Chapter 2
Geospatial Data for Hydrology

Data Sources and Structure

Abstract A key characteristic of distributed modeling is the spatially variable representation of the watershed in terms of topography, vegetative, or land use/cover, soils and impervious areas and the derivative model parameters that govern the hydrologic processes of infiltration, evapotranspiration, and runoff. Geospatial data exist that can be harnessed for model setup. Digital representation of topography, soils, land use/cover, and precipitation may be accomplished using widely available or special purpose GIS data sets. Each GIS data source has a characteristic data structure, which has implications for the hydrologic model. Major data structures are *raster* and *vector*. Raster data structures are characteristic of remotely sensed data with a single value representing a grid cell. Points, polygons, and lines are referred to generally as *vector* data. Multiple attributes can be associated with a point, line, area, or grid cell. Some data sources capture characteristics of the data in terms of measurement scale or sample volume. A rain gauge essentially measures rainfall at a point, whereas radar, satellites and other remote sensing techniques typically map the spatial variability over large geographic areas at resolutions ranging from meters to kilometers. Source data structures can have important consequences on the derived parameter and, therefore, model performance. Even after considerable processing, hydrologic parameters can continue to have some vestige of the original data structure, which is termed an *artifact*. This chapter addresses geospatial data structure, projection, scale, dimensionality, and source data for hydrologic applications.

2.1 Introduction

When setting up a distributed hydrologic model, a variety of data structures may be employed. Topography, for example, may be represented by a series of point elevation measurements or may exist as contour lines, grids, or triangular facets composing a triangular irregular network (TIN). Rainfall may be represented by a point when taken from rain gauges but exist as grids when measured by radar. Infiltration rates derived from soil maps are generalized over the polygon describing

B.E. Vieux, *Distributed Hydrologic Modeling Using GIS*,
Water Science and Technology Library 74, DOI 10.1007/978-94-024-0930-7_2

the soil mapping unit. Even though there exists variability within such mapping units, its properties are considered constant within the polygon enclosing a single soil type. Resulting infiltration modeled using this data source will likely show artifacts of the original soil map polygons such as abrupt changes at the boundary of adjacent soil mapping units. Land use/cover may be used to develop evapotranspiration rates or estimates of hydraulic roughness from polygonal areas or from a raster array of remotely sensed surrogate measures. In any case, the spatial variability of the parameters may be affected by the data structure of the source.

These examples illustrate two important points. First, in most cases a data source may be either a direct measure of the physical characteristic or an indirect (surrogate) measure requiring conversion or interpretation. Second, because of how the data are measured, each source has a characteristic structure including spatial and temporal dimensions as well as geometric character (points, lines, polygons, rasters, or polar arrays of radar data). Because the model may not be expecting data in one form or another, transformation is often necessary from one data structure to another. This necessity often arises because hydrologic processes/parameters are not directly observable at the scale expected by the model, or because the spatial or temporal scale of the measured parameter differs from that of the model. This issue can require transformations from one projection to another, from one structure to another (e.g., contours to TINs), interpolation of point values, or surface generation. Because the GIS data form the basis for numerical algorithms, hydrologic modeling requires more complex GIS analysis than simple geographical modeling using maps.

Typically, distributed hydrologic modeling divides a watershed or region into computational elements. Given that the numerical algorithms used to solve conservation of mass and momentum equations in hydrology may divide the domain into discrete elements, a parameter may be assumed constant within the computational element. At the sub-grid scale, parameter variation is present and may be represented in the model as statistical distribution known as sub-grid parameterization. Hydraulic roughness may be measured at a point by relating flow depth and velocity. In a distributed model, the roughness is assigned to a grid cell based on the dominant land use/cover classification. In most applications, the model computational element will not conform exactly to the measured parameter. Conformance of the data structure in the GIS map to the spatial pattern or scale of the process is a basic issue in GIS analysis for hydrology. Transformation from one data structure to another must be dealt with for effective use of GIS data in hydrology. The components of data structure are treated below.

2.2 Map Scale and Spatial Detail

A map of hydrography can be shown at any scale within a GIS. Once data are digitized and represented electronically in a GIS, resolution finer than that at which it was compiled is lost. A small-scale map is one in which features appear small, have few details, and cover large areas. An example of a small-scale map is one with a scale of 1:1,000,000. Conversely, large-scale maps have features that appear large and cover small areas. An example of a large-scale map is one with a scale of 1:2,000. A map compiled at 1:1,000,000 can easily be displayed in a GIS at a 1:2,000 scale, giving the false impression that the map contains more information than it really does.

Because GIS provides the ability to easily display data at any scale, we must distinguish between the compilation or native scale and the user-selected scale. The scale in a geographic context is large if features appear large (e.g., 1:1,000) and conversely small if features appear small (1:1,000,000). The scale and resolution at which the data are collected or measured is termed the **native** scale or resolution. If the spot or point elevations are surveyed in the field on a grid of 100 m, this is its native resolution. Once contours are interpolated between the points and plotted on a paper map at a scale of, for example, 1:25,000, we have introduced a scale to the data. Once the paper map is digitized, there will be little more information contained at a scale smaller than 1:25,000; enlarging to 1:1,000 would make little sense because detail was lost or never captured at 1:25,000. The importance to hydrologic modeling is that variations in landform, stream channels, or watershed boundaries may not be adequately captured at resolutions that are too coarse or at small scales greater than 1:100,000. Simply changing the scale of a map compiled at 1:25,000–1:1,000 could be misleading if small variations were lost or never captured when the map was compiled at the smaller scale of 1:25,000.

The hydrologist must decide what scale will best represent hydrologic processes. If micro-topography at the scale of rills or small rivulets controls the rate of erosion and sediment transport, then small-scale maps (e.g., 1:1,000,000) will contain little information relevant to the modeling of the process. However, such a map may contain sufficient detail for modeling the river basin hydrologic response.

2.3 Georeferenced Coordinate Systems

In geodesy, a georeferenced coordinate system is based on a vertical and horizontal *datum*. The classical datum is defined by five elements, which give the position of the origin (two elements), the orientation of the network (one element), and the parameters of a referenced ellipsoid (two elements). The World Geodetic System (WGS) is a geocentric system that provides a basic reference frame and geometric figure for the Earth, models the Earth gravimetrically and provides the means for relating positions on various data to an Earth-centered, Earth-fixed coordinate system. Even if two maps are in the same coordinate system, discrepancies may still be apparent due to a different datum or scale used to compile each map. This problem is common when a map is compiled with an older datum and then used with data compiled with an updated datum. The usual remedy is to adjust the older datum to bring it into alignment with the revised datum. Conversion routines exist to transform spatial data from one datum to another.

