CHAPTER 1

CONCEPTS OF SCALE AND SCALING

JIANGUO WU AND HARBIN LI

1.1 INTRODUCTION

The relationship between pattern and process is of great interest in all natural and social sciences, and scale is an integral part of this relationship. It is now well documented that biophysical and socioeconomic patterns and processes operate on a wide range of spatial and temporal scales. In particular, the scale multiplicity and scale dependence of pattern, process, and their relationships have become a central topic in ecology (Levin 1992, Wu and Loucks 1995, Peterson and Parker 1998). Perspectives centering on scale and scaling began to surge in the mid-1980’s and are pervasive in all areas of ecology today (Figure 1.1). A similar trend of increasing emphasis on scale and scaling is also evident in other natural and social sciences (e.g., Blöschl and Sivapalan 1995, Marceau 1999, Meadowcroft 2002).

Scale usually refers to the spatial or temporal dimension of a phenomenon, and scaling is the transfer of information between scales (more detail below). Three distinctive but interrelated issues of scale have frequently been discussed in the literature: (1) characteristic scales, (2) scale effects, and (3) scaling (and associated uncertainty analysis and accuracy assessment). The concept of characteristic scale implies that many, if not most, natural phenomena have their own distinctive scales (or ranges of scales) that characterize their behavior (e.g., typical spatial extent or event frequency). Characteristic scales are intrinsic to the phenomena of concern, but detected characteristic scales with the involvement of the observer may be tinted with subjectivity (Wu 1999). Conceptually, characteristic scales may be perceived as the levels in a hierarchy, and associated with scale breaks (O’Neill et al. 1991, Wu 1999). Ecological patterns and processes have been shown to have distinctive characteristic scales on which their dynamics can be most effectively described and understood (Clark 1985, Delcourt and Delcourt 1988, Wu 1999). Thus, properly identifying characteristic scales provides a key to profound understanding and enlightened scaling.

Scale effects usually refer to the changes in the result of a study due to a change
in the scale at which the study is conducted. Effects of changing scale on sampling and experimental design, statistical analyses, and modeling have been well documented in ecology and geography (e.g., Turner et al. 1989b, White and Running 1994, Wu and Levin 1994, Pierce and Running 1995, Jelinski and Wu 1999, Dunne et al. 2002, Wu 2004). In geography, scale effects have been studied for several decades in the context of the modifiable areal unit problem or MAUP (Openshaw 1984, Jelinski and Wu 1996, Marceau 1999). Scale effects may be explained in terms of scale-multiplicity, characteristic scales, and hierarchy, but may also be artifacts due to errors in sampling and measurements, distortions in data resampling, and flaws in statistical analysis and modeling (Jelinski and Wu 1996, Wu 2004). Characteristic scales and scale effects are inherently related to the issue of scaling. While characteristic scales provide a conceptual basis and practical guidelines for scaling, quantitative descriptions of scale effects can directly lead to scaling relations (Wu 2004).

![Figure 1.1. Rapid increase in the use of terms related to scale in the ecological literature. Based on an internet search using JSTOR (http://www.jstor.org), the number of articles containing words (scaling, hierarchy, hierarchies, hierarchical, hierarchy theory) shows a great increase in four major ecology journals in the last seven decades (gray line). The trend for the word scaling alone is similar (black line). The four journals are: Ecology and Ecological Monographs published by Ecological Society of America, and Journal of Ecology and Journal of Animal Ecology published by British Ecological Society. Note that the number of years for the 1990's was only seven (1990-1996).](image)

With the recent burst of interest in the issues of scale, the terms scale and scaling have become buzzwords in ecology. However, because these terms have been used in diverse disciplines, both have acquired a number of different connotations and expressions. Good science starts with clear definitions. The development of a
science of scale or scaling may be hampered if the concepts of scale and scaling are used without any consistence. In this section, we review the main usages of these terms, propose a three-tiered scale conceptualization framework, and discuss their relevance to the issue of ecological scaling.

