved under sediment or destroyed by erosion, and where? All these issues and more come into play in archaeological investigations; the range is brilliantly exemplified in Karl Butzer's study of the ancient site of Axum in Ethiopia (Butzer 1981a). Chapters in Bell and Boardman (1992) and Wagstaff (1987), displaying archaeology's capability to contribute to the history of landforms, should inspire archaeologists to master the concepts and language of geomorphology. The need is being met by new publications addressed to archaeologists (Herz and Garrison 1998; Rapp and Hill 1998).

PROCESSES AND CONCEPTS IN LANDFORM ANALYSES

Among the five spheres of the climate system, those most important in geomorphology are the physical, abiotic ones. Landforms result from the action of air, water, and ice on the geosphere. The biosphere responds to and modifies the products of the other spheres.

Geosphere

Chapter 3 introduced plate tectonics, orogeny, isostasy, and eustasy to explain the distribution of continental crust on the face of the planet, its elevation above sea level, and the location and form of mountain ranges. The crustal plates are subject to additional constructional processes at smaller scales, principally in the form of volcanism. Volcanoes, individually or clustered, form mountains very dissimilar to folded and faulted ranges. Smaller landforms related to eruptions of ash or lava may be prominent at meso- and micro-scales. Lava flows can extend over hundreds of square kilometers, inundating older landforms under a hardening sea of molten rock.

Starting with these constructional geological features in place, geomorphologists are fundamentally concerned with the resulting surface **relief** – a relative measure of surface ruggedness that expresses the height difference between the highest and lowest spots in a given unit of land: the vertical distance between hilltop and valley bottom. This fundamental expression of landform is analyzed in terms of smaller segments of form such as **slopes** (surfaces slanting at angles typically less than 45°), **scarps** (steep slopes or cliffs), **basins** (essentially concave landforms usually encompassing the catchment of a stream network), and **channels** (linear depressions; beds of streams or rivers). It should be clear from the definitions that basins incorporate most of the other forms; in fact, basins are considered the "basic geomorphological unit in many terrains" (Chorley et al. 1984: 316). It is at the scale of basins that the fundamental destructive processes of erosion, discussed below, are observed.

Atmosphere

The solidity attributed to rock is a prevailing simplification, even misrepresentation, of the fundamental material of the continental surfaces. Much rock is actually aggregates of crystalline matter or of small sedimentary elements such as microfossils or grains of sand or silt. Such aggregates are penetrated by joints and cracks that permit water vapor to infiltrate the mass and begin the process of disaggregation. Obsidian is supercooled liquid (glass), which tends to crystallize and weaken as it ages. The transition from rock to sediment begins when rock is exposed, at or near the surface, to the influence of the climate system. Cycles of warm and cold temperatures, high and low precipitation, and atmospheric pressures lower than those characteristic of environments where rock formed, result in physical and chemical changes in the minerals constituting the rock. All rock is susceptible to disaggregation by either mechanical or chemical **weathering**, most by both.

Mechanical weathering includes those processes that physically break up rock, by wedging it apart or reducing it by abrasion or glacial plucking. Mechanical weathering includes such processes as frost wedging, caused by the expansion of freezing interstitial water; differential thermal expansion and contraction of minerals subjected to extremes of temperature; salt wedging caused by the crystallization of saline interstitial water or chemically transformed minerals; and root wedging, caused by the growth expansion of rootlets, which prise apart minute cracks in rock. Chemical weathering is effected by water, in liquid or vapor form, which dissolves and ultimately carries away soluble minerals or mineral compounds. Rock disaggregates into its constituent minerals or insoluble residues, which may remain in place or be carried away by gravity or moving water, ice, or wind. Sediments created by redeposition of weathering products are discussed in Part V. The technical term for unconsolidated mineral matter lying above bedrock, whether remaining *in situ* or redeposited, is **regolith**. The erosion of regolith by wind, water, ice, or gravity results in destructional landforms; redeposition creates constructional landforms.

Still air, by itself, can do little work other than serving as a medium from which water vapor can be deposited onto surfaces. However, air in motion – wind – can carry mineral particles and be an effective erosional agent, shaping and even creating landforms. Wind is capable of transporting particles ranging in size from the finest silts and clays to small buildings (as in tornadoes). The stronger the wind speed, the larger the particles moved. The smaller the particles, the more likely they are to be transported long distances in suspension. Reduced wind speed deposits particles in order of their size: largest first. **Aeolian** deposits are therefore composed of materials well sorted in terms of particle size, although they may derive from different places and distances.

The largest aeolian landforms are **loess** plains: deposits of well-sorted medium- to fine-grained silt (2–64 µm) that may blanket prior landforms to great depths. They typically form in mid-continental areas downwind of sediment sources exposed by extreme aridity or cold. The largest formations are products of the cold, dry climates of glacial periods. The high winds characteristic of glacial margins mobilize the silts and fine sands and transport them varying distances until the wind velocity is slowed by distance from the glacier, by encountering vegetation, or by rising elevation. Being derived from freshly weathered bedrock, aeolian silts are typically calcareous and highly fertile if watered; in temperate climatic regimes, the silts readily support vegetation, stabilize, and develop soil profiles.

Dunes are smaller constructional forms, although dune fields may be of immense size. Dunes form where there is a source of medium to fine sand (64 µm–2 mm), wind velocities sufficient to move it, and some factor suppressing vegetation. When sand-laden winds are slowed by surface roughness, they deposit sand. These conditions typically occur in zones of arid climate, whether warm or cold, or along seashores where sand is plentiful and strong winds and salt air suppress vegetation. In temperate climatic zones, old relict dunefields may be remobilized and reshaped if the vegetation cover is destroyed by catastrophe or human abuse. Flat sand sheets form when winds carrying medium sands are slowed with less turbulence than that resulting in dunes.

The sediment sources from which winds remove materials are generally less dramatic than the depositional landforms. Winds move finer sediment grades from exposed deposits, leaving behind coarse **lag deposits**. Winds mobilizing previously sorted deposits, such as dry river channels, dunes, or sand sheets, create hollows called blowouts. Both lag deposits and blowouts are problematic for archaeologists since archaeological materials exposed in them lie on surfaces below those on which they were originally deposited and may be associated with materials deriving from more than one episode of deposition.

The intimate relationships between climate and landforms are the subjects of climatic geomorphology, a subdiscipline which, while losing favor as an approach to paleoclimatology, has much to offer for the understanding of geomorphological processes. While climate is undeniably a strong influence on landforms, it cannot be shown to be predominant. Landforms are polygenetic, the result of many different factors interacting over time spans of varying lengths with changing climatic regimes (Bull 1991). The number of relevant variables grows as measuring techniques and models are refined. Some correlations of landforms with climate are strong, as with the destructional landforms created by glaciation, the buttes and mesas of dry semitropical mid-continental areas, and the slopes and valleys of wet tropical areas thickly mantled with unconsolidated weathered sediments. However, no correlations between climate and landforms can be assumed strong enough to permit confident retrodiction of one from the other.

Hydrosphere

Water exists in the hydrosphere in liquid and gaseous states, with the liquid form episodically moving or depositing sedimentary matter. Increases or decreases in the annual amount or seasonal distribution of precipitation can initiate or reverse cycles of erosion and deposition, modifying landforms.

The potential of moving water to do work varies with the steepness of the land (slope or "gradient"), the compactness and roughness of the surface, and the volume of water; together these factors define the speed ("velocity") and thus the force of the water – its ability to erode materials and to carry them away. Moving water carries material in three forms: dissolved load, suspended load, and bed load. Dissolved load is derived mainly from products of chemical weathering of rocks and sediments, and secondarily from ions brought down by rainwater. Suspended load is the fine particles of clay and silt that remain in suspension in moving water, while bed load is the heavier particles that are slid, rolled, and bounced along the bottom of a stream. Bed load moves best when water is flowing at high volume and speed, and thus is transported episodically, while suspended and dissolved load may be carried almost continuously downstream. Both the amount and maximum size of material moved vary geometrically with velocity; small increases in velocity greatly increase a stream's ability to move material. That is why even very small and gentle summertime brooks may have bouldery beds that testify to powerful spring freshets.

Slope processes

Unconsolidated material on slopes moves under the influence of gravity, but most such moving material has been dislodged or lubricated by water. Mass wasting occurs in many forms and scales, from rapid massive landslide to slow soil creep; the latter is implicated as the dominant process in surficial geomorphology. In all cases, material is dislodged and moved down slope where it is more accessible to erosion by surficial water. Archaeological materials move in the same ways, being sorted by size as they are moved, separated and recombined in new associations, and suffering the destruction or even reversal of their stratigraphic relationships (Rick 1976). While the loss of archaeological integrity in landslides is rarely an issue, mudslides and **solifluction** (saturated sediments moving downhill), as well as surface displacement by creep, have presented daunting problems of archaeological interpretation and many opportunities for misinterpretation. The sediments resulting from slope processes are typically mixed, poorly sorted, and difficult to interpret when old. To express uncertainty responsibly, geomorphologists use the term **diamicton** for redeposited regolith which lacks sedimentary structure and sorting and for which, therefore, the generative processes cannot be demonstrated. The term carries no connotation of origin; it applies equally to landslides, some glacial deposits, volcanic mudflows, **gelifluction** (sediments moving over frozen substrates), and so on. Diamicton is an excellent example of an ambiguous word that increases precision of expression at the expense of accuracy; it perfectly expresses "I don't yet know what caused this mess."

Fluvial (river) processes

Water, being both heavy and shapeless, responds strongly to gravity, coalescing into streams that follow the lowest surface irregularities. Over the years and over most of the planet's surface, water has been the force shaping landforms. The water available for erosion and transportation of weathering products varies in space and time according to its receipt of precipitation and loss of water vapor into the air through evaporation and **transpiration** (exhalation of water by elements of the biosphere, mainly vegetation). As precipitation increases, or evapotranspiration diminishes, more water is free to run across and through sediments, redistributing and reshaping them (Fig. 3.6).

The structure, climate, and geological history of a physiographic area together define its unique properties. All of these factors influence the **hydrography** (the pattern of drainage and drainage basins). Streams flowing on the surface of the ground do not create an infinite variety of possible patterns; rather, a few generalized classes of patterns subsume the observed variation (Fig. 9.1). Stream patterns, observable on large-scale maps and aerial photos, are strong clues to fundamental

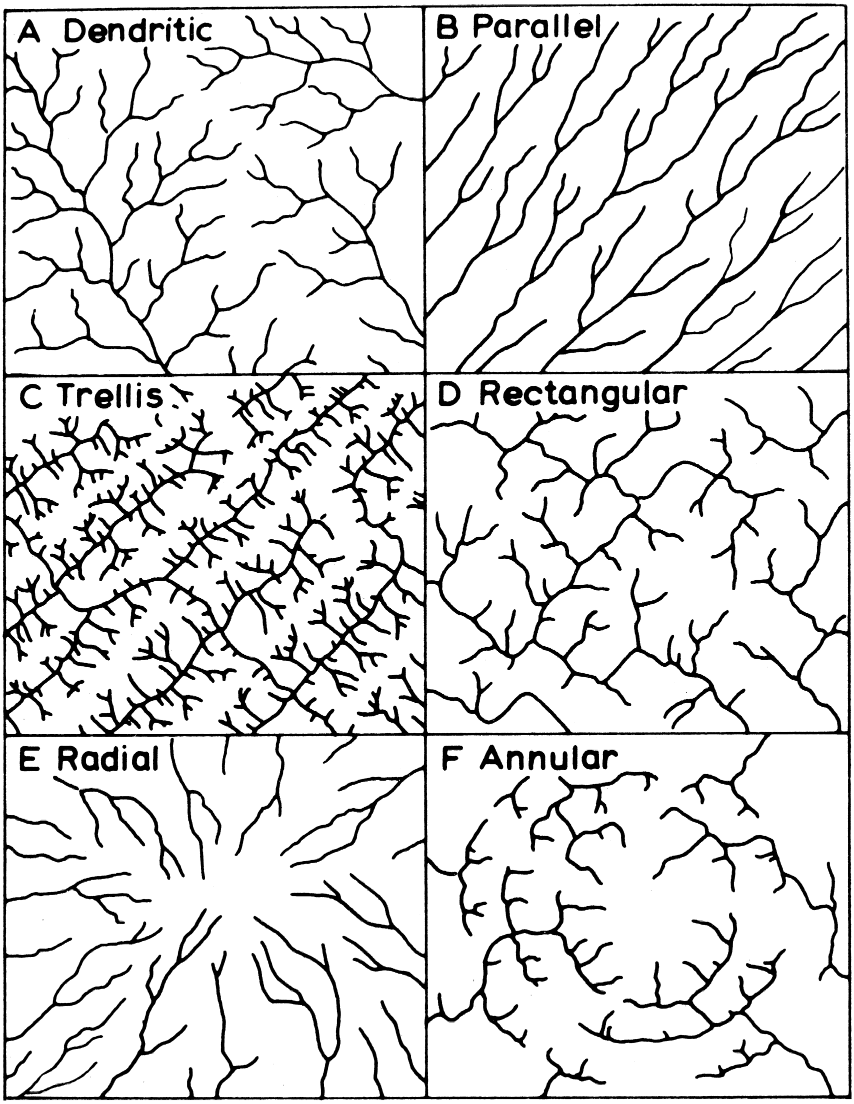


Figure 9.1 Selected stream pattern diagrams: dendritic, parallel, trellis, rectangular, radial, and annular. Dendritic patterns form on mildly sloping terrain; parallel on steeper slopes without strong bedrock controls. Trellis and rectangular patterns typify uplifted eroded folds (see Fig. 9.5, "fold mountains"). Radial patterns form on domes and volcanic cones, annular patterns on eroded domes of stratified bedrock (see Fig. 9.5). (Reproduced from Chorley et al. 1984: Fig. 13.3, with permission of Methuen & Co., publishers.)

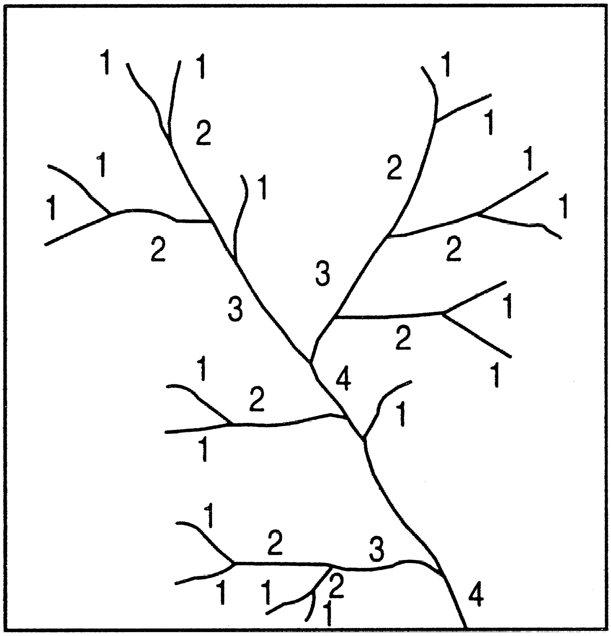


Figure 9.2 Strahler stream ordering begins with the streams initiating surface flow, which are "first order." Higher order streams are formed only by the confluence of equal orders. Where two first-order streams meet, they form a second-order stream. Two second-order streams form a third, and so on. (After Catt 1988: Fig. 2.33.)

underlying lithological structures at province and regional scales. Drainage that has been disrupted by glacial deposition comprises the major exception to the patterning rules, and may itself thus be readily interpreted.

The smallest, most numerous streams in a drainage basin, or catchment, tend to occur on the higher ground, while fewer large streams dominate the lowlands. This tendency can be quantified, and turns out to display mathematical regularities that are used to describe and compare different stream basins. The number of streams of each relative size class, or **order**, and the proportion of basin area that they drain, have regular relationships to the other orders, varying mainly with the surface relief. In this way, qualities of drainage basins which are significant to human settlement choices (slope, elevation, water availability) can be expressed as indices that are directly comparable from basin to basin. Among several available descriptive systems, the Strahler system is widely used and the most suitable for archaeology (Fig. 9.2).

Fluvial landforms are produced by dynamical responses of flowing water to changes in volume or to changed conditions of slope or sediment load (Leopold

1994). The fundamental forms are channels, the linear routes of normal river flow, and **floodplains** – the areas subject to periodic overflow and sediment deposition when there is more water than the channel can accommodate. Channels exist in three basic forms: straight, sinuous, and braided, the particular form determined by variables such as slope, water volume (discharge), load type, and bank sediment composition. Each looping curve of a sinuous channel is a **meander**. Erosion typifies the convex outer curves of sinuous channels. The inner curves are areas of deposition where **point bars**, relatively coarse fluvial deposits, form downstream from locations of channel cutting, as the water is slowed by turning. Braided streams are fluvial responses to steep gradients and heavy bed loads; they wind across broad beds, alternately depositing and remobilizing coarse load components (Chorley et al. 1984: 309, 349). **Levees** are linear fluvial deposits paralleling river channels; they form as floodwaters overflow banks and deposit their coarsest load where the velocity slows rapidly.

Channel gradient, the slope from the highest elevation of freely running water to base level, varies along a stream course but is continuous; water does not run uphill. Base level, the lowest point to which a channel segment can erode, is established by effective obstructions to erosion such as rock outcrops, standing water in ponds and lakes, or sea level. Obstructions in a channel cause a change in slope; the stream will pond and deposit sediment (aggrade) until it can flow over the obstruction. The new deposit decreases the gradient immediately upstream of the obstruction. Lowering of a base-level control can occur by tectonic adjustment, erosion of obstructions, or lowering of sea level. Rivers respond to base-level lowering by eroding (incising), which steepens the channel gradient upstream from the changed condition. Incision continues upstream until blocked by another base-level control. Aggrading or incising streams change the form and gradient of their channels as part of their response to new conditions of flow – faster and steeper or slower and flatter. The new conditions may destabilize river banks and adjacent slopes, bringing more sediment into the stream, thereby increasing the stream load or clogging the channel. The response of a stream to a change in base level at one place is expressed both upstream and downstream from the change; matters are further complicated by compensations triggered in tributary streams that join a stream under adjustment.

"Natural systems are inherently complex" (Schumm 1991: 85). River responses to base-level change can be extremely complex, with aggradation and incision cycling until a new state of stability is achieved. The concept of "complex response" in hydrography as developed by Schumm (1977) is crucial to understanding the dynamics of fluvial systems (Fig. 9.3). It is also a fundamental concept for natural systems of

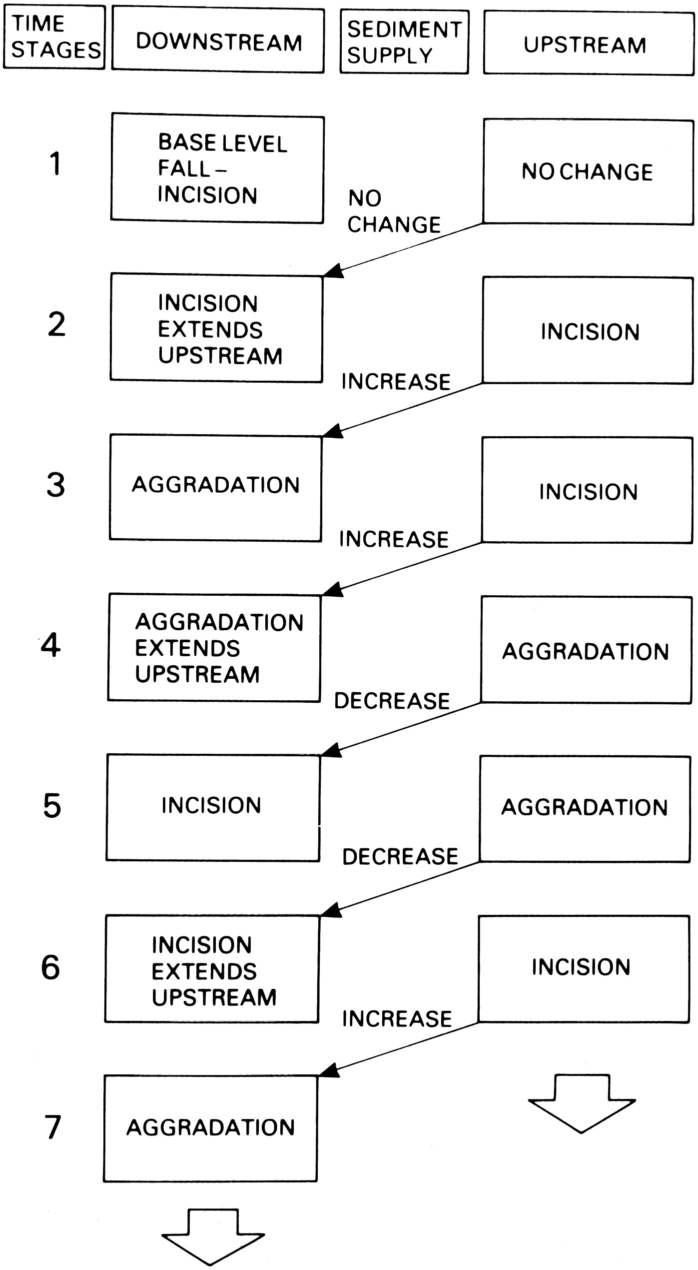


Figure 9.3 Complex response in a drainage system with fall of base level. Adjustments in regime work both upstream and downstream from the point of change, as the gradient steepens or flattens and the velocity of the flow is affected. (Reproduced from Summerfield 1991: Fig. 9.27, with permission of Addison, Wesley, Longman.)

all kinds. The complexity of response to perturbations is due to the combination of external and internal triggers of instability in a system. Systemic compensation for external triggers is likely to trip internal triggers, which in turn entail further changes within the system, and so on (Leopold 1994).

Stream responses to changes in gradient or water volume affect the width and the relief of valleys and the elevation of valley floors. Stream incision leaves floodplain fragments on valley walls above a new floodplain. For instance, a glacio-eustatic lowering of sea level will increase the gradients of all rivers running into the sea, causing them to incise their channels beginning at the mouths, leaving former floodplain segments elevated as **terraces**. Cycles of adjustment result in the creation of step-like terraces along valley walls – abandoned fragments of former floodplains adjusted to different conditions. Terraces may contain bedrock components, but most are composed of alluvial sequences, sometimes including older buried surfaces. Terraces are numbered up from the active floodplain, 1–*n*, from youngest to oldest. The sequences on facing sides of a valley may not be equivalent. Paired terraces form when incision dominates the valley-formation processes. Rivers meandering widely are likely to cut unpaired terraces at staggered levels on valley sides (Chorley et al. 1984; Summerfield 1991).

A stream adapting to an abrupt change in channel gradient will deposit some of its load at the point of change. The typical landform where gradient flattens abruptly is an alluvial **fan**, a triangular-shaped pile of complexly interbedded sediments over which the stream flows in braided channels. Alluvial fans are prominent features of semi-arid climates at the foot of mountain ranges, but they can form wherever gradient change is too abrupt for smooth transitions (Summerfield 1991: 224–225). In arid landscapes, fans may be important aquifers, with springs emerging at the margins or foot of the fan. Humans attracted to the springs are likely to leave archaeological deposits on the surfaces or edges of fans where they may be rapidly buried by sediments to form stratified sites.

In arid climates where precipitation is seasonal and episodic, river beds may be dry for large parts of the year. During the monsoon season, or following storms, these channels may carry impressive and destructive flows of water. Seasonally dry stream beds, often littered with boulders that look strangely out of place, can fill very suddenly with water, as many an unwary camper has learned the hard way. When active, these streams can incise or aggrade their beds very quickly, changing shape and sometimes course. In the Near East these features are called "wadis," in the Americas and Spain, "arroyos."

In contrast to superficial fluvial channels, water may flow underground through channels in bedrock. In areas of adequate rainfall and limestone bedrock, extensive river systems may be created by chemical dissolution of carbonates, running for miles through tunnels and caverns. When solution of the bedrock is far advanced, cavern and tunnel ceilings may collapse, opening the subterranean landscape to the surface and creating a **karst** topography, with or without underground caves. Maturely eroded karst landscapes are characterized by finely scaled and abrupt changes in relief, expressed either as depressions below the surface or as rock pinnacles rising above the levels of streams and floodplains. Caves, rockshelters, underground rivers, and dolines (closed depressions in a limestone surface, not components of fluvial networks) are typical features of karst terrains (White 1988). Limestone caves, which provide spectacular non-architectural archaeological sites, focus occupation and debris deposition spatially, and provide excellent preservation conditions for bone and antler artifacts and other remains. Repeated use of the same limited space forms stratified sites of great structural complexity.

Cryosphere

We are still in an ice age; at no time during the existence of the human species has the globe been free of ice. At the last glacial maximum, three times as much land was covered by ice as is now the case. Over a typical year today, snow and ice cover varies seasonally from ca. 8% to 16% of the planetary surface (Bradley 1985: 19).

The properties of glacial ice that make it an effective agent of erosion are its mass and its motion. The mass is the result of many years' accumulation of snow, compressed under its own weight and become ice. With continuing accumulation, the mass deforms by plastic flow, expanding outward toward areas of less pressure. The outward motion is channeled by any unevenness in the ground beneath the glacier. If snow continues to accumulate on it, the mass may grow upwards to the point where it constitutes an area of high elevation regardless of the relief or elevation of the land below; the ice may then flow "uphill" over opposing slopes. Water may exist under the ice, contained at high pressure. High pressure raises the freezing point and increases the erosional capabilities such that subglacial water is a significant agent of erosion. Mineral matter in glacial water and in moving ice increases the erosive capabilities of both. Gritty ice, moving over bedrock, grinds away rock surface, especially any that is weathered. Basal water infiltrates cracks in rock or unconsolidated sediment below the ice; if it freezes, mineral matter is incorporated into the moving ice, "plucked" up, and carried away. The leading edge of an ice sheet shoves unconsolidated sediments like a bulldozer, scouring the landscape down to solid rock. In ice, on the ice surface, in water, or in front of the moving edge, debris of all sizes is transported from its original location.

Rock debris and sediments incorporated into glaciers are eventually deposited as a glacier overrides its own excessive bed load or melts away. Material deposited from ice is called **till**, poorly sorted sediment of two basic types: till deposited directly at the base of a glacier is called lodgement till, while that deposited as a glacier melts away is ablation till. The latter may be subdivided on the basis of evidence for meltwater involvement. Other common names for till, especially when the transport and deposition processes are less than certain, include "boulder clay" and "drift," both now being replaced by diamicton. Lodgement till occurs in sheets or pockets, often beneath a deposit of ablation till, and more typically in ovate hills called "drumlins." Drumlins, which form in clusters ("swarms") when conditions are propitious, may be small enough to be mistaken for artificial mounds, but can be readily distinguished by their sedimentary fabric. **Moraines** of many depositional varieties are landforms of low to moderate relief composed of till deposits that may be highly controversial with respect to their origins, being typically composed of both sorted and unsorted debris.