Correction from one datum to another will not remove differences caused by compilation scale. If the aerial photography is collected at 1:25,000 but the hydrography was compiled at a smaller scale, then the streams will not line up with the photography. Registration can be a problem when using generally available geospatial data sets that have been compiled at disparate scales. A similar effect of misregistration may be observed when combining vector hydrography and raster elevation data because of differences in the scale at which the respective geospatial data were compiled.

Digital aerial photography is available for many parts of the US either through government-sponsored acquisition or on a project basis. A GIS should be able to properly overlay geospatial data that exist in two different georeferenced coordinate systems. For example, aerial photography that is georeferenced is called orthophotography or orthophoto. Because the aerial photo was adjusted through geometric transformations, it can be overlaid with other maps such as stream channels. An example of how georeferencing can produce obvious errors is seen in Fig. 2.1 where two streams are overlaid on top of a digital orthophoto. The orthophoto and stream shown in black were compiled in North American Datum 83 (NAD83), whereas the second stream (shown in white) is apparently displaced to the south and east, because it was compiled in the NAD27 datum. The NAD83 stream (red) matches well with stream channel features in the orthophoto because their data are consistent.

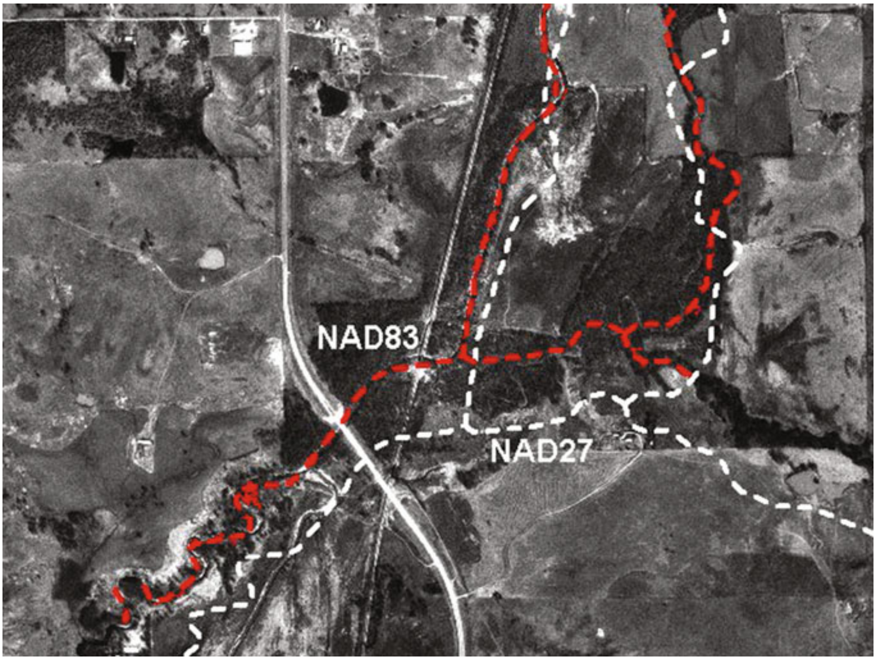

Fig. 2.1 Mismatched datum effects on location of hydrographic features (*red stream*, NAD83; *white stream*, NAD27)

The NAD27 stream (shown in white) is inconsistent with the orthophoto datum and like differs due to channel morphology since the original map compilation.

Georeferenced coordinate systems are developed to consistently map features on the Earth's surface. Each point on the Earth's surface may be located by a pair of latitude and longitude. Because the Earth is an oblate spheroid, the distance between two points on that surface depends on the assumed radius of the Earth at the particular location. Both spherical and ellipsoidal definitions of the Earth exist. The terms *spheroid* and *ellipsoid* refer to the definition of the dimensions of the Earth in terms of the radii along the equator and along a line joining the poles. As better definitions of the spheroid are obtained, the spheroid has been updated compared to historically older definitions that arose before the advent of satellite measurements such as the Clark spheroid developed in 1866. Recent spheroids derived from satellite measurements include the GRS1980 and WGS84. Periodically, the vertical or horizontal datum is updated with the aid of geodetic surveys or GPS measurements.

2.4 Map Projections

A variety of map projections exist that transform three-dimensional spherical/ellipsoidal coordinates, expressed in latitude and longitude, to two-dimensional planar coordinates. Together with a geoid and datum definition, the equations are called a *projection*. Depending on the source of the GIS data, one may encounter a variety of map projections. When distances or areas are needed from geospatial data, the data are almost always projected from latitude and longitude into a two-dimensional plane.

All projections introduce distortion, because the projection transforms positions located on a three-dimensional surface, i.e., spheroid, to a position located on a two-dimensional surface, called the projected surface. There are three main types of projections. *Conformal* projections by definition maintain local angles between the original decimal degree and the projected reference system. This means that if two lines intersect each other at an angle of 30° on the spheroid, then in a conformal projection, the angle is maintained on the projected surface only if the projection is *equal distance*. The stereographic projection is conformal but not equal area or equal distance. Because hydrology is often concerned with distances and areas, map projections that preserve these quantities find broadest usage.

The usefulness of maps for navigation has made geographic projections an important part of human history. Figure 2.2 shows the countries of the world projected onto a plane tangent to the North Pole using the stereographic projection. While this polar form was probably known by the Egyptians, the first Greek to use it was Hipparcus in the second century. In 1613, Francois d'Aiguillon was the first to name it *stereographic* (Snyder 1987).

While this projection has long been used for navigational purposes, it has been used more recently for hydrologic purposes. The US National Weather Service

