Scale and scaling: a cross-disciplinary perspective

7.1 Introduction

Scale and heterogeneity are two key concepts in landscape ecology which are inherently related. Scale would matter little in a world where entities and relationships remain invariant across space or time, or in a landscape that is spatially or temporally homogeneous (i.e., uniform or random). However, real landscapes are heterogeneous biophysically and socioeconomically, and they must be treated as such for most questions and problems that interest us as scientists or citizens. Spatial heterogeneity – the diversity of entities and their spatial arrangement – is one of the most essential and unifying features of all natural and anthropogenic systems. Landscape heterogeneity is the manifestation of patchiness (discrete patterns) and gradients (continuous variations) that are intertwined across multiple spatial scales. Thus, scale is indispensable for describing and understanding landscape pattern.

It is not surprising, therefore, that scale has become one of the most fundamental concepts in landscape ecology, a field that focuses prominently on spatial heterogeneity and its ecological consequences (Risser et al. 1984, Forman and Godron 1986, Forman 1995, Turner et al. 2001). In fact, landscape ecology has been widely recognized by biologists, geographers, and even social scientists for its leading role in studying scale issues (McBratney 1998, Marceau 1999, Withers and Meentemeyer 1999, Meadowcroft 2002, Sayre 2005). However, it was not until the 1980s that the notion of scale began to gain its prominence in landscape ecology (and in ecology in general). Also, landscape ecology is not the only discipline that deals with scale and spatial pattern. The goal of geographical research is to describe and explain the spatial patterns of natural and anthropogenic features on the Earth’s surface (Harvey 1968), and scale as a
The geographic variable is “almost as sacred as distance” (Watson 1978). However, geographers have long opted for single-scale studies without adequate justification (Watson 1978, Meentemeyer 1989).

Nevertheless, studies explicitly dealing with spatial scale in both ecology and geography date back to several decades ago. For example, plant community ecologists have used various block-variance methods to investigate multiple-scale patterns of vegetation since the 1950s (Greig-Smith 1952, Dale 1999). On the other hand, insightful discussions on the relationships among pattern, process, and scale were provided by several prominent geographers in the 1960s and the 1970s (e.g., Haggett 1965, Harvey 1968, Miller 1978), when the field of landscape ecology was still unknown to most ecologists around the world. The most notable research on scale issues in the geographic literature, however, is the study of the so-called “modifiable areal unit problem” or the MAUP (Openshaw 1984, Jelinski and Wu 1996). The MAUP is quite relevant to scale issues in landscape ecology and will be further discussed later.

Even those disciplines that do not focus explicitly on spatial patterns have not been able to completely ignore the role of scale. For example, economists have long made the distinction between microeconomics and macroeconomics that correspond to fine-scale and coarse-scale economic patterns and processes, whereas different levels of institutions or organizational hierarchies (e.g., household, community, regional, national, and international) often define the scope and objectives of sub-disciplines and research topics in social and political sciences. In these cases, however, scale has often been treated implicitly or rather coarsely. Although scale is as important in social sciences as in natural sciences, greater progress has been made in ecological and physical sciences in recent decades. To date, efforts to compare and integrate scale issues across disciplines are lacking, but urgently needed (Wu and Hobbs 2002, Sayre 2005).

The main goal of this chapter is to provide an overview of the key concepts, methods, and state-of-the-science of scale and scaling issues that are relevant to landscape ecology. Obviously, this is an extremely ambitious goal because of the enormous scope and complexity of this topic. I shall discuss both the conceptual and technical issues of scale and scaling, and identify major research questions and challenges in scaling across heterogeneous landscapes. Although the principal emphasis is placed upon spatial scale, most of the concepts and methods also apply to temporal scale.

### 7.2 Concepts of scale and scaling

The terms scale and scaling have acquired a number of connotations from various disciplines. No matter how it is defined, scale generally “implies a certain level of perceived detail” (Miller 1978), which most commonly pertains...
Scale definitions may be grouped into three classes: dimensions, kinds, and components of scale (Table 7.1). Space, time, and organizational hierarchies represent three primary dimensions of scale, of which space and time are most fundamental. Organizational hierarchies, when nested, generally follow the space–time correspondence principle: higher levels correspond to broader spatial and longer temporal scales, whereas lower levels are associated with finer spatial and shorter temporal scales (Simon 1962, Urban et al. 1987, Wu 1999). Within each scale dimension, one can distinguish between different kinds of scale: intrinsic scale, observation scale, experimental scale, analysis/modeling scale, and policy scale (see Table 7.1 for definitions). Except for intrinsic scale, all other types of scale are defined or imposed by the investigator. To quantify variations of a pattern or process across scale, however, one must specify scale components that are operational. Common components of scale include cartographic (map) scale, grain (resolution, support), extent, coverage (sampling density, intensity), and spacing (interval, lag).

While cartographic scale remains a fundamentally important concept in mapping science, grain and extent have firmly established themselves as the most frequently used, operational concepts of scale in ecology. Specifically, grain refers to the finest level of spatial or temporal resolution of a pattern or a data set, and extent is the spatial or temporal span of a phenomenon or a study (Allen et al. 1984, Turner et al. 1989a, Wiens 1989).

