Digital terrain analysis seeks to construct mathematical abstractions of the terrain surface (for example, Moore, Grayson, and Landson 1991, Florinsky 1998a) to delineate or stratify landscapes (for example, Hammond 1964, Dikau 1989, Burrough, Wilson, van Gaans, and Hanson 2001) and to examine or define the relationships between the terrain surface and various biophysical processes/patterns (for example, Moore, Gessler, Nielsen, and Peterson 1993, Franklin 1995, Beven 1997). The essential scientific value of these three tasks relies on three simple facts: (1) terrain poses significant control over other biophysical elements, (2) the former is much easier to measure than the latter; and (3) both tend to vary continuously over space in a correlated fashion (Burrough and McDonnell 1998). The computed terrain attributes often provide important, if not the only, clues indicating key biophysical patterns and processes, and sometimes serve as a spatial prediction tool directly (for example, Moore, Gessler, Nielsen, and Peterson 1993, Bell, Grigal, and Bates 2000). These roles of terrain analysis represent a bridge from the known to the unknown, and are often vital for resource inventory and environmental modeling, especially at topo- (that is, hillslopes of 50–200 m in length) to meso-scales (that is, watersheds of 10–100 km² in extent). They also point to the strong multi-disciplinary character of many terrain analysis applications.

Terrain analysis is nonetheless different from most scientific approaches to the study of the biophysical environment in that it is enabled by GIS and related computer technologies, and is supported more by digital terrain data (mostly gridded DEMs – digital elevation models) than by direct field observations or laboratory measurements of biophysical properties. Terrain analysis is quantitative, implying high precision in terms of outputs, but these results may simultaneously be plagued with uncertainties in terms of the relationship between terrain and biophysical attributes – implying the possibility of low accuracy. The common approaches employed in soil-landscape analyses (Park, McSweeney, and Lowery 2001) – statistical correlation (for example, Moore, Gessler, Nielsen, and Peterson 1993), classification of terrain attributes based on pre-defined criteria (for example, Zhu 1997), and statistical clustering of terrain indices (for example, Irvin, Ventura, and
Slater 1997, McBratney and Odeh 1997) – provide typical examples of how terrain analysis results tend to be used. They all deal primarily with the probability, instead of the certainty, that we can: (1) use knowledge of soil–landscape relationships to infer soil conditions from terrain properties, and (2) extrapolate the relationships to other places or interpolate them to other scales. This is fundamentally different from conventional sciences such as chemistry in which chemical reactions can both be reproduced and explained with certainty.

The relative ease with which terrain analysis can be performed points to numerous opportunities, but implies tremendous challenges because of these inherent uncertainties. In other words, terrain analysis is more a science dealing with uncertainty than with certainty. These uncertainties are often linked to issues such as terrain data quality (Adkins and Merry 1994, Bolstad and Stowe 1994, Hunter and Goodchild 1997, Krupnik 2000, Deng, Wilson, and Goodchild 2006), algorithm reliability (Skidmore 1989, Desmet and Govers 1996a, Florinsky 1998b, Quinn, Beven, Chevallier, and Planchnon 1991), spatial scale effects (Chang and Tsai 1991, Zhang and Montgomery 1994, Florinsky and Kuryakova 2000, Gertner, Wang, Fang, and Angerson 2002), objects with indeterminate boundaries (Burrough and Frank 1996), and ontological discrepancies regarding landform definitions (Hudson 1992, Zhu 1997, Burrough, Wilson, van Gaans, and Hanson 2001, Mark and Smith 2003). They signify the intrinsic complexity of terrain-environment relationships, as well as our relative lack of appreciation and understanding of these issues. The growth of new terrain data sources, terrain analysis programs, and terrain-based environmental models provide many new opportunities for biophysical study, although they should all be assessed carefully in terms of their scientific basis prior to widespread deployment. They may directly lead to an increase in precision, but do not necessarily imply a corresponding improvement in accuracy.

This chapter examines several essential aspects of terrain analysis based on the aforementioned general vision. Under the heading “Terrain Attributes – State of the Art” we describe terrain attributes as scale- and algorithm-dependent descriptions of the terrain surface and related biophysical processes. Three examples are used to identify the strengths and weaknesses of terrain-based environmental models and landscape stratifications in the second section, “Modeling and Synthesis.” The third section, “Enduring Challenges,” examines the effects of spatial scale and data quality on terrain analysis, and the final section highlights some of the developments that are likely to occur by the 2020s.