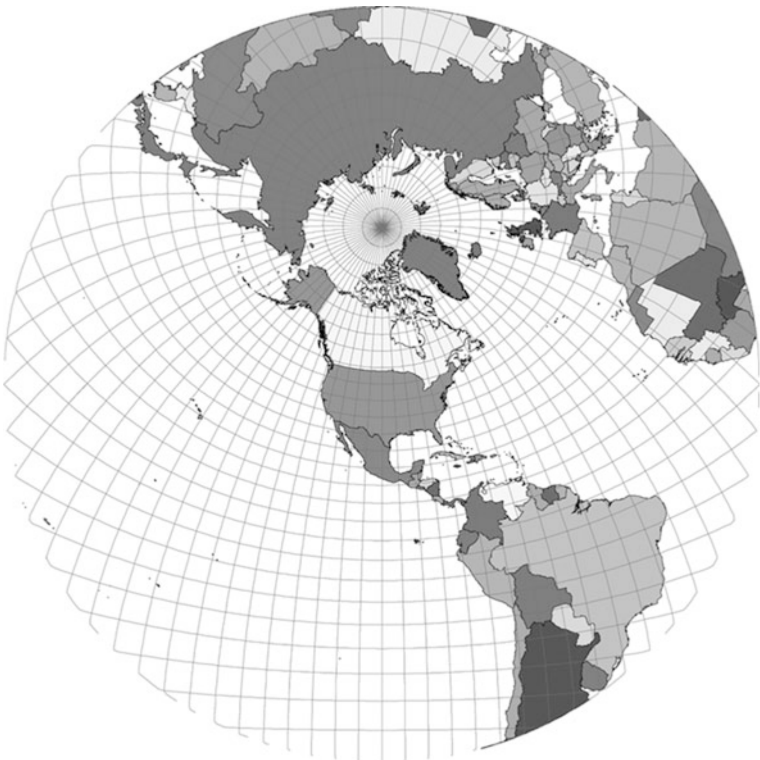

Fig. 2.2 Stereographic projection of countries together with *lines* of longitude and latitude

(NWS) uses it to map radar estimates of rainfall on a national grid called HRAP. Figure 2.3 shows the basic idea of the stereographic projection. The choice of map plane latitude, termed *reference latitude*, is a projection parameter that depends on the location and extent of features to be mapped. The distance between A′ and B′ is less on the map plane at 60 ºN latitude than the distance between A″ and B″ at 90 ºN latitude. Changes in distance on a given plane can constitute a distortion.

Parameters of a projection and the type of projection are important choices since the accuracy of mapped features depends on the selection. The spheroid may be represented with a single radius in the case of spherical definitions versus ellipsoidal definitions that contain a major and minor radius. Countries develop their own geodetic coordinate system. NAD 83 defined above is the horizontal datum on which many projections are based in North America. In other countries, a datum and coordinate system may also be defined, for example the Korean 1985 projection that consists in the Transverse Mercator projection (discussed below) and the Bessel 1941 spheroid with an equatorial radius of 6,377,397.155 m and a polar radius of 6,356,079.0, resulting in a flattening ratio of 299.1528128.

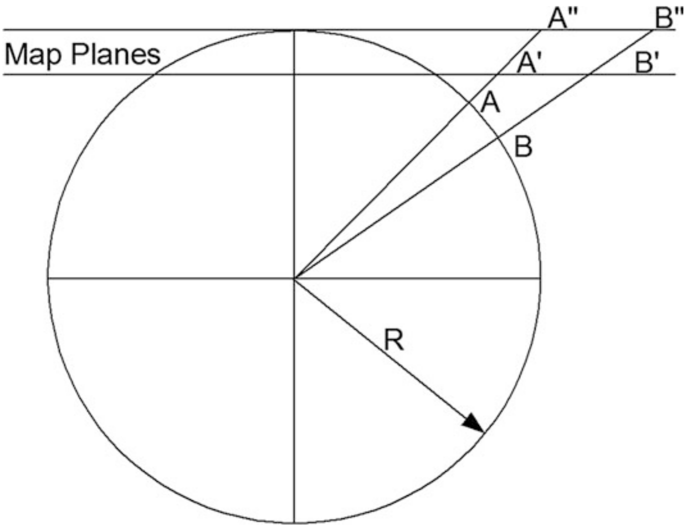

Fig. 2.3 Stereographic projection method for transforming the geographic location of *points A* and *B* to a map plane

In choosing a projection, we seek to minimize the distortion in angles, areas, or distances, depending on which aspect is more important. There is no projection that maintains all three characteristics because a geographic feature in three dimensions (elevation, latitude, and longitude) to two dimensions on a planar map always introduces some distortion. While mathematically there is no projection that simultaneously preserves local angles (conformal), or preserves area or distance, we can preserve two of the three quantities. For example, the Universal Transverse Mercator (UTM) projection is designed to be *conformal* and *equal area* though not *equal distance*. Figure 2.4 shows a transverse developable surface tangent along a meridian of longitude. To minimize the distortion in distance, the UTM projection divides the Earth's surface into 60 zones. In mid-latitudes around the world, identical projections are made in each zone of 6° longitude (360°/60). This means that the coordinates in the projected surface uniquely describe a point only within the zone. The projected coordinates are in meters with the *x*-coordinate (east-west) of 500,000 m being assigned to the central meridian of each zone. It is not enough to simply say that a particular point is located at $x = 500{,}000$ and $y = 2{,}000{,}000$ m, because this does not uniquely define the location in the UTM projection, we must also specify the zone.

Depending on the projection, distortions in either the east-west direction or the north-south direction may be minimized but we cannot have both. If the geographic feature spans more distance in the north-south direction, a projection that minimizes distortion in this direction is often used. The direction of least distortion is determined by the *developable surface* used to transform the coordinates on the spheroid

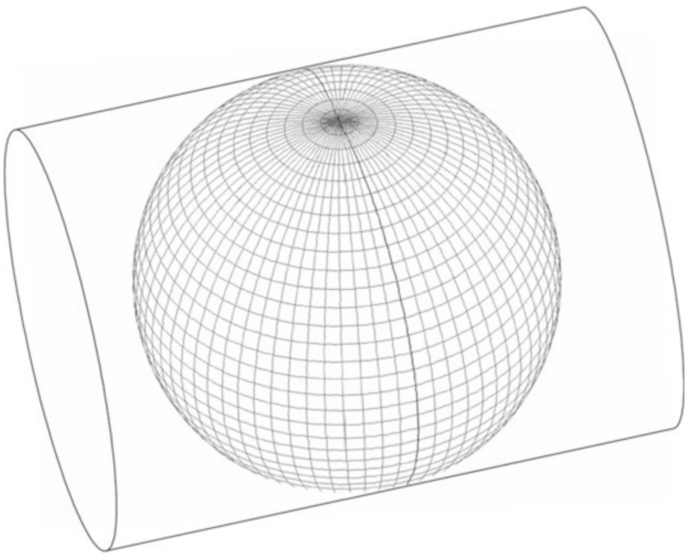

Fig. 2.4 Transverse cylinder developable surface used in UTM projection

to a two-dimensional surface. Typical developable surfaces are cylinders or cones. The orientation of the axis of the developable surface with respect to the Earth's axis determines whether the projection is termed transverse or oblique. A *transverse projection* would orient a cylinder whose axis is at right angles to the Earth's axis and is tangent to the Earth's surface along some meridian. The cylinder is then "unwrapped" to produce a two-dimensional surface with Cartesian coordinates. In an oblique projection, the axis of the developable surface and the Earth's axis form an oblique angle. Distortion in distance is minimized in the UTM projection by making this cylinder tangent at the central meridian of each zone and then unwrapping it to produce the projected map.