The term scaling is sometimes also known as scale transfer or scale transformation (Bloschl and Sivapalan 1995, Bierkens et al. 2000). In physical sciences, scaling has traditionally referred to the derivation of power laws, and this narrow definition has been adopted in biology and ecology for decades. In particular, biological allometry involves deriving power-law relationships between the size of organisms and biological processes (Schmidt-Nielsen 1984, Niklas 1994). Some researchers treat ecological scaling simply as the search for power laws in the biological world (e.g., Calder 1983, Brown and West 2000). However, a broader definition of scaling, i.e., the translation of information across scales or organizational levels, has widely been used in ecology, geography, and environmental sciences (Turner et al. 1989a, Wiens 1989, King 1991, Rastetter et al. 1992, Bloschl and Sivapalan 1995, Marceau 1999, Wu 1999, Bierkens et al. 2000). Accordingly, the process of transferring information from finer to broader scales is called scaling up or upscaling, whereas translating information from broader to finer scales is known as scaling down or downscaling. In general, scaling involves changing grain size, extent, or both (Allen et al. 1984, Turner et al. 1989a, King 1991, Wu 1999). Note that hierarchical levels and scales in time and space are different but closely related concepts. All levels can be characterized in terms of specific spatiotemporal scales, but not all scales represent organizational levels of hierarchical systems. Nevertheless,
<table>
<thead>
<tr>
<th>Dimensions of scale</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>A fundamental dimension that allows for fast or frequent events to be distinguished from those that are slow or infrequent</td>
</tr>
<tr>
<td>Space</td>
<td>A fundamental dimension whereby large and small entities can be distinguished and their configurations can be discerned</td>
</tr>
<tr>
<td>Organizational hierarchy</td>
<td>A directional ordering of interacting entities that have distinctive process rates, thus forming different levels. As ecological organizations exist in space and time, levels always correspond to certain spatial and temporal scales</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Kinds of scale</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrinsic scale</td>
<td>Scale at which a pattern or process actually operates</td>
</tr>
<tr>
<td>Observation scale</td>
<td>Scale at which measurements are made or sampling is conducted</td>
</tr>
<tr>
<td>Experimental scale</td>
<td>Scale at which an experiment is performed</td>
</tr>
<tr>
<td>Analysis/modeling scale</td>
<td>Scale at which an analysis is conducted or a model is constructed</td>
</tr>
<tr>
<td>Policy scale</td>
<td>Scale at which policies are intended to be implemented</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Components of scale</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Grain</td>
<td>Finest level of spatial or temporal resolution of a pattern or a data set; equivalent or similar to resolution, support, or minimum mapping unit (MMU)</td>
</tr>
<tr>
<td>Extent</td>
<td>Spatial or temporal span of a phenomenon or a study; equivalent to the study area or study duration</td>
</tr>
<tr>
<td>Coverage</td>
<td>Proportion of the study area or duration actually sampled; also called sampling density or intensity</td>
</tr>
<tr>
<td>Spacing</td>
<td>Distance between two neighboring sampling units; also called sampling interval or lag</td>
</tr>
<tr>
<td>Cartographic scale</td>
<td>Ratio of map distance to actual distance on the Earth’s surface; also called map scale</td>
</tr>
</tbody>
</table>
“a change in scale often necessitates consideration of new levels of organization” (O’Neill and King 1998).

7.3 Scale effects, MAUP, and “ecological fallacy”

Scale-related studies in landscape ecology during the past two decades have focused on three distinctive but intrinsically linked issues: characteristic scales, scale effects, and scaling. In this section I shall discuss the first two, with an emphasis on scale effects. Scaling approaches and methods will be the subject of the next section. In particular, this section makes a deliberate effort to compare and contrast scale effects in ecology with the MAUP and the so-called “ecological fallacy” in geography and the social sciences.

7.3.1 Characteristic scales and scale effects

The characteristic scale of an ecological phenomenon is the spatial and temporal scale on which the phenomenon principally operates and thus can be most appropriately studied. The background assumption of characteristic scales is that many, if not most, patterns and processes each take place on a finite range of scales (or scale domains), and thus different phenomena can be characterized by their distinctive scale domains. A number of empirically constructed space-time diagrams, in which phenomena are plotted against the space and time scales of their occurrences, corroborate this assumption (e.g., Stommel 1963, Clark 1985, Urban et al. 1987, Delcourt and Delcourt 1988, Blöschl and Sivapalan 1995). On the other hand, different phenomena may overlap in their scale domains to varying degrees, and this scale overlap can tell us the nature of the relationship between the different processes of interest. For example, processes operating on commensurate scales may interact frequently, whereas processes with disparate rates (e.g., a few orders of magnitude apart) may have no direct effect on each other. From this perspective, identifying the characteristic scales of relevant patterns and processes is a critical first step in designing a successful research project. Hierarchy theory has provided a conceptual framework as well as practical guidelines for the search of characteristic scales, whose detection is often associated with scale breaks (e.g., O’Neill et al. 1991, Cullinan et al. 1997, Wu 1999, Hay et al. 2001, Hall et al. 2004).

A phenomenon may not be observed or gauged properly if the scale of observation is not commensurate with the characteristic scale of the phenomenon. While the scale of observation is a choice by the observer, characteristic scales are intrinsic to that being observed. Across a landscape, changing the “lenses” of observation may lead to a series of different patterns, and the same phenomenon may be manifested differently on different scales. These are scale
effects, reflective of both the scale multiplicity of landscape structure and artifacts in pattern analyses (Wu 2004, Li and Wu, Chapter 2, this volume). More specifically, scale effects may occur in any statistical analyses or dynamic models that use area-based data when grain size or extent is changed. Although the effects of quadrat size and position on observed vegetation pattern were explicitly investigated in the 1950s by plant ecologists, it was not until the late 1980s that landscape ecologists began to investigate the various effects of changing grain size and extent on landscape pattern analysis and, to a lesser extent, on landscape modeling. Turner et al. (1989b) were among the first to systematically study how changing grain size and extent could affect three landscape indices (diversity, dominance, and contagion). Since then, numerous studies have examined scale effects in landscape pattern analysis (Benson and Mackenzie 1995, Wickham and Riitters 1995, Jelinski and Wu 1996, O’Neill et al. 1996, Saura 2004, Wu 2004) and spatial modeling (King et al. 1991, Wu and Levin 1994, Ciret and Henderson-Sellers 1998, Kersebaum and Wenkel 1998, Jenerette and Wu 2001).

