The scale issue in social and natural sciences
Danielle J. Marceau
Geocomputing Laboratory, Department of Geography, University of Montreal, C.P. 6128, Succ.
Centre-Ville, Montréal (Qué.), Canada, H3C 2J7, tel.: 514-343-8067, fax: 514-343-8008

Abstract
Spatial scale is a major concept in many sciences concerned with human activities and physical processes occurring at the Earth’s surface. In particular, geographic literature is rich in discussions about the importance of scale in both the scientific and the familiar representations of the world (Meentemeyer, 1989; Ferras, 1992). Scale is now increasingly recognized as a central concept in the description of the hierarchical organization of our world. Many environmental problems such as the impact of climate change on ecosystems require an understanding of how processes operate at different scales, and how they can be linked across scales. The necessity to predict and control the scale and aggregation effects on statistical results and modelling is also recognized as a major goal to achieve. This paper presents a review of the conceptual and methodological developments that have occurred within the social and natural sciences over the last four decades, regarding the importance of the scale issue and the solutions proposed to handle it. It is shown that while these developments evolved in parallel for several years, they are now converging to provide a strong theoretical framework and operational methods for explicitly dealing with the scale issue.

Keywords: Scale, scaling, MAUP (Modifiable Areal Unit Problem), Hierarchy theory

Résumé
L’échelle spatiale est un concept central dans les sciences qui s’intéressent aux activités anthropiques et aux processus physiques qui se produisent à la surface de la terre. La littérature géographique est particulièrement riche en discussions concernant l’importance de l’échelle dans les représentations scientifique et familière du monde (Meentemeyer, 1989; Ferras, 1992). Le concept d’échelle est maintenant largement reconnu comme étant au cœur de la description de l’organisation hiérarchique de l’espace géographique. Plusieurs problèmes environnementaux tels que l’impact des changements climatiques sur les écosystèmes requièrent une compréhension de la façon dont les processus opèrent à différentes échelles et comment ils peuvent être liés à travers plusieurs échelles. La nécessité de prédire et de contrôler l’effet d’échelle et d’agrégation spatiale sur les résultats d’analyses statistiques et de modélisation est aussi maintenant largement reconnue.
Cet article présente une synthèse des principaux développements théoriques et méthodologiques réalisés dans les sciences sociales et naturelles, au cours des quarante dernières années, concernant l’importance du problème d’échelle et les solutions proposées pour y faire face. Il est montré que bien que ces développements aient eu lieu en parallèle pendant plusieurs années, ils sont maintenant en voie de converger pour générer un solide cadre théorique et des méthodes opérationnelles pour appréhender le problème d’échelle.

Mots clés: Échelle, changement d’échelle, problème des unités spatiales modifiables, théorie de la hiérarchie
Introduction

The importance of scale has been recognized in sciences concerned with the spatial organization of human activities and physical processes on the Earth’s surface for more than four decades. The variability in statistical results originating from the use of different scales or aggregation levels was demonstrated by Gehlke and Biehl in 1934. Twenty years later, Mc Carthy et al. (1956) showed that conclusions derived at one scale are specific to that scale and should not be expected to be valid at another scale. In the fifties, ecologists were aware of the existence of prefered scales at which patterns in the landscape are revealed and associations with ecological processes could be determined (Greig-Smith, 1952). Since then, a considerable body of literature has been and is still currently published on the subject by geographers (Arbia, 1989), landscape ecologists (Turner et al., 1989a), geomorphologists (Clark, 1990), hydrologists (Blöschl and Sivapalan, 1995), climatologists (Raupach and Finnigan, 1995), and remote sensing specialists (Quattrochi and Goodchild, 1997; Marceau and Hay, 1999) among others. What justifies this widespread interest in scale?