Choosing the appropriate projection for the spatial extent of a hydrologic feature is important. Figure 2.5 shows the Arkansas-Red-White River basins and subbasins along with state boundaries. The graticule at 6º intervals in longitude corresponds to the UTM zones. Zones 13, 14, and 15 are indicated at the bottom of Fig. 2.5.

The watershed boundaries shown are in decimal degrees of latitude and longitude and not projected. A watershed that crosses a UTM zone cannot be mapped because the *x*-coordinate is nonunique. The origin starts over at 500,000 m at the center of each zone. The UTM projection is an acceptable projection for hydrologic analysis of limited spatial extent provided the area of interest does not intersect a UTM zone boundary. Many of the subbasins shown in Fig. 2.5 above could effectively be mapped and modeled using UTM coordinates, whereas the entire Arkansas-Red-White basin could not because it intersects two or more UTM zones. Selecting the appropriate projection and datum for hydrologic analysis depends on

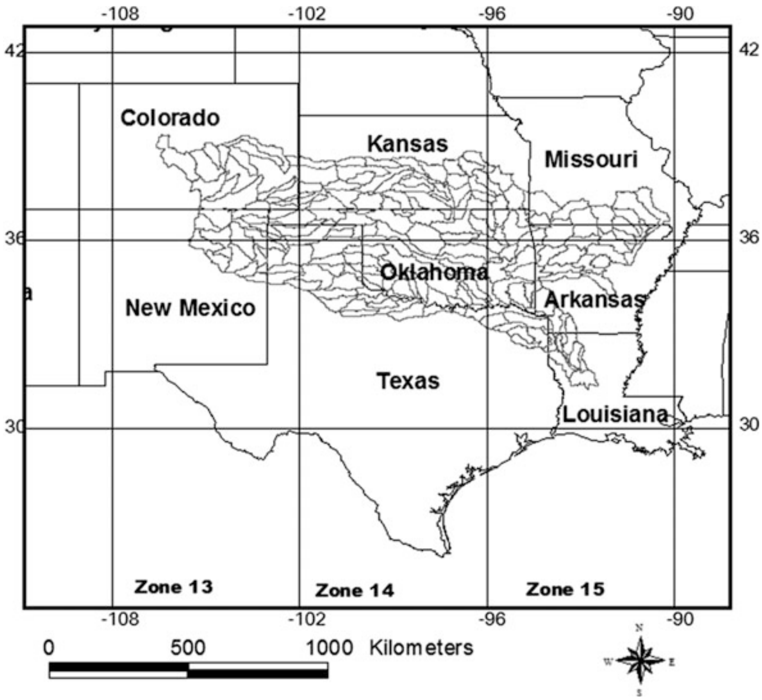

Fig. 2.5 Arkansas-*Red-White* River basin and UTM zones 13, 14 and 15

the extent of the watershed or region. If the distance spanned is large, then projection issues take on added importance. In small watersheds, the issue of projection is not as important because errors or distortions are on the order of parts per million and do not generally have a significant impact on the hydrologic simulation. In fact, for small watersheds we can use an assumed coordinate system that relates features to each other without regard to position on the Earth's surface. The following section deals with general concepts of geographic modeling along with the major types of data representation with specific examples related to hydrology.

2.5 Data Representation

2.5.1 Metadata

Knowing the origin, lineage, and other aspects of the data you are using is essential to understanding its limitations and getting the most usefulness from the data. This information is commonly referred to as data about the data, or *metadata*. For example, the fact that an elevation data set is recorded to the nearest meter could be discovered from the metadata. If the topography controlling the hydrologic process is controlled by surface features, e.g., rills and small drainage channels that physically have scales less than 1 m in elevation or spatial extent, then digital representation of the surface may not be useful in modeling the hydrologic process. An example of metadata documentation for GIS data is provided by the Federal Geographic Data Committee (FGDC) with content standards for digital geospatial metadata (FGDC 1998). The FGDC-compliant metadata files contain detailed descriptions of the data sets and narrative sections describing the procedures used to produce the data sets in digital form. Metadata can be indispensable for resolving problems with data. Understanding the origin of the data, the datum and scale at which it was originally compiled and the projection parameters are prerequisites for effective GIS analysis. The following section describes the major methods for representing topography and its use in automatic delineation of watershed areas. Types of surface representation can be extended to other attributes besides elevation such as rainfall.

2.5.2 Topographic Representation

A *digital elevation model* (DEM) consists in an ordered array of numbers representing the spatial distribution of elevations above some arbitrary datum in a landscape. It may consist in elevations sampled at discrete points or the average elevation over a specified segment of the landscape, although in most cases it is the former. DEMs are a subset of *digital terrain models* (DTMs), which can be defined as ordered arrays of numbers that represent the spatial distribution of terrain attributes, not just elevation. Over 20 topographic attributes can be used to describe landform. The slope and direction of principal land surface gradients are important and widely used terrain model attributes. Other primary topographic attributes include the specific catchment area and altitude. The terrain and land surface slope plays an important role in hydrologic processes, especially runoff. Reliable estimation of topographic parameters reflecting terrain geometry is necessary for geomorphological, hydrologic, and ecological studies, because terrain controls runoff, erosion, and sedimentation.

When choosing the particular method of representing the surface, it is important to consider the end use. The ideal structure for a DEM may be different if it is used as a structure for a distributed hydrologic model than if it is used to determine the topographic attributes of the landscape. There are three principal ways of structuring a network of elevation data for its acquisition and analysis

1. Contour
2. Raster
3. Triangular irregular network

With this introduction to the three basic types of surface representation, we now turn in more detail to each of the three.

2.5.2.1 Contour

A *contour* is an imaginary line on a surface showing the location of an equal level or value. For example, an isohyet is a line of equal rainfall accumulation. Representation of a surface using contours shows gradients and relative minima and maxima. The interval of the contour is important, particularly when deriving parameters. The difference between one level contour and another is referred to as *quantization*. The hydrologic parameter may be quantized at different intervals depending on the variability of the process and the scale at which the hydrologic process is controlled. Urban applications often utilize contours produced at 2-ft (0.61 m) intervals, whereas, USGS typically produces contours at 10-ft (3.048 m) intervals for general mapping purposes.