7.3.2 The MAUP

When landscape ecologists were busy “discovering” scale effects with new pattern metrics and remote sensing data in the 1990s, studies of the MAUP-related issues had existed for several decades in geography and the social sciences. The root of the MAUP, as the name suggests, is the use of areal units that are “modifiable” or arbitrary. Area-based data include census data, remote sensing data, and raster-based maps of soil, vegetation, land use, and other themes. The MAUP has two components: the scale effect and the zoning effect (Openshaw 1984, Jelinski and Wu 1996). The scale effect here refers to the variation in the results of statistical analysis caused by spatially aggregating data into fewer and larger areal units (i.e., reducing the spatial resolution or coarse-graining). This is equivalent to the effect of changing grain size in landscape ecology (Turner et al. 1989b, Wu 2004). The zoning effect is the variation in the results of statistical analysis due solely to different ways of aggregating areal units to a given scale of analysis (i.e., changing the boundaries and configurations of areal units at a given spatial resolution). In landscape ecology, variability in statistical results due to the aggregation of pixels along different directions is an example of the zoning effect (Jelinski and Wu 1996, Wu 2004).

The phenomenon of arbitrarily defined areal units affecting statistical results was first noticed in electoral geography more than 100 years ago when politicians purposefully manipulated the local boundaries of electoral districts to alter the outcome of an election without changing the individual votes themselves (gerrymandering). As the earliest study of the MAUP, Gehlke and Biehl
(1934) conducted a correlation analysis between male juvenile delinquency and median monthly income from 252 census tracts in Cleveland, USA, and found that the correlation coefficient increased considerably as the areal units were aggregated contiguously. While early studies were sporadic, the resurgence of interest in the MAUP in the 1980s was evident from the flurry of stimulating studies by Openshaw and his associates (e.g., Openshaw and Taylor 1979, Openshaw 1984). Numerous MAUP studies have been published ever since, most of which were concerned with correlation and regression analyses (Arbia 1989, Goodchild and Gopal 1989, Fotheringham and Wong 1991, Wrigley 1994, Amrhein 1995).

After decades of research, however, geographers still have different views on the nature and scope of the MAUP (Goodchild and Gopal 1989, Wrigley 1994, Jelinski and Wu 1996, Marceau 1999). One extreme view regards the MAUP as simply a consequence of using “bad” or improper methods, and thus the solution is to find “scale-independent” or “frame-independent” methods. But most other views recognize that the MAUP is a result of the interactions between the methods and the data used. That is, spatial effects are not just artifacts, and the MAUP can provide useful information on the multiple-scaled patterns embedded in the data (Jelinski and Wu 1996, Marceau 1999, Hay et al. 2001).

7.3.3 The “ecological fallacy”

The existence of the MAUP implies that statistical relationships from area-based data may change with the scale of analysis, and thus cross-scale inferences are unwarranted. This point was made clearly and loudly by Robinson (1950) when he introduced the distinction between “an individual correlation” and “an ecological correlation.” In individual correlations the variables are descriptive properties of indivisible individuals, whereas in ecological correlations the variables are descriptive properties of groups of individuals. A striking example in Robinson (1950) was the correlation between nativity and illiteracy for the USA in 1930. The analysis using individual-level data produced a positive correlation between foreign birth and illiteracy (i.e., the individual correlation $= 0.118$), supporting the common observation that the native-born generally had a better command of American English. However, the same analysis using the state-level aggregated data indicated that the percent illiterate was negatively correlated with percent foreign-born (the ecological correlation $= -0.619$). Apparently, this aggregate-level result could lead to a wrong inference at the individual level that the foreign-born were more likely to be literate of American English than the native-born. In reality, this
aggregate-level correlation was due largely to the fact that most foreign-born lived in states where the native-born were relatively literate (Freedman 2001).

Robinson (1950) concluded that individual and ecological correlations were almost always different in practice because the ecological correlations were usually stronger than the individual correlations. Since then, the phenomenon of improper inferences of individual behavior from an analysis of groups has been known as the “ecological fallacy” (Wrigley et al. 1996, King 1997). Note that the word, “ecological,” in this case means “of groups” or “of aggregates,” not really related to the interrelationship between organisms and their environment. Unfortunately, this connotation of “ecology,” a dangerously misleading distortion of the original meaning of the word, has long been used in the social and behavioral sciences, such as “ecological correlations,” “ecological regressions,” “ecological inferences,” and “ecological fallacies” (e.g., Dogan and Rokkan 1969, Poole 1994, Wrigley et al. 1996, King 1997, Freedman 2001). Alker (1969) attempted to develop “a typology of ecological fallacies” to include several types of inappropriate inferences from aggregated areal data. In particular, the individualistic fallacy referred to the improper generalization of aggregate-level relationships from individual-level results, a somewhat converse problem of the “ecological fallacy.” The rest of the “ecological fallacies” identified by Alker (1969) were related to different kinds of sampling and conceptual errors in statistical inferences. The “ecological” and individualistic fallacies are both cross-level inference fallacies, and really should have been called as such.

Robinson’s (1950) study has attracted a great deal of attention particularly because quantitative social and political studies (and thus policies and actions based on such studies) at the time were based primarily on aggregate areal data. It “startled, dismayed, and even infuriated many users” of areal data (Alker 1969), and “sent two shock waves through the social sciences that are still being felt, causing some scholarly pursuits to end and another to begin” (King 1997). Unfortunately, a number of unintended yet misleading consequences have resulted from Robinson’s (1950) study. In particular, the notion of “ecological fallacy” has led to several misguided conceptions: individual-level models are always better specified and more accurate than aggregate-level models, aggregate-level relationships are always intended as substitutes of individual-level relationships, and aggregate-level variables have no relevance to causal relationships and mechanistic explanations of individual-level activities (Allardt 1969, Schwartz 1994). In fact, aggregate-level relationships can be quite useful for defining the context, generating potential hypotheses, and identifying the relevance for studying individual-level phenomena. Frequently, aggregate-level variables may not only be constraints on, but also direct causes of, individual-level processes. For example, population-level
studies are crucial for identifying important public health problems, and certain risk factors for diseases genuinely operate at the population level (Pearce 2000).