**Terrain Attributes – State of the Art**

All three of the terrain analysis tasks listed at the start of this chapter rely on calculated terrain attributes. A distinction is generally drawn between primary attributes that are computed directly from the DEM and composite attributes that involve combinations of primary attributes (Moore, Lewis, and Gallant 1993, Florinsky 1998a, Wilson and Gallant 2000a). Elevation is unique because its computation does not rely on other points; however, we often make assumptions about the character of the land surface – in terms of its continuity and smoothness – to estimate elevation in a DEM using sparse source data in practice (for example, Hutchinson 1989).
Florinsky (1998a) also distinguished local primary attributes that are calculated as a function of their surroundings and non-local primary attributes that require the analysis of a larger, non-local land surface area from a computational perspective. Wilson and Burrough (1999) later explained this distinction between local versus non-local terrain attributes in terms of the existence of local interactions between neighboring points and “action-at-distance” forces (see Figure 23.1 for details).

Most primary attributes are calculated from the geometric derivatives of the terrain surface using either a second-order finite difference scheme (for example, Skidmore 1989, Moore, Lewis, and Gallant 1993b, Florinsky 1998b) or a bivariate interpolation function \( z = f(x,y) \) that has been fitted to the DEM (Mitasova, Hofierka, Zlocha, and Iverson 1996). Typical examples of local primary attributes include slope, aspect, and plan and profile curvatures; non-local primary attributes include flow path length, proximity to nearest ridgeline, dispersal area, and upslope contributing area. More complete lists can be found in Moore, Grayson, and Ladson (1991), Moore, Lewis, and Gallant (1993b), Florinsky (1998a), and Gallant and Wilson (2000).

By definition, the local terrain shape – which is usually thought of as the continuous variation of elevation values over the terrain surface from point to point – has an enormous impact on local terrain attributes, but this role is influenced by data and computational factors. Florinsky (1998b) suggested that local attributes, such as slope gradient, aspect, and curvatures, are mathematical variables rather than real-world values. This statement may be extended to all local terrain attributes for two reasons. First, local terrain shape can have different mathematical descriptions, so that the calculated local attributes depend on algorithm selection. Second, the

![Fig. 23.1 Schematic diagram showing site-specific, local, and regional interaction as a function of time](From Wilson and Burrough 1999)
terrain shape portrayed by DEMs is a function of scale, combining the complexity of the terrain, scale or resolution of data, and spatial scale at which the terrain surface is observed. Thus it is possible to use the same local attribute to describe terrain shape at different scales (resolutions). The special feature of non-local primary attributes is that they rely on the terrain shape of a larger, non-neighbor area and need to be defined with reference to other, non-local points. Therefore, calculating non-local attributes is more difficult because it incurs additional efforts in constructing point-to-point connections over the landscape and involves more complex algorithms (for example, Desmet and Govers 1996a, Gallant and Wilson 2000).

Secondary or composite attributes account for the spatial variability of biophysical processes as a function of topographic effects (Moore, Grayson, and Ladson 1991). They are often used to quantify the role played by the terrain surface in redistributing water and sediments over the landscape and in modifying the amount of solar radiation received at various surface locations. Wilson and Gallant (2000b), for example, described three sets of composite attributes – topographic wetness, sediment transport capacity, and solar radiation indices – and some of the ways they have been deployed to interpret selected hydrologic, geomorphic, and ecological processes and patterns. The topographic wetness index (W or $W_T$) is probably the most popular of these composite indices and is calculated using one of the following equations depending on whether uniform soil transmissivity ($T$) under saturation is assumed or not:

\[
W = \ln \left( \frac{a}{\tan \beta} \right) \tag{23.1}
\]

or

\[
W_T = \ln \left( \frac{a}{T \tan \beta} \right) \tag{23.2}
\]

where $a$ (m$^2$ m$^{-1}$) is the specific catchment area and $\beta$ (°) is the slope gradient (Kirkby 1976, Beven and Kirkby 1979, Moore, Grayson, and Ladson 1991). Provided certain conditions are met (Beven and Kirkby 1979), the wetness index describes the pattern of depth to water table in a catchment and hence the pattern of hydrologic response. It has frequently been used as an index of position in the landscape and of accumulation of materials for predicting soil properties.

**Modeling and Synthesis**

Both primary and composite attributes are frequently used to provide input data for various environmental models or landscape delineations based on attribute distributions. The outputs of these applications take the form of identified terrain–environment relationships, spatio-temporal predictions of environmental properties, and the delineation of meaningful spatial units over the landscape. Three examples are utilized below to demonstrate how new knowledge may be garnered through these modeling and synthesis activities. They also illustrate three enduring issues in
TERRAIN ANALYSIS

terrain analysis: (1) the impact of the choice of spatial scale or landscape unit on model predictions; (2) the difficulties encountered representing spatial continuity; and (3) the problem of “equifinality.”