1.2 CONCEPT OF SCALE

We propose a three-tiered conceptualization of scale, which organizes scale definitions into a conceptual hierarchy that consists of the dimensions, kinds, and components of scale (Figure 1.2). Dimensions of scale are most general, components of scale are most specific, and kinds of scale are in between. This three-tiered structure seems to provide a clearer picture of how various scale concepts differ from or relate to each other.

1.2.1 Dimensions of Scale

We distinguish three primary dimensions of scale: space, time, and organizational levels. Note that Dungan et al.’s (2002) three dimensions of scale (sampling, analysis, and phenomena) are commensurable with what we here call the kinds of scale (see below). Space and time are the two fundamental axes of scale, whereas organizational hierarchies are usually constructed by the observer (Figure 1.2a). Scale has commonly been defined in terms of time or space. In recent decades, the relationship between temporal and spatial scales has received increasing attention. It is well documented that the characteristic scales of many physical and ecological phenomena are related in space versus time, such that the ratio between spatial and time scales tends to be relatively invariant over a range of scales. This ratio is termed the characteristic velocity (Blöschl and Sivapalan 1995). The idea that spatial and temporal scales are fundamentally linked so that complex systems can be decomposed in time and space simultaneously is essential to hierarchy theory (Courtois 1985, Wu 1999). This space-time correspondence principle has been supported by a number of empirically constructed space-time scale diagrams (or Stommel diagrams) in the past two decades (Stommel 1963, Clark 1985, Urban et al. 1987, Delcourt and Delcourt 1988, Blöschl and Sivapalan 1995, Wu 1999). These studies have shown that, for a variety of physical, ecological, and socioeconomic phenomena, large-sized events tend to have slower rates and lower frequencies, whereas small things are faster and more frequent. However, one must recognize that not all natural phenomena strictly obey the space-time correspondence principle. Many temporally cyclic events, for example, take place over a wide range of spatial scales with a relatively constant frequency. In some other cases, scale variability of different sources may overwhelm the signal of scale correspondence. Furthermore, the space-time scale ratio of most ecological phenomena can surely be altered drastically by human modifications.

For the purpose of scaling, levels of organization or integration are most useful when they are consistent with spatial and temporal scales. Hierarchy theory states that higher levels are larger and slower than lower levels, which is consistent with
the space-time principle. This is generally true for nested hierarchies (i.e., systems in which small entities are contained by larger entities which in turn are contained by even larger entities), but not for non-nested hierarchies (Wu 1999). In this view, the three dimensions of scale – space, time and organizational or integrative levels – can be related to each other. When moving up the ladder of hierarchical levels, the characteristic scales of entities or events in both space and time also tend to change accordingly.

Figure 1.2. A hierarchy of scale concepts: (A) dimensions of scale, (B) kinds of scale, and (C) components of scale (A was modified from Dungan et al. 2002; B and C were based on Bierkens et al. 2000).
1.2.2 Kinds of Scale

Several kinds of scale can be distinguished based on any of the three dimensions of scale (Figure 1.2B). Intrinsic scale refers to the scale on which a pattern or process actually operates, which is similar to, but broader than, the concept of process scale, a term frequently used in earth sciences (e.g., Blöschl and Sivapalan 1995). Some may argue that there is no intrinsic scale in nature, and that scales or hierarchical levels are merely epistemological consequences of the observer (Allen and Starr 1992). We believe that the observed scale of a given phenomenon is the result of the interaction between the observer and the inherent scale of the phenomenon. Although the existence of intrinsic scales does not mean that they are always readily observable, a suite of methods, including spectral analysis, fractal analysis, wavelet analysis, scale variance, geostatistics, and multiscale object-specific analysis (e.g., Turner et al. 1991, Wu et al. 2000, Hay et al. 2001; Dale et al. 2002, Hall et al. 2004), have been used in detecting characteristic scales or scale breaks. Effective scale detection requires that the scale of analysis be commensurate with the intrinsic scale of the phenomenon under study (Blöschl and Sivapalan 1995, Wu and Loucks 1995, Dungan et al. 2002, Legendre et al. 2002). Because the latter is unknown a priori, multiple observation sets at different scales usually are necessary (Allen et al. 1984, Wu 1999).