7.3.4 Towards a more comprehensive understanding of scale effects

The numerous studies of the MAUP and cross-level fallacies are evidently relevant to understanding scale issues faced by landscape ecologists as well as other scientists. The literature of the social sciences on these issues is a rich source of information for learning how scale can help elucidate complex processes, identify hierarchical linkages, or create spurious patterns in human landscapes where social, economic, and political forces are dominant drivers. Findings of the effects of MAUP on correlation analysis, regression analysis, and geospatial models (Openshaw 1984, Arbia 1989, Goodchild and Gopal 1989, Fotheringham and Wong 1991, Amrhein 1995) should be relevant for similar types of landscape ecological analyses.

The “ecological fallacy” is a problem of disaggregation (or downscaling) in which inferences about a lower level are made from knowledge of an upper level. Ecologists are frequently faced with such challenges to predict the properties of “trees” using information on the “forest.” Developed in the social sciences over the past several decades, the various methods for solving the problem of cross-level fallacies may prove to be useful for solving genuinely ecological problems as well. These methods are collectively known as the “ecological inference” methods, including “ecological regression” (Goodman 1953, Freedman 2001), the neighborhood method (Freedman et al. 1991), and the EI method (King 1997). However, analogous to deciphering land cover composition within a pixel of a remote sensing image, inferring the behavior of lower levels from higher-level data is inherently difficult because: (1) aggregate data usually do not contain explicit information on subgroup behavior, and (2) the characteristics of aggregates may be outcomes of nonlinear interactions among subgroups or emergent properties, so that they cannot be simply “decomposed” using reductionist methods. Like other downscaling methods (more in the next section), none of the ecological inference methods can work well in all circumstances. Both impressive progress and thorny problems in cross-level inference research are evident in a series of exchanges between some leading scholars in this area (Freedman et al. 1998, Freedman et al. 1999, King 1999).

In spite of its relevance to ecology, the term MAUP seemed completely absent in the ecological literature until the mid-1990s when Jelinski and Wu (1996) discussed the implications of the MAUP for landscape ecology. Even today, the enormous literature on the MAUP and cross-level fallacies continues to
be ignored by biological and physical scientists, including the most scale-cognizant landscape ecologists. This situation is puzzling because ecological analyses frequently use area-based data and because landscape ecology is actually known for being highly interdisciplinary. In geography and the social sciences, on the other hand, even the recent literature on the MAUP and cross-level fallacies seldom cites any of the scale-related studies in ecology. This is equally disappointing given that geography and landscape ecology both emphasize spatial views and approaches.

7.4 Theory and methods of scaling

Spatial scaling is about translating information across heterogeneous landscapes. The significance and challenges for spatial scaling both reside in the fact that landscape patterns and processes are spatially heterogeneous, non-linearly interactive, and replete with feedbacks and threshold dynamics. Thus, to move from one scale to another in such complex landscapes, one has to either assume away heterogeneity, nonlinearity, and feedbacks, or deal with them explicitly and effectively. In practice, spatial scaling is done through seven basic operations (Bierkens et al. 2000). Changing extent, grain size, and coverage are the three basic operations, whereas the other four are different combinations of the three (Fig. 7.1). Strictly speaking, extrapolation is to increase the
extent of an observation set, while interpolation is to increase the coverage of a study area. In landscape ecology and geography, scaling frequently involves the change of extent (extrapolation) and grain size or resolution (fine-graining and coarse-graining). The key to spatial scaling is to figure out ways to implement these scaling operations, i.e., scaling approaches and methods.

Scaling methods may be grouped into two general approaches (Blöschl and Sivapalan 1995, Bierkens et al. 2000): the similarity-based scaling approach (SBS) and the dynamic model-based scaling approach (MBS). SBS is based on the principles of similarity, and often characterized by power-law scaling functions derived either analytically or empirically. In contrast, MBS transfers information between different scales through changing the input, parameters, and formulation of dynamic models. MBS tends to be more comprehensive, and usually does not lead to simple scaling functions as does SBS.

7.4.1 The SBS approach

Similarity has long been used as the background assumption in a number of scaling methods. Two systems are said to be similar if they share some properties that can be related across the systems by a simple conversion factor (Blöschl and Sivapalan 1995). These similarities can be of different kinds, including geometric, dynamic, and functional similarities. An important and relatively new concept in the SBS approach is self-similarity, which is the key idea in fractal geometry (Mandelbrot 1982, Hastings and Sugihara 1993). Self-similarity refers to the phenomenon that the whole is composed of smaller parts resembling the whole itself and that patterns remain similar at different scales. In the following, I discuss two commonly used SBS methods: similarity analysis and allometric scaling.

7.4.1.1 Similarity analysis

Similarity analysis aims to reduce dimensional quantities required for describing a phenomenon based on the known governing equations (Blöschl and Sivapalan 1995). Barenblatt (1996) provided a “general recipe” for similarity analysis that included seven steps: (1) to specify a system of governing variables that are necessary to describe the phenomenon of interest, such that a mathematical relation of the form, \(a = f(a_1, \ldots, a_k, b_1, \ldots, b_m)\), can be assumed to hold; (2) to determine the dimensions of variables and select those variables whose dimensions are independent of each other; (3) to represent or transform the relations under study as products of powers (or dimensionless ratios) of variables with independent dimensions; (4) to estimate the numerical values of the similarity parameters (dimensionless variables) using empirical data; (5) to formulate scaling laws (i.e., relationships between nondimensional groups) under
the assumption of complete similarity, and test them against empirical data; (6) if the test in step 5 fails, then formulate scaling laws under the assumption of incomplete similarity (or self-similarity) and test them against empirical data (in this case, scaling analysis cannot be completed using dimensional analysis because the power laws are fractal); and (7) to formulate similarity laws with as few similarity parameters as possible.