Periglacial landforms

Ice distorts sedimentary bodies, creating characteristic small-scale landforms that can be powerful indicators of past climates. The proximity of glaciers establishes extreme climates where frost remains perennially in the ground to great depths. Seasonal thawing at the surface mobilizes water into sediments; it expands as it freezes in cracks and pockets, forcing sedimentary material sideways and upwards. Such periglacial processes create networks of interconnected ice wedges that may be outlined by rocks forced to the surface by cycles of freezing and thawing. Water-saturated sediments on frozen ground may slide on even slight slopes (gelifluction), folding and overturning as they settle. The stratigraphic and associational relationships of any archaeological materials in gelifluicted sediments are likely to retain no integrity, although the deposits may mimic stratification by duplication of layers in folds. Archaeologists working in formerly glaciated terrain should be aware of the full range of periglacial landforms, and ready to deal with their interpretive problems (Clark 1988).

Biosphere

Living organisms influence landforms by controlling cycles of erosion and deposition both directly and indirectly. The type, density, and distribution of vegetation influence the exposure of sediments to erosive forces such as wind and water. Plants trap and hold sediments above and below the surface; their presence stabilizes slopes. The absence of vegetation, whether resulting from fire, herbivory, human land-use practices, or climate change, exposes sediments to erosion. In special cases such as the mounds and pits left by uprooted trees, plants create microrelief that itself mobilizes sediment. Small to medium-sized animals that live parts of their lives under the surface of the ground also disturb sediments and create microrelief features at the surface. Artificial landforms built by humans have increased the complexity of the surface for thousands of years. Depositional features such as middens, mounds, agricultural terraces, roads, and more recently cities, testify to the effectiveness of human construction of landforms. Erosional features such as canals, plowed fields, and artificial leveling have achieved scales matching those of natural forces.

BASIC GEOMORPHOLOGICAL METHODS

The fundamental descriptive field techniques in geomorphology are cartographic skills; it is on those, especially, that geomorphology and archaeology converge most closely. The basic mapping skills essential to good archaeological reconnaissance at local and regional scales are classics of geomorphology – both plane-table and transit surveys, photogrammetric mapping, air-photo interpretation, and the use and refinement of standard topographic and surficial maps of the sort that government agencies produce in many countries for many strategic purposes. The availability of maps of different kinds varies greatly from place to place. Most widely available are standard topographic contour maps at scales between 1:25,000 and 1:250,000; these are a great boon to archaeologists although they lack accuracy at site scales. Maps showing surficial geology or soils types also have immediate applications in archaeology. Archaeologists engaged in field work should be thoroughly familiar with the full range of map types and mapping conventions used in their area, and with the literature that supports their productive exploitation (e.g., mapping manuals supplied by the governmental agencies that publish the maps; Monmonier 1993). Although these maps are the first gift of geomorphologists to archaeologists, many of the specialized maps are still underutilized by the latter. The new technology of Global Positioning Systems (GPS), whereby suitably equipped persons on the ground can find their geographic position electronically from satellite coordinates, promises unprecedented precision for map-makers everywhere (Hofmann-Wellenhof et al. 1993; Leick 1995). Areas of the world not well served by conventional cartographic methods are now accessible by means of aerial photography and GPS (Maschner 1996).

Archaeologists have benefited from remote sensing of terrain as long as any such techniques have been in use. Aerial photography, beginning with primitive cameras held by balloonists, is now a highly technical specialty developed for military purposes but with immediate applicability for archaeological survey. While low-altitude photographs – oblique, mosaics, or orthogonally corrected – continue to be the most useful, photography from satellites and manned spacecraft has revealed classes of data previously undreamed of by archaeologists. The paleoenvironmental applications of aerial photography and other remote-sensing techniques extend far beyond the identification of landforms and vegetation suites. Infrared films record the growth rates of different types of vegetation that are sensitive, among other things, to the water content and mineral enrichment characteristic of the soils on archaeological sites. Normally invisible subsurface features are visible from the air under special conditions of ground cover or seasonal humidity. Satellite-borne cameras have transcended the traditional earthbound limits of spatial scales to give highly accurate records of current landforms and vegetation. Some archaeological applications are cited in chapters that follow. The technology is currently outpacing the literature (Foody and Curran 1994; Leick 1995; Scollar 1990).

Geophysical remote-sensing techniques such as radar, geomagnetic mapping, and electromagnetic and multispectral scanning are revealing new kinds of geomorphological data on the Earth's land surfaces, some of which have immediate archaeological relevance. Imaging radar that scans surface relief through dense vegetation and surficial sand sheets has revealed previously unsuspected agricultural systems in tropical jungles as well as ancient river networks buried under desert sands. Computer enhancement of remote-sensing data and photographs increases the resolution of data displays by orders of magnitude (Scollar 1990). Small-scale buried land surfaces and constructional features are mapped using subsurface radar scanners, sonic reflection techniques, proton magnetometers, and resistivity surveys – all techniques developed for geomorphological and engineering applications before being adapted for archaeological work (Clark 1990; Conyers and Goodman 1997). Data from many sources can now be formatted to a common scale by computer processing and combined into information-rich Geographic Information Systems (GIS) data displays (Fig. 9.4; Allen et al. 1990; Maschner 1996).

Large-scale buried landforms are sought and mapped by various satellite imaging techniques, ground-penetrating radar, and seismic refraction in addition to the older methods of drilling or trenching through layers of sediment. Buried soils revealed by digging, coring, or remote sensing trace former surfaces that may bear little resemblance to contemporary landforms in the same place. Former landforms may also be estimated by extrapolation from surface remnants across locations now eroded away or moved by faulting or isostatic adjustment (e.g., Bailey et al. 1993; Roberts 1987; Rolph et al. 1994).

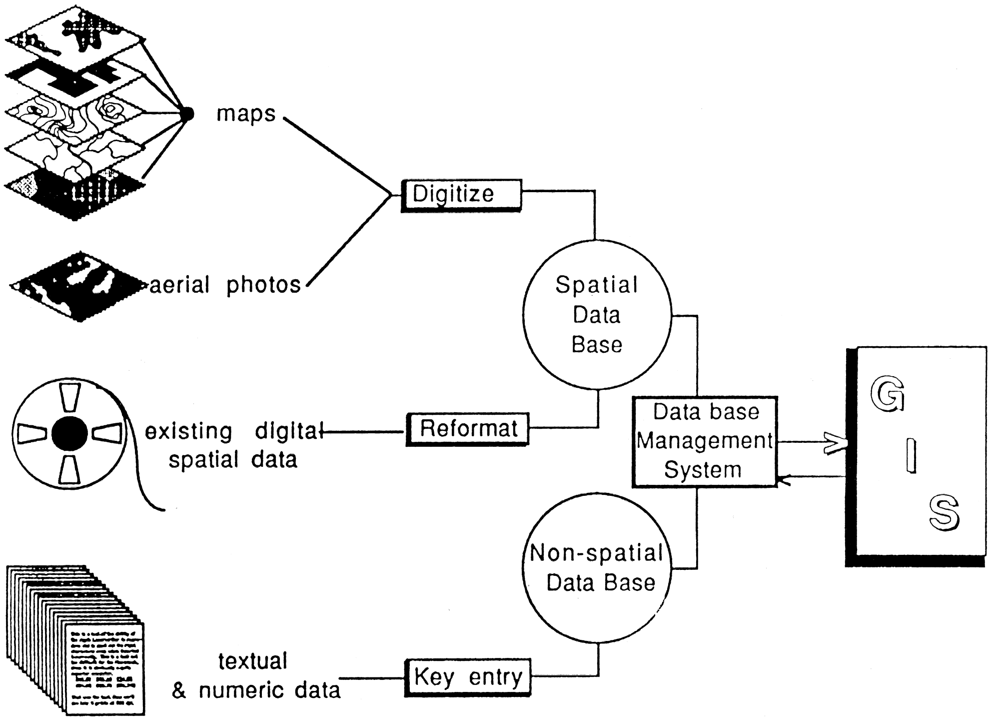


Figure 9.4 Geographic Information Systems equipment and process. (Reproduced from Madry 1990: Fig. 15.3, with permission of Taylor & Francis.)

Analytical techniques

Regardless of their sophistication, imaging techniques are inadequate for interpretation. Analytical description of bedrock and unconsolidated sediments (lithology) is the essential complement to the description of form. Aspects of sedimentology and **pedology** (soil science), particularly at archaeological scales, are the topics of Chapters 11 and 12. The order of presentation here, which places form before content, must not be interpreted as diminishing the importance of lithologic study in the field and laboratory. In most cases it is the sediments, rather than the landforms, which should engage the archaeologist's primary attention. The dominant analytical techniques in geomorphology are based on small-scale observations of contemporary conditions (Goudie 1981; Chorley et al. 1984; Thorn 1988: 110). Their relevance for the interpretation of larger, older, actual landforms may be problematic.

In contrast to the readiness with which archaeologists borrow techniques for discovery and description of landforms, geomorphological developments in analytical and interpretational techniques seem to be adopted slowly and reluctantly by archaeologists. Renewed awareness of site-formation processes and taphonomy (Schiffer 1987) is stimulating archaeologists to undertake field investigations into processes that affect unconsolidated sediments and the rates at which these act under different environmental conditions to modify or create landforms. The pace of innovation and development of analytical techniques in geomorphology defeats any effort to summarize here. The newest, fastest, most precise methods are best met in professional journals such as *Geomorphology*, *Geology*, *Quaternary Research*, *Holocene*, *Geoarchaeology*, and others more technical, where advances in mathematical, physical, and chemical methods are reported and evaluated.

Measurements of soil creep, sedimentation processes and rates, infiltration and chemical change, and bioturbation and cryoturbation are beginning to appear in archaeological publications (Chapters 11 and 12) although they are typically retarded in comparison to the capabilities displayed in the geomorphological literature. To a far greater degree than the tectonic and glacial processes that dominate the geomorphological literature, these small-scale processes are relevant at archaeological scales of space and time. It is at the small scale that human impacts on landforms are best revealed (Boardman and Bell 1992).

Constructional landforms are usually directly interpretable because of their characteristic forms and internal sedimentary structures. Destructional and erosional landforms are more problematic because the generating processes are subtractive, sometimes leaving no direct evidence of their presence, and because the original forms are modified or even lost during erosion. Lost landforms may be partly retrievable with expertise and appropriate techniques. Archaeologists who collaborate with geomorphologists find their interests very well served (e.g., Bettis 1995; Wagstaff 1987).

Geochronology: time concepts in geomorphology

Among the basic issues of world-view that complicate collaboration between archaeologists and geomorphologists, concepts of time and scale vie for first place. Archaeologists deal typically with radiocarbon time and local scales; geomorphologists with stratigraphical relative time and regional-scale process models. Professionals working on phenomena at one set of scales develop mind-sets that hinder their recognition of phenomena at other scales of space and/or time. These issues have been clearly addressed in the volume edited by Stein and Linse (1993), which defines the essential first steps toward awareness that will lead to resolution of some of the difficulties.

Earth-time and human-time are not easily correlated. Not only are the scales of temporal units different, but also the resolutions and comparability of the many chronometric methods available (Chapters 5 and 6). In addition to all the chronometric methods used by archaeologists, geomorphologists studying processes also utilize methods appropriate for very short time spans at the scales of decades (e.g., a lead isotope, ^{210}Pb, with a half-life of 22 years). As discussed in Chapters 5 and 6, different chronometric techniques employ distinct units, relevant time spans, and error ranges, all of which must be calibrated to some common scale before comparison. A continuing source of complications in interdisciplinary research is the uncritical correlation of incomparable time scales, such those of radiocarbon and orbital cycles (e.g., Braziunas 1994).

While archaeologists are comfortable assuming rough synchrony between events and even deposits separated in space when they contain similar artifact forms, geomorphologists are sensitive to equifinalities, complex responses, and process lags that produce superficially similar landforms independently in time and space. Productive collaboration across disciplines requires explicit evaluation of contrastive assumptions about phenomena and processes in culture and nature. "Because geologic processes and human behavioral processes commonly operate at different rates, the chronological information inherent in these processes is incapable of resolving the different scales" (Dean 1993: 59). Dealing with phenomena of mutual geological and archaeological interest within the classes of nested scales alleviates immediately some of the incongruity between them.

LANDFORMS AT MEGA- AND MACRO-SCALES ($>10^4$ KM^2; 10^{6-9} YEARS)

Anyone curious about the observable diversity of landforms has wondered about the large-scale processes that have shaped the surface of the Earth. Mountain ranges were thrust up and worn down by processes which we can observe in operation now, although we see and measure only brief segments of those long waves of change. Because of the large scales of these processes, we can measure them only at low resolution; error factors may exceed a million years. Furthermore, large landforms change relatively little in the time spans that are significant for archaeological studies; change at the scale of continents and provinces is rarely important even for Paleolithic archaeology. Therefore, interpretation at large scales rarely concerns archaeologists, who can adopt interpretive conclusions from responsible scientists without evaluating the details of methods and logic. Archaeologists do, however, need to understand the terms and concepts of structural geology sufficiently to interpret and adapt the descriptive literature of their research areas, because however it is explained, the gross shape of the landscape matters to people living on it (Skinner and Porter 1995).

Continents

We have seen above, in Chapters 3 and 7, that the latitudes and elevations of continental masses over the Earth are major determinants of climate. The geosphere thus sets the parameters defining the dynamics of the hydrosphere and atmosphere at large scales. Although the motions of the plates are very slow, and change is manifested only gradually in geological scales of time, at any one point in time the arrangement of oceans and continents, mountains and plains, is unique and determinative. The arrangement and timing of unique configurations cannot be ignored by the paleoenvironmentalist. For archaeology, however, continental-scale phenomena remain background.

Physiographic provinces

The continental masses are traditionally subdivided into more or less discrete **physiographic provinces** on the basis of distinct rock structures and relief. The boundaries of such units are generally quite obvious on the ground. Despite the expansion during the twentieth century of information about deep structures, age, and formation processes, early descriptive works on physiography retain a basic validity regardless of fundamental changes in interpretation required by the concepts of plate tectonics.

Provincial boundaries are typically defined by geologic **unconformities** created by plate movements and tectonic deformation, usually but not necessarily involving orogenies (mountain-building episodes). Figure 9.5 illustrates four types of orogenic structure, greatly simplified diagrammatically. Structures are best displayed in the youngest mountain ranges, among which are the Cascades and Andes ranges in North and South America and the Himalayas in Central Asia. All of those are situated on the margins of actively moving plates and are still being pushed up, even as they erode at the surface. Mountain ranges of very great age, such as the Appalachians of eastern North America and the Urals of European Russia, nevertheless form distinctive physiographic zones contrasting strongly with adjacent areas manifesting different origins and histories. The rugged relief of mountain ranges is a product of interactions among structure, lithology, and processes of erosion and transportation.

The interiors of continental plates are usually tectonically stable areas, typically characterized by essentially horizontal bedrock and by low relief. Physiographic provinces in such areas may be very large. Depending upon the dominant climatic regime, the interior provinces may be plains, steppes, deserts, or tropical lowlands.

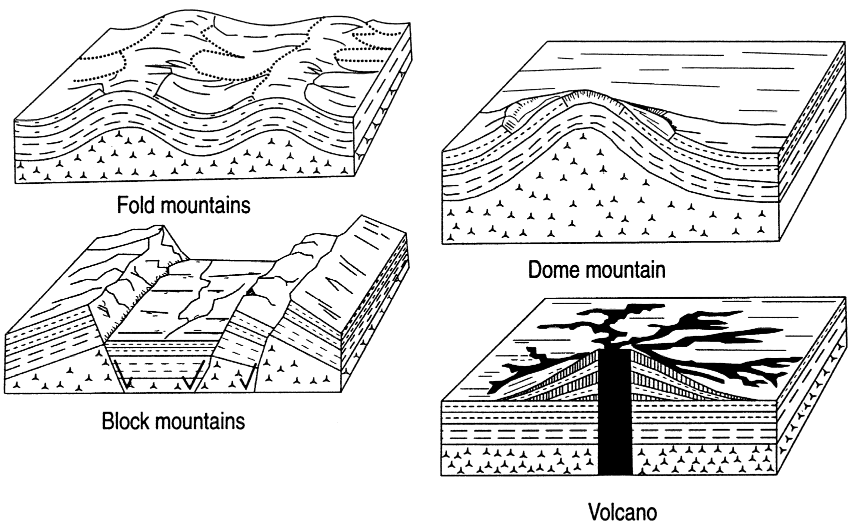


Figure 9.5 Four kinds of orogenic structures. The arrows showing relative movement in the block mountains are oversimplified. (After Hunt 1974: Fig. 3.5.)

Some are interrupted by upland areas contrasting in lithology and relief, such as the Ouachita Mountains in North America and the Urals of Eurasia. These represent very ancient plate boundaries, no longer active.

At the scale of physiographic provinces, climates are largely determined by the latitude, elevation, and orientation of landforms, but smaller-scale effects begin to be expressed. The location and size of mountain ranges, and their orientation in respect to prevailing atmospheric circulation, will affect precipitation patterns and the strength of seasonal differences, and thus define the dominant processes of erosion and deposition. Distance to oceans is also crucial. Landforms typical of physiographic provinces are well described in the literature of geology and geography and are readily available to archaeologists. When interpretation involves large time scales, the skills of a historical geologist or geomorphologist are essential to success.

LANDFORMS AT MESO-SCALES ($< 10^4$ KM^2)

Intermediate-scale landforms are affected by a wide range of processes, and by rates at many frequencies. The processes operative at the macro- and mega-scales set the parameters within which those of shorter wavelength can be expressed, while high-frequency processes may complicate analysis by imposing over the basic structures surficial landforms of short duration. Thus, there is practically no scale of landform, process, or rate that can be totally ignored during interpretation of features at meso-scales. The most immediately relevant processes, however, are those controlling weathering of rock and erosion and deposition of the resulting mineral grains. Translocation of mineral matter from higher to lower elevations, with consequent reduction of regional relief, is the business of these middle-scale processes, working through wind, water, and ice. Countering the prevailing elevational reduction are the geospheric processes expressed at meso-scales, principally volcanism and isostatic uplift.

Volcanism

Volcanic eruptions occur at or near boundaries of tectonic plates, or at intra-plate geothermal "hot-spots" such as the Hawaiian Islands. Zones of active rifting or of subduction display the conditions for volcanic eruptions, where liquid rock at very high temperatures rises from the mantle to emerge onto the surface of the Earth. The molten rock moves along pipes or fissures in the crust, and may emerge at the surface as flows of lava or as explosions of gas or steam with or without broken rock. The form of release varies with the chemical composition of the liquid rock, the form of the opening and other factors, not all of which are understood. The typical volcano is a conical pile of shattered rock and dust, often interlayered with solidified lava flows. The volcano grows by the ejection of new material onto its slopes, and can rise dramatically in short bursts of activity (Fig. 9.5). Volcanoes can also, as in the case of the Santorini volcano in the Aegean or the recent eruption of Mount St. Helens in Washington State, USA, destroy themselves in violent explosions and subsidence.

Volcanic eruptions can also create depositional landforms less dramatic than cones. Lava flows may fill preexisting valleys and blanket other topography, flowing for many kilometers from the fissure or pipe that released them. The rough, knotted surface of flows is inimical to most forms of life for a long time. Caves may be formed during lava flows when bubbles of gas create voids in cooling lava, or when the hot interior liquid flows away from the chilled, hardened surface. Clouds of fine volcanic ash (tephra) accompanying some explosions can smother both plants and animals. After settling and weathering, tephra forms a fertile substrate for new plant growth. Ash falls can be massive and destructive; very hot ash deposits may harden into rock (**tuff**) that defines a new surface and new landforms where it lies. If an explosion melts snowbanks or glaciers on the slopes of the volcano, or if it is accompanied by heavy rain, massive mudslides (lahars) may roll down the mountain and spread along the valleys, killing all life on their routes and changing the landform. Needless to say, lava flows, ash falls, and lahars are capable of creating or destroying archaeological sites almost instantaneously (Sheets 1992; Sheets and McKee 1994).

Isostasy

The relatively thin crust of rock that forms the continents and underlies the ocean basins rests on more viscous material of the mantle. When additional weight is imposed upon the crustal rocks, they deform elastically, sinking as pond ice may subside under the weight of a skater. The mantle material below slowly flows away from the depression, forcing crustal rock elsewhere to rise (Chapter 3). When weight is removed, as by the melting of a glacier, the draining or drying of a lake, the erosional removal of surficial sediments or rock, or the regression of the sea, isostatic crustal uplift occurs. The uplift is a subtle effect; however, fluvial responses to changed base levels and raised beaches of pluvial and glacial lakes indicate that the scale of landform change can be considerable. The seacoasts of Scandinavia and Labrador with their step-like raised beaches are dramatic demonstrations of the potential scale of isostatic landform change: 8000 years after the ice sheets melted, those northern coasts are still rising. Effective reconstruction of landforms deformed by isostasy requires modeling changes in elevation and slope over time.

Aeolian landforms

In arid lands, where vegetational cover is thin and surface sediments easily mobilized, wind erosion may be the dominant agent of surface modification. Air whirling at tornado speeds (hundreds of km/hr) picks up and transports objects as large as trees and houses, leaving behind both denuded spaces and new debris piles. Where large amounts of unconsolidated sediments are available to be transported by wind, aeolian landforms such as sand sheets, dune fields, and loess deposits may dominate the landscape, blanketing earlier landforms. Loess deposits of well-sorted silts sometimes cover thousands of square kilometers to depths of many meters, as in the American Midwest (≥30 m in Kansas), the South American pampas, Eastern Europe, and in Central and Eastern Asia (>100 m deep) (Lowe and Walker 1984: 112).

Dunes take many shapes, their forms responding to aerodynamic principles and reflecting prevailing or dominant wind directions and speeds as well as the supply of sand. They may be linear, crescentic or parabolic, star-shaped, or domed, and their orientation may vary from transverse to longitudinal in respect to the winds (Fig. 9.6). Every geomorphology text provides details about dunes.

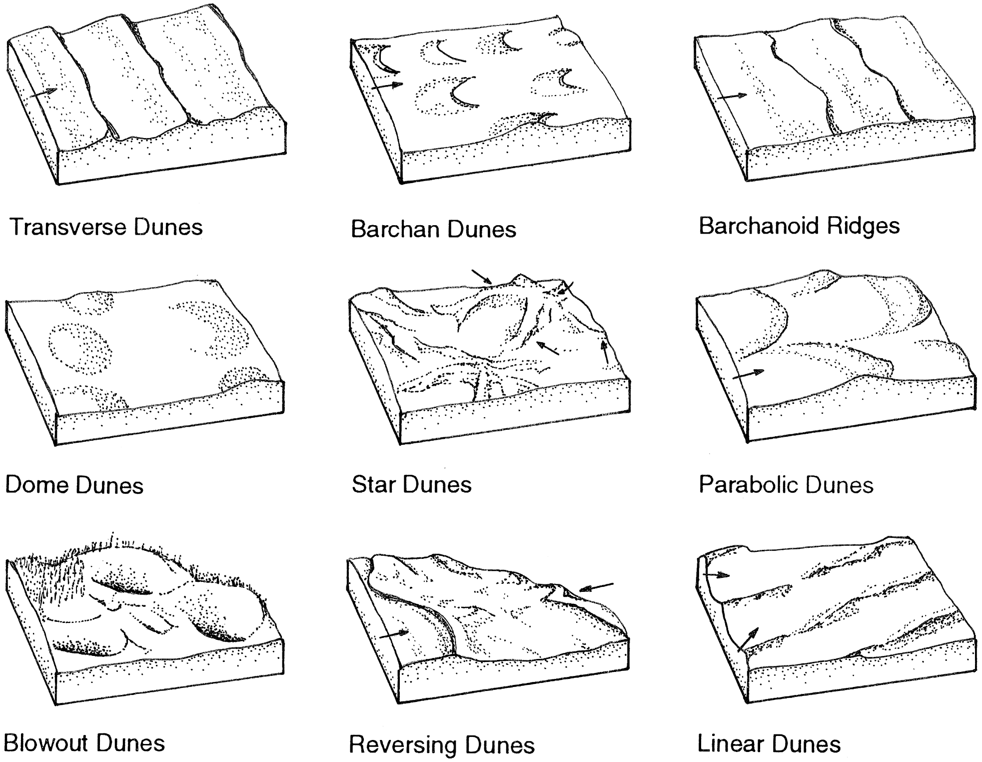


Figure 9.6 Typical dune forms; arrows indicate wind directions that influence the dune formations. (After Waters 1992: Fig. 4.4.)

Fluvial landforms

The most archaeologically relevant landforms created by moving water have been discussed above. Floodplains and terraces, river channels and channel forms are important elements of archaeological landscapes (Chapter 10). Terraces, levees, and the higher segments of floodplains have been for millennia favored places for human activities. As these landforms are all accretional, stratified sites may be found in them.

Glacial and glaciofluvial landforms

Landforms created by glacial erosion and deposition are easily identified, if not interpreted, but often difficult to date. Although archaeologists must be concerned far more with the dating of glacial features than with the details of their form and formation, it is helpful to understand some of the basic concepts and language of glacial geomorphology.

Landforms shaped by glaciers, and the low-temperature periglacial landforms that accompany them, are obvious at local and regional scales in the affected latitudes. For the most part, the features are local elements of limited size and extent, but they are typically distributed regionally. Features resulting from continental-scale ice sheets are rarely themselves continuous in space at comparable scales; they form, and are preserved, discontinuously, and must be interrelated by interpretation based on theories about glacial growth, flow, and disintegration. Most of what is known about glaciers has been learned from observation of currently active mountain glaciers and snow fields, features of much smaller size than the great ice sheets of the past. Research now under way in the high Arctic, in Greenland, and in Antarctica is producing the first significant sets of direct observations on large-scale ice bodies, and is making very clear how much is still to be to understood. The landforms remaining after the melting of ice sheets and valley glaciers, however, are available for direct observation; their forms and composition, therefore, are better understood than are the processes that shaped them (Bowen 1991).

