discussed shortly. *Conflict* results when, due to generalization, an inconsistency between or among features occurs. For instance, if generalization of a coastline eliminates a bay with a city located on it, either the city or the coastline must be moved to ensure that the urban area remains on the coast. Such spatial conflicts are difficult both to detect and to correct. The condition of *complication* is dependent on the specific spatial configurations that exist in a defined space. An example is a digital line that changes in complexity from one part to the next, such as a coastline (like Maine's) that progresses from very smooth to very crenulated. In this context, Barbara Buttenfield (1991) demonstrated the use of line-geometry-based structure signatures as a means of controlling the tolerance values, based on complexity, in the generalization process. Later, we provide details on other techniques for detecting changes in linear complexity.

One issue important to determining when to generalize is measuring the amount of generalization that either has been done or is planned. Two basic types of measures can be identified: procedural and quality assessment. *Procedural measures* are those needed to invoke and control the process of generalization. Such measures might include those that: (1) select a simplification algorithm, given a certain feature class; (2) modify a tolerance value along a feature as the complexity changes; (3) assess the density of a set of polygons being considered for agglomeration; (4) determine whether a feature should undergo a type change (e.g., area to point) due to scale modification; and (5) compute the curvature of a line segment to invoke a smoothing operation. *Quality assessment measures* evaluate both individual operations, such as the effect of simplification, and the overall quality of the generalization (e.g., poor, average, excellent). Several studies have discussed mathematical and geometric measures, including those by Buttenfield (1991), McMaster (1986; 1987), and Plazanet (1995).

6.4 THE FUNDAMENTAL OPERATIONS OF GENERALIZATION

6.4.1 A Framework for the Fundamental Operations

In the McMaster and Shea model discussed earlier, the third major component involves the fundamental operations, or how to generalize. Most of the research in generalization assumes that the process can be broken down into a series of logical operations that can be classified according to the type of geometry of the feature. For instance, a simplification operation is designed for linear features, whereas an amalgamation operator works on areal features. Table 6.1 provides a framework for the operations of generalization, dividing the process into those activities needed for raster- and vector-mode

TABLE 6.1 A framework for generalization operations. (After McMaster and Monmonier 1989, and McMaster 1989b.)

Raster-mode generalization	Vector-mode generalization
Structural generalization	Point feature generalization
Simple structural reduction	Aggregation
Resampling	Displacement
Numerical generalization	Line feature generalization
Low-pass filters	Simplification
High-pass filters	Smoothing
Compass gradient masks	Displacement
Vegetation indices	Merging
	Enhancement
Numerical categorization	Areal feature generalization
Minimum-distance to means	Amalgamation
Parallelopiped	Collapse
Maximum-likelihood classification	Displacement
Categorical generalization	Volume feature generalization
Merging (of categories)	Smoothing
Aggregation (of cells)	Enhancement
Nonweighted	Simplification
Category-weighted	Holistic generalization
Neighborhood-weighted	Refinement
Attribute change	

processing. Geographical features are normally represented in either a "vector" or a "raster" format inside of a computer. The vector representation uses x-y coordinate pairs to represent point features such as a house (a single x-y coordinate pair), a line feature such as a river (a string of connected x-y coordinate pairs), or an areal feature such as a park boundary (a string of x-y coordinate pairs in which the first pair matches the last pair). The raster approach uses a matrix of cells of a given resolution (e.g., 30 meters) to represent features. Many standard GIS books (e.g., Lo and Yeung 2007) describe these two data structures in more detail. Vector-based operators require more complicated strategies because they operate on strings of x-y coordinate pairs and require complex searching strategies. The next section will provide a more detailed discussion of individual vector-based operations (raster-based operations are not covered in this text). Figure 6.4 provides graphic depictions of some key operations.

6.4.2 Vector-Based Operations

Simplification

Simplification is the most commonly used generalization operator. The concept is relatively straightforward, because at its most basic level, it involves a "weeding" of unnecessary coordinate data. The goal is to retain as much of the geometry of the feature as possible, while eliminating the maximum number of coordinates. Below, we provide more detail on the simplification process.

Smoothing

Although often assumed to be identical to simplification, **smoothing** is actually a much different process. The