There are several other kinds of scale that are not intrinsic to the phenomenon of interest. Observational scale is the scale at which sampling or measurement is taken (also referred to as sampling scale or measurement scale). In experimentation, the spatial and temporal dimensions of the experimental system represent the experimental scale, which is the primary criterion for distinguishing among micro-, meso-, and macro-scale experiments. Similarly, the resolution and extent in space and time of statistical analyses and dynamic models define the analysis scale or modeling scale. In the context of environmental management and planning, local, regional, and national laws and regulations introduce another kind of scale – the policy scale, which is influenced by a suite of economic, political, and societal factors.

These different kinds of scales are related to each other in various ways (Figure 1.2B). In general, only when the scales of observation and analysis are properly chosen, may the characteristic scale of the phenomenon of interest be detected correctly; only when the scales of experiments and models are appropriate, may the results of experiments and models be relevant; only when the scale of implementation of policies is commensurate with the intrinsic scale of the problem under consideration, may the policies be effective. In reality, different kinds of scales may differ even for the same phenomenon, resulting in the problems of scale mismatch (or scale discordance). To rectify such scale mismatch or to relate one type of scale to the other usually involves scale transfer or scaling (Bierkens et al. 2000). An adequate understanding of the relationship among the different kinds of scale needs to invoke the definitions of scale components.
1.2.3 Components of Scale

Dimensions of scale and kinds of scale are useful concepts, but quantifying scale and developing scaling relations require more specific and measurable definitions. These are the components of scale, including cartographic scale, grain, extent, coverage, and spacing (Figure 1.2C). The traditional cartographic scale (or map scale) is the ratio of map distance to actual distance on the earth surface. A so-called large-scale map usually covers a smaller area with greater detail. Cartographic scale is essential for the creation and use of maps, but inadequate for studying the scale-dependent relationships between pattern and process in ecology because of its intended rigid connotation (Jenerette and Wu 2000).

In ecology and other earth sciences, scale most frequently refers to grain and extent – two primary components of scale. Grain is the finest resolution of a phenomenon or a data set in space or time within which homogeneity is assumed, whereas extent is the total spatial or temporal expanse of a study (Turner et al. 1989a, Wiens 1989). Grain may be considered as the pixel size for raster data, or the minimum mapping unit for vector data. A frequently used geostatistical term, support, refers to the smallest area or volume over which the average value of a variable is derived (Dungan et al. 2002). In most cases, grain and support have quite similar meanings, and thus have often been used interchangeably. However, support may differ from grain because support itself includes not only the size of an n-dimensional volume, but also its geometrical shape, size and orientation (Dungan et al. 2002). When the linear or areal dimension of grain is referred to, grain element or grain unit can be used, which corresponds to support unit in the literature. Note that soil scientists and hydrologists frequently use scale only to refer to support (e.g., Bierkens et al. 2000).

On the other hand, the concept of extent is less diversified than grain. A term equivalent to extent is geographic scale, which was defined by Lam and Quattrochi (1992) as the size of a particular map. Both grain and extent are of great importance to the study of heterogeneous landscapes (Turner 1989). Besides grain and extent, coverage and spacing, which are associated particularly with sampling, are also important in scaling. Coverage, not to be confused with extent, refers to sampling intensity in space or time (Bierkens et al. 2000), while spacing is the interval between two adjacent samples or lag. Spatial coverage can be represented as the ratio of the sampled area to the extent of a study, and spacing may be fixed or variable depending on the sampling scheme (Figure 1.2C). Support, extent, and spacing are sometimes called the scale triplet in hydrological literature, which highlights the importance of these three components in scaling (Blöschl and Sivapalan 1995).