Terminal moraines are ridges of unstratified material left at the outer edges of ice that has ceased to move. They may be traceable for long distances across terrain, but upon close examination are frequently found to be composed of deposits laid down by separate episodes of glacial-edge halts. Stratified debris, deposited from water, forms in cracks and at the edges of glaciers as well as in front of the ice margin. Within the ice zone the most common kinds of stratified deposits formed are small hills and mounds called “kames”; they may be of any shape, but are typically elongated because they formed as water-laid deposits in cracks or along the margins of ice and hillsides. Meltwater carrying debris away from glaciers as **outwash** may form extensive, level, sloping plains as it dumps its burden, the coarsest component near the ice, the finer farther away, in the manner of alluvial fans. Outwash plains and **deltas** are prominent features of postglacial landscapes at the edges of melting; their level surfaces and good drainage make them favorite places for human settlements and airfields. Outwash sediments are favored for sand and gravel quarries.

Like drumlins, small-scale glaciofluvial landforms such as kame deposits may be small enough to be mistaken for artificial landforms created by people. As discussed above, the distinction is easily made from observation of sedimentary structures in sections or cores. For further details about glacial landforms, consult any recent text on geomorphology or glaciology. As the field is developing quickly, texts more than 15 years old are no longer reliable, and the specialist literature is highly technical.

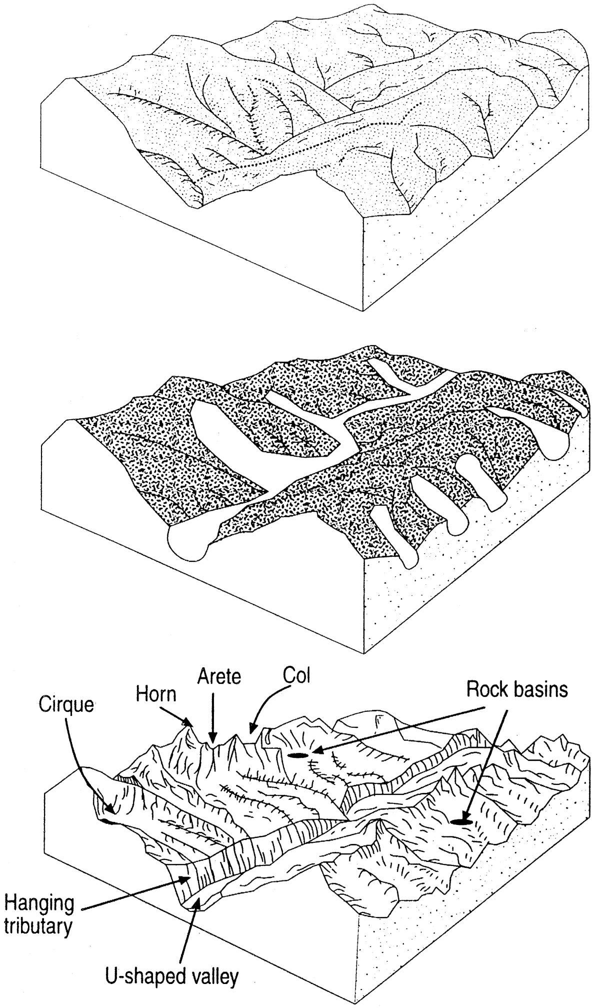


Figure 9.7 Stream-eroded uplands (top), subjected to alpine glacial erosion (middle). The bottom block diagram shows a U-shaped valley, cirques, hanging tributaries, and other classic valley glacial landforms. (After Flint and Skinner 1974: Fig. 11.12.)

Landforms shaped by glacial erosion are best developed in areas of high relief, where the flow of ice was strongly channeled and where bedrock was extensively exposed (Fig. 9.7). In areas of low relief and limited bedrock exposures, glacial erosional features are less available to observation, being typically covered by glacial or postglacial sediments. Striated and streamlined bedrock features are the most widely distributed erosional landforms, marking the passage of ice but usually affording no clues as to the time of the event.

Anthropogenic landforms

At the distance of orbiting spaceships, the view of Earth has been reassuring in that human impacts upon the continental surfaces are not immediately obvious, except as light patterns at night. However, aerial and satellite photographs demonstrate that at meso-scales roads, canals, cities, and the Great Wall of China are perceptible modifications of the planet's surface. Constructional and destructional forms such as irrigation systems, agricultural terraces, road and rail networks, dikes, landfills, airports, and urban sprawls are important geomorphological features. Humans have imposed on the Earth artificial landforms to which natural systems must adjust. Landforms influence the flows of air and water and, through them, local climate. The most pervasive effect of humans on landforms, however, is probably their "alteration in the rate at which geomorphic processes operate" (Summerfield 1991: 25). Anthropogenic vegetation clearance and regolith disturbance have effects far beyond the immediate places and times of the activities. The complexities of "downstream" effects of human meddling with geomorphological surfaces remain to be understood (Boardman and Bell 1992).

LANDFORMS AT MICRO-SCALE (< 1 KM²)

At the micro-scale, in spaces less than 1 km^2, geomorphological analysis emphasizes not only processes, but actual agents and small-scale variables such as soil temperature and moisture. Prominent among the effective agents are humans. Having reached the scale at which human activity is readily influential, we must contend again with the occasional difficulties of distinguishing purely natural (non-artifactual) phenomena from the products of human artifice.

Modern urban-dwellers may live their lives in casual disregard for the shape and composition of the landscape beneath their feet. This obliviousness is a fairly recent luxury for the species; the locational attributes of ancient dwelling and activity sites display the sensitivity of former societies to the topography and lithology of their habitats. Living sites and cemeteries are so consistently sited on topographic rises, even in areas of very subtle relief, as to leave no reasonable doubt that the founders selected for drainage and for visibility. Any settlement with pretensions for permanence needed a ready supply of potable water, as well as abundant water for industrial uses and, ideally, transportation. Every rocky outcrop on the surface of the globe has surely been scrutinized at some times in the past to ascertain its potential for raw materials; the known distribution of ancient quarries seems to be limited only by the abilities of field archaeologists to recognize them for what they are. Every cave and rockshelter on the face of the Earth has surely been evaluated as living space many times. In addition to analyzing the natural history of inhabited landscapes, archaeologists need to understand the advantages that landscape features offered to people in the past.

Among the many small-scale glacial landforms easily confused with results of human labor are perched rocks – boulders balanced precariously on smaller rocks or outcrops, which appear to some people as so completely unlikely natural phenomena that they "must" be artifacts. Large erratic boulders and boulder fields emplaced by glaciers have been similarly overinterpreted by people insistent on their mythic qualities. An informed appreciation of landscape, and the ability to test hypotheses about agents, are effective antidotes to such error.

On the other hand, human actions such as digging and piling and bringing and taking can reshape natural landforms until their original morphology is irrecoverable. Furthermore, the scale of some artificial platforms, created to elevate and support religious and secular structures of consequence, can easily compete with natural formations, and may be misunderstood in turn. The Eurasian **tells** composed of the debris of ancient cities rival hills in size, and the amount of earth movement involved in constructing some European Iron Age hillforts compares favorably with some terminal moraines. Indeed, many ancient urban areas, wholly artificial landforms, exceed the micro-scale here under consideration.

The variety of landforms at the micro-scale worldwide is very great, but within given regions is usually finite and even predictable. Archaeologists should familiarize themselves with the natural forms common in their areas, including the full range of variation involving frozen-ground phenomena, water-laid sediments, aeolian action, forest-floor morphology, animal burrows (from ants to rodents), and the products of mass movement on slopes. Observe small-scale processes such as rainwater splashes moving sand grains, water transporting material down slope in rills and gullies, and clay deposits forming in puddles. Many of these phenomena are ephemeral within archaeological time scales, even within seasonal durations, but familiarity with them in their many guises will sharpen awareness not only of the climatic and biotic agents involved in their formation and disappearance, but of the cultural agents from which they should be distinguished. Moreover, the scope and scale of the damage micro-scale processes can do to archaeological sites will become obvious.

BASIC PRINCIPLES OF SEDIMENTOLOGY AND SOILS SCIENCE

> To the field archaeologist the most obvious – and often the most abundant – constituent of a site is dirt . . . Dirt, properly called soil or sediment, is the subject matter of sedimentology.
>
> SCHIFFER 1987: 200

Sediments are composed variously of particles of disaggregated rock, dust from whatever source, bits of dead animals and plants, and chemical precipitates. Their deposition on the surface of the Earth or the bottom of lakes and seas creates three-dimensional sedimentary bodies (deposits) which are subsequently modified in characteristic ways by the five spheres of the climate system. In company with bedrock, sediments underlie the landforms on which life processes occur. For archaeologists, sediments are the enclosing medium and the environment for the physical and chemical remains that comprise archaeological sites.

INTRODUCTORY CONCEPTS

In contrast to the readiness with which archaeologists borrow geomorphological techniques for identification and description of landforms, developments in petrographic techniques seem to be adopted slowly and reluctantly by them. Methods for the technical description and interpretation of sediments and **soils**, particularly, need further development and more intensive application in archaeology. As with fish that cannot be expected to be aware of water, archaeologists often take for granted the materials within which their sites occur, rather than seeing them as problems and interpretive opportunities. Skilled geoarchaeological work remains, regrettably, a specialist domain instead of being incorporated as a matter of course in all field work.

Minerals are inorganic chemical compounds in crystalline form; rocks are composed primarily of minerals, sometimes accompanied by organic detritus and chemical precipitates. Mineral matter of the regolith recirculates through cycles of exposure, erosion, deposition, and burial at the surface of the Earth. Rocks and sediments are disaggregated by weathering; the resulting particles are eroded and transported by components of the atmosphere, hydrosphere, cryosphere, and biosphere that rearrange the surface of the Earth and remodel landforms. Sediments form when particles in transit are deposited, as transport media lose energy. The conditions of transport and deposition can often be inferred from attributes of the sediments, which are important sources of proxy data on former environmental states. These potentials are discussed below under the headings of "Lithology," the study of rocks and minerals, "Sediments as archaeological contexts," the study of sediments in sites, and "Pedology," the study of soils.

Lithology

The primeval rock of the planet cooled and crystallized from a molten state to form the continental and oceanic crusts of early geological time. Igneous rocks, cooled from the molten state called "magma," include such materials as granite, diorite, basalt, and obsidian. **Igneous** rock that flowed on the surface of the Earth as lava is "extrusive"; that which infiltrated under pressure into or between bodies of rock is called "intrusive." Igneous rock bodies may be very extensive; the massive "shield rocks" that form the stable ancient cores of the continents belong to this class. In our own time volcanic eruptions are common enough to remind us that molten material continues to emplace new rock on the crustal plates. The extrusion of lava along rift zones beneath the oceans is probably nearly constant if the rifts are considered as a whole.

Sedimentary rocks are composed of mineral matter transported by wind, water, or ice and consolidated after deposition. The consolidated material may be minerals weathered from older rocks, chemical precipitates from water, or detrital organic materials. Sediments are consolidated and lithified (transformed into rock) by cementation or compaction or both, resulting in such rocks as sandstones, limestones, and siltstones, all reflecting in their names their sedimentary origins.

Metamorphic rocks are modifications of lithic material from the other two groups that has been subjected to great heat, as from contact with magma, or both heat and pressure, as occurs during mountain building when rocks are folded and faulted. In such extreme conditions, both igneous and sedimentary rocks are modified by formation of new crystal states that are stable at high temperatures and pressures. For example, granite can be changed to gneiss, limestone to marble, siltstone to slate.

Before there can be sedimentary rocks there must be sediments, masses of unconsolidated particles of mineral or biogenic materials. Sediments are always deposited on geosphere surfaces, which may lie beneath air, water, or ice, with obviously different consequences for their archaeological relevance. Sediments are fundamentally important to archaeologists because they constitute the **matrix** for archaeological remains, the enclosing medium which keeps the remains in place and defines their immediate physical and chemical environment. Sediments constitute also the **context** of archaeological remains and sites, locking them into stratigraphic and locational relationships with other classes of data. The attention paid by environmental archaeologists to this second characteristic of sediments is what primarily distinguishes them from their non-environmental archaeologist colleagues, who may consider sediments no more than expendable "dirt." The amount of archaeologically relevant information that can be wrung from the analysis of sediments is limited only by our skill, imagination, and funding (Fritz and Moore 1988).

Sediments as archaeological contexts

In the absence of sediments there can be no typical archaeological sites. Piles of worked stones on bare bedrock can qualify by definition as archaeological sites but, by providing the investigator with the utter minimum of information about the piling event or accompanying conditions, they limit investigation to elements of form; they are artifact aggregations, rather than sites. Sediments provide context and structure. Sediments enclose artifacts and features, maintain relationships among objects, and protect buried materials from a range of disturbances. In their role as burial environments, sediments may also disturb and damage archaeological materials. Additionally, it is always worth asking whether or not the original depositional surface is preserved in the sediments.

Archaeological sites occur in most kinds of sediments. Sites in primary context may be buried by aeolian, alluvial, **colluvial**, volcanic, marine, or lacustrine sediments. Redeposited artifacts, provided they survive transportation, also occur in a great variety of sediments, although once moved from their primary locations they lose much of their strictly archaeological information and become a special class of geological phenomena. Each class presents its particular physical and chemical characteristics and interpretational problems, each in turn requiring a special set of analytical methods.

In order to understand the crucial relationships between a site or deposit of artifacts and its enclosing medium, archaeologists need to know as much as possible about five characteristics of sediments:

1. the source of the material, whether it is residual or derived, the nature of the parent rock, and, if derived, from which direction and what distance;
2. the transport medium which moved and deposited it;
3. the depositional environment in which it came to rest;
4. any subsequent natural transformations of the deposit, including mechanical or biological disturbances and chemical or physical changes such as soil formation, cementation, or compaction; and
5. any subsequent cultural transformations.

Hay's study of the Footprint Tuff in the East African Laetolil Beds (R. L. Hay 1981; Part IV Case Study) exemplifies the kinds of data and logic involved in determining these five characteristics. The fact that he was working in lithified sediments changes some of the techniques, but not the basic research strategy. The source (1) of the volcanic ash that comprised the tuff was traced to the extinct volcano Sadiman by observing the slope of the beds and by matching the chemical composition of the ash in the two locations. The transport medium (2) was the force of the volcanic explosion, with subsequent fall through air; after the initial deposition, the ash was determined to have been disturbed only by rainwater. These conclusions were reached by observing that the bedding of the dry-season tuffs was about the same thickness across the minor relief of the surface, indicating that, unlike other ash-falls in the series, it had been only lightly disturbed by wind. The upper tuff showed some redistribution from topographic highs to low points, and water transport was suggested for that. The depositional environment (3) for most of the Laetolil Beds was inferred to be a dry savanna, on the basis of the included fossils and some wind movement of the ash which indicated little vegetation to offer obstruction; the rainfall that disturbed the upper layers of the Footprint Tuff was interpreted as seasonal. Natural transformations (4) subsequent to deposition included the footprints, subsurface disturbances by termite colonies and rodents, the formation of a carbonate crust, the cracking of the crust by the movement of burrowing animals, and, ultimately, lithification. None of the investigators who have crawled over the Footprint Tuff in the years since its recognition have found any evidence of cultural transformations (5). While disappointing, this is not surprising given the great age of the tuff, which antedates by over a million years any cultural remains or behavior known anywhere.

Pedology

Pedology is the science of soils – chemically and mechanically altered terrestrial sediments. Soils form on and beneath the subaerial surfaces of sediments that are stable or only slowly aggrading. The formation of a soil requires above all else time; therefore, a soil represents a period in which deposition occurred only slowly if at all – a depositional hiatus and a time of relative stability. A surface that is rapidly building or rapidly eroding will not support the formation of a soil. While the deposition of a mass of sediment may be thought of as an event, with a beginning and an end, the formation of a soil is always a process, and soils must be understood in processual terms. The process has a beginning, which is usually coincident with the formation of a stable sedimentary surface. However, it is not known that soil formation as a process has an inherent endpoint; the process typically ceases with an environmental change that leads to burial or removal of the sediment supporting the soil. Archaeological materials, even entire sites, occur within soils, but the relationship of soil processes to a site or an artifact must be independently determined in each case. The soil may have been formed before archaeological materials were deposited on or in it; it may have formed after the creation of the site, or it may have been disturbed by human activity and then continued to develop with appropriate adjustment to the environmental change.

The primary literature on soils was developed within agronomy and is directed toward improving growing conditions for economically important plants. Geologists and geomorphologists, who also work in and around soils, have developed a literature that meets their special needs. The soils literature directed to the needs of archaeologists is limited (Cornwall 1958; Holliday 1990, 1992; Limbrey 1975; Waters 1992) and is not sufficiently detailed to be of substantive help in local situations. Archaeologists, therefore, need to use the soils literature as it exists, and build on it in consultation with local experts (Chapter 12).

There are at least three working definitions of the term “soil,” which can complicate communication. To agronomists, soils are surficial materials that support plant growth. Agronomists ignore buried soils, even to the point of not recognizing them, and tend to evaluate archaeological deposits (**anthropogenic soils**) in terms of their horticultural potential. Nevertheless, agronomists can be helpful, provided full communication about analytical goals is established before analyses are undertaken. To construction engineers, “soil” is all unconsolidated materials that can be “dug” rather than “blasted.” Their soils are our sediments and regolith; it is helpful for an archaeologist to know this before trying to interpret engineering drilling logs that may constitute the only preliminary glimpse of subsurface sediments in urban areas.

To geologists, "soil" is surficial sediment altered by weathering, whether buried or not. This definition is closest to the archaeological usage, and Quaternary geologists share with archaeologists an enthusiasm for buried soils and their information content (Holliday et al. 1993). However, geologists may include archaeological deposits in their concept of "soil," failing to distinguish them because geologists are trained to notice and interpret phenomena at scales larger than those typical of archaeology (Stein and Linse 1993). Archaeologists, who must communicate with all three groups of specialists, should be alert to the need for clear understanding of the language in use in particular situations.

Although sediments occur across the extent of the globe itself, and soils cover the continents, both sediments and soils are highly variable at the micro- and meso-scales. Generalization is difficult, since archaeologists need to apply soils analyses principally at the local and micro-scale. The text that follows here and in Chapter 12 constitutes the barest introduction to the complex world of surficial sediments and soils. Further reading could well begin with *Soils in archaeology* (Holliday 1992) and *Reconstructing Quaternary environments* (Lowe and Walker 1984: Ch. 3), progressing to chapters in specialized texts such as *Soils and geomorphology* (Birkeland 1984). The archaeological literature on soils analysis can be quite unforgiving to neophytes, presuming as it does a technical basic vocabulary. Ultimately, the descriptive soil surveys published by most national governments provide the most detailed information that is available outside of the analytical laboratory for the site and local scales.

STUDY TECHNIQUES IN SEDIMENTOLOGY

Sediments constitute the context in which archaeological materials are deposited and retained; their identification and analysis can inform about the history of the materials and the site itself, about agents of site burial, about the environment in which human behaviors that defined the site took place, and about chemical and physical conditions that determined the preservation of remains. Understanding of these processes properly begins with an understanding of sediment source and history.

Interpretation of sediments and soils requires an informed combination of field and laboratory techniques. Field observations are crucial; sedimentological consultants who are simply given bags of materials for laboratory analysis cannot be as fully supportive of archaeological investigations as their methods and expertise ideally allow. Sediments and soils are three-dimensional bodies, whose horizontal extents usually far exceed their vertical extent; they typically display variation in all directions. Variation necessarily raises issues of sampling adequacy that can only be resolved by field investigations. The discussion in this section is directed principally toward the investigation of sediments exclusive of soils.

Sources of sediments

With respect to origins, regolith (mineral aggregates) is either "residual" or "derived." Residual regolith (technically "saprolite") formed in place by disintegration of underlying bedrock, while derived regolith (sediment) has been deposited after transport. Weathering products may be fragments of rock, grains of mineral matter, chemical solutions, or all three. They are strongly determined by the mineral composition of the parent material and by the climatic conditions and weathering mechanisms that obtained during disintegration. For instance, granites subjected to extremes of heat and cold in very dry environments break up into their constituent mineral grains – quartz, feldspars, and micas – and undergo no further disruption. In moist climates, however, granites are reduced to grains of quartz and mica, while the feldspars weather into clays. If these residues are subjected to water transport, the grains of mica and to a lesser extent those of quartz will be further reduced by attrition; the clay will be carried off in suspension. The original three minerals, after transport, may be deposited in very different locations under distinct hydrodynamic regimes, because of their distinct specific gravities. While the minerals are all together, identifying them as the weathering products of a particular granite is fairly straightforward; once transported and separated, they cannot be traced to their origins. Rocks of more elaborate mineral composition, particularly those containing minerals less common than quartz and clay, may be recognized as the sources of mineral associations even after displacement over long distances.

Residual deposits form in place as bedrock is disaggregated by weathering agents (sun, water, ice, wind, salts, acids) attacking the inter-crystal bonds. If the loose material is not moved by wind, water, or ice, plants will colonize it and soils will begin to form. Without disturbance this process can continue, perhaps at slowing rates, to great depths. Organic detritus forms sediments in place, normally settling from water.

Derived sediments constitute the largest class of regolith, since the most likely fate of unconsolidated material is movement by wind, water, ice, and the force of gravity. Gravity, whether or not aided by ice and water, draws loose matter from cliff faces to form talus slopes; from cave roofs to floors; downhill to form colluvium; and through water to form subaqueous sediments in lakes and oceans. Moving ice carries materials on, in, and beneath it and deposits them as till. Moving water carries sediments away from ice sheets and other sources; as water slows or becomes otherwise overloaded, it drops its load to form any of a series of diverse fluvial deposits. Water currents moving along the shores of lakes and oceans sort and deposit coarse materials along beaches. Carbonate-rich water bubbling up in springs or flowing through limestone caves deposits calcareous "tufa" or travertine that hardens into rock. Fast-moving wind lifts and carries sand and silt-sized particles, sometimes to great distances, depositing them in dunes and sand sheets or in beds of loess. Wind also disperses volcanic ash after it has been exploded into the air.

Archaeological sites occur in all these kinds of deposit. Incorporation of archaeological materials into residual deposits occurs in the course of soil formation and the churning of sediments by organisms and ice. The mechanisms of incorporation into various derived sediments are research problems in each case; the challenge is to learn which of a large but not infinite set of agents and processes was involved.

Knowing the source of a sediment is a long step toward knowing its history. As the granite example above shows, tracing sediments to their sources is not a simple task with a guarantee of success. In cases where highly characteristic or unique suites of minerals survive transportation and deposition, their identification in a sediment can indicate the probable source or source area. It is self-evident that the source area for any sediments will be found in the direction from which the transporting agent came: upstream for wind, water, and ice. This rule of thumb can greatly simplify the task as long as non-human agents are at issue. Identifying the source of rock fragments, which by definition constitute a suite of minerals, is much more straightforward than is sourcing disaggregated minerals. Petrographic methods such as thin-section microscopy, X-ray diffraction of clay minerals, neutron activation for trace elements, and the set of spectroscopic methods all give good results when properly chosen. Many of these techniques are used by archaeometricians tracing the raw materials of lithic artifacts and clays; the principles and methods are the same for rocks in natural deposits.

Organic sediments composed of macrofossils, such as peats, generally accumulate where the plants grew, or in depositional basins very close by; vegetative detritus does not travel well. However, with the intervention of human agents, such materials (e.g., peat, lignite, and coal) may be transported for use over great distances. In such cases, sourcing is possible only when the material is clearly exotic and the original plant association is obvious, as might be the case for montane forest plants carried into a desert. Subaqueous sediments composed of microfossils such as diatomites and foraminiferal oozes can be traced only by identifying the characteristics of water bodies in which they formed, and in the case of ancient rocks, the known ages of the fossils. While such information may on occasion contribute to archaeological investigations, especially in submerged sites, it is mainly limnologists and paleoclimatologists who need to know about the temperature and salinity of water bodies. Such

Table 11.1 Velocity of transporting agents and sizes of particles moved (selected)

Water transport Stream velocity (miles per hour)	Sedimentary particles moved
⅙	clay
⅓	fine sand
1	gravel up to pea size
2	gravel up to thumb size
3	gravel up to size of hen's egg
Wind transport Wind velocity (meters per second)	Diameter of grains suspended (mm)
0.5	0.04
1.0	0.08
5.0	0.41
10.0	0.81

Source: Hunt 1974: 141.

determinations are made on the basis of assumed analogies with the modern habitats of organisms closely related to the fossil forms.

Transport media

Water, wind, and ice transport sedimentary materials more or less parallel to the surface within characteristic ranges of distance. These several media sort materials by size and shape, according to the energy in the system (Table 11.1). The coarsest materials are moved by high-energy systems such as glacial ice, fast-moving water as in rivers in flood and storm waves along a beach, wind in tornados and other cyclonic storms. The finest grades of materials, silt and clay-sized particles, may be carried by wind and water for great distances, from the centers of continents to the deep oceans. Gravity can be the agent of vertical transportation through characteristically short distances involving the settling of particles in place, or falling from a cliff or cave roof. Gravity acting alone does not sort particles, affecting all indiscriminately as Galileo demonstrated at the Leaning Tower.

Transport media are reflected primarily in the **texture** (particle size and sorting) and secondarily in the **structure** (bedding) of sediments, because of the close relationship between the energy in the system and its capacity for sorting the particles carried. In high-energy systems, in addition to sorting there may also be significant attrition of particles during transport, further reducing their size. The mill-like action of glaciers, which can reduce rock to fine "flour," is the extreme example, but given enough time media such as particle-loaded water and wind can work similar transformations. Unsorted and unstratified mixtures of coarse and fine particles may be difficult to interpret in terms of transport media. They are classed as diamicton unless there is some additional evidence indicating their origin as glacial till, landslides, or mudflows – all involving a significant component of gravity. Although no members of this confusing class of sediments are good preservers of archaeological sites, enough claims have been made for sites in and under such deposits that archaeologists need to be alerted to their characteristics and the potential confusions they bring (e.g., Shlemon and Budinger [1990]; see selected references in Dincauze [1984]). Alluvium, sediment moved by and deposited from rivers, is characteristically "graded," which means that the sediments deposited are sorted according to the speed of the water. Slowing water drops coarse materials first. Coarse materials typically initiate a graded depositional sequence, the characteristic "fining upward" signature of overbank floodwaters with finer materials toward the top. Such graded sequences may be repeatedly deposited as stratified sediments. Aeolian deposits (sand particles in dunes and silts in loesses) are usually well sorted by wind speed (Pye 1987; Pye and Tsoar 1990). Dune sands may show the characteristic cross-bedded structure.