Contour-based methods of representing the land surface elevations or other attributes have important advantages for hydrologic modeling because the structure of their elemental areas is based on the way in which water flows over the land surface (Moore et al. 1991). Lines orthogonal to the contours are streamlines, so the equations describing the flow of water can be reduced to a series of coupled one-dimensional equations. Many DEMs are derived from topographic maps, so their accuracy can never be greater than the original source of data. For example, the most accurate DEMs produced by the United States Geological Survey (USGS) are generated by linear interpolation of digitized contour maps and have a maximum root mean square error (RMSE) of one-half contour interval and an absolute error of no greater than two contour intervals in magnitude.

Developing contours in hydrologic applications carries more significance than merely representing the topography. Under certain assumptions, it is reasonable to assume that the contours of the land surface control the direction of flow. Thus, surface generation schemes that can extract contour lines and orthogonal streamlines at the same time from the elevation data have advantages of efficiency and consistency.

2.5.2.2 Raster

The raster data structure is perhaps one of the more familiar data structures in hydrology. Many types of data, especially remotely sensed information, are often measured and stored in raster format. The term *raster* derives from the technology developed for television in which an image is composed of an array of picture elements called pixels. This array or raster of pixels is also a useful format for representing geographical data, particularly remotely sensed data, which in its native format is a raster of pixels. Raster data are also referred to as grids. Because of the vast quantities of elevation data that are in raster format, it is commonly used for watershed delineation, deriving slope, and extracting drainage networks. Other surface attributes such as gradient and aspect may be derived from the DEM and stored in a *digital terrain model* (DTM). The term DTED stands for *digital terrain elevation data* to distinguish elevation data from other types of DTM attributes.

Raster DEMs are one of the most widely used data structures because of the ease with which computer algorithms are implemented. However, they do have several disadvantages:

- They cannot easily handle abrupt changes in elevation.
- The size of grid mesh affects the results obtained and the computational efficiency.
- The computed upslope flow paths used in hydrologic analyses tend to zig-zag and are therefore somewhat unrealistic.
- The delineation of catchment areas may be imprecise in flat areas.

Capturing surface elevation information in a digital form suitable to input into a computer involves sampling x, y, z (easting, northing, and elevation) points from a model representing the surface, such as a contour map, stereo-photographs, or other images. DEMs may be sampled using a variety of techniques. Manual sampling of DEMs involves overlaying a grid onto a topographic map and manually coding the elevation values directly into each cell. However, this is a very tedious and time-consuming operation suitable only for small areas. Alternatively, elevation data may be sampled by direct quantitative photogrammetric measurement from aerial photographs on an analytical stereo-plotter. More commonly, digital elevation data are sampled from contour maps using a digitizing table that translates the x, y, and z data values into digital files. Equipment for automatically scanning line maps has also been developed based on either laser-driven line-following devices or a raster scanning device, such as the drum scanner. However, automatic systems still require an operator to nominate the elevation values for contour data caused by poor line work, the intrusion of non-contour lines across the contour line being automatically scanned, or other inconsistencies. DEMs may be derived from overlapping remotely sensed digital data using automatic stereo-correlation techniques, thereby permitting the fast and accurate derivation of DEMs. With the increasing spatial accuracy of remotely sensed data, future DEMs will have increasingly higher accuracy. Van Zyl (2001) described the US Space Shuttle mission that mapped 80 % of the populated Earth surface to 30 m resolution. Because of *holes* in this data set, an improved data set was produced and made available by Jarvis et al. (2008).

The major disadvantage of grid DEMs is the large amount of data redundancy in areas of uniform terrain and the subsequent inability to change grid sizes to reflect areas of different complexity of relief (Burrough 1986). However, various techniques of data compaction have been proposed to reduce the severity of this problem, including quadtrees, freeman chaincodes, run-length codes, and block codes used to compress the raster data structure. Advantages of a regularly gridded

DEM are its easy integration with raster databases and remotely sensed digital data, the smoother, more natural appearance of contour maps and derived terrain features maps and the ability to change the scale of the grid cells rapidly.

2.5.2.3 Triangular Irregular Network

A *triangular irregular network* (TIN) is an irregular network of triangles representing a surface as a set of non-overlapping contiguous triangular facets of irregular sizes and shapes. TINs are more efficient at representing the surface than the uniformly dense raster representation. TINs have become increasingly popular because of their efficiency in storing data and their simple data structure for accommodating irregularly spaced elevation data. Advantages have also been found when TIN models are used in inter-visibility analysis on topographic surfaces, extraction of hydrologic terrain features, and other applications.

A TIN has several distinct advantages over contour and raster representations of surfaces. The primary advantage is that the size of each triangle may be varied such that broad flat areas are covered with a few large triangles, while highly variable or steeply sloping areas are covered with many smaller triangles. This provides some efficiency over raster data structures since the element may vary in size according to the variability of the surface. Given the advantages of TINs in representing data requiring variable resolution, we will examine the features of and methods for generating the TINs.

A TIN approximates a terrain surface by a set of triangular facets. Each triangle is defined by three edges and each edge is bound by two vertices. Most TIN models assume planar triangular facets for the purpose of simplifying interpolation or contouring. Vertices in TINs describe nodal terrain features, e.g., peaks, pits, or passes, while edges depict linear terrain features, e.g., break, ridge or channel lines. Building TINs from grid DEMs therefore involves some procedures for efficiently selecting the locations of TIN vertices for nodal terrain features or TIN edges for linear terrain features. Grid DEMs are widely available at a relatively low cost. Because of this increasing availability, the need for an efficient method to extract critical elevation points from grid DEMs to form TINs has increased. Some caution should be exercised to ensure that critical features are not lost in the conversion process. Figure 2.6 is a TIN created from a grid DEM. Notice the larger triangles representing flat areas and the smaller, more numerous, triangles in areas where there is more topographic relief. Using triangles of various sizes, where needed, reduces computer storage compared with raster formats.