Soil erosion/deposition modeling

The large number and complexity of factors influencing soil erosion rates coupled with the relative paucity of data at fine scales (that is, 5–10 meter grid cells) have slowed soil erosion model development since the 1950s. The first proposals for combining soil erosion models and Geographic Information Systems (GIS) were published nearly two decades ago (for example, Ventura, Chrisman, Connors, Gurda, and Martin 1988, Warren, Diersing, Thompson, and Goran 1989) and may be contrasted with several of the more recent models that have been implemented in GIS environments from the outset (for example, Mitasova, Hofierka, Zlocha, and Iverson 1996, Mitas and Mitasova 1998).

The most popular, and in many ways most important, soil erosion model up to this point in time is the Universal Soil Loss Equation (USLE), an empirical equation derived from observations of more than 10,000 plot-years on farmlands (Wischmeier and Smith 1978; see Wilson and Lorang 1999 for a detailed review). It incorporates six factors – rainfall–runoff, soil erodibility, slope length, slope gradient, crop management, and conservation practices – and calculates the mean erosion rate (t ha\(^{-1}\) yr\(^{-1}\)) by comparing the conditions of the target slope (slope length, gradient, erodibility, management, etc.) with a standard soil-loss plot that is 22.13 m long and has a uniform width and slope gradient (9 percent).

This model is intrinsically limited to: (1) landscapes in which erosion is detachment limited; (2) planar slopes (except where the special rules for irregular slopes proposed by Foster and Wischmeier (1974), which divided irregular slopes into a series of planar slope facets, are implemented); and (3) those parts of the landscape that experience net erosion over the long term (this requirement will often exclude footslopes and valley bottoms in semi-arid and humid areas for example; see Wilson and Lorang 1999 for additional discussion of this limitation). Two reasons explain these limits. First, the model uses the entire slope as the basic spatial unit, and does not incorporate within-slope change in runoff direction and speed (that is, convergence, divergence, acceleration, and deceleration). These changes are a function of surface shape (that is, curvatures along or perpendicular to the steepest slope direction) and they are key to the successful prediction of within-slope variation of sediment transport processes (detachment, transport, deposition, and detainment). Second, slopes are conceptualized as isolated spatial units, so that the impact of input sediment flow and the possibility of net deposition are not taken into account.

The topographic or length-slope factor (LS) for the USLE, originally computed as a function of overall slope length and average slope gradient (Wischmeier and Smith 1978), is primarily responsible for the aforementioned limits of the USLE (Wilson and Lorang 1999). Wilson (1986) proposed a way around these limits that involved sampling slopes in watersheds and using topographic map information along with the irregular slope estimation method of Foster and Wischmeier (1974) to generate frequency distributions of \(LS\) for specific watersheds (that is, catchments). This method facilitated watershed-level comparisons (for example, Wilson and Ryan 1988, Wilson
1989) but was not able to characterize the erosion hazard at finer scales. Griffin, Beasley, Fletcher, and Foster (1988) later generalized the topographic or length-slope factor so that the USLE was able to estimate the soil erosion potential at specific places (that is, points) in the landscape (so long as the original model assumptions noted earlier held true). However, this approach greatly increased the time and effort needed to implement the USLE and, as a consequence, it attracted little attention prior to the widespread adoption and use of GIS for natural resource assessment.

It was therefore not surprising when Desmet and Govers (1996b) proposed a GIS-based method to calculate the topographic factor over a two-dimensional landscape that automated the approach of Griffin, Beasley, Fletcher, and Foster (1988), although with one important modification. They utilized the upslope contributing area in place of upslope flowpath length and then concluded that the original (that is to say, manual) method leads to an underestimation of the erosion risk because the effect of flow convergence is not taken into account.

Moore and Wilson (1992) had several years earlier proposed a dimensionless, unit stream power-based sediment transport capacity index $T$ to replace $LS$ in certain landscape conditions:

$$T = \left( \frac{a}{22.13} \right)^m \left( \frac{\sin \beta}{0.0896} \right)^n$$

where $a$ is the specific catchment area ($m^2 m^{-1}$) and $\beta (^\circ)$ is the slope gradient. Foster (1994) criticized this approach because it relied on different assumptions (most notably that the erosion rate is transport- rather than detachment-limited) and attempted to modify just one of several components in this empirical model (the USLE is fundamentally a series of nested regression models and changing one component may necessitate changes to one or more other components). It is clear that these same criticisms would apply to the approach of Desmet and Govers (1996b) given the similarities between the two methods. Moore and Wilson (1994) subsequently acknowledged these shortcomings and went on to show that their equation produced similar results to the original USLE for certain slopes (that is, planar slopes with lengths $<100$ m and gradients $<14^\circ$) despite the fact their approach relied on different assumptions to the original USLE.