There have been a number of successful applications of similarity analysis in geophysical sciences. The Monin–Obukhov theory assumes that atmospheric boundary-layer flows can be viewed as being dynamically similar across scales and relates turbulent fluxes to a mean vertical gradient of wind, temperature, and specific humidity (Brutsaert 1982, Wu 1990). Thus, the gradient-diffusion theory (or K-theory), originally developed for molecular-level diffusion processes, has been used to estimate broader-scale turbulent transfer of heat and mass based on the small-eddy concept that treats turbulent transport as a result of local mixing by small eddies. In other words, even though the turbulent diffusivity (about 1 m² s⁻¹) is as much as 10⁵ times greater than the molecular diffusivity (about 10 to 20 mm² s⁻¹), turbulent transfer may still be treated as a dynamically similar process to molecular diffusion. Of course, this is not always a valid treatment. While the K-theory has been successful in modeling turbulent transfer for the boundary layer above vegetation, its success is limited within plant canopies where the small-eddy concept is less appropriate (Brutsaert 1982, Wu 1990).

Similarity analysis has widely been used in soil and hydrological sciences (Blöschl and Sivapalan 1995, Sposito 1998, Bierkens et al. 2000). A well-known example is the derivation of scaling equations for soil-water transport on the basis of the fine-scale similar-media concept known as the Miller–Miller similarity (Miller and Miller 1956, Sposito 1998). Similarity analysis has not been widely used in ecology maybe because the required governing equations for most ecological processes are either nonexistent or analytically intractable.

### 7.4.1.2 Allometric scaling

Allometry usually refers to the study of the relationship of biological form and process to the size of organisms (LaBarbera 1989, Niklas 1994, Brown and West 2000). The allometric scaling relations are usually based on assumptions of similarity (e.g., geometric similarity and self-similarity), and take the form of a power law: \( Y = Y_0 M^b \), where \( Y \) is some variable representing a pattern or process of interest, \( Y_0 \) is a normalization (or scaling) constant, \( M \) is some size-related variable (e.g., body mass), and \( b \) is the scaling exponent. There are two ways of obtaining allometric scaling relations: the analytical and empirical approaches. This dichotomy may be generalized for all SBS methods in physical and biological sciences.
The analytical approach derives scaling relations from the existing theory of similarity using techniques such as dimensional analysis, and thus has the ability to explain and predict cross-scale relationships. However, these analytically derived scaling relations must be tested against empirical observations for their validity. The empirical approach is descriptive and inductive, and usually employs two kinds of regression analysis. Ordinary least squares regression (OLS) can be used when the purpose of a study is only to predict one variable based on the other, or to find out if the relationship is statistically significant (in this case simple correlation analysis can also be used). However, if the purpose is to determine the exact value of the scaling exponent (i.e., the slope of the regression line in a log–log plot), OLS regression is generally inadequate especially when the coefficient of correlation is small (Niklas 1994). In this case, reduced major axis (RMA) regression is more appropriate (LaBarbera 1989, Niklas 1994) because it treats the two variables in the allometric equation in the same way (i.e., no “independent” variables in the regression equation and both variables have an error term).

Brown et al. (2002) discussed three classes of power laws that describe a variety of biological and ecological phenomena. Power laws of the first class have a rather limited range of variation in the scaling constant ($Y_0$) and the scaling exponent ($b$), and are mostly quarter-power laws (e.g., animal metabolic rates, developmental time, life span, maximum rate of population growth, and other organism-level allometric relations). The second class has a wide range of values of $Y_0$ and $b$ (e.g., population densities of different species). For the third class of power laws (e.g., species-area relationship, species-time relationship, and species-abundance distribution), not only are the scaling parameters not well constrained, but also the power laws themselves do not hold up over many orders of magnitude. Brown et al. (2002) asserted that the first class “apparently reflects the fractal-like designs of resources distribution networks,” whereas the third class “may not represent examples of self-similar behavior over a wide range of scales.” In the past decade, there has been a resurgence of interest in biological allometry which has generated much excitement and controversy (e.g., Dodds et al. 2001, Bokma 2004, Brown et al. 2004, Cyr and Walker 2004, Kozlowski and Konarzewsk 2004). More than 15 years ago, LaBarbera (1989) commented, “Whether a power law reflects a basic biological truth, the underlying structure of the universe we are embedded in, or whether it is simply fairly robust at approximating a variety of data relations is yet to be determined.” Alas, this statement still seems to hold true today.

Biological allometry is not always relevant to spatial scaling unless spatial scale is incorporated into the allometric equation. Schneider (2001, 2002) provided a number of examples of spatial allometry for lake ecosystems and aquatic mesocosms in terms of the geometric attributes of the systems (e.g.,
the volume, area, perimeter, and depth of lakes or mesocosms) and biological properties (e.g., fish catch and primary production). Landscape-scale studies using this approach have been increasing in recent years. For example, Hood (2002) identified several allometric scaling relations between slough attributes (e.g., area, outlet width, perimeter, and length) for rivers, and showed that detrital insect flotsam density was also allometrically related to slough perimeter. Similarly, Belyea and Lancaster (2002) found that the area, depth, width, and length of peatland bog pools were allometrically related. It is tempting to jump from empirically derived power-law relations to ecological explanations of underlying mechanisms by invoking the theory of self-similarity and self-organization. But this is unwarranted, be it in vogue. Nonetheless, spatial allometry provides a general method to summarize and extrapolate observed patterns over a range of scales, and to suggest underlying processes (Wu 2004).

7.4.2 The MBS approach

Unlike the SBS approach in which similarity goes both ways, MBS methods for upscaling versus downscaling differ in terms of both general perspectives and detailed procedural steps (Fig. 7.2). Thus, they are discussed separately here although both may be used interactively in a given scaling project.