Methods for constructing a TIN from a grid DEM were investigated by Lee (1991). In general, methods for creating TINs consist in selecting critical points from grid DEMs. Nonessential grid points are discarded in favor of representing the surface with fewer points linked by triangular facets. These conversion methods may be classified into four categories: (1) skeleton, (2) filter, (3) hierarchy, and (4) heuristic method. Each of these methods has its advantages and disadvantages,

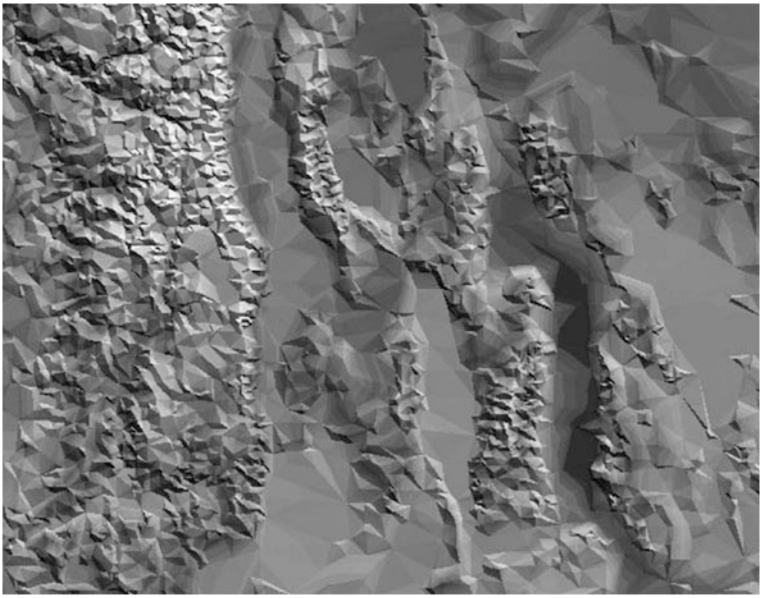

Fig. 2.6 TIN elevation model derived from a gridded LiDAR DEM

different in how they assess *point* importance and in their *stopping rules*, i.e., when to stop eliminating elevation points. The common property of all the methods is that the solution sets depend on some predetermined parameters; these are several tolerances in the skeleton method and either a tolerance or a prescribed number of output points in the other three methods.

The objective of any method of converting DEMs to TINs is to extract a set of irregularly spaced elevation points that are as few as possible while at the same time providing as much information as practical about topographic structures. Unfortunately, it is impossible to achieve both goals concurrently and difficult to balance the two effects. Surface geometry that is highly variable is best represented by a range of smaller to larger TINs where needed. Compared with raster, an advantage of the TIN format is that small variations in terrain such as hillslopes or road embankments can be modeled along with broad-scale features such as the floodplain along a river with triangular facets of varying sizes. The importance of digital elevation data and derivative products for distributed hydrologic modeling cannot be overstated. Cell resolution effects are discussed in Chap. 4, which deals with information content and spatial variability. Chapter 6 addresses in detail aspects of drainage networks derived from DEMs. Methods for automatic delineation of watershed boundaries from DEMs are discussed in the Sect. 2.6.

2.6 Watershed Delineation

Raster or TIN are the primary data structures used in the delineation of watershed boundaries by automated methods. Defining the watershed and the drainage network forms the basic framework for applying both lumped and distributed hydrologic models. Moore et al. (1991) discussed the major data structures for watershed delineation ranging from grid, TIN, and contour methods. Figures 2.7 and 2.8 show the Illinois River Basin delineated to just below Lake Tenkiller in Eastern Oklahoma. The watershed area is 4,211 km^2, delineated at a DEM resolution of 60 m (Fig. 2.7) and 1080 m (Fig. 2.8), respectively. At 60 m the number of cells is 1,169,811, while the delineated watershed at 1,080 m resolution has much fewer with only 3,690 cells. While the larger resolution reduces computer storage, sampling errors increase. The variability of the surface may not be adequately captured at coarse resolution, resulting in difficulties in automatic watershed delineation. The two delineated watersheds look similar but have slightly different watershed boundaries. The difference in boundary shape can be more severe where flatter slopes are not adequately resolved by the vertical resolution, in this case to the nearest 1 m.

Depressions may be natural features or simply a result of sampling an irregular surface with a regular sampling interval, i.e., grid resolution. The hydrologic significance of depressions depends on the type of landscape represented by the DEM.

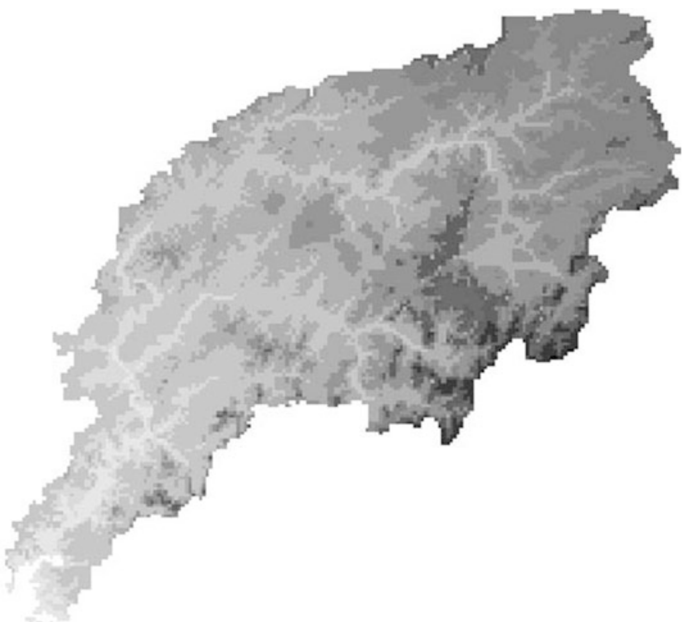

Fig. 2.7 Digital elevation map of the Illinois River Basin with 60 m resolution

Fig. 2.8 As in Fig. 2.7 but delineated at 1,080 m

In some areas, such as the prairie pothole region in the Upper Midwest (North Dakota) of the United States or in parts of the Sahel in Africa, surface depressions dominate and control the hydrologic processes. In areas with coordinated drainage, depressions are an artifact of the sampling and generation schemes used to produce the DEM. We should distinguish between real depressions and those that are artifacts introduced by the sampling scheme and data structure.

2.6.1 Algorithms for Delineating Watersheds

O'Callaghan and Mark (1984), Peucker and Douglas (1975) and Jenson (1987) proposed algorithms to produce depressionless DEMs from regularly spaced grid elevation data. Numerical filling of depressions, whether from artifacts or natural depressions, facilitates the automatic delineation of watersheds. By smoothing a DEM these methods are capable of determining flow paths iteratively, especially where there is more than one possible receiving cell and where flow must be routed across flat areas. The three main methods examined by Skidmore (1990) for calculating ridge and gully position in the terrain are

1. Peucker and Douglas algorithm—Maps ridges and valleys using a simple moving-window algorithm. The cell with the lowest elevation in a two-by-two moving window is flagged. Any unflagged cells remaining after the algorithm has passed over the DEM represent ridges. Similarly, the highest cell in the

window is flagged, with any unflagged cells in the DEM corresponding to valley lines.

2. O'Callaghan and Mark algorithm—Forms a DEM based on quantifying the drainage accumulation at each cell in the DEM. Cells that have a drainage accumulation above a user-specific threshold are considered to be on a drainage channel. Ridges are defined as cells with no drainage accumulation.
3. Band algorithm—Band (1986) proposed a method for identifying streamlines from a DEM, Rosenfeld and Kak (1982) thinning algorithm. The upstream and downstream nodes on each stream fragment are then flagged. Each downstream node is "drained" along the line of maximum descent until it connects with another streamline. The streams are again thinned to the final, one-cell wide, line representation of the stream network.