Moore and Wilson (1992, 1994) also suggested calculating a second index:

$$\Delta T_{cj} = \Phi[a_m^o \sin \beta_j]^n - a_m^o \sin \beta_j]^n$$

where $\Delta T_{cj}$ is the change in $T$ along the flow direction over a grid cell, $\Phi$ is a constant, $a_m$ is the specific catchment area ($m^2 m^{-1}$), subscript $j$ signifies the outlet of cell $j$ and $j^-$ signifies the inlet to cell $j$, and $\beta (^\circ)$ is the slope gradient (as in Equation 23.3). They proposed using this equation to distinguish those parts of the landscape likely to experience net erosion ($\Delta T > 0$) from those parts likely to experience net deposition ($\Delta T \leq 0$), although this would clearly only work for landscapes in which soil erosion is transport limited.

Mitasova, Hofierka, Zlocha, and Iverson (1996) and Mitas and Mitasova (1998) later incorporated some of these same ideas in a soil erosion model that relied on the solution of bivariate first principles water and sediment flow equations. These
equations can be used to characterize the relationship between erosion/deposition rates and terrain curvatures on slopes with varying soil and cover properties. Their models, which incorporated detachment- as well as transport-limited conditions and both profile and tangential curvatures, provide a sound theoretical explanation for the results of field experiments reported by Sutherland (1991), Busacca, Cook, and Mulla (1993), Quine, Desmet, Govers, Vandaele, and Walling (1994), and Heimsath, Deitrich, Nishiizumi, and Finkel (1997). The highest erosion rates were observed on divergent shoulder elements and deposition on convergent footslope elements in the first pair of studies, whereas the maximum soil loss was observed from the slope convexities and maximum gain in both the slope concavities and the main thalwegs in the final two studies. These results illustrate how small variations in terrain shape and soil and land cover can have a dramatic impact on the location and rates of soil erosion and deposition.

This discussion of terrain analysis and soil erosion models would not be complete without some mention of the Water Erosion Prediction Project (WEPP) model (Flanagan and Nearing 1995). The tremendous progress towards physically-based erosion models achieved within this project since at least the mid 1990s means that the USLE in its various forms is best suited to preliminary assessments and/or situations where data is limited nowadays. The WEPP model can be implemented at various levels and can predict erosion and deposition. The WEPP watershed model (Ascough, Baffaut, Nearing, and Liu 1997, Baffaut, Nearing, Ascough, and Liu 1997), for example, is an extension of the WEPP hillslope model and can be used for estimating watershed erosion and sediment yield. However, the application of WEPP to watersheds requires that hillslopes be delineated and channels identified (Figure 23.2). Each hillslope, represented as a rectangle in WEPP, must be assigned a representative length, width, and slope profile (as illustrated in the third part of Figure 23.2). Cochrane and Flanagan (1999) noted that GIS analysis using DEMs provides a useful tool for parameterization of hillslopes, channels, and representative slope profiles for WEPP simulations and set out to describe and evaluate three methods for integrating GIS and WEPP to facilitate watershed level applications that are often of interest to resource managers and policy analysts. This integration is relatively straightforward but it cannot overcome the fact that the WEPP family of models is based on a one-dimensional sediment routing over planar hillslopes (Foster, Flanagan, Nearing, et al. 1995) and that in many instances this approach will only partially explain the impact of terrain shape and the spatial variability of soil and land cover at the watershed scale (Mitas and Mitasova 1998).

Soil mapping and landform classification

Conventional soil–landscape analysis and soil mapping (Hudson 1992) are based on a crisp conceptual model that allows a data point to belong to only one class, and thus a sample location (such as a grid point) to fall in only one map unit. Soil variation only occurs across boundaries in geographic space and between the central concepts of prescribed soil classes in attribute space (Zhu 1997). From a utilitarian point of view, this leads to loss of soil information because: (1) minor deviations of local soil from the prescribed soil central concept may be known by local soil experts but cannot be included in a crisp soil map; and (2) soil bodies
smaller than the minimum map unit either will be ignored or combined into another soil class (Zhu, Hudson, Burt, Lubich, and Simonson 2001).