The relationship between intrinsic scale and other kinds of scales can be further elaborated in terms of scale components. Hierarchy theory suggests that the scale of observation must be commensurate with the scale of the phenomenon under consideration if the phenomenon is to be properly observed (Simon 1973, Allen et al. 1984, O’Neill et al. 1986, Wu 1999). On the one hand, processes larger than the extent of observation appear as trends or constants in the observation set; on the other hand, processes smaller than the grain size of observation become noise in the
data. Thus, the choice of a particular scale for observation, analysis and modeling in terms of grain size and extent directly influences whether or not the intrinsic pattern and scale of a phenomenon can be eventually revealed in the final analysis. The significance of the choice of scale has long been recognized in plant ecology (e.g. Greig-Smith 1983) and human geography (Openshaw 1984, Jelinski and Wu 1996).

In general, the grain size of sampling or observation should be smaller than the spatial or temporal dimension of the structures or patterns of interest, whereas it is desirable to have the sampling extent at least as large as the extent of the phenomenon under study (Dungan et al. 2002).

In addition, the concept of relative scale can be rather useful for comparative studies and scaling across different ecosystems or landscapes. Meentemeyer (1989) defined relative scale as the relationship between the smallest distinguishable unit and the extent of the map, which can be expressed simply as a ratio between grain and extent. Schneider (2001) used range to refer to extent, and defined scope as the ratio of the range to the resolution of a research design, a model, or a process. In principle, different phenomena and research designs can be compared on the basis of their scopes. Relative scale can also be defined by directly incorporating the ecological pattern and process under consideration. Such definition is rooted in the conceptualization of relative versus absolute space (Meentemeyer 1989, Marceau 1999). For example, Turner et al. (1989b) considered relative scale as “a transformation of absolute scale to a scale that describes the relative distance, direction, or geometry based on some functional relationship (e.g., the relative distance between two locations based on the effort required by an organism to move between them).”

1.3 CONCEPT OF SCALING

Scaling has been defined differently in various fields of study, and its meanings can be quite disparate. Scaling has long been associated with measurement that is “the assignment of numerals to objects or events according to rules” (Stevens 1946). In this case, scaling is a way of measuring the “unmeasurable” (Torgerson 1958). In multivariate statistics, scaling usually refers to a set of techniques for data reduction and detection of underlying relationships between variables. Multivariate statistical methods, such as polar ordination, multidimensional scaling, principal component analysis, and correspondence analysis, have been used extensively in vegetation classification and ordination to organize field plots (or community types) into some order according to their similarities (or dissimilarities) on the basis of species composition. Multidimensional scaling, in particular, is to represent similarities among objects of interest through visual representation of Euclidean space-based patterns, and has been widely used to analyze subjective evaluations of pairwise similarities of entities in a wide range of fields, including psychology, marketing, sociology, political science, and biology (Young and Hamer 1994). These multivariate statistical methods can be useful for relating patterns and processes across scales (e.g., multiscale ordination; ver Hoef and Glenn-Lewin 1989). However, the concept of scaling as either the assignment of numerical values to
qualitative variables or the reduction and ordination of data is not directly relevant to scaling as defined below.

In physical sciences, scaling usually refers to the study of how the structure and behavior of a system vary with its size, and this often amounts to the derivation of a power-law relationship. This notion of scaling has often been related to the concepts of similarity, fractals, or scale-invariance, all of which are associated with power laws. For example, a phenomenon or process is said to exhibit “scaling” if it does not have any characteristic length scale; that is, its behavior is independent of scale—i.e., a power law relationship (Wood 1998). This scaling definition has long been adopted by biologists in terms of allometry that primarily correlates the size of organisms with biological form and process (Wu and Li, Chapter 2). In this context, scale refers to “the proportion that a representation of an object or system bears to the prototype of the object or system” (Niklas 1994), and ecological scaling then becomes “the study of the influence of body size on form and function” (LaBarbera 1989). Thus, to some, ecological scaling is simply some form of biological allometry (e.g., Calder 1983, Schmidt-Nielsen 1984, LaBarbera 1989, Brown and West 2000).