There are several reasons which one can evoke to explain such an emphasis on scale. First, many scientists now recognize that scale is the central concept for describing and explaining the complex hierarchical organization of the geographic world. Scale can be defined as a continuum through which entities, patterns, and processes can be observed and linked. Empirical studies reveal that distinct breaks or thresholds can be detected within the scale continuum that correspond to specific levels of organization within a hierarchical system. Therefore, it is necessary to identify these scale thresholds, and to derive the appropriate laws governing the interactions occurring within and between the levels of organization. Second, many environmental problems, such as global warming, continental deforestation, and regional water management cannot be handled at a single scale of observation. An understanding of how processes operate at various spatial scales and how they can be linked across scales becomes a primary goal when investigating these complex phenomena. To do so necessitates understanding the scale and aggregation effects on statistical results and modelling, and requires the derivation of appropriate solutions to cope with these effects. Finally, the impressive developments in computational performance, the rapid growth of technologies for spatial data acquisition and analysis, such as remote sensing and geographic information systems (GIS), and the wide diffusion of digital databases of all sorts represent unprecedented favorable conditions for the development and testing of hypotheses and methods to fully address the many challenges related to scale.

Despite the abundant literature already available, there still remains confusion about certain scale concepts and their implications to scientific inquiry. Undoubtedly, this is due to the development and varied use of these concepts within a large spectrum of disciplines. The objective of this paper is to provide a comprehensive review of the main conceptual and methodological developments made about scale in the social and natural sciences over the last forty years. Definitions of major concepts are first provided, followed by descriptions of the principal approaches used to handle the questions related to scale. From this perspective, a clearer picture emerges about the nature and impact of what is often referred to as the scale problem or the scale issue. This problem can be expressed in terms of two complementary and fundamental questions: 1) what is the appropriate spatial scale for the study of a particular geographical phenomenon?, and 2) how can we adequately transfer information from one spatial scale to another?
To clarify the definition and the importance of the scale issue, it is first necessary to refer to the two representations of space that cohabit current scientific models and methods, namely the absolute and the relative points of view, since they both involve a different way of dealing with scale. It is also important to recognize the complementary perspective brought by scientists in the social and natural sciences. The first group is primarily concerned by the effects of scale on statistical inferences and models when using aggregated data, while the second group focuses on detecting and linking spatial patterns and processes at different scales. By appreciating both these perspectives, it can be seen that while ideas evolved in parallel for several years, they are now converging towards a solid theoretical framework to handle the scale issue.

Definitions of scale and related concepts

What is space?
The absolute representation of space dominated the scientific world until the beginning of this century when the theory of relativity was formulated in Einstein’s work. According to the absolute framework, space is a container that can exist independent of any matter. Its structure is rigid and is best described with the familiar three-dimensional Euclidean geometry in which a straight line may be drawn from any point to any other point, and in which trigonometric rules can be applied. Such a description is considered objective that is, independent of one’s viewpoint or frame of reference (Sack, 1980).

On the contrary, relative space is not immune to influence; its properties and descriptions are dependent on the distribution of mass and energy. From a geographical perspective, space exists only with reference to spatial entities and processes under consideration. While the absolute view focuses on space as the subject matter in which objects are located within a non-changing geometry, the relative view focuses on objects as the subject matter, and space is measured as relationships between objects. The former involves measurement referenced to some constant base, implying non-judgmental observations. The later implies interpretation of patterns and processes within specific contexts (Peuquet, 1994). In such a framework, space may be defined in non-Euclidean terms. For example, two areas separated by a barrier may be close in absolute space but very distant in relative space when time, rates, and interactions are considered (Meentemeyer, 1989).

What is scale?
In a general sense, scale refers to the spatial dimensions at which entities, patterns, and processes can be observed and characterized. Ecologists define scale as having two components: grain and extent. The former corresponds to the smallest spatial sampling units used to gather a series of observations. Extent is the total area over which observations of a particular grain are made (O’Neill and King, 1998). Cartographic scales represent the ratio of a distance on a map to the corresponding distance on the ground. In remote sensing, the spatial resolution of a digital image represents the surface on the ground, or the spatial sampling increment, from which average values are collected and registered by the sensor.

The two representations of space bring different meaning to scale. In the absolute framework, scale can be defined as operational and refers to a practical, standard system used to partition geographical space into operational spatial units. Examples are provided by the use of standard cartographic scales, predefined digital remote sensing resolutions, census and administrative units, and any zoning or spatial partitioning system defined for a particular study. Such scales constitute
reference schemes that are used to precisely locate objects in space and to express their topological relationships.