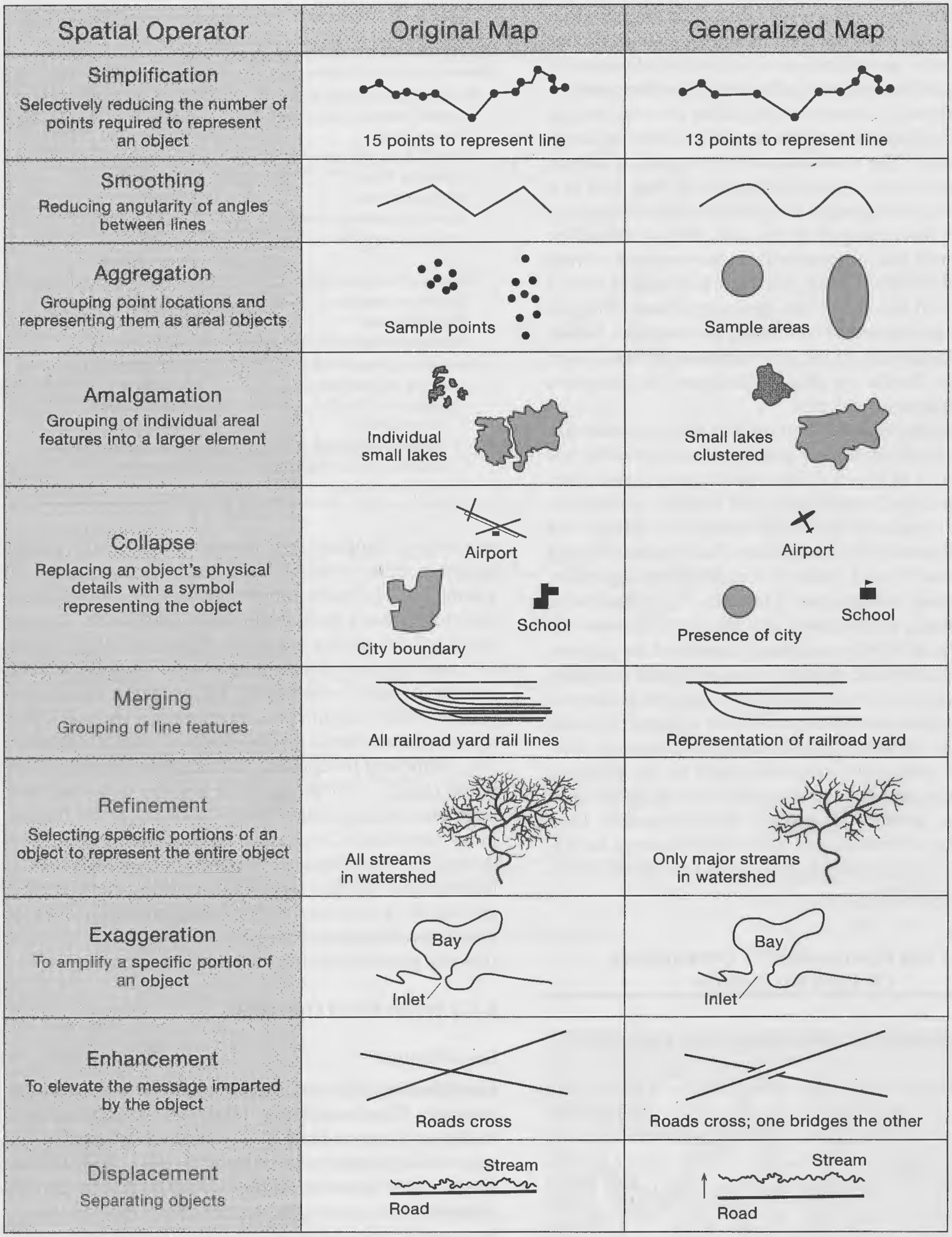

FIGURE 6.4 Fundamental operations of generalization. (Courtesy of Philippe Thibault.)

smoothing operation shifts the position of points to improve the appearance of the feature (Figure 6.4). Smoothing algorithms relocate points in an attempt to plane away small perturbations and capture only the most significant trends of the line (McMaster and Shea 1992). As with simplification, there are many approaches for the process—a simple classification is provided in Table 6.2. Research has shown that a careful integration of simplification and smoothing routines can produce a simplified, yet aesthetically acceptable, result (McMaster 1989a).

Aggregation

As depicted in Figure 6.4, **aggregation** involves merging multiple point features, such as a cluster of buildings. This process involves grouping point locations and representing them as areal units. The critical problems in this operation are determining the density of points needed to identify a cluster to be aggregated and specifying the boundary around the resulting cluster. The most common approach is to triangulate the points (i.e., create triangles among neighboring points) and determine the density of the triangles (e.g., a grouping of smaller triangles might represent a cluster for aggregation) (Jones et al. 1995).

Amalgamation

Amalgamation is the process of fusing together nearby polygons, a process needed for both noncontinuous and continuous areal data (Figure 6.4). A noncontinuous example is a series of small islands in close proximity that have size and detail that cannot be depicted at the smaller scale. A continuous example is census tract data, where several tracts with similar statistical attributes can be joined together.

Collapse

The **collapse** operation involves the conversion of geometry. For instance, it might be that a complex urban area is collapsed to a point due to scale change, and is resymbolized with a geometric form such as a circle. A complex set of buildings might be replaced by a simple rectangle—which might also involve amalgamation.