However, a more general and widely accepted definition of scaling in ecology and earth sciences is the translation of information between or across spatial and temporal scales or organizational levels (Turner et al. 1989a, King 1991, Blöschl and Sivapalan 1995, Stewart et al. 1996, van Gardingen et al. 1998, Wu 1999, Bierkens et al. 2000, Gardner et al. 2001). In some cases, this across-scale translation of information can be done through explicit mathematical expressions and statistical relationships (scaling equations), whereas in many other cases process-based simulation models are necessary. This definition of scaling is also referred to as scale transfer or scale transformation (Blöschl and Sivapalan 1995, Bierkens et al. 2000). This broadly defined scaling concept neither implies that scaling relations must be power-laws, nor that ecological patterns and processes must show scale-independent properties in order to “scale” or to be “scaled.” In this case, allometric scaling is but only one special case of scaling. Based on the directionality of the scaling operation, two kinds of scaling can be further distinguished: (1) scaling up or upscaling which is translating information from finer scales (smaller grain sizes or extents) to broader scales (large grain sizes and extents), and (2) scaling down or downscaling which is translating information from broader scales to finer scales.

Several other terms are closely related to, but not the same as, scaling. These terms are associated with three basic scaling operations: changing extent, changing grain size, and changing coverage. Extrapolation is transferring information from smaller to larger extents, coarse-graining transferring information with increasing grain size, and fine-graining transferring information with decreasing grain size. Sometimes, upscaling and downscaling refer specifically to coarse-graining and fine-graining, respectively (e.g., Bierkens et al. 2000). When dealing with spatial data that do not have 100% coverage, one may need to estimate the values of unmeasured spatial locations using information from measured sites—a process called interpolation. The reverse process of interpolation is sampling. In practice, the three basic operations may all be needed in a single study. That is, different
CONCEPTS OF SCALE AND SCALING

methods for interpolation, sampling, coarse-graining, fine-graining, and extrapolation may be used together to achieve the overall goal of scaling. In general, to make the concept of scale operational, one needs to be specific about the scale components (e.g., grain, extent, coverage, spacing); to put the concept of scaling into action, one has to invoke specific scaling operations (e.g., extrapolation, coarse-graining, fine-graining, interpolation). Any spatial scaling approach or method will inevitably involve one or more of the basic scaling operations.

Note that the definition of extrapolation given above is quite specific and unequivocal. However, in the literature, extrapolation in space has been used in at least four distinct ways: (1) using known data acquired from certain locations to estimate unknown values or draw inferences at other locations, (2) estimating values or drawing inferences about things that fall outside the study area, (3) transferring information from one scale to another in terms of either extent or grain, and (4) transferring information between different systems at the same spatial scale (Turner et al. 1989a, Wu 1999). The multiple meanings of extrapolation may cause confusions. For example, the first usage is simply spatial interpolation. The second is consistent with the definition of spatial extrapolation as information transfer with increasing extent. The third is extremely broad and may refer to coarse-graining, fine-graining, or scaling in general. The fourth usage makes sense with regard to the literal meaning of the word, but it does not fit the definition of scaling because scaling has to involve at least two or more scales. Hence, the term extrapolation should be used with caution.

1.4 WHY SCALING AND HOW?

Simply put, scaling is the essence of prediction and understanding, and is at the heart of ecological theory and application (Levin 1992, Levin and Pacala 1997, Wu 1999, Chave and Levin 2003). More specifically, two main reasons are commonly recognized. First, scaling is inevitable in research and practice whenever predictions need to be made at a scale that is different from the scale where data are acquired. In general, whenever information is averaged over space or time, scaling is at work. For example, the sampling plots that ecologists usually use for determining the distribution of organisms or the stocks and fluxes of materials are only a small portion of the spatial extent of ecological systems of interest. Thus, system-level descriptions dictate the translation of information from these small plots to much larger areas. Also, while most ecological studies traditionally have been conducted on local scales, environmental and resource management problems often have to be dealt with on much broader scales (i.e., landscapes, regions, or the entire globe). To bridge such scale gaps requires scaling.