The problems with this soil–landscape model follow from the fact that natural soils often exist as spatial continua and natural soil boundaries only exist under special circumstances (Burrough 1993, Burrough and Frank 1996, McBratney and Odeh 1997, Zhu 1997). It is also debatable whether soil should always be viewed as spatial objects or as the surrogate of aggregated soil properties, because there is “no agreement on what a basic or fundamental unit of soil is” (Arnold 1983). Hence, there could be endless combinations of dynamic soil conditions (properties), although individual soil properties are usually the major concern in applications such as soil and water conservation, non-point source pollution control, and precision agriculture (for example, Burrough 1993, Indorante, et al. 1996, Berry, Delgado, Khosla, and Pierce 2003). These two “discontinuity” and “object” problems in conventional soil mapping often occur simultaneously and help to explain the limitations of knowledge-based interpretations of soil–landscape relationships (Zhu, Hudson, Burt, Lubich, and Simonson 2001), which can be more precisely captured as correlations between soil properties and landscape attributes (for example, Burrough 1993, Moore, Gessler, Nielsen, and Peterson 1993, Florinsky and Kuryakova 2000).

Simonson (2001) provided a solution for both sets of above-mentioned problems using a soil similarity model. This approach describes local soil as a similarity vector to prescribed soil classes (central concepts) through three steps: (1) identification of a set of central concepts according to existing soil classifications or expert knowledge; (2) definition of the linkages between these central concepts (or soil classes) and higher resolution landscape properties; and (3) calculation of the similarity (for example, values from 0 to 1) of each data point to each central concept through a comparison of their landscape properties. The soil properties of a point can then be derived by combining its similarity vector with the soil properties of the central concepts (Zhu, Band, Vertessy, and Dutton 1997, Zhu, Hudson, Burt, Lubich, and Simonson 2001). Terrain analysis offers several useful inputs for describing environmental conditions and constructing the soil–landscape models in this approach (Zhu, Hudson, Burt, Lubich, and Simonson 2001). The net effect of using terrain attributes is to incorporate continuous spatial variations of the biophysical environment into the output fuzzy soil classes at a higher resolution than conventional soil maps.

Soil mapping in the US, nevertheless, is a special case of natural resource inventory because it has a long history and has involved large human investments to provide rich resources for the identification of central soil concepts. When sufficient mapping resources and/or expert knowledge do not exist to support the central concepts (for example, Franklin 1995) the fuzzy $k$-means landform classification method will usually provide a better solution for the continuous delineation of the biophysical environment. This approach uses terrain attributes as input data to define the most representative clusters of the data following an iterative, unsupervised clustering procedure that can differentiate the data to a maximum extent (McBratney and Odeh 1997, Irvin, Ventura, and Slater 1997, Burrough and McDonnell 1998, Burrough, Wilson, van Gaans, and Hanson 2001). Class centers so defined are similar to central concepts in soil classification. The membership (similarity) of each data point to each class center is defined as the attribute distance between the point and the class center in the attribute space, which is calculated using a selected distance function (Irvin, Ventura, and Slater 1997). The biophysical meanings of fuzzy $k$-means landform class centers must be post-interpreted, rather than pre-defined, based on the terrain attributes that were used. Furthermore, some additional work will usually be required to select the terrain attributes (and weights) that will be used to identify the specific biophysical pattern(s) of interest.

Terrain analysis might also be used to create explicit environmental stratifications for survey design and to provide quantitative spatial predictions of individual soil properties. McKenzie, Gessler, Ryan, and O’Connell (2000), for example, described a two-step stratified random sampling strategy for the Bago-Maragle study area in New South Wales, Australia that combined geologic information (that is, published geologic map units supported by airborne gamma radiometric remote sensing) with the Prescott Index (which is a function of mean monthly precipitation and potential evapotranspiration), and the topographic wetness index. A field survey of 144 selected sample sites was then conducted and the data used for quantitative spatial prediction of key land qualities, including soil erodibility, nutrient status, and the soil–water regime (McKenzie and Ryan 1999). Terrain analysis variables can be used in conjunction with these other variables (climate geology, remote sensing, etc.) to extend point observations of individual soil properties using statistical models so long as the
explanatory variables are easier to obtain than soil variables. Hence, Gessler (1996) used a regression tree approach to predict solum depth in the above-mentioned study area, and Bell, Grigal, and Bates (2000) used a series of linear and exponential statistical models to predict soil organic carbon in the Cedar Creek Natural History Area in Minnesota. Interested readers can learn more about the challenges and subtleties of the statistical methods employed in these types of modeling applications in McKenzie and Austin (1993), Gessler, Moore, McKenzie, and Ryan (1995), Gessler (1996), and McKenzie and Ryan (1999).