Merging

Merging involves fusing together groups of linear features, such as parallel railway lines or edges of a river or stream (Figure 6.4). Merging can be viewed as a form of collapse in which an areal feature is converted to a line. A simple solution is to average the two or multiple sides of a feature, and to use this average to calculate the new feature's position.

Refinement

Refinement is another form of resymbolization that is much like collapse (Figure 6.4). However, refinement is an operation that involves reducing a multiple set of features such as roads, buildings, and other types of urban structures to a simplified representation, rather than a conversion of geometry. The key to refinement is that complex geometries are resymbolized to a simpler form, thus creating a "typification" of the objects. The example

TABLE 6.2 A classification of algorithms used to smooth cartographic features. (After McMaster and Shea 1992, *Generalization in Digital Cartography*, pp. 86–87, copyright Association of American Geographers.)

Category 1:	Weighted averaging Calculates an average value based on the positions of existing points and neighboring points, with only the endpoints remaining the same; maintains the same number of points as the original line; algorithms can be easily adapted to different smoothing conditions by adjusting tolerance values (e.g., the number of points used in smoothing); all algorithms use local or extended processors.	
	Examples:	Three-point moving average Five-point moving average Other moving average methods Distance-weighted averaging Slide averaging
Category 2:	Epsilon filtering Algorithm uses certain geometrical relationships between the points and a user-defined tolerance to smooth the cartographic line; endpoints are retained, but the absolute number of points generated for the smoothed line is algorithm-dependent; approaches are local, extended local, and global.	
	Examples:	Epsilon filtering Brophy algorithm
Category 3:	Mathematical approximation Develop a mathematical function or series of mathematical functions that describe the geometrical nature of the line; the number of points on the smoothed line is variable and is controlled by the user; retention of the endpoints and of the points on the original line is dependent on the choice of algorithms and tolerances; function parameters can be stored and used later to regenerate the line at the required point density; approaches are local, extended local, and global.	
	Examples:	Local processing: cubic spline Extended local processing: b-spline Global processing: bezier curve

of refinement shown in Figure 6.4 is a selection of part of a stream network that depicts the "essence" of the distribution in a simplified form.

Exaggeration

Exaggeration is one of the more commonly applied generalization operations. Often it is necessary to amplify a specific part of an object to maintain clarity in scale reduction. The example in Figure 6.4 depicts the exaggeration of the mouth of a bay, which would close under scale reduction.

Enhancement

Enhancement is a symbolization change that emphasizes the importance of a particular object. For instance, the delineation of a bridge under an existing road is often portrayed as a series of cased lines, which assists in emphasizing that feature over another.

Displacement

Displacement is perhaps the most difficult of the generalization operations because it requires complex measurement (Figure 6.4). The problem can be illustrated with a series of cultural features in close proximity to a complex coastline. Assume, for example, that a highway and a railroad follow a coastline in close proximity, with a series of smaller islands offshore. In the process of scale reduction, all features would tend to coalesce. The operation of displacement would pull the features apart to prevent this coalescence. What is critical in the displacement operation is the calculation of a displacement hierarchy because one feature will likely have to be shifted away from another (Nickerson and Freeman 1986; Monmonier and McMaster 1990). A description of the mathematics involved in displacement can be found in McMaster and Shea (1992).

The Simplification Process

Most simplification routines utilize complex geometrical criteria (distance and angular measurements) to select **critical points,** those points that are significant in defining the structure of a linear or areal feature. A general classification of simplification methods consists of five approaches: independent point routines, local processing routines, constrained extended local processing routines, unconstrained extended local processing routines, and global methods. **Independent point routines** select coordinates based on their positions along a line, nothing more. For instance, a typical *n*th point routine might select every third point to quickly weed out unnecessary coordinate data. Although computationally efficient, these algorithms are crude in that they do not account for the

TABLE 6.3 A classification of algorithms used to simplify cartographic features. (After McMaster and Shea 1992, *Generalization in Digital Cartography,* p. 73, copyright Association of American Geographers.)

Category 1:	Independent point algorithms Do not account for the mathematical relationships with the neighboring pairs; operate independently of topology.	
	Examples:	*n*th point routine Random selection of points
Category 2:	Local processing routines Utilize the characteristics of the immediate neighboring points in determining significance.	
	Examples:	Distance between points Angular change between points Jenks's algorithm (distance and angular change)
Category 3:	Constrained extended local processing routines Search continues beyond immediate coordinate neighbors and evaluates sections of a line. Extent of search depends on distance, angular, or number of points criterion.	
	Examples:	Lang algorithm Opheim algorithm Johannsen algorithm Deveau algorithm Roberge algorithm Visvalingam algorithm
Category 4:	Unconstrained extended local processing routines Search continues beyond immediate coordinate neighbors and evaluates sections of a line. Extent of the search is constrained by the geomorphological complexity of the line, not by algorithmic criterion.	
	Example:	Reumann–Witkam algorithm
Category 5:	Global routines Considers the entire line, or specified line segment; iteratively selects critical points.	