More recent research has improved upon these models and many GIS modules exist for processing DEMs for purposes of extracting stream networks and delineating watersheds as discussed in Chap. 7. The decision on which algorithm to use depends largely on the resolution of indeterminate flow directions caused by flat slopes.

2.6.2 Problems with Flat Areas

Truly flat landscapes, or zero slope, seldom occur in nature. Yet, when a landscape is represented by a DEM, areas of low relief can translate into flat surfaces. This flatness may also be a result of vertical quantization (precision) of the elevation data. This occurs when the topographic variation is less than 1 m yet the elevation data are reported with a precision to the nearest meter. Flat surfaces are typically the result of inadequate vertical DEM resolution, which can be further worsened by a lack of horizontal resolution. Such flat surfaces are also generated when depressions in the digital landscape are removed by raising the elevations within the depressions to the level of their lowest flow.

A variety of methods has been proposed to address the problem of drainage analysis over flat surfaces. Methods range from simple DEM smoothing to arbitrary flow direction assignment. However, these methods have limitations. DEM smoothing introduces loss of information to the already approximate digital elevations, while arbitrary flow direction assignment can produce patterns that reflect the underlying assignment scheme, which are not necessarily realistic or topographically consistent. Given these limitations, the application of automated DEM processing is often restricted to landscapes with well-defined topographic features that can be resolved and represented by the DEM. Improved drainage identification is needed over flat surfaces to extend the capabilities and usefulness of automated DEM processing for drainage analysis.

Garbrecht and Martz (1997) presented an approach that produced more realistic and topographically consistent drainage patterns than those provided by earlier methods. The algorithm increments cell elevations of the flat surface to include information on the terrain configuration surrounding the flat surface. As a result, two independent gradients are imposed on the flat surface: one away from the higher terrain into the flat surface and the other out of the flat surface towards lower terrain. The linear combination of both gradients, with localized corrections, is sufficient to identify the drainage pattern while at the same time satisfying all boundary conditions of the flat surface. Imposed gradients lead to more realistic and topographically consistent drainage over flat surfaces. The shape of the flat surface, the number of outlets on its edge, and the complexity of the surrounding topography apparently do not restrict the proposed approach. A comparison with the drainage pattern of an established method that displays the "parallel flow" problem shows significant improvements in producing realistic drainage patterns.

One of the most satisfactory methods for assigning drainage directions on flat areas is that of Jenson and Dominique (1988). The Jenson and Domingue (JD) algorithm is useful over most of the DEM but does not produce satisfactory results in areas of drainage lines because it causes these lines to be parallel. The JD algorithm assigns drainage directions to flat areas in valleys and drainage lines such that the flow is concentrated into single lines and it uses the JD method over the rest of the DEM where less convergent flow is more realistic. Automated valley and drainage network delineation seeks to produce a fully connected drainage network of single cell width because this is what is required for applications such as hydrologic modeling. No automatically delineated drainage network is likely to be very accurate in flat areas, because drainage directions across these areas are not assigned using information directly held in the DEM.

One method for enforcing flow direction is to use an auxiliary map to restrict drainage direction where a mapped stream channel exists. Turcotte et al. (2001) described a method for incorporating a river and lake network into the delineation process that yields a more satisfying result where large flat areas or lakes form a part of the drainage network. If an accurate river or stream vector map is available, it may be used to *burn in* the elevations, forcing the drainage network to coincide with the vector map depicting the desired drainage network. By burning in, i.e., artificially lowering the DEM at the location of mapped streams, the correct location of the automatically delineated watershed and corresponding stream network is preserved. Watersheds and stream networks delineated from a DEM become a type of data structure for organizing lumped and distributed hydrologic model computations. The following section deals with another type of map representation useful in soil mapping or other thematic representations.

2.7 Soil Classification

An important source of hydrologic modeling parameters is a map soils for simulating infiltration. Mapping soils usually involves delineating soil types that have identifiable characteristics. The delineation is based on many factors germane to soil science, such as geomorphologic origin and conditions under which the soil formed, e.g., grassland or forest. Regardless of the purpose or method of delineation, there will be a range of soil properties within each mapping unit. This variation may stem from inclusions of other soil types too small to map and from natural variability.

The primary hydrologic interest in soil maps is the modeling of infiltration as a function of soil properties associated with mapped polygons. Adequate measurement of infiltration directly over an entire watershed is impractical. A *soil mapping unit* is the smallest unit on a soil map that can be assigned a set of representative properties. The soil properties are stated in terms of layers. At a particular location on the map, because the properties of the soil vary with depth, some infiltration scheme is adopted for representing an essentially one-dimensional (vertical) process. The infiltration model representation may not include all layers used in the soil classification. Soil maps and the associated soil properties form a major source of data for estimating infiltration. A map originally compiled for agricultural purposes may be reclassified into infiltration parameter maps for hydrologic modeling. As such, the parameter map takes on a data structure characteristic of the original soil map. Figure 2.9 shows mapping units as polygons each with an identifier (map unit symbol, MUSYM). Associated soil properties within each polygon delienated are useful for modeling infiltration.

Estimating infiltration parameters from soil mapping units is introduced in Chap. 5. Other readily available digital maps may be used to derive parameters as discussed in the following section.