TOPMODEL

The original TOPMODEL introduced by Beven and Kirkby (1979) estimates overland runoff by integrating a spatially variable contributing area model with a simple, lumped soil water response (storage) model (Kirkby 1976). It defines the saturated area on the landscape \( A_c \) as the area where:

\[
W > S_T/m - S_d/m + \lambda
\]  

(23.5)

where \( W \) is the topographic wetness index (Equation 23.1), \( S_T \) is the local maximum water storage (sum of surface interception, depression, near surface infiltration, and subsurface storage), and \( m \) and \( \lambda \) are constants. Given the assumption of a time-independent steady state rainfall rate, overland flow \( (q_{ol}) \) is estimated as:

\[
q_{ol} = iA_c
\]  

(23.6)

where \( i \) is an instantaneous rainfall intensity and \( A_c \) is the saturated area calculated with Equation 23.5.

\( W \) and \( A_c \) were the only spatial variables used in the original model. All of the other variables – such as flow velocity, interception, infiltration, subsurface storage, and channel routing – were treated as lumped parameters and had to be measured or estimated. Beven (1997) discussed the equifinality problem in his critique of TOPMODEL and noted that many different sets of parameter values can simulate observed data (that is, the hydrograph) almost equally well in terms of some quantitative goodness-to-fit measure. Beven and Binley (1992) attributed the difficulty of finding a global optimum parameter set to the complexity of the multi-dimensional attribute space involved in hydrological modeling that is a function of the use of threshold parameters, intercorrelation between parameters, autocorrelation and heteroscedascity in the residuals, and inclusion of insensitive parameters. Beven (1993) and Savenije (2001) also linked this equifinality problem to scales of hydrological processes and “laws,” as well as to the effects of aggregation and averaging (lumping) across scales during modeling.

Beven and Binley (1992), Beven (1997) and Beven and Freer (2001) have advocated using the GLUE (generalized likelihood uncertainty estimation) method to manage the equifinality problem. This method is based on Monte Carlo simulations in which the predictions of each parameter set realization are given a likelihood weighting according to how well that model fits the observed data (Beven 1997). The likelihood weights of the parameter sets can be updated when more data (that is, observations)
become available. With the use of the GLUE method, the uncertainty of models can be defined more precisely and TOPMODEL applications can potentially become an iterative process of model selection, rejection, and optimization. Approaches similar to GLUE may see potential use in terrain analysis because the selection of terrain attributes and terrain analysis scales is currently based on data availability and it is likely that different combinations of scales (resolutions) and attributes (as well as their weights) may achieve the same terrain analysis goal. The advent of finer resolution DEMs may improve the definition of uncertainty in models using these types of uncertainty estimation methods.

**Enduring Challenges**

*Data quality*

The key role of terrain shape in terrain analysis indicates that the distribution, instead of the magnitude, of elevation errors should be the primary focus of DEM quality assessments (Hunter and Goodchild 1997; Burrough and McDonnell 1998, pp. 244–7; Heuvelink 1998; see Hutchinson, Chapter 8 of this volume, for additional discussion of the role of shape in DEM quality assessments). However, most DEM producers only provide aggregated error indicators (for example, RMSE or root mean squared error) that are calculated based on more accurate elevations of a few control points to report the mean magnitude of elevation errors over one tile (or map area) of the DEM. Some DEM users have utilized mutually independent sample points to link point elevation errors to the accuracy of calculated point terrain attributes (for example, Isaacson and Ripple 1990, Adkins and Merry 1994, Bolstad and Stowe 1994). The reliance on isolated points in both instances precludes the assessment of local distribution of error and the impact of these errors on terrain shape (see Wise 2000 for an extensive review of some of the key issues here). Table 23.1 shows that terrain shape may be severely distorted by local errors in DEMs that have a small RMSE and vice versa, and that the local standard deviation of errors calculated with a moving window is a better statistic to describe local distortions of terrain shape, given the availability of an error surface.