In the relative framework, scale becomes a variable intrinsically linked to the spatial entities, patterns, forms, functions, processes, and rates under investigation. Scale can then be defined as the window through which the investigator chooses to view the world. It is a particular point of view, among many others that are possible, about geographic reality (Levin, 1993). Therefore, spatial scales within relative space are more difficult to define than those within the absolute space as used in cartography and remote sensing (Meentemeyer, 1989). Intuitively, we know that the choice of scale must be related to the phenomenon under observation and the questions being posed about it. However, there is no obvious method to determine this ideal window, nor is a single scale sufficient to investigate phenomena that are inherently hierarchical in space (Marceau et al., 1994; Peuquet, 1994; Hay et al., 1997).

**What is scaling?**
Scaling means transferring data or information from one scale to another. Upscaling consists of taking information at smaller scales to derive processes at larger scales, while downscaling consists of decomposing information at one scale into its constituents at smaller scales (Jarvis, 1995).

In the context of absolute space, scaling primarily involves a change in the geometric structure of the data and their corresponding attributes. The procedures currently used are resampling and cartographic generalization. Resampling implies a change in the geometric coordinate systems and attributes used to represent geographical entities or phenomena. It is often used when it is desired to compare two or more spatial datasets available at different scales, or to bring several datasets to a common denominator (Tobler, 1988). Cartographic generalization is the process of extracting the nature and amount of information about phenomena that can be represented on a map at a specified scale. Several methods of automated cartographic generalization have been proposed during the last twenty years, mostly for linear entities (McMaster, 1987). Recent studies have emphasized the need to develop more consistent logic rules to relate to a structural hierarchy based on geographical characteristic and meaning (Weibel, 1992; Wang and Muller, 1993).

From the relative perspective, scaling becomes a more complex task than from the absolute framework. It represents the transcending concepts that link processes at different levels of space. It entails a change in scale that identifies major factors operational on a given scale of observation, their congruency with those on the lower and higher scales, and the constraints and feedbacks on those factors (Caldwell et al., 1993). As expressed by Jarvis (1995), what makes scaling a real challenge is the non-linearity between processes and variables, and the heterogeneity in properties that determines the rates of processes. Therefore, scaling requires an understanding of the complex hierarchical organization of the geographic world where different patterns and processes are linked to specific scales of observation, and where transitions across scales are based on geographically meaningful rules.

While the concepts of absolute and relative space help to clarify the definition of scale, it is important to note that these two representations must not be seen as contradictory, but rather as complementary. Sciences dealing with geographic space require a fixed geometric structure, with standard coordinate systems and convenient scales to clearly define and accurately measure space.
However, to detect phenomena occurring at specific levels of organization in a hierarchy, it is necessary to have access not only to arbitrary or pre-imposed scales, but to a range of scales from which significant structures and relationship between variables can emerge. This implies a shift in paradigm where entities, patterns and processes are considered as intrinsically linked to the particular scale at which they can be distinguished and defined. Such a shift has gradually developed through a series of studies conducted in social and natural sciences during the last four decades; they are presented in the following sections.

**Developments in the social sciences**

**Importance of the scale problem**

The seriousness of the scale problem was first underlined by human geographers and social scientists. Gehlke and Biehl (1934), followed by Yule and Kendall (1950), demonstrated that correlation coefficients could greatly vary according to the number and size of areal units used to describe the phenomenon under investigation. They concluded that their correlation coefficients only measure the relationship between the variates for the specified units chosen for the study, and that they have no absolute validity independently of these units. Robinson (1950) introduced the term *ecological fallacy* to describe the error resulting from making statistical inferences from aggregate relationships to individual relationships.

Another major aspect of the scale issue was formulated by McCarthy et al. (1956) in their work about association in industries. They wrote: "In geographic investigation it is apparent that conclusions derived from studies made at one scale should not be expected to apply to problems whose data are expressed at other scales. Every change in scale will bring about the statement of a new problem, and there is no basis for presuming that associations existing at one scale will also exist at another." The work of these pioneers was followed by Blalock (1964) and Clark and Avery (1976), among others, who illustrated the significant consequences resulting from the spatial aggregation of data, particularly in the correlation and regression analysis of areally distributed phenomena.