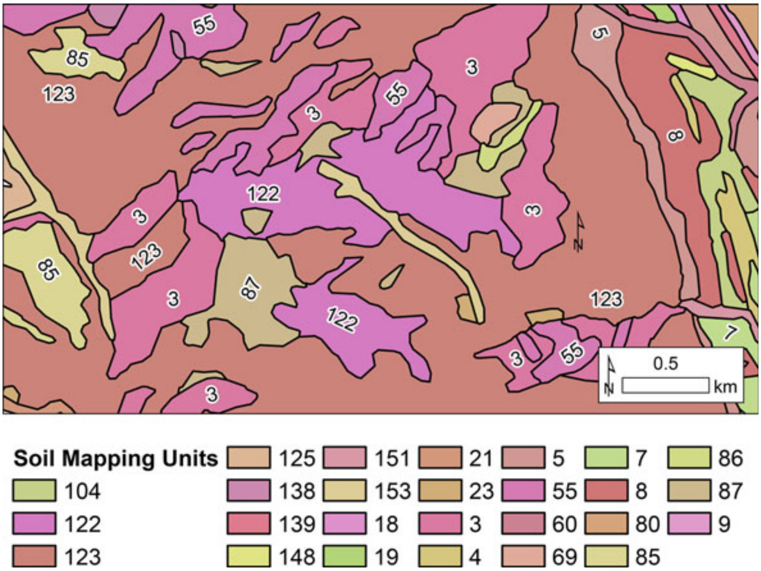

Fig. 2.9 Soil mapping units mapped as *polygons* and labeled with MUSYM

2.8 Land Use/Cover Classification

It is well known that land use, vegetative cover, and urbanization affect the runoff characteristics of the land surface. The combination of the land use and the cover, termed land use/cover, is sometimes available as geospatial data derived from aerial photography or satellite imagery. To be useful, this land use/cover must be reclassified into parameters that are representative of the hydrologic processes. Examples of reclassification from a land use/cover map into hydrologic parameters include hydraulic roughness, surface roughness heights affecting evapotranspiration and impervious areas that limit soil infiltration capacity. The data structure, raster or polygon, of the parent land use/cover map will carry into the model parameter map similar to how soil mapping units define the spatial variation of infiltration parameters.

Maps derived from remote sensing of vegetative cover affect the peak discharge and timing of the hydrograph in response to rainfall input. The hydraulic roughness parameter map takes on a data structure characteristic of the original land use/cover map. If derived from remotely sensed data, a raster data structure results rather than a polygonal structure. Figure 2.10 shows the polygonal data structure of a land use/cover map. This map is derived from the National Land Cover Database (NLCD) data updated in 2011 as described by Homer et al. (2012, 2015) and analysis of impervious cover presented by Xian et al. (2011). The unsupervised

Fig. 2.10 Land use/cover for areas surrounding Lake Tenkiller according to the NLCD

classification scheme is derived from Landsat Thematic Mapper, producing the delineations shown in Fig. 2.10. The major categories are Open Water (11); Developed Low Density Residential (22); Developed Open Space (21); and Cultivated Crops (82). These land use classifications can be used to derive initial estimates of overland hydraulic roughness as discussed in Chap. 6.

2.9 Summary

Effective use of GIS data in distributed hydrologic modeling requires understanding of the type, structure, and scale of geospatial data used to represent watershed and runoff processes. GIS data often lack sufficient detail, space-time resolution, attributes, or differ in some fundamental way from how the model expects the character of the parameter or how the parameter is measured. Existing data sources are often surrogate measures that attempt to represent a particular category with direct or indirect relation to the parameter or physical characteristic. Generation of a surface from measurements taken at points, reclassification of generalized map categories into parameters and extraction of terrain attributes from digital elevation data are important operations in the preparation of a distributed hydrologic model using GIS. Having transformed the original map into usable parameter maps, the parameter takes on the characteristic data structure of the original map. Thus, the data structure inherent in the original GIS map has a lasting influence on the hydrologic process simulation using the derived parameters. In the following chapter, we turn to the generation of one data structure to another, surfaces from point data.

References

Band, L.E. 1986. Topographic partition of watershed with digital elevation models. *Water Resources Research* 22(1): 15–24.

Burrough, P.A. 1986. *Principles of Geographic Information Systems for Land Resources Assessment.* Monographs on Soil and Resources Survey, No. 12, 103–135. Oxford Science Publications.

FGDC. 1998. *Content Standard for Digital Geospatial Metadata.* rev. June 1998 Metadata Ad Hoc Working Group, Federal Geographic Data Committee, 1–78. Reston, Virginia.

Garbrecht, J., and L.W. Martz. 1997. The assignment of drainage direction over flat surfaces in raster digital elevation models. *Journal of Hydrologiques* 193: 204–213.

Homer, C.H., J.A. Fry, and C.A. Barnes. 2012. The national land cover database. *US Geological Survey Fact Sheet* 3020(4): 1–4.

Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., and Megown, K. 2015. Completion of the 2011. National land cover database for the conterminous United States-Representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing* 81(5):345–354.

Jarvis, A., Reuter, H.I., Nelson, A. and Guevara, E. 2008. Hole-filled SRTM for the globe Version 4. *available from the CGIAR-CSI SRTM 90 m Database.* http://srtm.csi.cgiar.org.

Jenson, S.K., 1987. Methods and applications in Surface depression analysis. *Proceedings of Auto-Carto*: 8, 137–144. Baltimore, Maryland.

Jenson, S.K., and J.O. Dominique. 1988. Extracting Topographic Structure from Digital Elevation Data for Geographic Information System Analysis. *Photogrammetric Engineering and Remote Sensing* 54(11): 1593–1600.

Lee, J. 1991. Comparison of existing methods for building triangular irregular network models from grid digital elevation models. *International Journal Geographical Information Systems* 5(3): 267–285.

Moore, I.D., R.B. Grayson, and A.R. Ladson. 1991. Digital terrain modeling: A review of hydrological, geomorphological and biological applications. *Journal of Hydrologiques Process* 5: 3–30.

O'Callaghan, J.F., and D.M. Mark. 1984. The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics and Image Processing* 28: 323–344.

Peucker, T.K., and D.H. Douglas. 1975. Detection of surface specific points by local parallel processing of discrete terrain elevation data. *Computer Vision, Graphics and Image Processing* 4(4): 375–387.

Rosenfeld, A., and A.C. Kak. 1982. *Digital picture Processing*. New York: Academic Press. No. 2.

Skidmore, A.K. 1990. Terrain position as mapped from a gridded digital elevation model. *International Journal of Geographical Information Systems* 4(1): 33–49.

Snyder, J.P. 1987. Map projections—A working manual. *U.S. Geological Survey Professional Paper* 1395: 383.

Turcotte, R.J.-P., A.N. Fortin, S. Rousseau, and J.-P.Villeneuve Massicotte. 2001. Determination of the drainage structure of a watershed using a digital elevation model and a digital river and lake network. *Journal of Hydrology* 240(3–4): 225–242.

Van Zyl, J.J. 2001. The shuttle radar topography mission (SRTM): a breakthrough in remote sensing of topography. *Acta Astronautica* 48(5): 559–565.

Xian, G., C. Homer, J. Dewitz, J. Fry, N. Hossain, and J. Wickham. 2011. Change of impervious surface area between 2001 and 2006 in the conterminous United States. *Photogrammetric Engineering and Remote Sensing* 77(8): 758–762.

http://www.springer.com/978-94-024-0928-4

Distributed Hydrologic Modeling Using GIS
Vieux, B.E.
2016, XVI, 262 p. 124 illus., 60 illus. in color., Hardcover
ISBN: 978-94-024-0928-4