Both the topography and internal structure of sedimentary bodies provide evidence of the transport media that deposited them. Terminal moraines, river terraces, and dunes normally testify by their scale and form to deposition from ice, water, and wind, respectively. The sorting and grading that are typical of variable speeds of transport by wind and water result in bedded (stratified) sediments, the individual beds of which may be internally complex, thus providing a number of analytical criteria for discrimination of the transport medium (Fig. 11.1). Alignment of particles within a sediment body is also indicative, being best developed in water-laid materials where elongated particles tend to align with the axis of movement. Magnetic alignments are best developed in particles settling from still water, as in deep lakes. Unconsolidated sediments moved downslope under the force of gravity (colluviation, solifluction, gelifluction) are poorly stratified, if at all, but may be roughly aligned.

Depositional environments

When sedimentary bodies enclose or bracket archaeological materials, inferences made about the paleoenvironmental conditions at their deposition can be used to

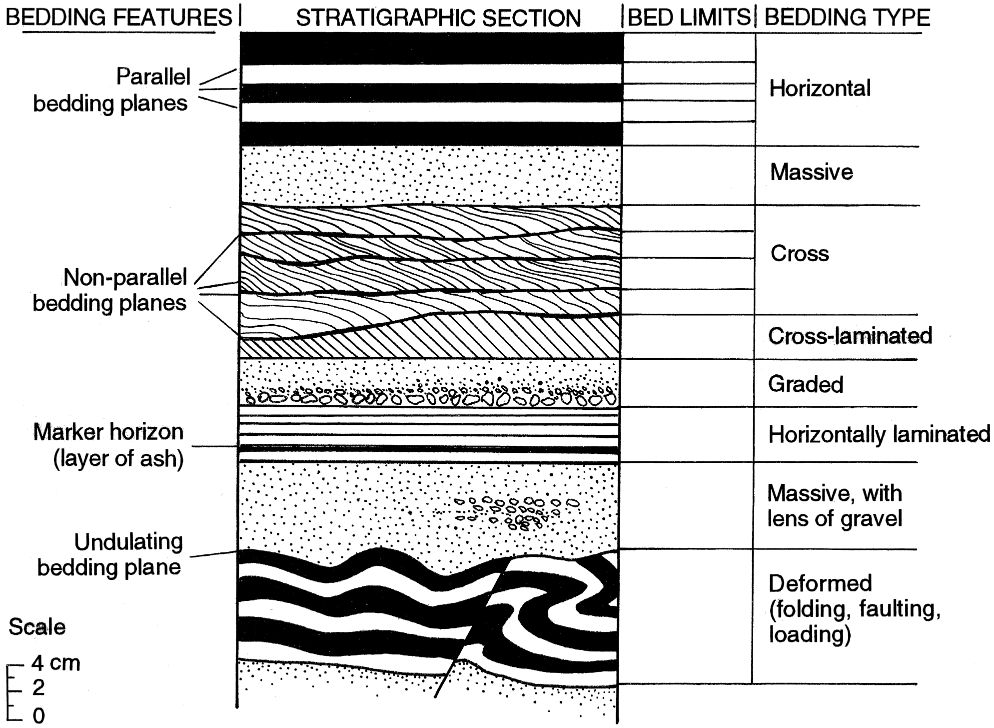


Figure 11.1 Terminology for stratification and bedding structures. The cross-bedding structure forms as wind or water deposits sediments over ripples and down the farther slopes, with the ripples moving downstream. (After Waters 1992: Fig. 2.12.)

understand the environments of the materials or sites. In this section, we address the *environments of deposition*, related essentially to sediments themselves. For archaeological interpretations, depositional environments are only part of the story, since we also need to understand the *environments of incorporation* – how the archaeological materials came to be associated with the sediments and the environments of burial. Those issues are developed in Chapter 12.

Materials mobilized by ice, water, or wind are carried as long as the transporting agent maintains sufficient energy to move them. Deposition, therefore, represents a change of environment for both the materials and the transporting medium. This change is recorded, more or less clearly, in the structure of the deposit itself, whether stratified, graded, sorted, or not. Variation in physical characteristics within sediment bodies, in either the vertical or horizontal dimension, indicates variation in the immediate environments of deposition. Vertical differences reveal the stratification of a sediment mass, expressing change in time. Horizontal differences in sediment bodies (facies) reveal environmental differences in space.

Depositional environments are interpreted primarily from structural evidence, secondarily from textural characteristics. Structural evidence includes both sedimentary structures and landforms. Sedimentological concepts important here are: bedding, the vertical contrasts in sediment bodies that define stratification; **interface**, the conformable contact surface between two different beds; and unconformity, an erosional surface between two beds. Particle surface textures that inform about transport agents are also relevant for depositional contexts, as are textural qualities of sediments such as sorting and grading. The degree of particle packing, measured as the frequency of interstitial air spaces in a deposit, can also be informative about sediment history.

Sedimentary structures (bedding, lenses) incorporate evidence about the scale, frequency, and rates of depositional and erosional events. The major determinant of sedimentary accumulation rates everywhere is climate: the scale and seasonal distribution of precipitation and the frequency of storms determine the prevailing depositional agents, the stability of sedimentary systems, and the frequency and amplitude of interruptions. For these reasons, sedimentary landforms are employed as climate proxies. It is crucial to realize, however, that the stability of a landform or sediment body is a function not only of the strengths and frequency of external perturbations but also of the internal state and condition of the system. The internal characteristics control response rates and the energy available for response to stimuli. Even when perturbing factors are similar, responses in one system (e.g., a fluvial basin) may be different from those in other systems.

Subaerial deposits

Sediments deposited in glacial and periglacial environments dominate major portions of the northern hemisphere, but their direct relevance to archaeological sites is limited. Obviously, no archaeological sites are contemporary with the deposition of till or glaciofluvial (outwash) deposits. Sites associated with such deposits are centuries or more older or younger than the formations themselves, and by their very existence reflect less extreme environments.

Periglacial deposits, on the other hand, may be intimately involved with human settlement and activities, now as in the past, but they have never been environments conducive to large concentrations of human beings. Modern periglacial areas, where large-scale freeze–thaw processes are studied, are mainly far removed in time and space from the ice age periglacial environments occupied by Paleolithic communities in the Old and New Worlds and are likely to be significantly different in the details of their seasonality and temperature ranges. Therefore, paleoclimatic and paleoenvironmental studies based on periglacial characteristics in sediments must be undertaken with informed caution (Clark 1988; Washburn 1980). Active periglacial environments are characterized by gelifluction and subsurface disturbances related to permafrost, by wind erosion and deposition, and by seasonal fluvial action of typically overloaded, braided streams; all of these have characteristic sedimentological structures (Chorley et al. 1984; Summerfield 1991; Washburn 1980). Dramatic secondary sedimentary structures such as ice and sand wedges, soil involutions, and gelifluction lobes (Fig. 11.2) may be encountered by archaeologists working far from the Arctic. Interpretation of such features requires the involvement of experts who can scrupulously evaluate their relevance for any associated archaeological remains which, outside of the Arctic, may postdate them by large spans of time.

Strong periglacial winds that lift the finer grades of sediment from exposed outwash and till deposits carry silt-sized materials far downwind from the periglacial environments themselves, to deposit them as loess in the cool, usually dry, steppe and grassland environments of continental interiors. Loesses are typically massive deposits with strongly vertical structure and with bedding rare or only subtly developed. Because pollen is unreliably preserved in calcareous loesses, terrestrial gastropods constitute the major source of incorporated evidence about climate and vegetation. Buried soils mark depositional interruptions and provide evidence for temperature and precipitation cycles. Sequences of buried soils in mid-continental loesses have been shown to correlate well with the glacial–interglacial cycles identified in deep-sea sediment cores (Kukla and An 1989).

Temperate sedimentary environments, on the other hand, are dominated by fluvial erosion and deposition by perennial streams (Table 11.2). Terrestrial fluvial deposits are typically local in scale and therefore both highly sensitive to local conditions and variable in time and space. Unless disturbed, abundant vegetation in temperate environments reduces the effectiveness of erosional forces, maintaining relatively stable sediment bodies. When surfaces or slopes are destabilized, either by denudation of vegetation or by water saturation, sediments may move quickly, forming colluvial deposits or alluvial fans, making their particles available for further fluvial transport. Otherwise, periodic deposition on floodplains is the sedimentary environment most typical of temperate zones. Archaeological sites oriented toward rivers and streams may be preserved within or under floodplain deposits.

The complexity characteristic of fluvial deposits challenges generalization. Responses to changed conditions vary according to the state of the system involved; they may be complex or simple, massive or very limited in area. A change in either climate or base level can evoke cyclical erosional and depositional responses in a stream network, varying in space and time (Chapter 9). Since archaeological scales of observation are essentially finer than geological scales, the complexity of fluvial

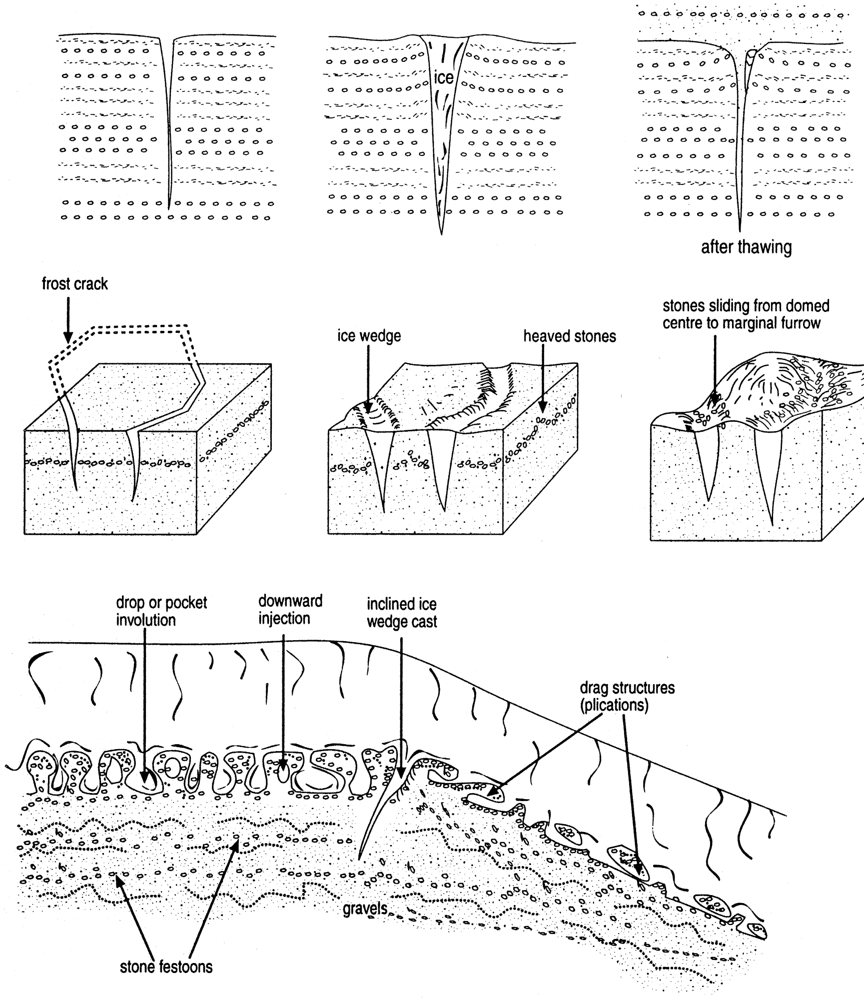


Figure 11.2 Schematic of the formation and form of selected periglacial sedimentological features: ice wedges, sorted polygons, and involutions. The ice wedges contribute to the polygonal cracking patterns that may result in polygons. Involutions sliding downslope complicate solifluction structures. Note that, to the unwary, each structure can resemble archaeological features such as postholes, hearth circles, and stratigraphy. (After Catt 1988: Figs. 2.49, 2.51, and 2.52.)

systems can seriously mislead archaeological efforts at understanding environments of deposition and the triggers of change. Extrapolation from archaeological-scale observations of fluvial deposits to regional processes is not recommended.

Sediments in subtropical arid environments reflect the special processes and landforms of those areas. Precipitation is strongly seasonal and episodic; vegetation cover is consequently sparse. These circumstances are conducive to episodic floods

Table 11.2 Flow words

The following related words are based on -*luv*-, a root meaning "flowing," derived from Latin *fluere*, to flow. The prefixes change the meaning in obvious ways. The adjectival form is given for regularity.

f-	luvial	*fluvial:* flowing in rivers, or pertaining to a river
al-	luvial	*alluvial:* of sediments deposited by rivers (flowed to)
col-	luvial	*colluvial:* of sediments sliding down hill, usually lubricated by water (hill flow)
e-	luvial	*eluvial:* of solutes and fines removed by water in suspension or solution. The A or E zone of a soil (flowed out)
il-	luvial	*illuvial:* of solutes and fines precipitated or deposited by groundwater in the B zone of a soil (flowed in)
p-	luvial	*pluvial:* of abundant rain; a rainy season or period

with large erosional and depositional events, as well as to large-scale aeolian activity moving and depositing sediments. Arid subtropical fluvial sediments are typically poorly sorted, as in alluvial fans and gravel spreads on pediments and in arroyos (Chapter 9). Fine-grained sediments at the downstream limits of fans and arroyos provide material to winds that carry away sands, silts, and dust. Winds removing sands and silts leave behind lag deposits forming gravelly desert pavements. Ephemeral shallow bodies of water dissolve salts and carbonates from the dust deposited in them and, on drying, leave mineral-rich crusts on the surface of playas (dry lake beds).

In rainy tropical environments, dense vegetation stabilizes slopes and retards erosion of sediments. In such climates, chemical weathering dominates and chemical erosion and eluviation of the finest sedimentary particles are characteristic. Consequently, slopes and elevated surfaces are deeply mantled with residual regolith. Episodically, the loose material becomes unstable, usually when saturated, and slumps or slides downslope where it becomes accessible to mobilization by rivers. Tropical rivers carry mainly clay-sized particles in suspension and colloids and ions in solution. Broad floodplains are seasonally inundated and typically swampy between floods. Archaeological sites in the tropics tend to be situated on the more stable, raised landforms, but they are subject to burial under colluvium and dense, clay-rich alluvium.

Deposits in caves and rockshelters vary with the specific conditions and climates of their host landforms. Stringently localized, speleological deposits are nevertheless an especially diverse group. The open fronts of caves and cliff shelters receive particulates from running water that ponds inside and from winds that are stilled there. Alluvial and aeolian sediments create a mixed record of very local sequences of changing depositional environments related to external conditions. Caves and shelters form by the loss of rock mass from cliffs and roofs; their floors accumulate rock shatter (clasts) and finer particles from those sources. Deposition of shattered rock is episodic, partly reflecting cyclic extremes of atmospheric temperature or moisture, partly responding to conditions on the surface of the rock mass itself. For a long time the degrees of sorting and of angularity of rock waste in limestone caves was interpreted as direct proxies of climates, especially of glacial–interglacial cycles. The history of such deposits is now recognized as being far more complex (Butzer 1981b; Straus 1990). Limestone caves, the surficial areas of underground drainage systems, are characterized by carbonate deposits called speleothems (stalactites, stalagmites, and travertine), which contribute essentially to the special strangeness of these interior spaces as well as to the preservation of diverse materials deposited on the floors and subsequently sealed under crystallized crusts. Because they lie underneath a stretch of surface, caves and shelters may also receive groundwater from cracks or openings in their roofs and back walls. Water coming thus from outside may either erode or deposit sediment on the floors of enclosures.

Subaqueous deposits

Lake and pond deposits constitute special classes of organic sediments. Their importance in paleoenvironmental reconstruction (see Berglund 1986) gives them a role far in excess of their frequency on the face of the Earth. Being subaqueous, lake and pond sediments do not form soils and do not support archaeological sites, although they may incorporate archaeological materials. Their importance to archaeologists lies in their centrality for paleoenvironmental and paleoclimatic reconstructions utilizing the organic particles such as pollen, diatoms, and macrofossils included in them. Ponds and lakes of temperate zones collect heavily biogenic sediments in typical postglacial sequences running upward from abiotic clay to **gyttja** (algal-rich organic detrital mud) and/or mud, and finally peat, as the water body is filled by sediments and plants. In basins so deep that they lack oxygen at the base of the water column, and thus do not support life in their depths, sediments may show strongly seasonal variation in texture or color, forming sequences of annual deposits (**varves** or rhythmites) that may be counted like tree rings (Chapter 5). Large deep lakes in tectonically active areas are prized for their thick, stratified, climatically informative sediment bodies. Shallow saline lakes in arid lands form less diverse paleoclimatic records, and have less archaeological relevance.

Near-shore marine sediments are of interest to archaeologists as the locations of inundated terrestrial landscapes or as platforms for shipwrecks. They merit special expertise, but the principles for their study are novel only in scale. The floors of lagoons, estuaries, and coastal ponds are mixtures of alluvial, organic, and marine current deposits, reflecting the energy levels of the local aquatic systems. Offshore, inundated terrestrial and coastal landforms may complicate matters at the landward edge of continental shelves. Postglacial marine transgression over the shelves typically planed off preexisting sedimentary landforms and deposited a broad sand sheet as the surf zone moved landward (Belknap and Kraft 1981).

Organic inclusions

Plant and animal remains included in sediments are among the richest evidence for depositional environments provided they are interpreted with care and informed imagination. Inclusions can range in size from the bones of whales or mammoths to grains of pollen. Not all organic materials relate to depositional environments. The bulk of the inclusions may derive from animals and plants living in or on the sediments subsequent to deposition (remnant), while others may be elements of communities that were carried along with the sediments, finally coming to rest far from their native habitats (redeposited). The distinction is, of course, crucial to the interpretation of environments relevant to any archaeological materials involved, which must themselves be evaluated for their status as remnant or redeposited materials. Remnant (**autochthonous**) fossils are problematic as to their time of introduction into a body of sediment, belonging normally to times following the subaerial depositional event itself. Their environmental signals must be evaluated for their chronological relationships to the depositional event and to the archaeological events under investigation. Naturally redeposited (**allochthonous**) materials belong to earlier times and distant space in relation to any deposit that contains them. As elements of sedimentary history, they represent environmental conditions at their source. They may, consequently, either complement or contradict the autochthonous evidence. How much time and space separates them from the deposit itself is to be determined in each case; it is never irrelevant to interpretation of the deposit. Organic materials introduced to a deposit by people may be exceptions to both these relational rules.

Bogs, fens, marshes, and swamps, as depositional environments, are intermediate between subaerial and subaqueous environments – more ambiguous even than fluvial deposits. A globally useful typology of wetlands is provided by Retallack (1990: 213–214). Wetlands attract people, not as comfortable places to be, but because of their rich biotic resources. Archaeological sites were rarely formed on such surfaces; more often they are underneath them, having been incorporated by wetlands expanding beyond their margins. Occasionally sites were built above them, as in the cases of trackways, refuge villages, and ritual sites, famous for their degree of organic preservation. Because the sediments enclosing such materials are composed predominantly of vegetation, they are discussed in Part VI.

Field methods

Observation of sediments in the field depends upon visual access, which is achieved by seeking out natural exposures of subsurface materials (scarps, eroded surfaces), by creating exposures (usually vertical) excavated for the purpose, or by using special equipment to pull cores from beneath a surface that is not more directly accessible. For geological and archaeological investigations, the vertical exposure, a "section" or "profile," is preferred since it minimizes distortion and provides good two-dimensional access. The larger the exposure, the better the investigator may observe spatial variation in the sediment body.

Particles that comprise a sediment may vary in composition, size, shape, orientation, sorting, grading, packing, and cementation. In combination, these characteristics create the texture and structure of the sediment. Texture is defined by the combined attributes of particle size, shape, and sorting, which together determine whether a sediment is fine or coarse grained, homogeneous or heterogeneous. Structure, on the other hand, is the result of the manner of transport and deposition of the particles constituting a sediment and its stratification, and may also reflect subsequent transformation processes. Materials deposited from wind, moving water, still water, ice, or other agents vary in characteristic arrangements of particles within the sediment in their orientation, packing, and size grading. Postdepositional disturbance of sediments by ice, plants, or animals will be recorded in structural attributes. Structural and stratigraphic observations in the field are essential preconditions to adequate sampling, since samples must be obtained from each discrete stratigraphic unit.

Description of sediments in the field must be undertaken with awareness that sediments are *both* matrix and context. Therefore, description of observed materials and structures should be accompanied by active questioning and hypothesis formulation, to collect data for alternative interpretations. Once the number and boundaries of lithostratigraphic units in a study section have been defined, each unit may be sampled. Sampling locations should be selected to represent the full diversity of the sediments, and each sample should normally include material from only one unit. If the full diversity cannot be sampled along a single transect, additional sampling locations should be selected as complements (Fig. 5.2). Sedimentary samples from exposures may be taken in bags, tubes, or sampling boxes; the choice of container and sample sizes depends on the research questions to be addressed and therefore the kinds of analyses that are anticipated (Catt and Weir 1976; Courty et al. 1989; Stein 1987). Sampling of inaccessible sediments by coring presents a special range of technical problems and challenges. Sediment sampling and analysis is distinct from sampling for particular kinds of archaeological data; it should be pursued by lithological methods (e.g., Aaby and Digerfeldt 1986; Gale and Hoare 1991; Reineck and Singh 1980; Stein 1987).

Lithostratigraphy

In a stratigraphic section, interruptions in texture, structure, or mineralogy of the sediment column are the criteria for defining sedimentary bedding units. Comparisons with sedimentary sequences beyond the immediate locality are aided by reference to a formally defined hierarchy of lithostratigraphic units. Lithostratigraphic units ("formations," "members," and "beds" in declining order of scale) are defined and described at type localities that are standards for comparison and formal naming. The local scale of archaeological sites places them most directly in a member or bed of a regional lithostratigraphical sequence. Deposits at archaeological sites of Holocene age are rarely classified into formal lithostratigraphic units, those being merely implied by the surficial deposits involved (e.g., alluvium, loess). The assignment of older sites to their correct lithostratigraphic member or formation can be a difficult endeavor but rewarding for chronological clarity (e.g., the Laetolil case study). Hence, archaeologists should be aware of the formalities of lithostratigraphic classification. Although not needed every day, the rules of stratigraphic nomenclature and classification are important for comparison and correlation of deposits at regional and larger scales. The intricacies of the formal system are well defined and accessible to non-experts in sources such as Ager (1993), Catt (1986: Ch. 4), Farrand (1984), Stein (1987), and Waters (1992: 60–88). The official international standards and nomenclature are set out in the periodically revised reports of the North American Commission on Stratigraphic Nomenclature (NACSN 1983).

Lithostratigraphic analysis and interpretation require *scrupulous distinction between description and interpretation* of observed phenomena (Stein 1990; see also Barker 1993). In the field, description should be as precise and interpretation-neutral as possible. Such care will keep the description useful as interpretive hypotheses are subsequently tested against analytical data. When interpretation is kept separate from description, multiple analytical techniques and their different kinds of data can more effectively be brought to bear on problems. Only the simplest of historical questions is likely to be answered definitively by a single analytical technique. The distinction between field description and interpretation is clearly demonstrated in Figure 11.3, a diagram from a complex English site of Middle Pleistocene age.

Laboratory methods for determining composition and structure

There is no ideal or required list of laboratory analyses that can be mechanically invoked to satisfy the demands of a scientific approach. The methods selected and the results obtained should have direct relevance for research questions, especially those that arise during field examination of sediments. No analysis can be any better than the quality and appropriateness of the samples available and the research questions asked. Archaeologists must, therefore, ensure that there is effective communication and coordination between personnel in the field and in the laboratory; the best situation is for sedimentological consultants to be involved in both places (Rapp 1975).

The physical and chemical properties of sediments and particles are informative about the sources and transport media of particles and the depositional environments and subsequent histories of the sediments. The methods cited here are for illustrative purposes only; they are definitely not exhaustive. Anyone planning to undertake or to utilize such analyses should obtain the necessary information and instruction from qualified practitioners. There is a wide range of choice among methods, even for a particular class of information; the choice will vary with the information required, the nature of the sample, the resolution desired, and the equipment available (Holliday and Stein 1989). Applications of these methods will be referenced in discussions that follow in this and later chapters. Fuller discussion of these methods and their limitations, and references to the primary literature, may be found in Gale and Hoare (1991).

Sedimentary particles and cementing minerals can be identified by a variety of chemical and mineralogical techniques, starting with simple chemical tests for a key element. The choice among mineralogical techniques, such as examination of thin sections under polarized light, atomic absorption spectrometry, electron microprobe, X-ray fluorescence, and X-ray diffraction, will depend on whether heavy minerals or clay minerals are involved, and whether information on mineral concentrations is needed. Analysis of the organic content of sediments is traditionally done by combustion and reweighing, but wet methods may be better choices in some instances (e.g., Lowe and Walker 1984: 93; Waters 1992: Ch. 2). Quantitative description of sediments in terms of particle size (**granulometry**), important for a variety of interpretive issues, is done by dry or wet sieving, hydrometer, pipette, sedimentation column (Gale and Hoare 1991; Goudie 1981: Pt. 3), or one of the newer electronic

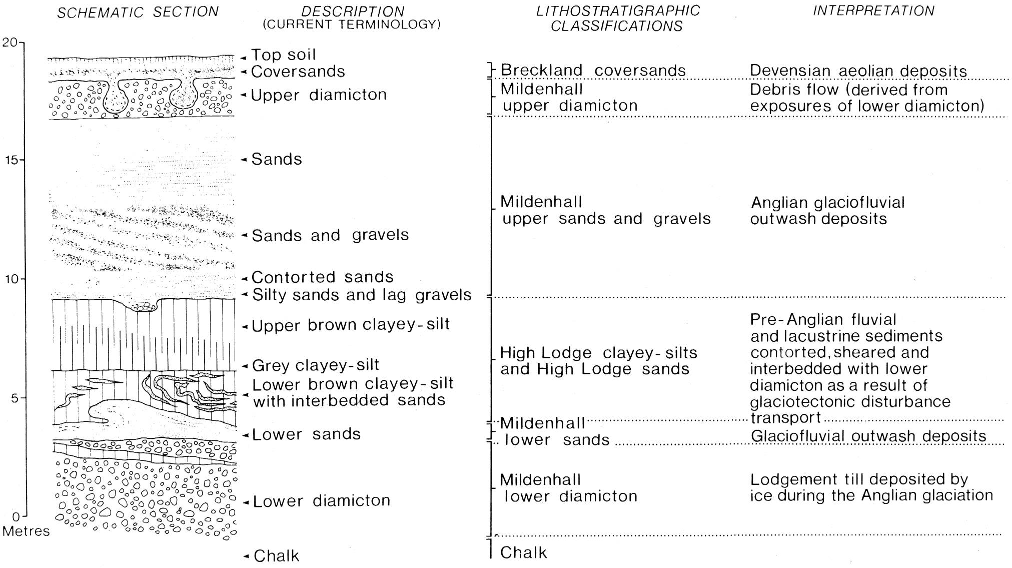


Figure 11.3 Schematic section through deposits at High Lodge, with the field description, lithostratigraphic classification, and interpretation of the units. (Reproduced from Ashton et al. 1992: Fig. 3.1, with permission; original caption, © the Trustees of the British Museum, British Museum Press.)