The next significant step in the recognition of the scale problem was made by Openshaw (1977, 1978, 1981, 1984a, b) and Openshaw and Taylor (1979, 1981) who fully described what they called the *modifiable areal unit problem* (MAUP). The MAUP originates from the fact that a tremendously large number of different ways exists by which a geographical study area can be divided into non-overlapping areal units for the purpose of spatial analysis. Usually, the principal criteria used in the definition of these units are the operational requirements of the study. As a result, none of these spatial units have any intrinsic geographical meaning. Therefore, if the areal units are arbitrary and modifiable, then the value of any work based upon them may not possess any validity independent of the units which are being studied. The MAUP encompasses two components called the *scale problem* and the *aggregation problem*. The former represents the variation in results that can be obtained when data acquired from areal units are progressively aggregated into fewer, larger units for analysis; the latter refers to the variation in results produced by the use of alternative combinations of areal units at similar scales.

A series of studies were undertaken to illustrate the scale effect in various statistical models generally applied in human geography, such as location-allocation modeling (Goodchild, 1979;
Bach, 1981), input-output analysis (Blair and Miller, 1983), spatial interaction modeling (Putman and Chung, 1989), traditional statistical analysis (Dudley, 1991), and multivariate statistical analysis (Fotheringham and Wong, 1991). In this later study, the authors demonstrated that modification of the areal units from which data are collected create a severe problem for parameter estimation in multiple regression models. They concluded: "The results of this analysis are rather depressing in that they provide strong evidence of the unreliability of any multivariate analysis undertaken with data from areal units. Given that such analyses can only be expected to increase ... with the current proliferation of GIS technology which permits even more access to aggregated data, this paper serves as a topical warning."

A substantial contribution was also made by Arbia (1989) in a book that presented, for the first time, an integrated theoretical presentation of the consequences of the scale and aggregation problems in spatial data and their influence on the statistical properties of estimation and significance testing. More recently, the implications of the MAUP were also investigated in the classification of remote sensing imagery (Marceau et al., 1994; Arbia et al., 1996), on the analysis of landscape structure (Jelinski and Wu, 1996), and on the principal axis factoring technique (Hunt and Boots, 1996). In all cases, the studies illustrate the severity of the MAUP effects on the analysis results.

**Solutions proposed to overcome the MAUP**

One of the first solutions proposed to overcome the MAUP effects was the application of clustering techniques for grouping a set of areal units used to partition a study area in such a way that their number and configuration reveal as much information as possible about the phenomenon of interest. These techniques are based on the use of an objective *a priori* criterion and the results are conventionally represented by a dendrogram which shows the stages in grouping and the loss of detail that occurs during the process of aggregation (Masser et al., 1975; Batty, 1976; Sammons, 1976, 1979).

This strategy was criticized by Larson (1986) and more particularly by Openshaw (1977, 1978, 1984a) who rejected the premise of objectivity in the design of zoning systems and proposed at the time a radically different approach in which the inevitable dependence of results on the areal units used is recognized as an intrinsic valuable information source about the spatial entities under study. His methodology starts by formulating an hypothesis concerning the expected result for a given model or method of analysis, and then aggregating areal units to the point where the target result is attained. The originality of this approach is to state that the definition of an optimal zoning system changes with the kind of problems under investigation. A zoning system that is optimal for one variable may not be optimal for another. Such an approach emphasizes one of the fundamental aspects involved in the scale issue, i.e. the dependency between a variable and the particular scale at which it is measured.

Other solutions were investigated by Fotheringham (1989) and Visvalingam (1991) who respectively suggested the identification of basic geographical entities, and the use of basic spatial units that define the spatial primitives of the phenomenon under study and for which information could be collected and analyzed. The practicality of this approach is not obvious since a universal applicable set of areal units is almost impossible to attain due to the diversity of geographical problems and strategies that can possibly be addressed and implemented. However, the
identification of basic attributes defining the objects under investigation is a fundamental practical exercise that must be done to ensure the validity of any scientific enquiry. This question is at the core of the scale issue. If entities and relationships between variables emerge at specific scales, there must be a way to define them and to relate them across discrete levels of organization. This practical problem is raised when building a digital database where individual objects must be identified, and where their spatial and temporal topology must be built. This idea is also at the foundation of new sets of techniques for features extraction from digital imagery referred to as multi-scale object-specific techniques, where the focus is on the detection and identification of individual objects appearing on an image when the appropriate scale is reached (Hay et al., 1997).