7.4.2.1 Upscaling methods

Upscaling with dynamic models typically consists of two major steps: characterizing heterogeneity, and aggregating information by scaling up local (or patch-level) models (Fig. 7.2). Characterizing heterogeneity usually involves the classification and quantification of spatial patterns, which is a way of simplifying the complexity of scaling by partitioning the heterogeneous landscape into a limited number of relatively homogeneous patches. The second step is to aggregate information from the finer to the broader (target) scale through manipulating the input and parameters or altering the formulation of the local-scale model. Depending on the scaling context, this process may correspond to one of two basic scaling operations: coarse-graining (increasing grain size) or extrapolation (increasing extent). A number of upscaling methods have been developed in geophysical and biological sciences during the past decades. King (1991) presented four methods: extrapolation by lumping (EL), direct extrapolation (DE), extrapolation by expected value (EEV), and explicit integration (EI). This list can be expanded to include additional methods, such as extrapolation by effective parameters (EEP), spatially interactive modeling (SIM), and the scaling ladder method (SL). Each method is described below.
Figure 7.2
Illustration of general procedures in spatial scaling using dynamic models. (A) Flow diagrams for upscaling by aggregating model input and parameters (upper panel), and by aggregating model output (lower panel; adapted from Bierkens et al. 2000); (B) A two-step scheme for both upscaling and downscaling (modified from Blöschl and Sivapalan 1995); and (C) A schematic representation of the essence of downscaling involving disaggregation and singling out (adapted from Bierkens et al. 2000).
Extrapolation by lumping is to estimate the target-scale result by running the local-scale model with the mean values of parameters and inputs averaged across the entire landscape. The major procedural steps of this scaling method can be depicted in a “scaling flow diagram” (see the upper panel of Fig. 7.2A). EL does not deal explicitly with spatial heterogeneity; rather, it suppresses it in average values of model arguments. EL is the simplest and most error-prone upscaling method. In theory, it only works well when the local model is linear and still valid at the target scale, and when horizontal interactions between patches are weak and symmetric.

Instead of averaging parameters and inputs before running the local model as in EL, DE obtains the target-scale results by averaging the outputs of the local model that is run, with spatially varying parameters and inputs, for all patches of the entire landscape (King 1991). The scaling flow diagram of DE is in sharp contrast with that of EL (see the lower panel of Fig. 7.2A). Averaging the outputs rather than inputs of the local model can significantly reduce scaling errors due to the nonlinearity in the model (Bierkens et al. 2000), and eliminates the need to apply the local model directly at the landscape scale. DE treats spatial heterogeneity explicitly but not interactively, assuming that horizontal interactions and feedbacks are negligible or at steady state. Typically, DE does not consider any processes that operate at scales larger than the patch on which the local model is developed. DE is data-demanding and computationally intensive, and thus may not be feasible when the landscape is too large.

EEV obtains the target-scale results by deriving the expected value of the outputs from the local model, which is run based on joint probability density functions or a sampling approach (e.g., Monte Carlo simulation) to account for spatial heterogeneity (King 1991, Rastetter et al. 1992). The scaling flow diagrams of EEV and DE are the same in terms of the general steps, but differ in the specifics of how to go from one step to the next. EEV does not treat spatial heterogeneity explicitly, but in statistical terms. By so doing, EEV alleviates the problems of excessive computational and data demands that DE may suffer, and is amenable to uncertainty analysis (Rastetter et al. 1992, Li and Wu 2005). As with DE, EEV neither explicitly considers the patch configuration nor feedbacks and interactions among patches.

EI refers to directly integrating the local-scale model to the landscape scale analytically or numerically based on explicit mathematical formulations (King 1991). In this case, the spatial heterogeneity of the landscape must be represented as mathematical functions of space in closed forms, and the indefinite integral of the local model with respect to space must be obtainable. When all of its requirements are met, EI is the most elegant, efficient, and accurate upscaling method. Unfortunately, this is rarely the case with real landscapes.
Similar to EL, EEP also assumes that the local model applies to the target scale, but uses “effective” or “representative” parameters, instead of spatial averages, to produce the target-scale estimates (Blöschl and Sivapalan 1995, Bierkens et al. 2000). Because both methods run the local-scale model with landscape-scale input and parameters, EEP and EL share the same scaling flow diagram (Fig. 7.2A). EEP has been widely used in soil physics, hydrology, and micrometeorology, and finding effective parameters can be quite difficult for nonlinear models (Blöschl and Sivapalan 1995, Bierkens et al. 2000).

When horizontal or lateral interactions must be considered explicitly (e.g., metapopulation processes, disturbance spread, and land–water interactions), spatially interactive modeling seems to be the only option (Judson 1994, Wu and Levin 1997, Tenhunen and Kabat 1999, Rastetter et al. 2003, Peters et al. 2004). SIM is able to incorporate feedbacks, time delays, and new features on larger scales. Spatially interactive models include variables, parameters, and input at multiple scales. Thus, the scaling flow diagram of SIM would be different from those in Fig. 7.2A; rather it needs to reflect the multi-scaled nature of the models themselves (e.g., Fig. 2 in Wu and Levin 1997). Such models can easily become ecologically too complex and computationally overwhelming (Levin et al. 1997, Levin and Pacala 1997). This is particularly true when the number of scales becomes more than just a few. In this case, a hierarchical scaling scheme is useful to simplify complexity and reduce aggregation errors.