Table 11.3 Particle size classes: the Wentworth scale and phi (ϕ) units

Wentworth scale	phi (ϕ) units	mm equivalents
Boulder	− 8.0 and larger	>256.0
Cobble	− 6.0 to − 8.0	64.0 to 256.0
Pebble	− 2.0 to − 6.0	4.0 to 64.0
Granule	− 1.0 to − 2.0	2.0 to 4.0
Very coarse sand	0.0 to − 1.0	1.0 to 2.0
Coarse sand	1.0 to 0.0	0.5 to 1.0
Medium sand	2.0 to 1.0	0.25 to 0.5
Fine sand	3.0 to 2.0	0.125 to 0.25
Very fine sand	4.0 to 3.0	0.0625 to 0.125
Coarse silt	5.0 to 4.0	0.0312 to 0.0625
Medium silt	6.0 to 5.0	0.0156 to 0.0312
Fine silt	7.0 to 6.0	0.0078 to 0.0156
Very fine silt	8.0 to 7.0	0.0039 to 0.0078
Clay	8.0 to 12.0	0.00024 to 0.0039
Colloid	<12	<0.00024

Note: The Wentworth scale is quantified in mm units (3rd column); the categories granule to cobble are subdivisions of gravel. The phi scale is logarithmic, based on 0 = 1mm. The conversion is $-\log_2$ of the diameter in mm; the advantage is that the scale is then based on whole numbers. Negative numbers represent particle classes larger than coarse sand; positive numbers represent the progressively finer categories.
Sources: Adapted from Lincoln et al. 1982, Lowe and Walker 1984, and Waters 1992.

sensing devices employing laser beams. Particle size definition can be accomplished by direct measurement or by reticules used with microscopes, except for the small size ranges (Table 11.3).

The analysis of particle shape is more problematic than that of size, since methods for rapid and objective determination of shape classes are still being developed. In cases where particle shape is a critical element for interpretation of sediments (uncommon in archaeology), it is best to consult someone working actively on this subject. The surface textures of sedimentary particles themselves contain useful information about the transport and depositional environments of particles. Viewed under high magnification in a scanning electron microscope (SEM), quartz grains, especially, display a wide range of surface textures and microtopography developed in different environments (Krinsley and Doornkamp 1973). While the interpretation of these forms is less than direct and obvious (Brown 1973), there are occasions when this method in the hands of an experienced researcher can distinguish between otherwise ambiguous possibilities: e.g., the distinction between beach and dune sands.

Box samples of sediments retain the fragile three-dimensional relationships among the mineral, air, and water contents of a sediment. Measurements of mass, density, porosity, and moisture content, quickly and reliably achieved, may refine description and interpretation of the sediment. Analyses of the "fabric" of a sediment involve measurements of the preferred orientation and dip of major particle classes, and of the proportion and distribution of air spaces ("packing") among the particles. Orientation studies, traditionally important in the analysis of glacial deposits, are also relevant to the study of alluvial material, in which context these studies may have archaeological applications. Widely applicable in sedimentology but still underutilized in archaeology, is X-radiography (Butler 1992). Radiographs illustrate even small structures in aqueous and aeolian deposits, such as ripples that indicate the direction and velocity of the transporting water or wind, as well as a range of inclusions. Thin sections of impregnated sediments studied under magnification reveal fine details of structure and content not otherwise visible (Bullock et al. 1985; Catt 1986: 180–181; Courty et al. 1989).

PEDOGENESIS AND DIAGENESIS

> Soil is never truly in equilibrium with its environment although we often assume an equilibrium state in order to develop an understanding of processes.
>
> WILD 1993: 90

We have seen that rocks at the junction of geosphere and atmosphere are modified by weathering. Similar chemical and physical changes modify sediments after deposition, during periods of relative stability. Groundwater acidified by carbon dioxide dissolves and redeposits salts and oxides to begin the long process of diagenesis – turning sediments into rock. Physical churning and chemical changes induced by plant and animal life on and in deposits, aided by moisture and temperature changes, begin almost immediately to form soils, a process called **pedogenesis**. Humans, also, have been active agents in soil formation and degradation for a very long time. The overriding difference between diagenesis and pedogenesis is that the latter process is dependent upon living organisms; organic matter and biological activity are essential to the transformation of mineral deposits into soil.

Natural transformation near the surface

Pedogenesis is a continuous, reversible, and interactive process. It works progressively from the surface down into underlying sediments. Organic matter collecting on the surface releases acid compounds that, carried down in water, begin the chemical transformation of mineral matter through differential leaching and deposition. Classic definitions of pedogenesis recognize five "soil-forming factors" (Jenny 1941):

- climate
- biota
- topography
- parent material
- time.

Within each factor many different processes react with sediments to produce soils that vary strongly but systematically in space and time (Brady 1990; Catt 1986; Johnson et al. 1990). Because the five factors cannot be quantified, are not independent, may vary through time, and operate at different scales, soils scientists seem rather embarrassed by them (Birkeland 1984: 162–168). However, they comprise a useful mnemonic for the major contributors to pedogenesis and the environmental constituents that may be studied through soils.

The climate of any particular place is a product initially of the moisture and temperature ranges defined by atmospheric circulation interacting with the elevation, slope, and aspect of landforms (Fig. 11.4). Extreme climatic components interrupt or delay pedogenesis. For example, wet sediments slide downslope, interrupting pedogenesis. Sediments below the water table resist the normal oxidizing reactions that typify active soils. Permanently frozen sediments do not form soils, although they develop characteristic structural features that record environmental conditions (Washburn 1980). Very arid climates delay soil development and produce typical desert soil characteristics, such as subsurface carbonate concentrations. In temperate climates, on level to moderately sloping ground, vegetation strongly influences both temperature and moisture at the ground surface. High temperatures in the tropics accelerate oxidation and leaching, causing soils to mature rapidly.

Biota in life and death contribute the essential organic matter from which pedogenic chemicals are derived. The physical churning of sediments and soils caused by animals and plants is collectively termed **bioturbation** (disturbance by living things). "The soil is clicking, turning, and changing with the energies of a fantastic variety of occupants" (J. Hay 1981: 38). Bacteria, and animals on a scale from minute mollusks through ants, earthworms, and larvae to insectivores, rodents, and lagomorphs (rabbits, hares), live within or actually derive their food from the soil. Some prey upon each other, entirely underground. In the course of their daily routines, some of these animals displace or digest significant masses of sediment, some of which is then redeposited on the surface. Plants, which we tend to think of as static creatures, disrupt the soil that supports and nourishes them as they expand in

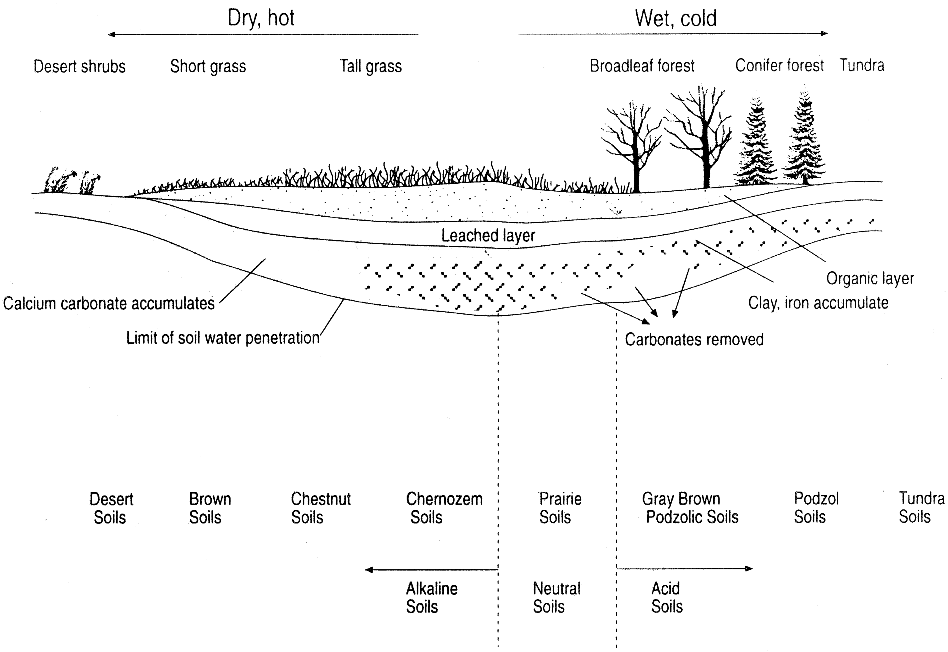


Figure 11.4 Diagrammatic transect through North America from southwest to northeast, showing pedogenesis varying with vegetation and climate. (After Hunt 1974: Fig. 6.6.)

size and increase in numbers, prying apart the sediments to make space for their roots. When roots die and rot, sediment drops into the hollow spaces left. Trees, when mature or aged, may be toppled by storms or their own failing strength; root masses tear out sediment which then gradually falls back, destroying the original structure.

Topography's influence on soil formation derives from the effects of slope, relief, elevation, and aspect. Any slope at all permits movement of unconsolidated materials to lower elevations. The steeper the slope, the thinner is soil likely to be because of the shorter residence time of particles in any one place. Elevation and aspect influence microclimate, and thus the biological activity in sediments. Even adjacent slopes on the same parent material will support soil differences according to the direction faced and solar energy received.

"Parent material" is simply the original deposit of sediment, which determines some of the soil's potential for development in terms of chemistry, grain sizes, compaction or permeability, and other characteristics that set parameters for interactions with groundwater and air. Over time, soil-forming processes will modify all such characteristics by the cumulative effects of chemical action and physical churning (Wilding et al. 1983).

Despite the diversity of factors and components, some basic processes characterize all soil development. Rainwater is acidified by carbonic acid and humic acid, the latter from the decomposition of organic matter. Acidic water leaches salts, oxides, and clays from the upper range of sediment, the **eluvial** zone, and carries the solutes and fine particles deeper into the sediment column where they may be deposited in the **illuvial** zone (Table 11.2). The subtraction of fine materials from the top and their deposition below create soil **horizons** within the sediment body: contrastive zones parallel to the surface. In vertical section, such bands constitute a soil profile (Table 11.4; Fig. 11.5). The color, texture, and properties of successive zones are changed by the pedogenic processes acting in place, progressively deeper with time.

Soil horizons are formed by change *in situ*, subsequent to the deposition of sediment. They are not lithostratigraphic units, not layers, beds, or strata. In fact, the color and textural changes of horizonation eventually destroy original stratification by homogenization and chemical modification. Therefore, *soil horizons are not stratigraphic markers.* Although they are occasionally so used in archaeological field work, such usage seriously distorts the concept of stratification and may preclude any clear understanding of depositional events at a site.

All the soil factors play out their roles, harmoniously or competitively, as long as a sediment body remains in place. The relative influence of any factor varies with conditions, and the others adjust in turn. Soils are ecosystems; pedogenesis is nothing if

Table 11.4 Soil horizon nomenclature[a]
Master horizons recorded in field studies

O horizon	Organic material accumulated on the surface.
A horizon	Humified organic matter mixed with mineral substrate near surface; typically dark-colored.
E horizon	Light-colored mineral horizon from which oxides, clays, and organic matter have been chemically leached (eluviation zone).
B horizon	Mineral horizon underlying O, A, or E horizon with little evidence of original sediment structure. May be zone of accumulation (illuviation) of sesquioxides, carbonates, and/or clays. Typically red in color.
K horizon	Subsurface horizon impregnated with carbonate so that carbonate dominates the structure. Typical of soils in arid climates.
C horizon	Parent material of the soil, only minimally transformed by pedogenesis, underlying A and B horizons. May show some evidence of weathering.
R horizon	Hard bedrock.

Note: [a] Soils nomenclature, even for horizons, is more detailed than this, and varies geographically.
Source: Adapted and simplified from Birkeland (1984: 7), which see for details.

not dynamic. Over time, barring interference, a soil develops from rawness to a maturity that supports a richly diverse biota within and above the soil.

Soils are considered immature or mature, according to the extent to which they inhibit or permit the full climatic potential of vegetation. However, maturity is not stasis; soils continue to change until they are destroyed. Pedogenesis is reversible, usually by burial or climate change, but soils can also be exhausted by vegetation demands. Eventually, the natural permeability of any soil is compromised by increasing clay concentration, compaction at depth, or development of carbonaceous or mineral hardpans, depending on the type of soil. The amount of time required for such degradation varies with everything that influences pedogenesis, and so cannot be predicted closely (Johnson et al. 1990). Birkeland (1984: 204–220) suggests that an organic A horizon develops to a steady state within a century or so, whereas a B horizon takes thousands of years. Local factors will override these generalities, and it is worth noting that soils utilized for farming or grazing do not have natural histories. Old unburied soils, called relict soils, continue to evolve and change as long as they remain at the surface; their histories are partly recorded in their chemical and physical compositions.

The B horizons of soils are the most dynamic areas in profiles. Soils taxonomies offer a large set of labels for different kinds of B horizons (e.g., Holliday 1990), the

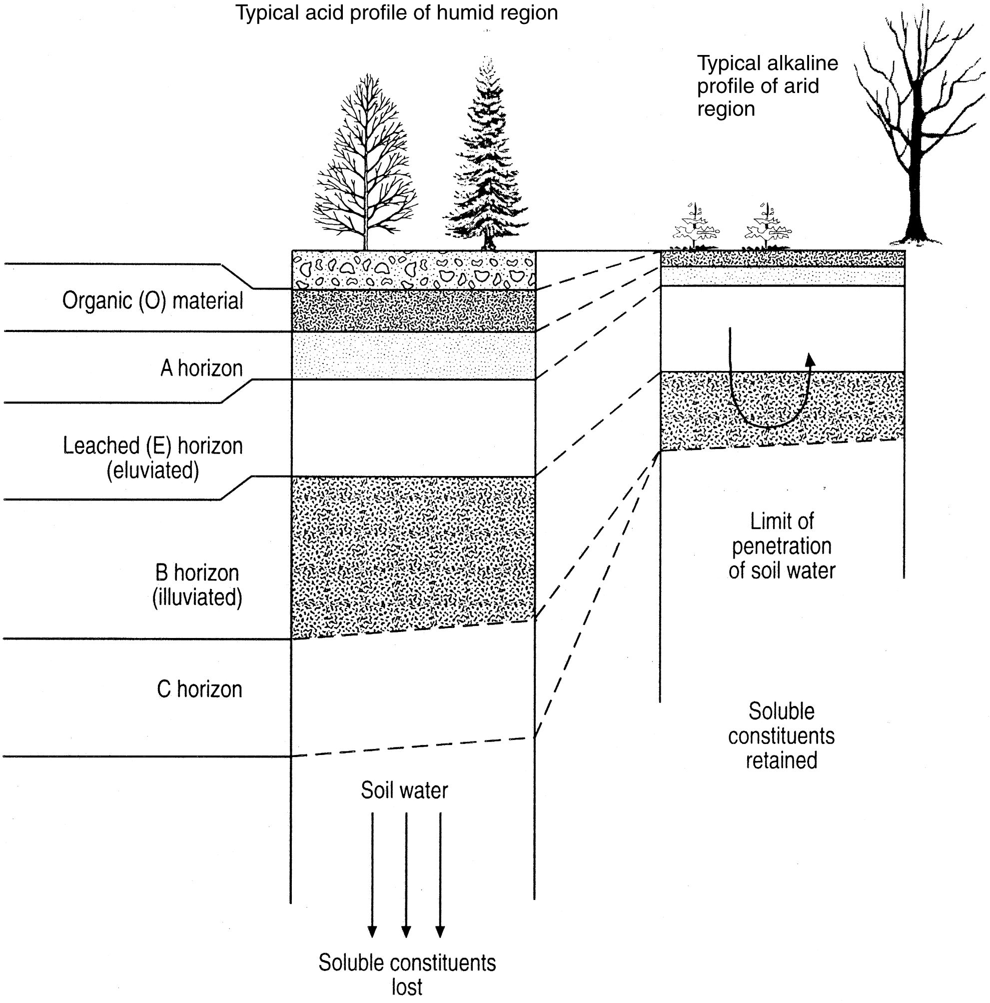


Figure 11.5 Soil horizons under two different climatic regimes. (After Hunt 1974: Fig. 6.4.)

most used of which are Bt for clay accumulations and Bk for calcareous concentrations. The former are typical of soils in moist environments; the latter, in arid climates. Both clays and carbonates accumulate in soils from two sources: chemical changes within sediments due to weathering, or the addition of fine particles from airborne dust. In both cases, the fine particles accumulate in the B zone where they may ultimately dominate completely over original constituents. Soils with very high clay contents, from either original deposition or illuviation, shrink when drying and swell on wetting (Vertisols). The deep cracks that can form when such materials dry out cause slumping and the introduction of surficial materials at depths. The intricate cracking of sun-dried mudflats is a familiar small-scale example of such activity. Carbonate concentrations in soils range from fine thread-like features to massive, cement-like secondary deposits called **caliche**. Discussions of B horizons particularly useful for archaeologists are found in the volume by Retallack (1990: 264–276).

Cultural transformations of soils

Human activities strongly influence soil development by disturbing superficial sediments and by changing their chemical composition. Humans are major bioturbators of soils and agents of deposition. Their dominance today relates to sheer numbers as well as to the efficacy of their tools for digging, transporting, and depositing (Schiffer 1987). Plow zones are such ubiquitous phenomena that they have a soil horizon designation of their own (Ap). In the US soil taxonomy, surficial horizons chemically enriched by human wastes and food debris have been given a special taxon: anthropic epipedon. **Epipedon** is the technical name for horizons near the surface (O, A, E, B); the meaning of "anthropic" should be self-evident. Archaeologists speak of "**anthrosols**," but this term has not established itself in soil science. It is only a matter of time before it must be recognized as another class of **azonal** soils (Chapter 12). Anthrosols are characterized by physical disturbance, high organic content, and phosphate enrichment typically due to high concentrations of animal wastes (Eidt 1985). Exotic sedimentary particles (e.g., artifacts) are diagnostic of anthrosols only in company with chemical and structural changes.

Analysis of ancient agricultural soils is a developing methodology. Chemical analyses and micromorphological (thin-section) methods have been most successful in identifying the changes in soil chemistry and structure attendant upon long-term plowing, fertilization, and irrigation (Artzy and Hillel 1988; Groenman-van Waateringe and Robinson 1988; Sandor 1992). Less attention has been paid to hoe horticulture and agriculture without animal wastes as those affect soils, but interest and a literature are growing (e.g., Denevan et al. 1987). Paleolimnology also has provided insights into the scale of soils transformation and disruption due to hand cultivation (Binford et al. 1987; Deevey et al. 1979; O'Hara et al. 1993).

Transformations at depth

In the C horizon, below the surface levels at which soils are forming, sediments may be compacted, cemented, churned, and chemically changed by processes involving acids and ions carried in groundwater. Such subsurface processes collectively are termed diagenesis, and although they resemble some of the processes that form soils, their distinctiveness is important to archaeologists because they indicate different aspects of subsurface environments in which archaeological remains may lie. Because many diagenetic processes take place at the water table or below it, they can be useful indicators of the water table's dynamics. Oxides and salts carried in groundwater tend to be deposited at the water table. Iron oxides may accumulate, singly or in sets, as wavy red lines that superficially resemble either bedding planes or genuine soil horizons, but are very different from both. This red color may also be confused with burning. Hydrated manganese and iron compounds may impart distinctive gray colors to sediments at depth, creating mottling or the more pervasive *gleying* that can puzzle the uninformed by suggestive resemblance to buried soils. Oxides deposited from groundwater have been involved in many disputed archaeological interpretations. Massive accumulations of oxides and carbonates create "hardpans," cemented rock-like layers that have no history related to surface exposure.

Clays leached from surficial zones of soils and redeposited at depth resemble units of stratification. Nodules of clay or oxides resembling small stones may form within sediments at depths determined by groundwater; their presence must not be interpreted as a "stony layer" indicative of stratification. The gradual settling and compaction under pressure of overburden experienced by water-saturated sediments such as pond muds and peat deposits is an aspect of diagenesis with special implications for any included archaeological materials.

Soils formed originally at the surface may be later buried. After burial these **paleosols** cease to develop as soils and become subject to subsurface modifications by any of the processes active at their location. They may undergo eluviation, illuviation, cementation, compaction, or other processes that change their characteristics (Retallack 1990). Buried soils retain for some time the attributes and inclusions they acquired while at the surface and, as a class, are among the most valued repositories of paleoenvironmental data. At the time of burial they cease to be organically active and thus, unless truncated, may be more readily dated by radiocarbon than their equivalents at the surface, which actively incorporate new organic matter throughout the zone of bioturbation. In time, organic materials and the A zone are lost to diagenetic changes that transform soils properties (Brady 1990).

Soils chemistry

The destructiveness of soils acids has long been recognized as limiting preservation within the archaeological record. Such recognition has established the notation of soil pH (hydrogen ion concentration) as a standard item in archaeological field reports, typically with no further discussion. As with the phi scale of particle size (Table 11.3), pH is a negative logarithm such that the higher the hydrogen concentration, the lower the index number. On a scale from 0 to 14, 7 is neutral, lower numbers are increasingly acidic, and higher numbers are increasingly alkaline. In highly acid soils, organic macrofossils are rapidly destroyed; however, the inhibition of oxygenating bacteria in such soils results in good preservation conditions for pollen grains. Soils with high pH are poor environments for the preservation of pollen, but are excellent for preserving bone and shell. Consequently, carbonate crusts in archaeological sediments are worth a close look; they may contain important organic materials.

Water-saturated soils effectively exclude oxygen and thus preserve from decomposers otherwise fragile organic matter, including artifact classes rare in other environments (Coles 1984). Damp or wet soils have another property significant for archaeological materials: they support the transfer of ions between buried organic materials and their soils matrices. The resultant chemical changes, tending toward homogenization, complicate chemical analyses of archaeological materials (Lambert et al. 1984; White and Hannus 1983). Chemical analyses can be useful complements to other environmental study of soils formation and history, helping to test alternative interpretations (McBride 1994).

Pedogenesis and the archaeological record

The dynamism of pedogenesis has implications for archaeology and for the condition of archaeological sites that are only recently being appropriately exploited. Pedogenesis as a process affecting the environment of burial is progressive, continuous, variable, and reversible. Because soil, as matrix and environment of burial, affects what is available for archaeological study and the integrity of observed associations, as well as the condition in which materials are recovered, a basic awareness of pedogenesis is essential to successful field work.

Postdepositional disturbance of sediments and soils by the churning actions of organisms and/or ice has dismayed many observant archaeologists. Artifact associations and features underground can be disaggregated or rearranged by such mechanisms, eliminating parts of the record and creating false associations (Johnson and Watson-Stegner 1990; Wood and Johnson 1978). Such destructiveness is inherent in pedogenesis; it is nearly ubiquitous and must be anticipated by responsible archaeologists (Barker 1993).

Regrettably for archaeologists, archaeological sites are not exempt from the natural processes of erosion and deposition that ceaselessly reshape the surface of the

Earth. Erosional processes destroy landforms and displace everything on them. With each loss, the fabric of ancient landscapes is torn. Depositional processes bury surfaces, subjecting them to compression, deformation, and diagenesis underground. Overburden not only changes the environment of sediments, it also introduces mechanical stresses that result in compaction, displacement, and particulate sorting within them. Pedogenesis and diagenesis, erosion and deposition, all restructure the archaeological and paleoenvironmental records, requiring of interpreters precise, thoughtful observation and application of informed imagination (Schiffer 1987).

SOIL SCIENCE

There are many reasons why archaeologists must know something about soil science; in fact, the more the better. However, it is a discipline of its own, not something one can pick up as a sideline. Here, with the emphasis on paleoenvironmental reconstruction, a few matters are presented with the hope that they can ease readers into the sometimes arcane literature of an important discipline (Fanning and Fanning 1989).

The analytical language of soil science is probably the major obstacle to its use in archaeology. Contemplating the formal soil taxonomy used in the United States and increasingly elsewhere, Birkeland notes (1984: 42) that it "carries such exotic combinations of Latin and Greek as Cryaqueptic Haplaquoll, Aquic Ustochrept, and Natraqualfic Mazaquerts." No matter; for the archaeologist the real problem with the US soil taxonomy is that the criteria for classifying soils do not include historical (genetic) concepts (Guthrie and Witty 1982; Hallberg 1985; Soil Survey Staff 1975). Furthermore, and perhaps for this reason, in practice the soil units shown on local-scale soils maps bear only a tenuous relationship with the geological sediments of the parent materials. The Canadian system includes more genetic information and is therefore more immediately applicable to archaeological and paleoenvironmental uses (Canada Soil Survey Committee 1978). Most industrialized countries have a system of their own (e.g., Avery [1980] for Britain, Stace et al. [1968] for Australia). UNESCO is developing international conventions for soils maps (Fitzpatrick 1980), but in the short run, archaeologists must familiarize themselves with the system in use locally, and learn the limits of its reliability and applicability (e.g., Catt 1986; further discussion in Chapter 12). Beyond that, active collaboration with a soil scientist is the best approach.

Because of the way that they form, soils on contiguous landform surfaces vary with the topography, vegetation, parent material, and microclimates. Soils, therefore, provide information on the development and environmental history of landforms and surfaces. **Catenas** are sequences of soils profiles varying downslope. At the top of a slope soils are typically well drained, but subject to erosion and therefore relatively thin. At intermediate elevations slopes may be more gentle and soils development deeper. At the foot of a slope, sediments accumulate; soils may be deep, but any drainage limitations will directly affect the soils on such landforms. Geomorphologists exploit catenas to study the subtle diversity of landform histories and the complex play of the soils-forming factors (Birkeland 1984: 238–254; Daniels and Hammer 1992; Gerrard 1992; Knuepfer and McFadden 1990).