Another approach is to systematically perform sensitivity analysis in order to provide the range of results obtained when different areal units are used. As an example, Knudsen (1987) proposed the adoption of standard sets of significance testing procedures to verify the appropriateness of traditional statistical methods applied on spatial data and to check the robustness of the results to variations in scale and level of aggregation. It has been argued that if the number of variables, scales, and zoning alternatives are large, the amount of work to perform an exhaustive sensitivity analysis may be too onerous to be considered as an operational solution (Jelinski and Wu, 1996). However, such analysis should always be done, even on a limited set of cases, in order to produce an estimate of the error resulting from the use of different areal units. In some applications, like the modelling of flood risks in a given area, showing only the results originating from the use of one set of areal units can hide the most significant aspects of the problem and jeopardize the validity of the conclusions that are presented.

Recent studies addressing the MAUP demonstrated the possibility of controlling and eventually predicting its impact to a certain extent. In their systematic study of the MAUP in principal axis factor analysis of spatial data, Hunt and Boots (1996) revealed that specific MAUP effects are strongly influenced by the presence of spatial autocorrelation in the data. Similarly, Amrhein and Reynolds (1996), using spatial statistics to assess aggregation effects for a set of socio-economic variables, concluded that the modified Getis correlation statistic may provide a reliable diagnostic to estimate the possible aggregation effects imbedded in a given set of data. Holt et al. (1996) also demonstrated that scale effects can be explained in terms of factors inducing correlation between individuals within the same geographic area.

While generally applicable solutions to overcome the MAUP are not yet available, these contributions provide new hope for managing the problem at an acceptable level. For many years, the research about the MAUP was clouded by a pessimistic view that its effects were intractable (Wong and Amrhein, 1996). In contrast, these recent studies illustrate that it is possible to control and predict the MAUP effects to some extent. But, most of all, they represent further steps towards the derivation of theoretically sound and operationally practical methods to deal with the issue.

Developments in the natural sciences

Importance of the scale problem

One of the major contributions in the field of natural sciences was to acknowledge the existence of natural scales at which ecological processes and physical characteristics occur within the landscape. This fact was revealed by a series of studies oriented toward the choice of an
appropriate sampling unit size for analyzing ecological phenomena, particularly to detect spatial patterns in plant communities (Greig-Smith, 1952, 1961; Kershaw, 1957; Usher, 1969, 1975; Hill, 1973; Mead, 1974; Carpenter and Chaney, 1983; O’Neill et al., 1988; Carlile et al., 1989). Their central conclusion was that the scale of investigation determines the range of patterns and processes that can be detected and, therefore, that the most appropriate levels of resolution for study of these processes should be identified. This conclusion was reinforced by further studies undertaken to evaluate the impact of spatial scales on the analysis of landscape (Turner et al., 1989b; Moody and Woodcock, 1995). As an example, Benson and McKenzie (1995) examined the effects of increasing spatial resolution from 20 m to 1100 m on landscape parameters characterizing spatial structures. Their results indicated that most measures were sensitive to changes in spatial resolution, some of them decreasing, others increasing with coarser resolutions.

An additional step towards understanding the effects of scale was revealed in the emergence of two complementary concepts: domain of scale and scale threshold. The former corresponds to regions of the scale spectrum over which, for a particular phenomenon, patterns do not change or change monotonically with changes in scale. Such domains are separated by thresholds, that are relatively sharp transitions or critical points along the spatial scale continuum where a shift in the relative importance of variables influencing a process occur (Meentemeyer, 1989; Wiens, 1989). The importance of these concepts was emphasized through a series of studies (presented below) undertaken to address three related issues, namely: the identification of relevant domains of scale and scale thresholds when studying a particular phenomenon, the evaluation of the scale effect on the explanatory role of sets of variables, and the determination of scaling laws to adequately link information across scales.