All the upscaling methods discussed above typically are of “short-range” because the assumptions behind them are less likely to be satisfied over a broad range of scales and because they become technically less feasible when multiple scale breaks (or thresholds) are encountered. In these cases, the scaling ladder method may be used (Wu 1999). SL is based on the hierarchical patch dynamics (HPD) paradigm, which integrates hierarchy theory and patch dynamics (Wu and Loucks 1995). The basic idea is to establish a spatial patch hierarchy consisting of a series of nested scale domains, and then use it as a scaling ladder to move information between two adjacent scales one step at a time (Wu 1999, Wu and David 2002). Thus, the short-range scaling methods discussed above can all be used in a hierarchical scaling framework. Examples of patch hierarchies for upscaling purposes include levels of biological organization (e.g., leaf–plant–stand as in Reynolds et al. 1993) and different types of nested landscape units (e.g., Reynolds and Wu 1999, Wu and David 2002, Hall et al. 2004).

7.4.2.2 Downscaling methods
Downscaling also consists of two general steps: disaggregating information and singling out (Fig. 7.2). The goal of disaggregating coarse-grained information is to derive the fine-scale pattern within a given areal unit (e.g., pixel
or patch), a process also known as fine-graining (Blöschl and Sivapalan 1995, Bierkens et al. 2000). Downscaling often uses stochastic or probabilistic methods with auxiliary information on the finer scale. Singling out is simply to locate the site of interest in the disaggregated pattern. Much of the research on downscaling in the past few decades has been done in the context of global climate change, and the primary goal is to translate general circulation model (GCM) output into regional-scale predictions for scientific research as well as decision-making purposes. These methods are usually classified into two general categories: empirically based statistical and process model-based downscaling approaches (Hewitson and Crane 1996, Wilby and Wigley 1997, Kidson and Thompson 1998, Wilby et al. 1998, Murphy 1999, 2000).

The empirically based statistical downscaling approach aims to derive regional climate conditions (e.g., temperature, precipitation, and wind velocity) from large-scale synoptic circulation features (e.g., upper-level winds, geopotential heights, and sea-level pressure) predicted by GCMs. This is usually done through “transfer functions” which are obtained through multiple linear regression, artificial neural networks, classification and regression trees, or other statistical methods (Hewitson and Crane 1996, Wilby et al. 1998, Li and Sailor 2000, Crane et al. 2002). The empirical downscaling approach works well for temporally continuous variables such as temperature, but much less effectively for temporally discontinuous and highly intermittent variables such as precipitation (Li and Sailor 2000).

The process model-based downscaling approach, on the other hand, uses nested dynamic models of different scales to disaggregate information downward. For example, a higher-resolution regional climate model may be embedded within a global GCM, so that the GCM output drives the regional model which in turn produces downscaled results. The models are coupled either through one-way or two-way nesting schemes. In a two-way nesting scheme, GCM and the embedded regional climate model are run simultaneously and interact with each other across scales (Hewitson and Crane 1996, Kidson and Thompson 1998, Murphy 1999).

While the current literature on downscaling is dominated by meteorological and climatologic studies, other methods exist in hydrological and soil research that focus on regional down to local scales. These methods may also be grouped into the two general downscaling approaches discussed above. Examples of disaggregating information on soil properties, hydrological time series, and precipitation patterns are abundant (Blöschl and Sivapalan 1995, Bierkens et al. 2000). Also, in the social sciences, as mentioned earlier in this chapter, the methods of “ecological inferences,” including “ecological regression,” the neighborhood model, and the EI model (Freedman et al. 1998, Freedman et al. 1999, King 1999), may also be used for ecological downscaling, particularly,
when the research goal is to decipher the behavior of lower-level elements from higher-level aggregate relationships.

In addition, the common problem of “pixel mixing” in remote sensing arises from the fact that a single pixel is often a mixture of multiple spectrally unique land cover types (i.e., the “endmembers”), which leads to errors in image classification. Remote-sensing scientists have developed a series of subpixel analysis methods to “unmix” individual pixels to estimate the relative areal proportions of different land-cover types within a pixel. The most widely used has been the linear spectral unmixing model, which assumes that the reflectance spectrum of any pixel is the result of linear combinations of the spectra of all constituent land-cover types within that pixel (Rosin 2001, Song 2005). The relative abundance of each land-cover type within a pixel is obtained by solving a closed system of \( n \) linear equations where \( n \) is the number of bands in an image. In recent years, a number of other methods have been developed for pixel unmixing, including fuzzy membership functions (Foody 2000), the least median of squares method (Rosin 2001), and wavelet and neural network-based methods (e.g., Mertens et al. 2004). The potential of these pixel unmixing methods for ecological downscaling studies is yet to be explored.

7.4.3 Uncertainty analysis

Scaling practices always come with uncertainties because of spatial heterogeneity, nonlinearity, data inadequacy, and problems with scaling techniques. The main purposes of uncertainty analysis (or error propagation analysis) are to identify the various sources of uncertainties and quantify their effects on scaling results (Rastetter et al. 1992, Heuvelink 1998a, 1998b). Different scaling methods are amenable to different uncertainty analysis techniques. For example, many empirically based statistical scaling methods produce scaling results with some relevant information on uncertainty (e.g., variance, confidence intervals, and regression or correlation coefficients). Monte Carlo techniques may be used with dynamic modeling methods, such as extrapolation by expected value and other stochastic models, to estimate confidence intervals. While uncertainty analysis can be quite challenging, a number of methods have been developed in recent years (Rastetter et al. 1992, Heuvelink 1998a).

Li and Wu (2005) reviewed different aspects of uncertainty analysis, including sources of uncertainty in scaling, evaluation of scaling algorithms, error propagation from parameters and input data to scaling results, and presentation of prediction accuracy and error partitioning. Several techniques for uncertainty analysis have been used in ecology and environmental sciences, including probability theory, Taylor series expansion, Monte Carlo simulation, generalized likelihood uncertainty estimation, Bayesian statistics, and
sequential partitioning. They recommended the following desirable outputs of uncertainty analysis: (1) measures of model adequacy, (2) full probability distributions of model outputs (e.g., density function and probability-weighted values), (3) reliability of model results (e.g., accuracy, confidence level, and error), (4) relative contribution or importance of each factor as an error source to total uncertainty, (5) the likelihood of different scenarios (probability or ranking), and (6) identification of the least understood or critical components of the model.