The focus on point or average errors dominated error assessments provided with DEMs of coarser resolutions (for example, 30 m or 3 arc second US Geological Survey DEMs), which are aggregated representations of toposcale surface conditions (that is, they are capable of identifying and characterizing 50–150 m long slopes). This focus presents even larger problems when applied to new DEM data sources of higher resolutions, because 10 m or finer DEMs can delineate within-slope variations much more precisely and accurately than previous 30 m or 3 arc second DEMs. For example, the subtle variations of landform structures on a slope (topographic hollows or local convexities) may be contained, or hidden, in the width of one 30 × 30 m DEM cell, but be manifest on a 10 m or finer resolution DEM. In the meantime, a 0.2 m elevation error of one point on a 10 percent uniform slope may cause the estimate of slope gradient to change from 10 percent to 10.7 percent on a 30 m DEM, but from 10 percent to 12 percent on a 10 m DEM, and from 10 percent to 30 percent on a 1 m DEM. These two examples indicate that the emergence
of 10 m, 5 m, or even 1 m DEMs implies more than increasing detail in terrain surface modeling. Instead, these higher resolutions, compared to previous 30 m or 3 arc second ones, may represent a major shift of spatial scales incorporated in the DEMs, thus a more imminent need to evaluate the effects of DEM error distributions. In other words, we may be simultaneously facing more precise (in terms of horizontal resolution) and more erroneous (in terms of terrain shape) terrain surface depictions, given the same magnitude of RMSE of elevations reported for these high resolution DEMs as for the previous coarser resolution ones.

New methods are needed to circumvent the unavailability of a true error surface. Hunter and Goodchild (1997) proposed that, instead of dealing with error itself, we could define data or model uncertainty, or the extent to which we are uncon- fident in the obtained results. Specifically, they suggested that a worst-case scenario could be identified by introducing into DEMs a set of error fields that incorporate different degrees of spatial autocorrelation. All possible uncertainties caused by DEM errors could only occur within a range defined by this worst case scenario. This approach thus directs the research focus from the DEM errors themselves to the possible effects of DEM errors. A similar approach was adopted by Ehlschlaeger, Shortridge, and Goodchild (1997), who added a series of error surfaces with various autocorrelation and disturbance variables to an interpolation process and generated animated visualizations of data uncertainty with 250 realizations of interpolated 30 m DEMs (from a 3 arc second source DEM). Deng, Wilson, and Goodchild (n.d.) argue that it is also necessary to adopt a spatially explicit view to define the differences between various DEM sources and suggest that a DEM difference surface calculated from two DEMs can be used to develop spatial tools to estimate either the DEM errors themselves or the effects of terrain shape distortion on calculated terrain attributes.

Table 23.1 DEM errors, shape representation, and appropriate/inappropriate statistics (adapted from Deng, Wilson, and Goodchild n.d.)

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution of point elevation errors (imagined units)</td>
<td>2 2 2 2</td>
<td>8 8 8 8</td>
<td>2 0 4 0</td>
<td>8 0 16 0</td>
</tr>
<tr>
<td>Local mean error</td>
<td>2</td>
<td>8</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Local RMSE based on RMSE</td>
<td>2</td>
<td>8</td>
<td>√(20/3)</td>
<td>√(320/3)</td>
</tr>
<tr>
<td>Error description</td>
<td>low</td>
<td>high</td>
<td>low</td>
<td>high</td>
</tr>
<tr>
<td>Local standard deviation of errors</td>
<td>0</td>
<td>2/√3</td>
<td>8/√3</td>
<td></td>
</tr>
<tr>
<td>Distortion of terrain shape</td>
<td>none</td>
<td>none</td>
<td>low</td>
<td>high</td>
</tr>
</tbody>
</table>

Note: The word “local” refers to the 3 × 3 window shown in the four scenarios. It also implies that a 3 × 3 moving window can be applied to the entire area of interest to generate distributed error statistics.
Spatial scale

Moore, Lewis, and Gallant (1993) identified four scale-related issues in terrain analysis – basic element size, choice of attribute algorithms, merging of data sources, and scale differences between model and data – that still resonate today. Indeed, two additional issues should be added to this list given the recent emergence of numerous high-resolution DEMs; the need to: (1) define geomorphic units at various scales; and (2) calculate terrain attributes at appropriate scales with high-resolution data. The DEM spatial resolution was linked to issues of data quality and modeling uncertainty in the previous section – a fact that has been long observed (for example, Chang and Tsai 1991, Wolock and Price 1994, Zhang and Montgomery 1994, Mitasova, Hofierka, Zlocha, and Iverson 1996, Bian 1997, Wilson, Spangrud, Nielsen, Jacobsen, and Tyler 1998, Hutchinson and Gallant 2000, Gertner, Wang, Fang, and Angerson 2002) but has increased in importance in recent years. All these perspectives indicate the need to interpret the spatial scale in terrain analysis as an independent dimension that is related to all terrain analysis practices in either an explicit or implicit manner.