Bruneau et al. (1995) conducted a sensitivity analysis on the space and time resolutions of a hydrological model called TOPMODEL, a model based on the contributing area concept. Their analysis clearly showed that the modeling efficiency is fairly high and constant inside a relevant domain of space and time resolutions and that working outside this domain induces a strong decrease of modeling efficiency. Qi and Wu (1996) investigated how changing scale might affect the results of landscape pattern analysis using three commonly adopted spatial autocorrelation indices. Their results revealed that all three indices were scale-dependent, that is the degree of spatial autocorrelation measured by these indices varies with the spatial scale on which analysis was performed. All three indices showed a rather distinctive change around a specific spatial resolution suggesting the existence of an abrupt change in spatial patchiness around that scale for the landscape under investigation.

Further studies were conducted to demonstrate the effects of scale change on the explanatory power of different sets of variables. For example, at the scale of individual leaf surfaces it has been shown that transpiration is regulated by stomatal mechanisms while, at a broader scale of vegetation, climate becomes the controlling factor (Jarvis and McNaughton, 1986). As reported by Wiens (1989), the distribution of phytoplankton in marine systems is dominated by different factors according to the scale of investigation. At scales up to about 1 km, the controlling factor is the horizontal turbulent diffusion (Platt, 1972); at broader scales, zooplankton grazing and vertical mixing override the local effects (Lekan and Wilson, 1978); at scales greater than 5 km, phytoplankton distribution is governed by eddies and upwelling occurring over areas of 1 to 100 km (Gower et al., 1980).
Finally, since it was recognized that ecological and physical processes operate at different spatial scales, the next step in the understanding of the scale problem was to emphasize the need for appropriate scaling laws in order to relate information across a wide range of scales. As pointed out by some authors (Wiens, 1989; Levin, 1993), a variety of statistical and mathematical tools can be used for scaling, such as correlation and extrapolation. However, it appears that such techniques can be appropriate only when applied for short-term or small-scale predictions or, in other words, within the relevant domain of scale for the phenomenon under investigation. Extension across scale thresholds may be hazardous because of the instability in the dynamics of the transition zone between two domains of scale. The major steps required in order to make predictions across scales were best summarized by Turner et al. (1989a). They imply: 1) the identification of the processes of interest and parameters that affect this process at different scales, 2) the development of rules to translate information across scales, and 3) the ability to test these predictions at the relevant spatial and temporal scales.

Solutions proposed to the scale problem

The first step in the analysis of the scale problem in the natural sciences was the development of appropriate quantitative methods for detecting scales or discrete levels at which regular and irregular patterns occur in the landscape. Among the first techniques used are the blocking techniques that involve iterative aggregation of contiguous quadrats, and analysis of within and between block size variance performed at each aggregation level (Wiegert, 1962; Goodall, 1974; Ludwig and Goodall, 1978; Ludwig, 1979). Spectral analysis is also used to identify cyclic patterns in a sequence of equally spaced data according to the length of the intervals within which variation occurs (Ripley, 1978; Renshaw and Ford, 1984; Mulla, 1988). Autocorrelation functions, such as correlograms and semivariograms, are also commonly applied to measure the degree of spatial dependence between a sequence of data as a function of the distance separating them (Legendre and Fortin, 1989; Carlile et al., 1989; Oliver and Webster, 1986; McBratney and Webster, 1986; Mulla, 1988). Edge detection methods have also been used in various applications, such as analysis of edges along altitude gradients (Beals, 1969), species distribution and abundance (Schuerholz, 1974), composition gradients (Wilson and Mohler, 1983), and ecological gradients (Ludwig and Cornelius, 1987). The principle of these techniques is to search for edges, or discontinuities, by breaking the transect or the neighborhood of adjacent samples into sub-groups that are as internally homogeneous as possible and as distinct as possible from other groups. The between-within variance ratio is often used as the measure to detect significant discontinuities.