From the above discussion it is clear that uncertainty analysis should be regarded as an essential part of the scaling process. But this has not been the case in ecological studies. Given the increasing importance of cross-scale studies in today’s scientific research and environmental decision-making, it is crucial to properly quantify and report uncertainties with scaling results.

7.5 Discussion and conclusions

The increasing prominence of scale issues in ecology and other sciences since the 1980s seems inevitable for several reasons. First, ecology as a science has become progressively more explanatory, and mechanistic explanations inevitably invoke multiple scales in space and time as well as multiple levels of organization. Second, for the increasing need to understand and solve broad-scale environmental problems, scientists have to translate information across spatial and temporal scales or organizational hierarchies. Third, the past two decades have witnessed significant advances in theory and methodology for tackling the complexity of spatially extended, heterogeneous systems such as landscapes. Important theories and methods for scaling include hierarchy theory (Allen and Starr 1982, O’Neill et al. 1986), fractal geometry (Mandelbrot 1982), phase transition and percolation theory (Gardner et al. 1987, Milne 1992), cellular automata (Wolfram 1984), self-organized criticality (Bak 1996), and complex adaptive systems (Cowan et al. 1994, Levin 1999).

Fourth, recent advances in remote sensing, geographic information systems (GIS), and computing technologies have equipped scientists with powerful tools for dealing with issues of heterogeneity and scale. In addition, the rapid development of landscape ecology since the 1980s has certainly contributed to the widespread recognition of the importance of scale within ecology and beyond.

Today, landscape ecologists are generally aware that scale may directly influence the results of a study whenever spatial heterogeneity cannot, or should not, be assumed away. Heterogeneity makes no sense without the explicit consideration of scale, and scale matters little without heterogeneity. There seems to be a consensus among landscape ecologists today that, whenever possible, a
multiple-scale or hierarchical approach is preferable to a single-scale approach. Scaling, as the process of translating information across spatiotemporal scales and organizational levels, has been increasingly emphasized in ecological studies. Indeed, scaling is the essence of understanding and prediction, and has a central role in ecological theory and application (Levin 1992, Levin and Pacala 1997).

There are still numerous problems and challenges in dealing with scale and scaling issues across disciplines. I conclude this chapter by highlighting several of them as follows:

- First, scale and scaling are unifying concepts that cut across all disciplines in both natural and social sciences, and the diversity of connotations presents both problems and opportunities. To avoid unnecessary confusion, these terms should always be specified when they are used. Beyond that, to take a leading role in developing a science of scale, landscape ecologists must familiarize themselves with the scale-related terminology and methods developed in other fields, such as geography, soil science, hydrology, and the social sciences.

- Second, while scale effects are pervasive in the study of heterogeneous landscapes, we must move beyond simply reporting the occurrences of scale effects, which would be an endless effort. Instead, the emphasis should be placed on the search for scaling relations that can be used to identify underlying processes and translate information across scales. A straightforward and powerful approach is to construct empirical scalograms in which variations of patterns, processes, and their relationships are plotted directly against scale (Turner et al. 1989b, Ludwig et al. 2000, Wu 2004). Such scalograms provide not only direct evidence to test scaling theories, but also a simple yet reliable way of scaling up and down information across landscapes.

- Third, the two general scaling approaches, similarity-based and dynamic modeling, need to be better understood and integrated in ecological studies. No matter how authentic it may sound, the classic definition of scaling that hinges on power laws is not adequate; it only covers part of what has actually taken place in ecological scaling. The two scaling approaches are not contradictory, but complementary to each other. Scale-invariance theory and hierarchy theory may seem at odds, but they are simply different perspectives on the same multi-scaled world. A hierarchical system is composed of a number of scale domains within which scale-invariance may well exist. Similarly, a hierarchical scaling scheme may include similarity-based methods. Future scaling studies in landscape ecology should clearly recognize the pros
and cons of both approaches, and emphasize the integration between the two whenever necessary. Neither brutal forces with overwhelmingly complex models nor scale-free power laws with elegantly simplistic equations alone would be adequate for understanding and predicting the dynamics of landscapes.

- Fourth, for both approaches it is important to properly identify scaling thresholds at which scaling relations change abruptly. These thresholds suggest fundamental shifts in underlying processes or controlling factors (Gardner et al. 1989, Turner et al. 1989a, King et al. 1991, Wu and Loucks 1995), and define the domains of applicability of the various scaling methods.

- Fifth, one of the greatest challenges for scaling in real landscapes is to integrate biophysical with socioeconomic processes. This is especially true for human-dominated landscapes (e.g., agricultural and urban landscapes) where natural and anthropogenic processes are intertwined and often operate on different scales. The mechanics and rules of scaling for different processes may also vary dramatically. When it comes to the practice of scaling, universality is elegant, but more of utopia; idiosyncrasy is torturous, but more of reality. Complex interdisciplinary issues call for a hierarchical, pluralistic scaling strategy that integrates both empirical statistical and dynamic modeling methods.

- Finally, scaling without known accuracy is unreliable, and uncertainty analysis needs to be an integral part of the scaling process. Ecological scaling, especially with dynamic models, has rarely been done with rigorous accuracy assessment. While it is challenging, uncertainty analysis should be emphasized in future scaling studies because it provides critical information about the accuracy of scaling results. This uncertainty issue of scaling becomes particularly important when scaling results are expected to be used for management and policy-making purposes.

**Acknowledgments**

I sincerely thank Helene Wagner, Darrel Jenerette, Richard Hobbs, and Alex Buyantuyev for their comments on an earlier version of the chapter. My research on scaling has been supported by US EPA, US NSF, USDA, and NSF of China.

**References**