Second, because ecological phenomena occur over a wide range of scales and because there are often hierarchical linkages among them, relating information across scales as well as levels of biological organization is an essential part of ecological understanding. For example, the dynamics of sub-watershed units and their interactions are crucial to understanding the hydrological and biogeochemical cycles of the whole watershed ecosystem (Wickham et al., Chapter 12). The dynamics of local populations and their interpatch interactions are crucial to
understanding population dynamics at the landscape scale. In a similar vein, understanding the primary productivity of the whole ecosystem requires knowledge of photosynthesis at the individual leaf level.

While it is imperative in almost all ecological studies, spatial scaling can also be extremely challenging in theory and practice. Spatial heterogeneity can greatly complicate the scaling process. Spatial heterogeneity may manifest itself in terms of various patterns of land use and land cover, topography, hydrology, soils, climatic conditions, and biological factors. For example, extrapolation of plot-scale data to the landscape or regional scale is a trivial matter in a spatially homogeneous (uniform or random) environment. In a heterogeneous landscape, however, simply multiplying the plot-scale average with the total study area usually provides a rather poor estimate at the landscape scale (Li and Wu, Chapter 3). When ecological relationships are translated across scales in heterogeneous environments, they often become distorted—a phenomenon known as “spatial transmutation” (sensu O’Neill 1979, King et al. 1991, Wu and Levin 1994).

Also, as scale changes, new patterns and processes may emerge, and controlling factors may shift even for the same phenomena. Thus, observations made at fine scales may miss important patterns and processes operating on broader scales. Conversely, broad-scale observations may not have enough details necessary to understand fine-scale dynamics. Third, nonlinear interactions, time delays, feedbacks, and legacies in ecological systems may impose formidable challenges for translating information across scales or levels of organization (O’Neill and Rust 1979, Wu 1999). Therefore, on the one hand, spatial heterogeneity, scale multiplicity, and nonlinearity are important sources of biodiversity and ecological complexity; on the other hand, they are major hurdles for successful scaling.

Given the various obstacles, how should we proceed with scaling? This is the focus of our next chapter, where we will discuss two general scaling approaches: similarity-based and dynamic model-based scaling. A dozen specific scaling methods will also be examined in terms of their assumptions, ways of dealing with spatial heterogeneity and nonlinear interactions, and accuracy of scaling results. No matter which approaches is used, a theoretically and practically important concept in scaling up and down is scaling thresholds or scaling breaks, which are scale ranges at which scaling relations change abruptly. A scaling threshold may also be understood as a critical scale of a phenomenon where emergent properties due to nonlinear interactions and spatial heterogeneity come into effect. Thus, scaling thresholds, when properly identified, may reflect fundamental shifts in underlying processes or controlling factors, and can be used to define the domains of applicability of specific scaling methods.

1.5 DISCUSSION

In this chapter, we have discussed and clarified a number of concepts related to scale and scaling as used in a variety of fields of study. We propose a hierarchical framework in which the different connotations of scale can be organized with clarity and consistency. The three-tiered definitional hierarchy, consisting of the dimensions, kinds, and components of scale, shows both the diversity and
interrelatedness of the concepts of scale. In the practice of scaling, the components of scale (most frequently extent, grain, and coverage) must be invoked. Indeed, scaling methods are often designed to capture and deal with the change in these scale components singularly or in concert (see Wu and Li, Chapter 2 for details).

Clarification of key concepts is the first step towards a science of scale. The three-tiered definitional hierarchy seems to serve this purpose well even though it is only one of many possible ways of organizing these concepts. It is crucial for ecologists to recognize the different usages of scale and scaling, and to adopt a system of definitions that are consistent, clear, and accommodating to the development of quantitative methods. The science of scale will certainly benefit from clear concepts and definitions, which are essential for the development of effective methods and sound theories of scaling.

ACKNOWLEDGEMENTS

We would like to thank Geoffrey Hay, Fangliang He, Bruce Jones, and Simon Levin for their comments on an earlier version of the chapter. JW’s research on scaling has been supported in part by grants from U.S. Environmental Protection Agency’s Science to Achieve Results (STAR) Program (R827676-01-0) and National Science Foundation (DEB 9714833 for Central Arizona-Phoenix Long-Term Ecological Research).

REFERENCES