Description of other methods, mostly applied in landscape ecology, can be found in Turner et al. (1991) and Fortin (1999). These techniques represent a sample among the available set of statistical and mathematical tools that can be effectively used to detect spatial patterns and to establish the degree of relationship between those patterns and underlying processes. While they are often strongly related, each of them has requirements and restrictions that must be taken into account in relation to the data set in use.

The next step in fully understanding the scale issue is to explain and predict how patterns and processes change across scales. This involves the development of a theoretical framework, and the implementation of quantitative methods for scaling. Hierarchy theory was first developed by general systems theorists (Simon, 1962; Koestler, 1969), and then applied in ecology as a framework for linking information across spatial scales (Allen and Star, 1982; O’Neill et al.,
According to this theory, a landscape can be described as an aggregate entity composed of a triad of contiguous hierarchical levels of organization. The phenomenon of interest is represented at Level 0 as a component of a higher level (Level +1). Level 0 is made of components forming the lower level (Level -1). Dynamics of the upper level appear as driving or controlling variables of Level 0; they govern from among possible behaviors the results of Level 0 processes. On the other hand, components at Level -1 can be used to explain the mechanisms operating at Level 0. In this way, the theory focuses on a particular subset of spatial scales and enables systematic scientific study of very complex systems (Salthe, 1985; O’Neill, 1988; O’Neill et al., 1989).

A good example of the use of hierarchy theory for downscaling is provided by the aquatic production relationship demonstrated by Vollenweider (1975) and Schindler (1977). Their study relies on the fact that in nutrient-limited fresh water systems, annual production is closely related to phosphorus loading. By applying the principle of hierarchy theory, which is that dynamics of the level of interest can be determined by knowing the higher level constraints, they were able to predict productivity simply by the phosphorus loading, without information about the species of phytoplankton involved in the process. A downsampling approach was also applied to derive transfer functions from large-scale climatic general circulation models to local-scale systems (Gates, 1985; Cavazos, 1997; Wilby, 1997). Methods for upscaling have also been developed to extrapolate information across hierarchical levels. Such an approach was applied to evaluate the response of northern North-American forests to climatic variations (Pastor and Post, 1988), to extrapolate individual tree water use to stand water use (Hatton and Wu, 1995), and to examine factors controlling the absorption of photosynthetically active radiation (PAR) at the leaf, canopy, and landscape levels (Asner and Wessman, 1997). In a complementary study, Cullinan et al. (1997) compared four models, based on postulates of hierarchy theory, for scaling up variability in plant cover from field to remotely sensed data.

Recently, a more encompassing framework for scaling was proposed, called the hierarchical patch dynamics paradigm (HPDP) (Wu and Loucks, 1995; Wu and Levin, 1998), that integrates hierarchy theory and the theory of patch dynamics (Pickett and White, 1985). The central concept underlying the theory of patch dynamics is the patch, defined as a spatial unit differing from its surroundings in nature or appearance (Kotliar and Wiens, 1990). The patch is the fundamental structural and functional unit of a landscape, and is scale and context-dependent. While hierarchy theory focuses on the vertical structure of the landscape composed of a limited number of discrete hierarchical levels, patch dynamics theory explicitly deals with spatial heterogeneity and hierarchical interactions among system components in a horizontal way. When merging the two theories, ecological systems can be described as hierarchical systems of patches that differ in composition and spatial configuration at particular scales. The dynamics of ecological systems can be explained by the interactions of constituent patches present at contiguous hierarchical levels of organization in the landscape. As noted by Wu and Loucks (1995), the hierarchical patch dynamic paradigm represents a robust theoretical framework for studying how spatio-temporal heterogeneity, scale, and hierarchical organization influence the structure and dynamics of ecological systems. This framework was applied to model the patch disturbance in an annual grassland in California (Wu and Levin, 1994; 1997).

Substantial contributions to the scaling problem were also made by hydrologists and meteorologists. Two comprehensive reviews can be found in Blöschl and Sivapalan (1995), and Raupach and Finnigan (1995). These authors describe conceptual guidelines for scale translations
and a series of approaches developed to perform the linkages across scales in the particular contexts of catchment hydrology and boundary-layer meteorology. The most significant contributions to the scale problem achieved in these fields are through the use and improvement of distributed physical models at different spatial and temporal scales. These models enable the testing of hypothesis about the dominance of physical parameters at specific scales and their linkages across different scales through various upscaling and downscaling strategies.