Standard analytical methods that provide clues to soils' histories include the conventional chemical spectra, grain-size analysis (granulometry), and percentage of organics (e.g., Holliday 1990; Macphail 1987: 361–363), which provide information about the parent material and the transformations it has experienced. The transformations are sometimes interpretable in terms of the soil-forming factors that dominated in the past. Analysis of the relative degree of eluviation and illuviation amongst horizons provides information about relative ages of soils and the contributions of the five factors of soil formation. Soil thin-sections reveal evidence for disturbances and microstructures in the soil that are directly relevant to environmental history (Bullock et al. 1985; Goldberg 1992; Limbrey 1992).

Paleoenvironmental information is derivable from soils once one learns to elicit it analytically. In a volume addressed to paleopedology at geological scales, Retallack (1990) presents particularly clear discussions of the potentials and limits of soils for paleoenvironmental inferences. Fine resolution is rarely possible, but if, say, the climate during soil formation differed significantly from that at the time of observation, some evidence of that difference may survive. Changes in the height of the water table can be read in some subsoils. Resolving the differences in terms of time, however, is difficult. Soils contain biological residues such as pollen, charcoal, and microorganisms that serve as climate proxies when they can be dated. Buried soils are the best sources of paleoenvironmental data; when the time of burial can be specified, they can be eloquent. A promising technique for reading past environmental states directly from stable carbon isotopes in soils is based on the differences in carbon metabolism between tropical grasses and other vegetation (C_4/C_3 ratios). In suitable areas, vegetation community successions, interpreted as changes in microclimates, have been tracked by carbon-isotope ratios in organic residues in soils (Ambrose and Sikes 1991). Biological data derived from soils are considered further in Parts VI and VII.

There is a growing, tantalizing literature on direct dating of soils (Andersen 1986; Matthews 1993; Scharpenseel and Becker-Heidmann 1992). Archaeologists cannot help but be attracted, even tempted, by the preliminary results and claims.

Nevertheless, this subject must be approached with critical awareness. Organic matter is certainly available in soils and can be extracted for dating. The difficulty is in knowing which organic matter is equivalent in age to the archaeological event to be dated. Soils development is a time-transgressive process. Organic matter cycles in and out of soil throughout the active life of the soil; roots and burrowing animals penetrate deeply into the B zone. Bulk sampling of soil carbon in the A zone will give ages averaged over the duration of the soil with bias toward the youthful side; technically, what is dated is the "average residence time" of the organic matter sampled. Now that AMS dating permits the selection of definable components of soil (charcoal fragments, humates), sample selection will define the age determined. Soluble humates tend to be younger than bulk charcoal, but they may sometimes be older. Dating of close-interval samples through a sequence of soil horizons often demonstrates that soils are frequently churned, although the progressive translocation of older carbon compounds into the B zone may impart some semblance of stratigraphic order to the sample ages. Soils with archaeological materials include organic matter both younger and older than the anthropogenic materials. Soils processes are continuous; archaeological deposition is episodic.

CODA

Sediments and soils are the essential contexts of field archaeology; on their appropriate interpretation rests all understanding of the relationships among artifacts and aspects of environments, past and present, as well as understanding of relative ages. The results of pedogenesis must be correctly distinguished from variation in sediments. The structural priority of sediments over soils developed in them is a fundamental tenet of analysis. The field relationships of sediments and soils must be demonstrated convincingly before interpretation of included archaeological remains can begin.

ARCHAEOLOGICAL MATRICES

> The excavator's aim should be to explain the origin of every layer and feature he encounters whether it be structural or natural; made by man, animal or insect, accidental or purposeful.
>
> BARKER 1982: 68

Every surface on which humans lay foot or artifact is a potential archaeological site, requiring only that subsequent processes not dislodge and transport the surficial deposits. Of course, disturbance of surficial sediments of every kind is the normal case. This vulnerability ensures that archaeological sites are neither ubiquitous nor permanent.

The focus of this chapter is on sediments and soils as matrices of archaeological sites, at local and micro-scales. We occasionally lift our eyes to regional-scale phenomena, as in considering the information potentials of widespread deposits of loess or volcanic ash, but we pay no attention here to the mega- and macro-scales of phenomena or to regional-scale interpretations.

MESSAGES IN THE MATRIX

Sedimentological analyses are undertaken to learn about the sources, transportation agents, depositional and transformational history of the materials comprising deposits (Chapter 11). Although archaeologists typically treat that information as background, environmental archaeology must begin with the environments in which materials, whether cultural or natural sedimentary particles, were brought to a site, deposited, and affected by postdepositional processes including pedogenesis and diagenesis. The enclosing matrix is the fundamental source of information about all the processes essential to understanding the context of human behavior at a site. Not all evidence is visible, and not all is extractable by techniques currently known. However, for sites lacking written evidence the matrix is the only source of non-artifactual information; for sites with written histories, the matrix will variously confirm, expand, or contradict elements of that record.

Archaeological matrices are very complex deposits. Altogether they represent a significant subset of the geological diversity on the surface of the Earth, compounding that diversity with incorporated cultural debris. The complex chemical and compositional attributes of archaeological matrices are worthy of more intense analytical scrutiny by geologists and archaeologists than they have typically received (Stein 1987).

Burial and incorporation

Archaeological materials and structures go underground within a wide range of environments of burial or incorporation, but in any single instance the range is finite and potentially determinable. Archaeological materials deposited on surfaces may be (1) buried by sediments added on top or (2) incorporated into existing sediments by pedoturbation or other disturbances. Many different materials and processes are capable of burying surfaces and whatever is at rest on them. Overbank floodwaters, rising lakes and seas, wind-borne particulates, expanding peat bogs, volcanic ash, colluvium, and carbonates precipitating from solution may all bury sites, gently or violently. On the other hand, everything that disturbs the soil, from ice crystals through burrowing animals to mechanized excavating tools, may contribute to the incorporation of archaeological materials into regolith. The burial and incorporation processes, whether episodic or continuous, are always potentially reversible: deeper is not always older.

Environments of incorporation determine the diversity of environmental data that is included and preserved in matrices. The full range of matrix constituents is almost never buried or incorporated at the same time; that is, incorporation is nonsynchronous for the artifacts, ecofacts, and other materials included. Whatever archaeologists retrieve must be evaluated for its relevance and synchroneity to the cultural constituents.

Transformations

As the discussion in Chapter 11 showed, materials under the surface are not at rest. Rather, they are subject to all the equifinalities of pedogenesis and diagenesis that keep the regolith in a dynamic state. Materials close to the surface are affected by pedogenesis at rates varying with climate and other soil-forming factors. The chemical changes of soil formation and diagenesis influence the relative preservation of ma-

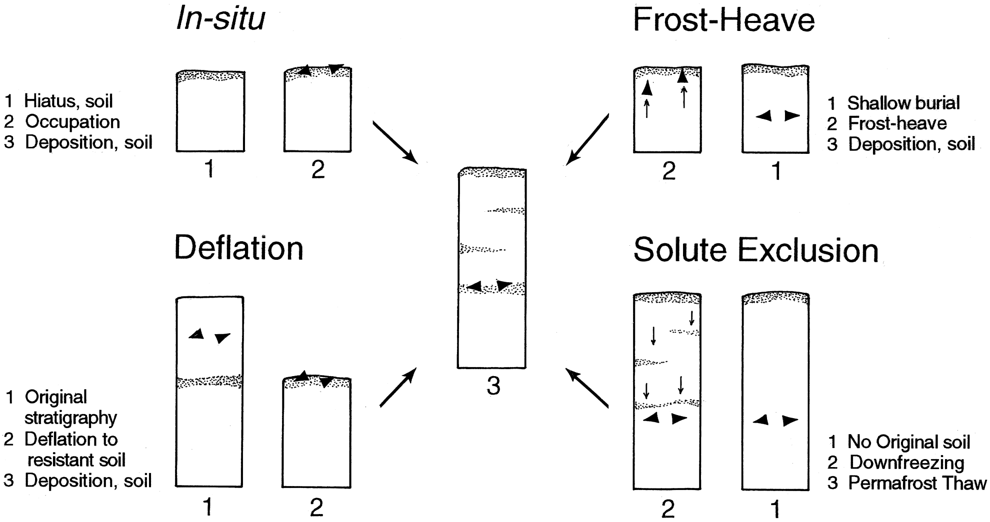


Figure 12.1 Four hypothetical original stratigraphic states resulting in the situation shown in the central column. The numbers indicate the sequence of states and transformations in the four cases. Black triangles represent artifacts. "Solute exclusion" is cases where permafrost or hardpans stop groundwater above the base of sediments, concentrating leachate after deposition. (After Thorson 1990a: Fig. 3.)

terials. Bioturbation by resident organisms modifies original spatial relationships. Objects below the surface – pebbles, cobbles, and boulders no less than archaeological artifacts and ecofacts – will be moved up, down, and sideways by living creatures and by ice wedging when the climate supports that (Fig. 12.1; Thorson 1990a; Wood and Johnson 1978).

No degree of exquisite excavation technique in archaeological matrices is likely to reveal absolutely pristine spatial associations; careless excavators unconcerned about the messages in the matrix recover only inadequate approximations to the relationships there (Barker 1993). Everything about the matrix – color, texture, structure, microstructure, water content, pH, biota, particles, and so forth – has a message to tell about the history of the materials it encloses. The history is analyzed first at the site and more intensively in the laboratory; the two opportunities are complements, never substitutes. Post-excavation analyses cannot overcome deficiencies in excavation observations, recording, or sampling.

Integrity

Archaeological accumulations buried gently under fine-grained materials, ideally either totally dry or anaerobic and permanently wet or frozen, may retain much of their original integrity of materials and associations. This is the ideal of archaeological context. In contrast, materials deposited in high-energy environments such as fast-moving rivers, glacial ice, dunes, beach deposits, or any of the range of diamictons are transferred from archaeological into geological context; their spatial relationships are defined by natural forces rather than human behavior (Binford 1981). In such cases, the sites must be considered geological phenomena with the artifacts being simply "rather specialized particles" among many others subjected to transport and deposition (Macklin and Needham 1992: 10).

The structural integrity of archaeological matrices can be evaluated on the basis of diverse criteria, some of which are likely to be available in any given situation. The first line of attack is the soil profile: are horizons comparable in development and condition to those within the same substrate away from the archaeological deposits? If not, it is necessary to decide whether the differences can be explained by anthropic enrichment of site soils, by age differences, or by disturbances (Johnson and Watson-Stegner 1990). Are features such as hearths and pits visibly intact in the substrate? Are artifacts in resting positions, not turned on their edges by cryoturbation (Wood and Johnson 1978)? Do the artifacts appear to be sorted by size or weight, as would happen to those affected by clay turbulence in Vertisols, by cryoturbation, or by slippage of water-saturated sediments on even moderate slopes? Are animal burrows evident or likely given the nature of the deposit (Erlandson 1984)? Are earthworms numerous (Armour-Chelu and Andrews 1994; Stein 1983)? Are artifacts of recent age present at depth in the sediments? Small objects such as the once ubiquitous pop-tops for aluminum cans, coins, nails, broken glass, and such can be informative about soils disturbances. Alertness to all these matters as the trowel moves is crucial to an ultimately successful interpretation of samples of any kind from the matrix. Without field records explicitly addressing such issues in terms of presence, absence, and patterning, *post hoc* claims for integrity of association or its antithesis can never be more than hypotheses beyond the reach of testing (Barker 1993; Schiffer 1987; Waters 1992: Ch. 7).

The chemical integrity of soils and sediments, especially, cannot be assumed. Recent and ancient anthropogenic contaminants need bear no relationship to any archaeological materials included in a matrix. Acid rain and other modern pollutants similarly need have no great age to be significantly misleading to analysis. Even beyond the site area itself, modern or ancient contaminants may thwart the search for analogs and comparative samples needed for interpretation of soils histories. Soils with a history of deforestation and agriculture are distinct from forest soils, in ways that change their chemistry to the detriment of archaeological analyses (Kaiser 1996; Trumbore et al. 1996).

Paleoenvironmental data

The enormous range and diversity of paleoenvironmental data in archaeological sites exceeds the capacity of this volume to inventory. Evolving knowledge of sediments, soil processes, and living things, complemented by improvements and innovations in methods of examination, outpaces the ability of archaeologists to encompass them in investigations. A brief review here may help establish the value of expanding awareness. Additional information about these techniques is in Chapter 11, Parts VI and VII, and sources noted.

In sediments

Laboratory analyses for interpreting environments of deposition or incorporation of archaeological remains will normally begin with the sediments. As techniques are being developed all the time, there is no need here to recommend specifics, only to review some of the currently most informative approaches.

In most situations granulometric methods based on the Wentworth or phi scales (Table 11.3) suffice for interpretation of depositional agents; at the least, they can indicate situations in which further analysis is necessary for resolution. The major caveat for direct interpretation of depositional agents on the basis of granulometry is the fact that archaeological sediments typically include materials added by human activity – sediments tracked or carried onto living floors, debitage and trash accumulated during processing, damage to exposed bedrock on site, erosional byproducts of mudbrick or fired clay, plant and animal refuse. All such additives must be eliminated from samples processed for interpretation of natural agents but, of course, they are informative about the full suite of formation processes at a site. Organic detritus observable in screening, flotation samples, or thin-sections will be informative about biogenic disturbances as well as environments during and following deposition.

Identification of the mineral suite in a deposit may help to define the source of the sediments; at the least, it can be the basis for distinguishing successive depositional events should those require clarification. Clay mineralogy reflects processes of pedogenesis and diagenesis. Micromorphological analysis of thin-sections can be very informative about the history of deposits, especially the illuvial (B) zones of soil horizons where disturbances of many sorts leave characteristic signatures (Courty et al. 1989; Fitzpatrick 1993). Organic remains and signals of bioturbation present problems for synchroneity, but offer information worth reaching for.

In soils

Information is sought in soils about former environmental conditions near the surfaces on which they formed. The proxies required are those properties or inclusions of soils that reflect conditions of precipitation and temperature prevalent at times of interest in the past. The purely pedological proxies relate to transformations of sediments in the epipedon – chemical and mineralogical changes brought about by weathering, eluviation, and illuviation. The distributions and states of minerals and simple compounds, as well as small structural changes within the epipedon, can provide information about climates: warm or cold, wet or dry episodes at scales above the seasonal (e.g., McBride 1994).

Paleoenvironmental inferences based on soils data are less reliable for archaeology than is often assumed, because of the coarse chronology and multiple sources and ages of soils constituents. Nonetheless, coarse as the resolution of soils data may be, they have important roles in environmental archaeology (Holliday 1992; Matthews 1993; Waters 1992). The Comprehensive Soil Classification System of the United States Department of Agriculture (Soil Survey Staff 1975) offers an internationally applicable, high-level classification of soils. The great soil Orders, at the top of the hierarchical taxonomy, are defined in terms of selected attributes of environments in which they formed: precipitation, temperature ranges, seasonality, and some chemical attributes. Six of the ten Orders are zonal soils classes, whose distributions reflect the influence of zonal climates. In Figure 12.2 these are shown as they would be distributed on a hypothetical continent in the northern hemisphere lacking major contrasts in relief. The remaining four azonal soil Orders are less constrained by climate, but emphasize attributes of soils relevant for archaeology. Table 12.1 lists the ten Orders, their major characteristics, and the suffixes that identify them in creating names for "Suborders" in the taxonomy.

The limitations of soils data for reconstructing environments at archaeological scales are less constraining in the geosciences, where most paleoenvironmental research has been done (Lowe and Walker 1984; Retallack 1990). The older the archaeological site and the more geological the context, the more useful will be geological data. For archaeology, the optimal contexts for recovering environmental data are those rapidly deposited and sealed by human agency, not soils of any kind. Deposits at intimate scales of time and space, effectively protected from pedogenesis, retain integrity. Surfaces buried immediately under structures, microstratified deposits such as middens, and the contents of deep pits are the treasuries of paleoenvironmental data. Their abundance in urban contexts justifies recent excitement about paleoenvironmental research in urban and town sites (Deagan 1996).

TERRESTRIAL MATRICES

The range of sediments in which archaeological materials occur is essentially bounded only by the range of unconsolidated materials on the planet (Chapter 11).

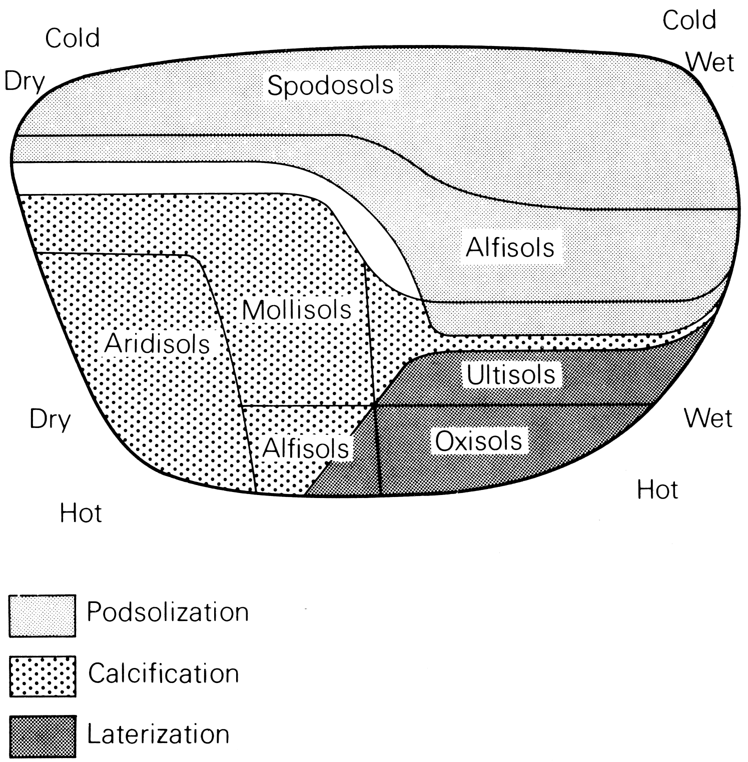


Figure 12.2 Schematic representation of the great soil Orders on a hypothetical continent in the northern hemisphere. (Reproduced with permission of Blackwell Publishers from Goudie 1993: Fig. 2.11. Original caption.)

Here we consider some typical sediments, classed by depositional environments, to emphasize characteristics of subaerial matrices, including soils, that set boundaries for paleoenvironmental investigations in archaeology.

Alluvial sedimentary matrices

Recent geoarchaeological literature presents us with contrastive statements such as this pair. "Holocene alluvial sequences are arguably unique in the way that they integrate and record environmental change (natural and anthropogenic) over a wide range of spatial and temporal scales" (Macklin and Needham 1992: 20, working with alluvium in the British Isles). On the other hand, Ferring (1992: 15), working in the arid American Southwest, claims: "In too many cases . . . direct evidence for paleoenvironments, such as pollen, snails or plant macrofossils, is poorly preserved in alluvial sediments." Both statements have merit; they highlight the importance of context, contingency, and multidisciplinary research designs suited to specific cases.

Table 12.1 Ten soil Orders of the USDA soils taxonomy[a]

Order	Suffix	Characteristics
Alfisols	-alfs	soils with argillic horizon and moderate base content; woodland and forest soils of temperate regions
Aridisols	-ids	soils of deserts and semi-arid areas; likely to have calcified B horizons
Entisols	-ents	new soils, slight development; strongly expressing sedimentary features
Histosols	-ists	soils composed of plant tissue accumulated in waterlogged areas; peats
Inceptisols	-epts	young soils, esp. on alluvium; precursors of other Orders
Mollisols	-olls	soils with dark, loose A horizon and high base content; usually grassland soils
Oxisols	-ox	soils with strongly oxidized B horizon; typically formed in hot, moist climates; laterites
Spodosols	-ods	podzols; soils rich with iron and aluminum compounds in the B horizon; cool moist climates
Ultisols	-ults	forest soils with argillic horizon but low base content; pronounced seasonal rainfall
Vertisols	-erts	soils with high clay content subject to cracking and shrink–swell turbulence

Note: [a] The ten soil Orders are the highest classificatory categories in the USDA taxonomy. The suffixes combine with additional syllables to form the Suborders, which are more specific, and more numerous, categories. Six of the Orders are zonal categories, identified with zonal climates. The other four (Entisols, Inceptisols, Histosols, and Vertisols) are azonal soils that form under special conditions in any zone.
Sources: Definitions after Birkeland 1984, Goudie 1984, Holliday 1992, Retallack 1990: 107–111, and Waters 1992.

Alluvial sediments and soils are polygenic, complex, and potentially very informative matrices for paleoenvironmental study that must be addressed appropriately, in recognition of their sensitivity to climate.

The biological necessity for water explains the frequent association of archaeological sites with alluvial deposits on floodplains and terrace surfaces. The dynamism of riverside environments brings a range of special problems and opportunities for the archaeologist. Old floodplain landforms are typically discontinuous, difficult to find or trace over distance even when not buried, and sensitive to disturbances from tectonics, climate change, human activities, or fluvial dynamics. Alluvial deposits originally laid down on floodplains are eventually buried by aggrading rivers or isolated as components of terraces after river incision. Archaeological sites encountered on fluvial terraces, therefore, may have been established originally on a floodplain, a very different environment. Artifacts in alluvial sediments unaffected by pedogenesis are most likely to have been deposited from water, whether or not they display transport damage, because they are not associated with surfaces. In such cases the only environmental issues that can be investigated are those of deposition of the displaced artifacts. Artifact clusters in soils, particularly if heterogeneous in the sizes and shapes represented, are likely to have been introduced by people occupying a stable land surface. If the depositional agents were people, and the matrix retains integrity, the way is open for investigation of environments prior to, during, and after human use of the location (Rapp and Hill 1998).

The mechanisms that bury archaeological sites in or on alluvium vary with the original topographical location of the site. Active floodplains can be used by humans only seasonally; even so, people are likely to favor elevated landforms. On meander stretches or large deltas, sites tend to be located preferentially on levees near channels or oxbow lakes. High-energy floods that overtop or break through levees may destroy such sites. Quieter floodwaters (slackwater) simply inundate a location and deposit a fresh layer of silt. Slackwater flooding occurs typically upstream of bedrock constrictions in valley walls, and inland from active channels. Quiet flooding, whether away from the channel or simply due to a minor overbank episode, may create stratified sites or merely "overthickened" A horizons in soils. Sites on terraces above active floodplains may be buried by hillslope processes such as mass-wasting or colluviation, or by windblown sands and silts. Burial processes differentially influence the integrity and preservation of sites.

For most of the twentieth century, paleoenvironmental studies in alluvium emphasized interpreting alluvial deposits directly in terms of climate change. The idea, unadorned, was that extensive gravel deposits represented cool wet periods, and soil development represented warm periods. The case study of alternative interpretations of Greek valley fills (pp. 320–325) introduces the importance of chronology in such studies, to test synchroneity of fluvial sequences from valley to valley (review Fig. 9.3). Clarity about the causes and timing of valley alluviation is important well beyond archaeology, because paleoenvironmental inferences find applications in modern land-use policies. In a thoughtful review of recent studies of human influences on alluvial processes, Bell (1992b) concludes that both climate and culture contribute to cycles of valley erosion and deposition in temperate latitudes. He notes differences in fluvial responses according to latitude and climatic regime, especially related to continentality, a geographical complication not yet well developed in the wider debate. Human behavior is a significant component of geomorphological systems worldwide but neither its influence nor the reflexive effects on human culture are well understood.

Aeolian sedimentary matrices

Archaeological sites are deposited on the surfaces of aeolian sediments, but they are often recovered within them. Sites may be incorporated within aeolian sediments during accumulation episodes, may be buried by them, and may be exposed, disturbed, sorted, and reburied by wind action. Beaches and dunefields occupy the high end of the energy continuum, with loess deposits at the other; archaeological sites in loess are less likely to be damaged than are those in beach or dune sands. As dunes often form by aeolian reworking of beach deposits (see below), determination of the environment of deposition can also help clarify the relative ages of sediment bodies.

At local scales, aeolian deposits can be subtly misleading. For example, in 1968 excavation began in Manchester, New Hampshire, USA on a river terrace high above a major waterfall. The soils survey had mapped fine-grained deposits there as “alluvium,” implying overbank flooding at a considerable elevation above the modern river. Archaeological testing revealed nearly 2 meters of black anthrosol, rich with artifacts. Excavation exposed a late Holocene sequence of artifacts, hearths, and pits, underlain by artifacts not recognized in the taxonomies of the time. Senior archaeologists dismissed the deposit as colluvium displaced from a famous site on the bluff above. Geologists recognized that hearths in place belied that interpretation, but thought the dark soil represented an old bog. Granulometry indicated that the sediments were sandy aeolian deposits on a truncated alluvial terrace, supporting interpretation that human use of the surfaces was synchronous with their accretion since the early Holocene (Dincauze 1976).

Dunes and sand sheets

Dunes, forming downwind of extensive sources of lightly vegetated sand, represent and perpetuate microclimates. Dunes do not themselves constitute full climate proxies since they can form in temperate climates where vegetation is stripped, as well as in regimes unsupportive of vegetation. Absence of vegetation reduces the surface friction that would interfere with wind velocity, and increases albedo-caused temperature differentials at the land surface. Dune sands are not homogeneous; they may display strong cross-bedding produced by variable winds during deposition. The challenge of dunefields to archaeologists is to determine whether archaeological sites found in and among dunes were deposited before, during, or after dune formation. Each situation has a different implication for environmental contexts of the sites.

Archaeological deposits overtaken by developing or moving dunefields will likely be disturbed. Unless protected by a vegetated soil and rapidly buried, archaeological deposits will be exposed by wind deflation of finer sediments enclosing them. Wind erosion leaves particles heavier than sand grains in more or less their original locale as lag deposits, perhaps with surfaces polished, pitted or striated, and effectively isolated from their original matrices. Such lag deposits can comprise artificial associations of materials of different ages sorted only by their relative weights, lying on surfaces created by processes younger than themselves. The underlying surfaces, whether sand, bedrock, or even paleosols, cannot be assumed to bear any original relationship with the artifacts resting on them.

Dune sands do not hold water; rapid percolation and oxidation of the sediments militates against the preservation of organic matter. Stabilized dunefields colonized by vegetation support slow pedogenesis and may be chemically altered by both pedogenesis and diagenesis. Soils typical of stabilized dunefields include Entisols and Aridisols. Stabilized dunefields may support inter-dunal pools if the concave surfaces intersect the local water table or if water is held on pedogenic hardpans. Such pools may become foci for human activity or residence, and ultimately the locations for archaeological sites younger than the period of dune stabilization. In such cases, only the environmental evidence associated with the ponds, and possibly the period of soil formation, is relevant to the human activities. Humans, fires, and climate change all destroy dune-stabilizing vegetation, initiating periods of remobilization that in turn can destroy human settlements and agricultural fields as well as any archaeological sites in the vicinity. The dynamism of active dunefields equals that of floodplains.