The basic (spatial) element size has attracted the most attention in the study or treatment of scale-dependencies of terrain analysis. For example, the topographic analysis for the first applications of TOPMODEL were based on a set of uniform areal elements that were approximately $5,500 \text{ m}^2$ in extent and delineated by dividing the basin according to flow lines, contour lines, and steepest slope lines (Beven and Kirkby 1979). Wood, Sivapalan, Beven, and Band (1988) adopted a similar approach based on the concept of representative elementary areas to define the scale effects of hydrological modeling. The consideration of landscape features (for example, flow lines, uniform slopes, etc.) in these applications produced irregular subdivisions that may provide a higher “actual” resolution than regular grid cells of the same size (Florinsky 1998a).

Most of these studies have examined the effects of the selected DEM resolution on calculated terrain attributes and modeling results given the widespread use of gridded DEMs. Isaacson and Ripple (1990), for example, observed very low correspondence between grid point slope gradient and aspect values calculated from 30 m and 3 arc second (roughly $65 \times 92 \text{ m}$) US Geological Survey DEMs respectively. Chang and Tsai (1991) concluded that the accuracy of the same two attributes decreased with the increase of DEM cell size from 8 m to 80 m. Zhang and Montgomery (1994), Mitasova, Hofierka, Zlocha, and Iverson (1996), and Florinsky and Kuryakova (2000) identified threshold DEM resolutions for the modeling of soil moisture, overland flow, and erosion processes. Florinsky and Kuryakova (2000) interpreted the regular grid cell size, or DEM resolution, in terms of its adequacy for the description of specific landscape properties. Various statistical measures (for example, means, standard deviations, terrain–environment correlation coefficients, fractal dimensions, etc.) have been used in these types of studies to characterize the effects of spatial scale on computed terrain attributes (for example, Florinsky and Kuryakova 2000; see Moore, Lewis, and Gallant 1993 for a comprehensive review).

Figure 23.3 shows that spatially aggregated statistical analysis may not sufficiently capture the impact of DEM resolution on calculated terrain attributes. Hence, the effect of resolution variation varies from place to place and a simple assessment
of attribute value change in magnitude may hide the fact that \( \beta_1, \beta_2, \) and \( \beta_3 \) in Figure 23.3 have different topographic, as well as biophysical, meanings. A more dramatic impact could be reasonably expected with greater change of spatial resolutions and/or in more complex terrain. Therefore, a spatially explicit approach that incorporates more complete interpretations of terrain attributes (for example, combining slope gradient with aspect) may need to be developed to account for the scale effects of terrain analysis.

Several other scale issues warrant further investigation as well. One is the potential problem of using a single-sized neighborhood window to estimate terrain characteristics. Multi-scale terrain analysis – the use of expanding neighborhood windows to calculate and compare terrain attributes – can potentially identify threshold window sizes across which the attribute values change abruptly to help clarify the meaning of different attributes and delineate natural landform boundaries. Gallant and Dowling (2003) demonstrate a method that combines terrain attributes at different scales into a single multi-scale attribute to represent geomorphic objects (valley bottoms) that occur at a range of scales.

Other key questions regarding spatial scales in terrain analysis that also require answers include:

- How do the model and attribute scale interact with one another?
- How can the various spatial resolutions of different attributes be incorporated into the same model according to their different process scales (or scales at work, see Bian 1997)?
- How do different terrain analysis algorithms behave with the change of scales?
• What modeling effects should be expected (that is to say anticipated) when combining data sources that incorporate different spatial scales?

SUMMARY

We briefly reviewed several key characteristics of terrain analysis and discussed several emerging perspectives, including the role of fuzzy logic, equifinality, shape-based data quality evaluations, and multi-scale terrain analysis. Several soil erosion models were reviewed to demonstrate the importance and great difficulties that are encountered delineating landscape units in these types of modeling applications. The traditional approach to soil–landscape analysis was described to portray how digital terrain analysis, together with new methodologies such as fuzzy logic, can improve soil classification and support a shift from crisp to continuous paradigms in various environmental modeling domains. The guiding principles of TOPMODEL were briefly discussed and used as an example to explain the equifinality problem and its potential significance to terrain analysis. The limitations encountered in managing spatial scale and data quality problems were identified as enduring challenges to terrain analysis and our ability to use computed terrain attributes to describe environmental patterns and processes of interest.

REFERENCES