**Conclusion**

When examining the evolution of scale contributions made in the social and natural sciences, it can be seen that ideas are converging towards the edification of a solid theoretical framework and operational methods to deal with the scale issue. The first common step towards the achievement of that goal was to recognize the *scale dependence* effect, which is the combined influence of the sampling unit size and the spatial extent on the information collected about the phenomenon of interest. The scale dependence effect is manifest in situations where spatial patterns and processes may be observable only at certain specific scales, and vary when observed at different scales, and where relationships between variables are affected by the scale of observation (Walsh, 1997). If ignored, the scale dependence might jeopardize the validity of any spatial analysis results.

In the social sciences, the most significant contribution was to formally identify the modifiable areal unit problem (MAUP), and to demonstrate the importance of its effects in statistical analysis and modelling. The effects of MAUP are not restricted to the social sciences; instead, they may be manifest in any study involving the use of spatial units upon which data are collected and analyzed. Fortunately, the MAUP is also becoming understood in the natural sciences. Recently, authors have pointed out its implications in landscape analysis (Jelinski and Wu, 1996). Others have demonstrated that remote sensing data represents a particular case of the MAUP, and that it can significantly affect the results of any analysis generated upon them (Marceau, 1992; Marceau et al., 1994; Arbia et al., 1996). In the natural sciences, the development of concepts such as *domain of scales* and *scale threshold* is crucial to the understanding of the hierarchical organization of the geographic world. Such concepts can also be applied in the social sciences to explain the strength of a relationship between specific variables at one scale and their disappearance at another, or the dominance of one variable at a specific scale only.

The second major step in the understanding of the scale issue in both the social and natural sciences is the development of quantitative methods to predict and control the MAUP effects, and to explain how entities, patterns, and processes are linked across scales. The advanced capabilities of computers, and the widespread use of technologies for spatial data acquisition and analysis, such as remote sensing and geographic information systems (GIS), provide unprecedented means to achieve significant progress in relation to the scale issue. Current remote sensing systems offer unique methods for detecting patterns at the surface of the earth, and for acquiring data about the underlying processes at a variety of spatial scales, ranging from centimeters to kilometers. A comprehensive review of the main contributions of remote sensing to the scale issue can be found in Marceau and Hay (1999). In a complementary fashion, GIS provide opportunities to create multiscale representations by incorporating and linking digital maps at different scales, and through the development of statistical and mathematical functions to deal with scale as a generic issue. GIS are particularly beneficial for analyzing the relationships between variables at different scales, and for assessing the impact of scale in modelling.
While there is still progress to be made, these methodological developments are certainly contributing to the emergence of a new paradigm: a science of scale (Meentemeyer and Box, 1987). In such a science, scale must be an explicitly stated variable in the analysis, and the following key topics should be formally addressed: the role of scale in the detection of patterns and processes, the scale impact on modelling, the identification of scale thresholds, and the derivation of scaling laws (Goodchild and Quattrochi, 1997).

Finally, the last step in the evolution of understanding the scale issue is to build a solid unified theoretical framework from which hypotheses can be derived and tested, and generalizations achieved. Recent studies addressing the MAUP (Hunt and Boots, 1996; Amrhein and Reynolds, 1996; Holt et al., 1996) demonstrate the possibility to explain scale effects and hold the prospect to derive theoretical foundations and operational methods to control them. At the same time, hierarchical patch dynamics derived from complex system theory is emerging as a strong framework to understand the hierarchical organization of ecological systems in which scale plays a fundamental role. The convergence of these ideas certainly holds a promising future for the next generation of scientists to come.

Acknowledgments

This work has been supported by grants awarded to D.J. Marceau by the Natural Science and Engineering Research Council of Canada (NSERC) and by the Fonds pour la Formation de Chercheurs et l’Aide à la Recherche (FCAR) of Quebec. I am very grateful to the anonymous reviewers of this paper, and I thank Geoffrey Hay for the stimulating discussions.

References