Notable paleoenvironmental work in desert sandsheets and dunefields has been accomplished in northwestern India and in northeastern Africa (Allchin et al. 1978; Misra and Rajaguru 1989; Wendorf et al. 1993). Research in arid zones elsewhere, particularly in the American Southwest and Australia, is elucidating the diversity and challenges presented by such archaeological locations.

Loess deposits

Loess covers thousands of square miles in the centers of continents. The largest deposits formed during periods of glacial retreat, when meltwater deposits of rock flour were exposed prior to colonization by vegetation and carried away by strong winds blowing from the glaciers. Silt-grade particles are transported to great distances because they are deposited only from winds slowed to near-exhaustion.

No matrix is more conducive to a pleasant excavational experience than loess, which is easily disaggregated and maintains vertical faces because of its homogeneity and compactness. Because of their high pH, loesses retain for millennia carbonaceous organic matter such as bones and mollusk shells. In addition, sites on loess were buried gently and therefore usually retain significant spatial integrity. Loess sheets typically accumulate episodically, with increments separated by paleosols. Mollusks in loesses add their oxygen isotopes and racemized amino-acids to the available climate proxies (McCoy 1987), offering alternatives to the missing, mixed, and damaged pollen. Because loess deposition is a Late Glacial process, cryoturbation of loesses can be extreme, roiling and even overturning original stratification.

Littoral sedimentary matrices

Studies of archaeological sites on beaches typically emphasize landforms and RSL. Archaeologists have only recently confronted the challenges of understanding processes of incorporation of archaeological sites into active beach sediments, and explicating methods that will advance such understanding (e.g., Bailey et al. 1994; Thorson 1990a; Johnson and Stright 1992). They have been rather complacent about these issues, overlooking what sedimentologists could tell them and being content with very little ("beach" does not even appear in the index to Schiffer's survey of formation processes [1987]). Research such as that of Kirch and Hunt on To'aga (1993), summarized in Chapter 10, shows the advantages of paying attention to beach formation processes. The deceptive dynamism of beach sands was dramatically revealed at the Lower Paleolithic site of Terra Amata in southern France, where artifact refitting exercises showed that what appeared to be a series of stratigraphically separated living surfaces was belied by extreme vertical distributions of matching fragments (Villa 1982). Such special archaeological analyses of sediment integrity are invaluable in deposits formed by waves or wind, subject to *in situ* reworking.

Beaches are narrow, linear, geographically constrained landforms marked by dynamic sedimentary histories and locational shifts over time. Only the backshore zone and inter-ridge swales (troughs between ridges), where deposition by storm waves or aeolian sand is episodic and erosion minimal, are amenable to the creation and preservation of archaeological sites (Chapter 10). Occupation on beaches, spits and barrier beaches, or islands is usually seasonal, represented by middens and lenses of anthropic soil within sorted aeolian or poorly sorted storm deposits. Careful discrimination during excavation may reveal microstratigraphic sequences, which, nevertheless, must be evaluated for their agents, periodicity, and integrity before being overinterpreted as "annual" or other regular deposits. Thorson offers a sobering discussion of formative and destructive processes on Arctic beaches, where tsunamis and ice dynamics further complicate matters (1990a: 413).

Beach sites are best preserved where they are raised above sea level; there, their special littoral characteristics may be overlooked. On sinking coasts, old beaches are subject to shore erosion, a typical mechanism exposing ancient shell middens. The erosion and reworking of coastal sites present important problems for regional-scale settlement patterns and seasonality studies, and for estimates of paleodemography. Improved understanding of coastal use throughout prehistory requires that more attention be paid to the dynamics of coastal site formation and deformation. Evaluation of beach environments should include consideration of offshore conditions in both water and foreshore, which may be represented in shore deposits of driftwood, molluskan and vertebrate remains, and seaweeds, occurring in cultural deposits and as wrack in storm beaches.

Shell middens, special kinds of beach deposits, deserve mention here; they are further discussed among anthrosols below. As with other shore landforms, middens may be destroyed and redeposited by storms and transgressing seas, appearing then as artifact-rich versions of cheniers (Fig. 12.3).

Volcanic sedimentary matrices

Explosive volcanic eruptions throw fine rock "ash" (tephra) into the wind and the stratosphere. The fine particles return at varying rates and distances according to their size, wind speed, and altitude. Many tephra deposits are identifiable chemically to their sources and may be dated by organic materials on the surfaces on which they fell. Tephra falls provide the best indicators of synchronic, long-distance land surfaces available to the geosciences. Tephra chronostratigraphic markers have been recovered from cave deposits, alluvium and lakes, buried land surfaces (soils and peats), and marine cores (e.g., Dugmore and Newton 1992; Sheets and McKee 1994).

Volcanic ash deposits are extraordinary matrices for paleoenvironmental study, as excavations at Pompeii demonstrated in the eighteenth century. Thick layers of volcanic ash seal living surfaces quickly, holding intact much of their microrelief and associated objects. Analysis of ancient gardens in Pompeii revealed patterns of planting and, in some cases, the species involved, by analysis of root casts (Jashemski 1979). Research at Ceren in Central America set new standards for archaeological data recovery and paleoenvironmental interpretation under tephra (Sheets 1992).

Rockshelters and caves

Rockshelters and caves, special sheltered environments, are only partially open to the weather. They have their own microclimates and localized sediment sources, and may offer unusually favorable preservation conditions. They concentrate and circumscribe activities within them, and the deposits that result. They are occupied,

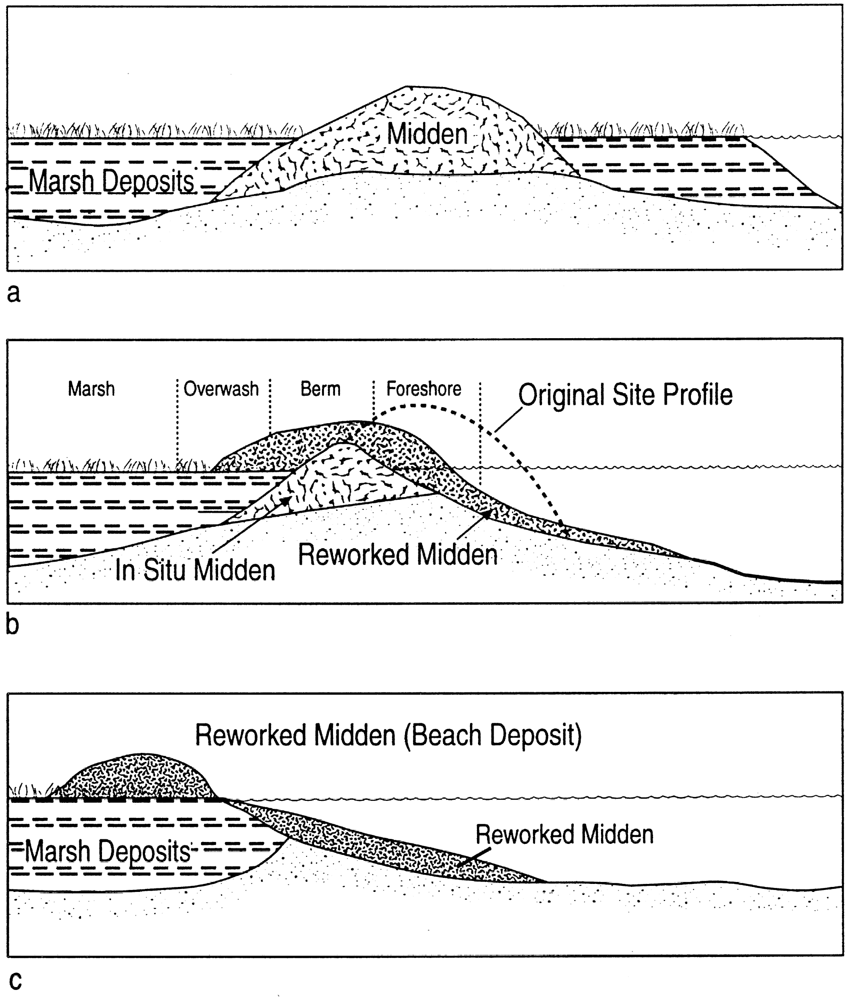


Figure 12.3 Redeposition of a coastal midden. Preserved by marsh overgrowth as water rose, the midden is destroyed and redeposited when attacked by waves. It is transformed by wave action from an archaeological to a geological deposit. (After Waters 1992: Fig. 6.15.)

synchronously or serially, by many different species including microorganisms. Cave and shelter deposits, once heavily relied upon for cultural sequence data, are now recognized as dense taphonomic challenges that yield data on sequence and environments only to critical, disciplined analysis.

Beyond their natural fascination and rich archaeological remains, caves and rockshelters with deep stratigraphic sequences have been especially attractive to archaeologists seeking definitive paleoclimatic data from sediments and inclusions (e.g., Laville 1976). However, tracing deposits from outside into shelters, to support chronological comparisons, is difficult. Many deposits are endogenous (originating inside), or polygenetic. Sediments in caves and rockshelters do not support soils development comparable to that on sediments outside. Organic remains and some sediments are brought into caves by creatures other than humans (e.g., Andrews 1990).

Whether synchronous with human occupations or not, environmental change within caves and shelters can rework deposits, erode sediments, and move, damage, or destroy artifacts (e.g., Bar-Yosef 1993: 19–22). The sequence of sediment types, as well as their integrity, must be evaluated as part of the environmental investigation: do stratigraphic interfaces represent old surfaces or erosional truncation? Spatially concentrated and repetitious occupation by humans and animals results in physical disturbance and chemical changes in the sediments. The environmental information potential varies with the area's access to external sediment carriers – wind, water, people, animals. Over time, the entrance may be blocked and reopened, according to very local circumstances.

Rockshelters

Shelters are typically formed by erosional undercuts in cliff faces, usually alluvial or marine erosion into the face of a sandstone or limestone cliff. They may also be formed by the collapse of lava tubes or karstic structures, and by rock layers of differing integrity and resistance to erosion. In glaciated regions, on the slopes of volcanoes, and at high-energy shores, sheltered spaces may be created by rough piles of boulders. The semi-enclosed environments of rockshelters form a continuum between open sites and cave mouths.

Rockshelter floors have a characteristic microrelief, dominated by a linear pile of sediments at the dripline directly below the overhanging cliff edge (Fig. 12.4). Water and sediments sliding down the cliff hit ground at the dripline, which forms a watershed directing falling sediments and water both outside and inside the sheltered space. Large rock fragments dropping from the cliff face usually form a talus pile outside the dripline. Rock fragments falling from the roof or back wall land inside the shelter, where they create topographic relief. Humans using rockshelters typically concentrate domestic activities in the sheltered space between the dripline and back wall, where they receive protection from precipitation and strong sunlight. If there is a ledge or level ground beyond the dripline, which is not always the case, people may situate hearths and activity areas there in fair weather.

Sediments within shelters include both exogenous (from outside) and endogenous elements. Typical of exogenous materials are fluvial, lacustrine, or marine

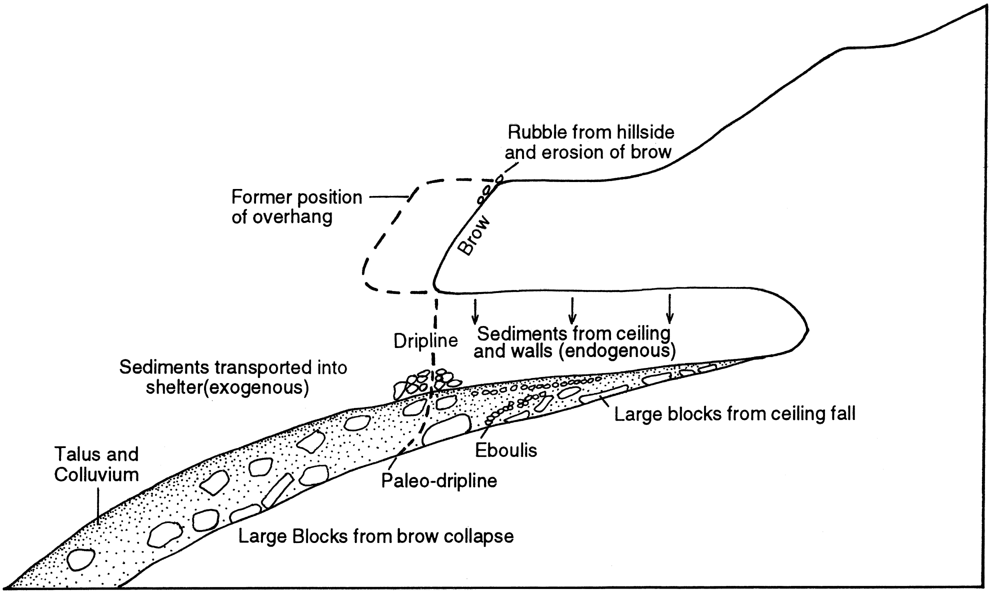


Figure 12.4 Cross-section of a rockshelter, showing retreating cliff face, overhang, dripline, roof-fall, talus. (After Waters 1992: Fig. 5.13.)

deposits related to the original formation of the undercut, colluvium washing or sliding in from the sides or over the top, materials falling from above the cliff onto the dripline, aeolian sediments deposited by winds slowed by the cliff, and materials of all kinds brought in by users of a shelter – humans, denning carnivores, small mammals, and birds. Endogenous materials include rock fragments and minerals dislodged from the shelter walls and roof by weathering, precipitates from springs and groundwater, and organic debris created by resident fauna and plants. The sedimentary sources and environments vary with distance from the entrance, and through time.

The dripline, therefore, separates contrastive sedimentary environments as well as microclimates; humans may enhance its effectiveness in the task with structural elements such as walls. The interior space may be damp or dry; the outside is usually more variable in temperature, precipitation, and windiness. Bedrock bases of shelters formed by erosion are often close to the local water table. Because of this, springs and seeps commonly form within shelters. Sheet wash on surfaces above the cliff may flow along the wall to drain into the back, marking its passage with flowing stains or deposits on the wall. On the other hand, deep shelters and those in arid environments with appropriate protection may be dry year-round; such places offer comfortable living space with good preservation for organic debris. As with all landforms based on relief, the orientation of a shelter is crucial to its efficacy as living space.

Shelters opening to the Sun were utilized much more often than those that face away. Shelters open to seasonal storms have seasonally constrained habitation potential.

Caves

Caves proper differ from shelters in being enclosed at the sides as well as back and ceiling. They are most typically formed in karstified limestone or dolomitic bedrock, where they are the surficial openings of large solution cavities. Special cases of cave-like enclosures include large lava tubes and bubbles in extensive basaltic flows (e.g., Barba et al. 1990). Because caves are more fully shielded from exterior climatic conditions than are rockshelters, their sediments contain larger proportions of endogenous materials. Sediment sources such as roof-fall (rock fragments and fine particles), reprecipitated limestones (speleothems: crusts, flowstone, stalagmites, stalactites, tufas/travertine), residual minerals left after solution of carbonates (silicas, metal compounds, and clays), and fluvial deposits originating within the karst drainage system reflect in their content and structure the disparate environmental conditions within the cave space. Behind the entrance, aeolian and cultural sources of sediment diminish rapidly, temperature stabilizes, and light dims. Beyond the reach of daylight interior cave environments are used by humans only for special ritual purposes or exploratory visits. Preservation of organic materials such as bone, coprolites, and plant materials is enhanced in stable interior environments, both damp and dry. Dryness and high pH are especially favorable factors in preservation, as are layers sealed under flowstone (Straus 1990; Waters 1992: 240–247).

External environments are represented in caves by aeolian materials, drainage into fissures, slope wash and colluvium, and particles deposited by high water reaching the cave mouth. In brighter and drier parts of a cave are concentrated hearths and other cultural deposits created by human activities. Along with those, and sometimes extending deeper into the enclosed space, are materials brought in and deposited or tossed by humans and animals, including bones and other food wastes, dung, seed caches, and bedding. Packrat middens are paleoenvironmental features of caves and fissures in the American West.

Postdepositional changes in cave deposits were ignored for a long time while cave strata were confidently interpreted as climate proxies. *In situ* weathering and carbonate impregnation of sediments transform chemical and physical attributes of the deposit, reduce mass, and translocate elements after deposition. Deposits may be eroded away at the surface and undermined at depth (e.g., Tankard and Schweitzer 1976). Trampling by occupants in the confined spaces of caves, and subfloor digging, add cultural disturbances to the list. When these possibilities are considered with awareness that microclimates inside caves are distinct from those of the exterior, it is clear that the paleoenvironmental interpretation of cave sediments is no straightforward task. Careful discrimination of sediment sources and depositional agents is essential to identify the several source environments. Even so, the potential for equifinality is high; many variables at play in complicated interactions make the interpretation of cave deposits tenuous at best (Barton and Clark 1993; Butzer 1982: 77–87; Straus 1990).

With regional-scale climates so imperfectly represented in caves, the potential of cave sediments to serve as proxies at that scale is drastically reduced. Any respectable effort will require minute attention to the structure and stratification of deposits in place, followed by laboratory analyses selected to reveal attributes relevant to the interpretation. Increasingly, analysts are devoting attention to the smaller classes of sediments – moving down the phi scale (Table 11.3). Sediment components attributable to human activity must be removed prior to textural analyses intended to illuminate climatic factors (Butzer 1981b). Micromorphological study of thin-sections has proven its worth in cave studies, where events as intimate as a gust of wind stirring the ashes of a Paleolithic hearth buried thousands of years ago may be interpreted (Courty et al. 1989: 214).

There is now less reason to stretch the study of cave deposits so far beyond their best use in search of paleoclimatic data. Regional-scale climatic proxies are multiplying all the time, as Chapters 7 and 8 show. Cave sediments are most appropriate for interpreting cave environments, which can change without external influences as floors build up, roofs fall, and entrances are sealed or opened. Caves inhabited by people qualify as domestic space, subject to modification by humans, whose hearth fires, large bodies, and interior constructions change interior climates. The study of cave microclimates can reveal much about living conditions at intimate scales if we relax our grander expectations of them.

Soils as matrices

Soils form from the stable surfaces of sediments, where human lives are lived and biological wastes are deposited. Soils are the typical matrices into which archaeological data are incorporated, and in which such data are preserved or transformed. The archaeological literature shows very plainly how variable is the preservation potential of diverse soils, subject to both pedogenesis and diagenesis. The soils Orders (Table 12.1) provide a basis for discussing some regularities. For further specifics on soil Orders see Brady (1990) and Fanning and Fanning (1989).

The Spodosols, formerly “podzols” as their name suggests, develop on sandy parent material in cool, moist climates in the middle and high latitudes. They are

Table 12.2 Abbreviations for some common chemical elements

Al	aluminum	N	nitrogen
C	carbon	Na	sodium
Ca	calcium	O	oxygen
Cu	copper	P	phosphorus
Fe	iron	Pb	lead
H	hydrogen	S	sulphur
I	iodine	Si	silicon
K	potassium	Sr	strontium
Mg	magnesium	U	uranium
Mn	manganese	Zn	zinc

characterized by low pH, concentrations of Al and Fe oxides in the B horizon (Table 12.2), and bio- and cryoturbation. They cycle organic matter rapidly into humus and leachates, and yield to archaeologists only calcined bone, stone, ceramics, and some oxidized metals. The acidity may be tempered locally by concentrated organic wastes or wood ash, permitting preservation of some bone or shell materials. Ants, rather than earthworms, are the typical bioturbators, cycling particles from within the B and C horizons and redepositing them on the surface. The churning moves artifact-sized particles ever deeper into the soil profiles. Cryoturbation may bring them up again. Very small-scale, intensely colored Spodosols can form under anthropogenically or biologically enriched deposits, such as pits and ditches, in areas of abundant rainfall. Hardpan development in Spodosols can raise water tables and support bog formation.

Alfisols form mainly under woodlands on base-rich parent materials, and develop a B horizon rich in clay. Their carbonate content makes them less corrosive for organic matter than the Spodosols, so that they are likely to yield more paleoenvironmental proxy data. Mid-continental grasslands are the typical location of Mollisols – base-rich, organic, dark-colored fertile soils. They form in a wide range of climates, and in drier areas can preserve bone and shell constituents very well. Aridisols, the soils of dry climates, typically have calcic cemented B horizons because rainfall is inadequate to leach away soluble salts and oxides. They are not strongly weathered, but loose in texture and light in color. Their preservative qualities vary widely with rainfall and seasonality.

Ultisols are typically ancient forest soils of humid warm climates. They also form a clay-rich B horizon, but are low in bases and therefore more deeply weathered than Alfisols. Tropical rainforests are the usual vegetation on Oxisols – deeply leached, typically red clayey lateritic soils. Oxisols preserve for archaeologists neither organic materials nor spatial relationships.

Azonal soils (Table 12.1) include the Histosols, composed of plant remains with minimal mineral matter (e.g., peats). Whether acidic, as is typical, or carbonaceous, their preservative properties are legendary (see below). Vertisols are clay-rich soils that shrink and swell, crack and churn, preserving no spatial relationships for any inclusions. They form in seasonally dry climates on flat terrain in clay-rich parent materials. The microtopography characteristic of Vertisols complicated the interpretation of Maya agricultural techniques in wet lowlands of Central America, creating decades of controversy that has been productive of alternative hypotheses and ultimately enhanced understanding (Jacobs 1995).

Entisols are too immature to be classified as any of the zonal soils. They are found on recently deposited materials, excessively drained substrates such as dune sands, and newly exposed eroded surfaces. The parent material is modified very little. Archaeological sites are rarely found in Entisols because of their recency and instability. Inceptisols show the beginnings of soil horizonation on parent material of no great age, typically alluvium or colluvium. They form under almost any climatic regime on parent materials of diverse sorts.

A model of conditions conducive to the preservation of various kinds of organic matter was developed for paleosols on the basis of "Eh," a measure of "the extent of oxidation and reduction reactions in . . . soils" that varies with water content and the soil pH (Retallack 1990: 218). The model elegantly predicts the kinds of organic remains that will be preserved under diverse conditions of pH and drainage (Figure 12.5), and appears to be appropriate for archaeological matrices.

Soils on alluvium

Alluvial contexts and matrices for archaeological sites, especially where sequences of buried soils are involved, offer opportunities for the study of environments at microscales. Changes in microclimates at the immediate site location may be directly recorded in the relative intensity of processes such as leaching, gleying, illuviation, or subsurface deposition of carbonates – all of which reflect temperature and precipitation regimes that may differ from those at the time of observation. When additional climate proxy evidence is preserved in alluvial sediments, especially near or below water tables, the circumstances may be excellent for paleoenvironmental work. Of course, recognition of changed water tables offers no direct evidence for the causes or times of such change. Refined stratigraphical and chronometric data are required to show (1) whether materials associated in alluvial deposits are truly synchronous and not simply mixed together during fluvial deposition or pedogenesis, and (2) how

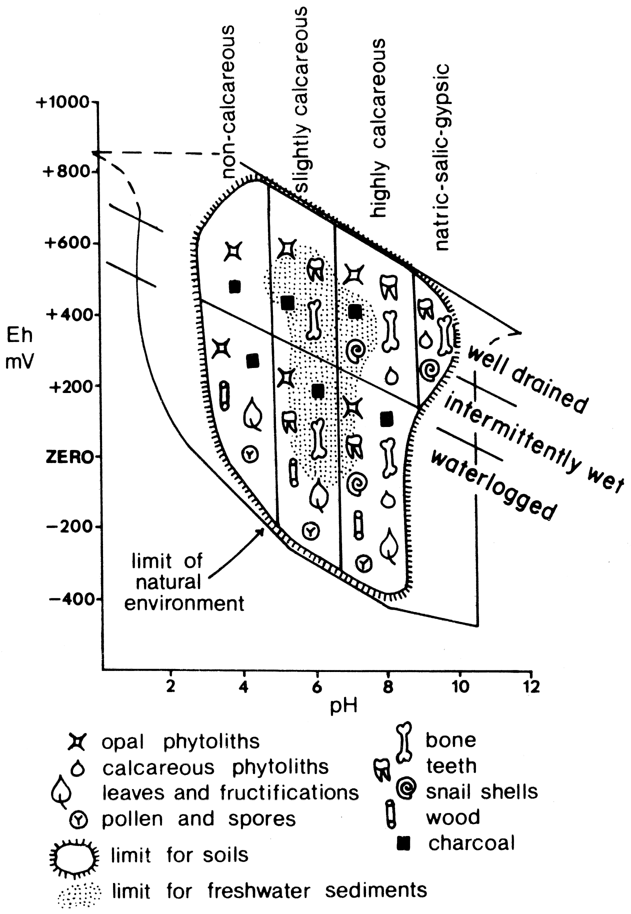


Figure 12.5 Theoretical Eh–pH stability fields of common kinds of terrestrial fossils preserved in paleosols. (Reproduced from Retallack 1984: Fig. 11, with permission of *Paleobiology*; original caption.)

changes in one place relate to similar or even opposite changes spatially distant. Successful paleoenvironmental reconstruction at scales above that of the site depends entirely upon control of time at fine resolution, coupled with evidence permitting extension of the local paleotopography to the regional scale.

The surfaces of active floodplains are neither very stable nor likely to be old. Consequently, they are characterized by Inceptisols and Entisols that carry little paleoenvironmental information, and may be kept immature by periodic additions of sediment (Table 12.1). Terrace soils are likely to be more useful, supporting Zonal soils which, although offering no more than confirmation of regional soils types, may include proxy data of some antiquity. Sealed paleosols in floodplain deposits are likely to be below the modern water table; on terraces, they may be important clues to old landforms now fragmented (Ferring 1992).

The notorious dynamism of fluvial environments alerts us to the fact that soils may vary locally in mineralogy, biogenic climate proxies, and archaeological deposits. Long-distance correlations are hindered by such variability. Paleosols on alluvium also share many textural and structural characteristics with organically rich flood deposits, which require very different interpretations (Brown 1997). Micromorphology will be helpful in making such difficult distinctions.

Anthrosols

Anthrosols (anthropogenic soils; anthropic epipedons) are as much cultural features as are pits, hearths, and house foundations. They are sedimentary bodies enriched by chemicals, mineral particles including artifacts, and biogenic materials collected and discarded by human beings. Anthrosols deserve analysis at least as thorough as that devoted to other complex culturogenic compounds such as ceramics, mortar, and metal slags. The disdain for sedimentary analysis of site matrices expressed by some archaeologists even today is comparable to the dismissal of bones, seeds, and charcoal earlier in this century – deplorable destruction of evidence. Techniques of micromorphology promise to revolutionize the study of anthropic soils, for which they seem unusually appropriate (e.g., Macphail 1994).

The dark color and greasy feel of developed anthrosols derive from concentrated organic wastes, especially carbon, nitrogen compounds, calcium, phosphates, amino acids, lipids, and diverse colloids. Spatial variation within anthrosols follows from different activity distributions, which can be plotted from on-site mapping of phosphates (Eidt 1985) and other constituents (Dillehay 1997: Ch. 10). Research refining the extraction and interpretation of organic compounds from soils, wet sites, and cave earth is regularly reported in such journals as *Journal of Archaeological Science*, *Archaeometry*, *Nature*, and *Science*. Organic particles typical of anthrosols and middens, such as pollen, insect parts, mollusk shells, and animal bones contribute to the definition of micro-scale site environments.

Most middens qualify as landforms composed of anthrosols and rubbish (Kolb et al. 1990; Lawrence 1988). Urban landforms, including tells, enclose anthrosol lenses. Ancient and modern urban backyards and privies share many characteristics and information content with more conventional anthrosols (Macphail 1981; Reinhard et al. 1986). The living surfaces of non-urban sites are likely to display some chemical traits typical of anthrosols. Another signal of concentrated human activity in such soils is enhanced magnetic susceptibility induced by iron oxides modified by burning (Catt 1986: 177; see also Bellomo 1993).

Anthropogenic deposits in dry climates, such as the Near East and southwestern North America, may be treasure troves of paleoenvironmental data, with no pedogenic modifications at all. Such sediments have traditionally been separated by sieving or flotation, and the constituent elements inventoried in sets by size or kind. Micromorphological techniques permit analyses of actual associational sets, with behavioral interpretations supported as well as paleoenvironmental ones. Box samples from a Mesopotamian city, transformed into large thin-sections and examined microscopically, revealed details about living conditions within dwellings, brick-making, and insect infestations in food and other trash (Matthews and Postgate 1994).

Shell middens may be distinguished from natural shore features such as shelly cheniers, all else failing, on the basis of their anthrosol component, which may be discontinuously distributed. Normally, the artifact and feature contents of shell middens are adequate to distinguish them from other kinds of shell concentrations (e.g., Bailey et al. 1994; Sullivan and O'Connor 1993), but in cases where structure has been damaged, soils chemistry may be definitive. Molluskan shells, large calcareous particles, give midden soils typically high pH which preserves bones and bone artifacts. Shell middens are not choice sediments for pollen preservation; pollen analyses should be undertaken only when local questions can be answered by special situations.

Culturally modified soils include "agric" and **plaggen soils**; the first are plowzones, the second heavily manured soils such as those of gardens, fields, and pastures (Behre and Jacomet 1991; Simpson 1997). These are best developed in areas of population concentration and intensive cultivation, such as urban gardens and terraced fields. Both are identifiable by their chemistry and structure (e.g., Groenman-van Waateringe and Robinson 1988; Macphail 1994; Macphail et al. 1990; Sandor 1992). Sandor reports that the obvious changes in Peruvian terraced soils are thickened A horizons enriched with organic matter, nitrogen, and phosphorus; they develop distinctive soil structures and pores and become more acidic (Sandor 1992: 237). Limbrey (1992) reports successful discrimination in thin-sections between tilled and untilled sediment bodies. Agricultural soils lose humus and carbon to a degree detectable by radiocarbon analysis (Harrison et al. 1993; Trumbore et al. 1996). Enhanced erosion of tilled fields, now an international concern, began in antiquity (Bell and Boardman 1992).

Soils culturally modified for other purposes are receiving attention. Depletion of soil carbon, nitrogen, and fungal biomass following clear-cutting of timber retards growth of new seedlings (Moffat 1993). Such depletion reverses pedogenesis in formerly forested tracts, and might be detectable in isotopic ratios. Pastured tracts can be identified by suites of phytoliths characteristic of animal foods, by changes in soil mollusks, and by symbionts such as dung beetles (e.g., Powers-Jones 1994; Robinson 1988).

WET AND FROZEN SITE ENVIRONMENTS

> Excavating frozen ground is a tediously slow process that can make young knees into old knees in a single field season.
>
> SCHWEGER 1985: 128

The special preservation conditions of archaeological sites at the interface of land and water have long been appreciated. **Anaerobic** burial conditions impede bacterial decomposition and preserve archaeological deposits rich in organic materials. Anaerobic sediments, saturated or frozen as well as special cases such as tar pits, are important sources of paleoenvironmental data well represented in the archaeological literature (Coles 1984; Purdy 1988). Exploitation of such situations is increasing as constraints on excavation are overcome by technology (e.g., Dillehay 1997).

Sites and objects preserved in wet or frozen contexts were either originally subaerial or were deposited intentionally into water or wet media. For a proper understanding of environments of incorporation, it is obviously essential to determine which situation obtained in each case. The long, contentious history of reinterpretation of the Swiss Neolithic “Lake Villages” and their associated middens indicates that the issue is not always as unambiguous as a shipwreck. For excellent preservation, either rapid inundation or original submergence are the best circumstances. Since organic materials are subject to rapid decomposition by bacteria and fungi in air, their condition on recovery can indicate the relative duration of exposure prior to submergence, and thus the rapidity of incorporation and the integrity of associations in the matrix.

Histosols as matrices

The natural continuum of wet sites begins on land, with waterlogged depositional environments such as bogs, fens and **carrs,** swamps and marshes. Definitions are helpful, because the environments differ, as does the likelihood of archaeological deposits in each. **Ombrotrophic** (rainwater-fed) bogs covering uplands are the classic archaeological “blanket bog” sites, best developed in northwest Europe and the British Isles, and less extensive elsewhere. Other terrestrial wetlands form where the water table intersects the surface of the ground. Sphagnum peats growing in acidic groundwater pools or shallow ponds create the classic bogs. Quaking bogs are peat mats floating on water but attached to shore; cherished by frogs, they are not suitable archaeological environments. Fens are boggy landscapes formed in alkaline or neutral groundwater; carrs are variants supporting woody swamp vegetation in addition to peat. Swamps are terrestrial habitats formed where woody vegetation alternates with stretches of open water. Marshes typically have grassy herbaceous vegetation that may be partly submerged in water. Coastal marshes in brackish water are further characterized as low-energy depositional environments; with rising sea levels archaeological sites may be protected under the vegetation mats of coastal marshes transgressing inland. Fens, ombrotrophic bogs, and coastal marshes are of greatest interest here, as most likely to cover terrestrial archaeological sites. In each case, the matrix will be a Histosol, with the characteristic good preservation of organic remains.

Acid Histosols preserve plant materials including pollen, and animal soft tissues, very well; they are less kind to bone (Clymo 1984). In contrast, alkaline groundwater is credited with remarkable soft-tissue preservation at the Windover site in Florida, where early Holocene burials recovered from peat retained brain tissue in a shallow pond with neutral pH (Doran and Dickel 1988: 285). The preservation of bog bodies in sphagnum bogs depends on more than acidic suppression of bacterial action; complex chemical reactions in humates derived from sphagnum explain the loss of bone mineral components and the preservation of soft tissues (Painter 1994). Acidic Histosols of fens, carrs, and ombrotrophic bogs have preserved the renowned peat sites of Great Britain and northwestern Europe, where environmental archaeology began. The preservation of plant macrofossils and pollen along with insect remains, occasionally bones, and abundant wooden structures and artifacts supported multidisciplinary studies beginning in the 1930s (Coles 1992). Excavation in wet peats is an exacting and expensive undertaking, which yields its extraordinary surprises slowly (e.g., Pryor 1991, 1995).

Nearshore matrices

Sites below the strandline in lacustrine and marine sediments, exclusive of shipwrecks and such material deposited from the surface of the water, represent times of lower water levels. By their very existence, they implicate different environmental conditions. Typically they exist in a matrix of peat or silt, permanently or mainly waterlogged. Saturation and the lack of pedogenesis usually permit better preservation of organic materials than in terrestrial sites. Nearshore sites occur at the edges of lakes with climatically fluctuating water levels, in estuaries and shallow sea coasts with barrier beaches, and wherever rising sea levels lift water tables on land. The postglacial rise in oceans inundated numberless sites, a few of which have been investigated, most of them only a few thousand years old, visible under water or exposed by tides or during periodic low-water stages in lakes. Tectonic elevational changes, particularly in the eastern Mediterranean and western North America, have added others. Areas inundated early in the Holocene are, for the most part, lost at depth and probably essentially destroyed.

The shallow foreshore or intertidal sites that are available to archaeologists lie on terrestrial sediments but are enclosed in subaqueous sediments. Each of these deposit classes offers a different suite of potential paleoenvironmental data. Investigation of such contrastive and complementary deposits demands special expertise for interpretation of the transport and deposition agents, the range of biological inclusions, the chronology of inundation, and the best techniques for disaggregation. Excavating by pumped air or water, common enough in underwater archaeology, puts special responsibility on managers for the collection of crucial environmental data, which may take second place to the spectacular and valuable artifacts typical of such sites.

Sites under fresh water

Freshwater deposits enclosing archaeological sites and data are very diverse. The simple cases are those (e.g., fishweirs, quays) under rivers, which are covered by sands and silts if protected at all, and those under lake waters in sand, silt, clay or marls, or peats. Sites of the early Holocene or older also occur in deposits of various salt playas, and in tufas and travertines near springs.

These several matrices preserve organic materials suitable for paleoenvironmental studies; some offer unique opportunities. For example, investigation of travertine deposits in western North America yielded diatom evidence, for the period of human use (1500–1000 B.P.), of organic enrichment of water sources that disappeared when the area was subsequently depopulated (Blinn et al. 1994). Excavation at a wet site in Japan revealed “a series of rice fields of the Yayoi, sealed and preserved by flooding silts, with, in places, the individual rice plants still visible in the fields, planted 2000 years ago” (Coles and Coles 1994: 5). Large lakes with fluctuating water levels have revealed sites under water throughout Europe, but few have been investigated elsewhere. The Swiss lakes are among the most famous, with houses, furnishings, and food remains receiving most of the publicity. An Upper Paleolithic campsite, left on a Late Glacial shore of Lake Neuchatel, was inundated gently enough to preserve hearths and activity areas. A rescue dig, mounted in 1985, explored the site after pumping out the shallow water within a cofferdam (Stickel and Garrison 1988). The paleoenvironmental potential of lake sites remains to be exploited appropriately.

Shallow marine sites

Seacoasts worldwide offer opportunities for wet-site archaeology, especially within the tidal zone. Beyond that, maritime archaeological methods and problem foci dominate. Where shores offer a wide expanse between high and low tide there are opportunities to excavate or at least sample archaeological sites. Estuaries and lagoons, semi-enclosed near-shore fringes of the sea, are among the most attractive locales for human exploitation and habitation (Chapter 10); their brackish mixture of fresh and salt waters has a high biotic potential, with characteristic fauna including shellfish. Archaeological sites are often found near or under intertidal swamps on such shores.

The potential for site integrity in such locations can be estimated in preparation for searches (Belknap and Kraft 1981). Site preservation potential will vary with the slope of the inundated surface, the depth of water and tidal range, wave energy, abandonment time in relation to rising sea level, and other local factors. Preservation of paleoenvironmental data will, of course, depend on the integrity of the site itself. Environments during active use of sites may have differed significantly from those at the time of abandonment, as well as from the environment at inundation. Reasonably intact sites near shore should include preserved wood and other organics. Protection under mucks in zones of low wave energy, especially lagoons, may be optimal. In silts, environments of deposition, incorporation, and inundation may be studied in sedimentary textures and structures, and in included fauna, microfauna, and pollen. Sites are destroyed if the surf zone reaches them. Sites in deeper water offshore will have been subjected to compaction of the enclosing sediments, with resulting distortion of organic materials and disruption of associations.

Permafrost and periglacial matrices

No soils form on frozen ground, but ice is a more powerful disrupter of sediments than is pedogenesis. Periglacial and Arctic cycles of freezing and thawing may preserve incorporated organic materials, but disrupt associations and structures (Washburn 1980). Permafrost holds water tables near the surface during periods of thaw. Archaeological materials deposited on surfaces may sink into muck during thaws, to be mixed with whatever other discrete particles they join there. Excavation in permafrost requires removal of small thawed increments, to facilitate further thawing. Thawed ground is little different from waterlogged muck.

The superb preservation of organic materials and artifacts made from them recompenses excavators for the discomforts and challenges of digging in permanently frozen ground, but in sediments with histories of alternate freezing and thawing, observation and interpretation of archaeological associations and structures is at or beyond the limits of feasibility. Thus, the recovery of a broad suite of paleoenvironmental data from Arctic sites is dependable, but the interpretation of associations and synchronies is not (Waters 1992: 292–299, 302–304). Cryoturbation in sediments subject to deep freezing results in churning of sediments and sorting and displacing of particles that are comparable to those of Vertisols, or worse (Schweger 1985; Thorson 1990a). Included objects are displaced both upward and downward; sediment layers slide and fold. Prior to widespread awareness of these processes, false "stratigraphic" relationships were reported from Arctic sites on the basis of superposed folds of geliflucted sediments. The chronological problems created by churning cannot be solved, but they are now widely recognized.

Paleolithic sites in areas not now subject to frost disruption may be misunderstood if cryoturbation is not recognized. The dramatic disruptions at the High Lodge site in England were not disentangled for years (Ashton et al. 1992). Figure 11.3 indicates some of the difficulties ultimately overcome through a harrowing investigative process.

Seasonally frozen ground in middle and high latitudes and high altitudes suffers a related set of disruptions and displacements, varying with the depth and duration of freezing and the frequency of thaw cycles. Archaeological sites in ground subject to freeze–thaw cycles will have heavy particles shoved toward the surface, flat objects rotated vertically, and everything subjected to lifting by ice expansion from beneath and to dropping into frost cracks. In northern North America, sites of Paleoindian age (twelfth to eleventh millennium B.P.) rarely offer any evidence of intact fireplaces or postmolds, and rarely preserve even charcoal for radiocarbon dating. The implication, which must be tested, is that the sites were subject to severe freeze–thaw cycles in the distant past, if not currently.

Permanently frozen archaeological sites, on the other hand, are among the most complete, as is evident in the case of the Eneolithic fugitive in the Tyrol (Spindler 1993), the Pazyryk graves from Central Asia, and Arctic sites with intact house interiors. Environmental archaeology is never easier than with such complete site preservation.

CODA

The matrix of an archaeological site is the environment of burial, the source of diverse information, and a labile, corrosive storage medium. Archaeological matrices vary with climate, sediment class, history, disturbance processes, and original contents. Retrieval of paleoenvironmental information requires that the excavator be informed, alert, and innovative in observing and recovering data classes. Sampling schemes should be defined to catch the entire range of variation within a site and to permit comparisons to off-site contexts that are similar except for the cultural materials.

DID THE CLASSICAL CIVILIZATIONS DESTROY THEIR OWN AGRICULTURAL LANDS?

The Classical lands of the Mediterranean present the thoughtful observer with the paradox of the homelands of great early civilizations in landscapes now characterized by limited and discontinuous arable soils, bare rocky hillsides, and silted harbors. Already in late Classical times writers speculated about the destruction of formerly richer landscapes by abusive land-use practices. Early environmentalists used the Mediterranean case as a moral lesson, threatening similar impoverishment to heedless peoples elsewhere (e.g., Marsh 1965). This view of things is necessarily based on the assumption that the damage had been done during classical times and that later populations simply endured the burden of their poor inheritance, which doomed them to economic marginality in the modern world.

By the decade of the 1960s, informed observers had noticed that the massive alluvial deposits in circum-Mediterranean valleys contained Roman and younger sherds, and that in some instances they buried Classical and Byzantine sites (e.g., Judson 1963). These observations particularly impressed Claudio Vita-Finzi, who inspected valley fills around the Mediterranean and published in 1969 a monograph on his investigations.

Vita-Finzi observed two major episodes of Mediterranean valley fills, which he called the "Older" and "Younger" Fills. The older and more massive was very rocky in places, was typically a deep red color, and had been deeply incised by stream-cutting before the deposition of the Younger Fill that was "nested" within it. Along the coasts, the Older Fill extended seaward, as if emplaced during times of lower sea level, and was overlain by more recent beach deposits of the postglacial transgression. Interpreting the stony component in the fill as evidence for frost weathering, Vita-Finzi concluded that the Older Fill was a product of the last glacial age. He saw that, in the twentieth century, streams in the valleys in which the Younger Fill lay were not depositing sediments but were instead incising and carrying seaward some of the Younger Fill. The Younger Fill itself, gray or brown in color, usually finer in texture and far less massive than the Older Fill, and apparently graded to modern beaches, he interpreted as a Holocene deposit created under environmental conditions different from those of the present.

His archaeological observations led him to conclude that the younger fills were of medieval age everywhere he looked. At the same time, climatologists were demonstrating that a Little Ice Age had begun in northern Europe and North America in late medieval times, and Vita-Finzi tentatively suggested that such a climatic change might have brought heavier or less episodic rainfall to the Mediterranean lands, thus reversing regional stream regimes and reinstating the erosion-and-deposition cycle that he claimed had emplaced the Older Fill during the last glaciation. He was convinced that only climatic change could have caused such a reversal over the entire region within the limited time span of the last two millennia. Moreover, he assumed that the fill sequences were synchronous over the entire Mediterranean basin and therefore represented only two discrete cycles of events. He rejected agriculture and husbandry as explanations of the change because they had characterized the region for a far longer time span.

Vita-Finzi's claim for large-scale Late Holocene landscape remodeling, developed and expounded strictly at a geological rather than archaeological scale of resolution, influenced thinking about classical and modern economies alike. Vita-Finzi concluded that Classical land-use practices had not been ruinous, that climate change was responsible for the redistribution of unconsolidated sediments from hill slopes to valley bottoms, and that this redistribution, painful in its immediate consequences, ultimately provided modern farmers with excellent and accessible arable soils.

The hypothesis was embraced and expanded by John Bintliff (1977) who claimed that field situations in many areas of Greece supported Vita-Finzi's conclusions, although he did change the dating of the Older Fill, assigning it to the early part of the last glaciation rather than the height of the ice age.

Attractive as the Vita-Finzi hypothesis was in some quarters, it was immediately challenged. Karl Butzer (1969), reviewing Vita-Finzi's book, noted that the mechanisms by which climatic change could trigger reversals in stream regimes were not demonstrated and that the data presented were inadequate to eliminate the alternative hypothesis of abusive human land-use practices as local causes. Moreover, he stated that the argument failed to account for complexities recognized even in the Older Fill components. Butzer later demonstrated that several different modes of land use influence stream regimes and hill-slope erosion, and that interactions of climate change and human land use may be so complex that interpreters cannot emphasize either as the critical factor (Butzer 1969, 1981a, 1982).

The assumptions on which the climatic conclusions were based came under closer scrutiny, as Davidson (1980) challenged both the generality and the synchroneity of the Younger Fill. He thereby weakened the necessity for invoking a universal large-scale mechanism such as climate change. Wagstaff (1981), citing geomorphological and sedimentological research on the behavior of intermittent streams in dry climates, argued that the Mediterranean field data of Vita-Finzi and Bintliff were not only inadequate to refute the hypothesis of human agency, but also were themselves not strongly supportive of the climate-change conclusions (see also Bell 1982). Basing his argument partly on complex-response theory in hydrology (see Chapter 9 and Figure 9.3), Wagstaff claimed that the alternation of cut and fill regimes, as well as local differences in mode of sediment transport and deposition, implied far more complex processes operating at local, not regional, scales. "Variation [in stream activity] could result from differences in the location of the site studied, local topography, soil quality, vegetation and climate, as well as from diachronic alternations in human activity" (Wagstaff 1981: 253). The modern range of variation in rainfall, in both time and space and on seasonal, annual, and larger scales, could easily account for the variability cited in the stream regimes. Furthermore, there was then little independent evidence supporting the inference of climatic change of the scale and age necessary for the hypothesis. Wagstaff's forceful argument for the testing of alternative hypotheses closed with a call for a technical interdisciplinary study involving geomorphologists, climatologists, and archaeologists, with detailed attention to variation and correlation through a range of temporal and spatial scales.

An interdisciplinary study meeting some of the stated requirements began about the same time in the Southern Argolid area of Greece, a small, seagirt peninsula which incorporates within its limited area much of the native landform diversity of modern Greece. Geological field work was carried out in conjunction with an intensive archaeological survey, and data on regional population densities, land-use practices, and economic conditions were gathered from administrative and literary sources (Pope and van Andel 1984; van Andel et al. 1986). The results gave strong new credibility to the anthropogenic erosional hypothesis, but still fell short of settling the matter, even locally. A closer look at the methods and conclusions will help explain the complexities involved.

Examination of multiple exposures in most of the valleys of the Southern Argolid revealed that both the Older and Younger Fill bodies were the products of multiple episodes of valley deposition, and thus of hill-slope erosion. The major episodes of valley fill were themselves internally diverse, some beginning with debris flows followed by stream channel deposits and ending with overbank deposits of fine silty alluvium, others lacking the channel deposits, still others involving only the channel and overbank deposits. For all but the most recent sequences of deposition, active sedimentation was followed by episodes of local stability within which soils formed on the exposed silts. The soils, being progressively younger through the sequence, provided a reasonably reliable means of correlating distinct episodes of deposition between adjacent valleys, a correlation that could not be made on the basis of the deposits alone except when appropriate archaeological inclusions helped to refine the chronology.

As many as four distinct episodes of deposition followed by soil development were discerned within some of the Holocene fills in the Argolid, each dated by archaeological criteria or radiometric dating. The oldest Holocene unit recognized was assignable to the Early Bronze Age (Early Helladic), about 4000 years ago. It characteristically began with debris flows indicative of catastrophic hill-slope erosion. Debris flows occurred again in deposits assignable to medieval times (Middle Byzantine/Frankish period). Both of these events occurred during periods of major population expansion, and in each case are thought to have been triggered by extensive land clearance lacking effective conservation measures. The earlier Neolithic land clearance is interpreted, on the basis of the small numbers of sites involved and the absence of evidence of erosion, as localized and probably based on long fallow rotation, so that no great amounts of sediment were at risk at any one time.

Following the Bronze Age soil losses, a long period of stability ensued in which streams neither aggraded nor gullied and little soil was lost from the hills. The stability lasted through the time of high population associated with the Mycenaean civilization; the investigators propose that the landforms and agricultural surfaces were stabilized by effective conservation tactics that, given the high population, are more likely to have been terraces and check dams than long fallow intervals. The depopulation that followed the collapse of Mycenaean civilization was not accompanied by massive soil mobilization. The investigators, noting that today the natural scrub vegetation quickly stabilizes the soil of abandoned agricultural terraces, posit a similar natural healing to explain the continued stability of the valley floors. The population expansion that preceded the prosperity of Classical Greece did not trigger soil losses either, presumably because terracing technology was understood and properly employed.

The political and economic upheavals following the Alexandrian empire and Roman conquests led again to reduced population in the Argolid, and for a second time slopewash from abandoned fields carried sediments into the valleys, where they were deposited as channel gravels. In this case, it is thought that the hillside terraced fields were used as pasture for sheep and goats, preventing the natural vegetation from asserting itself, while the terrace walls were allowed to disintegrate. While the same effects could have been achieved by increased rainfall, there is no other evidence for such a change, and the brief duration of the event, dated by archaeological means to the Hellenistic period, makes the case for neglect much the stronger. With economic recovery the soils were stabilized again; the stability continued through another depopulation, when evidence from pollen analysis indicates the regeneration of the native scrub vegetation. The last sporadic episode of valley deposition, dating to the most recent couple of centuries, correlates with observable neglect of terrace walls and upland fields as economic attention shifted from agriculture to herding or tourism.

The Southern Argolid investigators concluded, therefore, that cycles of economic and population expansion and contraction, in conjunction with changing land-use practices, account effectively for most observed Late Holocene cycles of slope erosion and valley deposition (van Andel et al. 1986). Pointing out that either increased rainfall or terrace neglect can trigger slope erosion, and that either terracing or natural vegetation cover can stabilize the slopes, they rely on historical evidence and archaeological and radiometric dating to discriminate among the possibilities. However, the absence of field evidence for significant climatic change within the last 5000 years raises unanswered questions, given evidence accumulated recently for climatic reversals at the end of the fifth millennium and during the Little Ice Age of late medieval time.

The reinstated anthropogenic hypothesis is convincing because it is based on more observations, finer chronological control, greater analytical detail, more fully developed theories of stream processes, fuller historical and prehistorical evidence for the human factor, and sounder arguments than were lavished on the climatic hypothesis (Zangger 1992). It is incomplete because the complexities of the natural and cultural records could not be revealed or dated to the degree required for assessment of the relative contributions of the several possible agents involved in the scenarios (Bintliff 1992).

Whenever there are equifinalities involved in explanations, as with the choice above between terrace neglect and greater rainfall, it behooves the investigator to identify the crucial variables in the instances under investigation. In the Southern Argolid, it was not possible (1) to measure past rainfall amounts, (2) to find evidence of the seasonality of precipitation, (3) to establish the nature or abundance of the local vegetation in anything but gross approximations, (4) to date closely either episodes of soil displacement or agricultural terracing, or (5) to reconstruct in detail the ancient stream gradients and their sediment sources. Neither are the antique agricultural systems well understood. Without such detail, neither the sensitivity of the stream basins to local changes nor the contributions of the several relevant trigger factors can be realistically assessed. Note, for instance, that a change to greater precipitation accompanied by less marked seasonality of rainfall might allow more robust vegetation cover in the region, making soil erosion far less likely rather than the reverse. We see in later chapters that study of molluskan microfaunas in the soils could provide relevant information on past vegetation covers, even where pollen is poorly preserved. On the other hand, stream basin regimes can change longitudinally because of basin morphology and sediment yield without any necessity for external triggers (e.g., Patton and Schumm 1981).

A revised climatic argument draws strength from the neo-catastrophism of recent climatology, which recognizes the episodic and rapid nature of short-term climate change, exemplified by the extraordinary storms of the past three decades (Bintliff 1992). The argument is tempered by new data that suggest that the dramatic depopulation cycles following economic expansions in ancient Greece may have been triggered by the literal collapse of the agricultural base. The weakness of linear causation models based on too few key variables once again is exposed. As the debate continues in the Mediterranean, a growing literature on Holocene valley fills in temperate Europe and the British Isles is revising the historical geomorphology and land-use history of those regions. The potentials and problems of such research agendas are well discussed by John Boardman and Martin Bell (1992). Bell summarizes cogently: "environment is the product, not only of natural factors, but the history of their interrelationship with human activity" (Bell 1992a: 21).