Peter Dale

MATHEMATICAL TECHNIQUES in GIS

SECOND EDITION

CRC Press
Taylor & Francis Group

MATHEMATICAL TECHNIQUES in GIS

SECOND EDITION

Peter Dale

MATHEMATICAL TECHNIQUES in GIS

SECOND EDITION

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20140306

International Standard Book Number-13: 978-1-4665-9555-2 (eBook - PDF)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

Preface to Second Edition

Many people wishing to make use of geographic information systems (GIS) start from a limited mathematical background. In *Mathematical Techniques in GIS*, Second Edition, the text as before focuses on those who are unfamiliar with mathematics and need to understand the principles behind the manipulation of spatial data. The new text adds additional material. The first nine chapters lay out the basic foundations and introduce the reader to the relevant techniques and shorthand notations that frequently occur in mathematics; the remaining five chapters build on the earlier material. These later chapters place particular emphasis on the use of the techniques in GIS and geomatics. Throughout the text there are a number of examples shown in boxes and there is a summary at the end of each chapter listing all the important ideas that have been introduced.

Preface to First Edition

This book has been written for nonmathematicians who wish to understand some of the assumptions that underlie the manipulation and display of geographic information. It assumes very little basic knowledge of mathematics but moves rapidly through a wide range of data transformations, outlining the techniques involved. Many of these are precise, building logically from certain underlying assumptions; others are based on statistical analysis and the pursuit of the optimum rather than the perfect and definite solution.

Mathematics has its own form of shorthand that often gets in the way of understanding what is going on. For those who are unfamiliar with mathematical notation this can be daunting; but it cannot be avoided. It can in many cases be kept to a minimum and in what follows, the derivation of some of the formulae is placed in boxes that can be digested at leisure without interrupting the narrative. But at the end of the day, compromise has had to be made and as the text progresses there is an increasing use of symbols.

This spirit of compromise is most apparent in the selection of topics discussed. Many things have had to be left out—indeed every chapter could be expanded to a full book and most would require several volumes in order to do their subject justice. *Introduction to Mathematical Techniques Used in GIS* is therefore a book that allows the reader to get started and then to turn to the many more informative texts that are available.

The text begins with an introduction to geographic data but soon focuses on the "where" rather than the "what." It assumes that the data have been measured and refrains from discussing the techniques of measurement science, other than to recognize that measurement is prone to error. Pure mathematics, even when dealing with vague concepts, provides precise answers that can be verified by anyone. Even statistical analysis uses processes that can be programmed into a computer to give a consistent answer, even when the underlying assumptions are not met or the hypothesis has been incorrectly formulated. The apparent exactness of an answer does not mean that it is correct. To understand the output from, for example, a geographic information system, one needs to understand the quality of the data that are entered into the system, the algorithms behind the data processing, and the limitations of the graphic displays.

This book deals with only part of the bigger picture. It focuses on the basic mathematical techniques, building the whole of mathematics in a series of steps that are the foundations for a deeper understanding. It seeks to lay the foundations for the more complex forms of manipulation that arise in the handling of spatially related data.

The technology behind geographic information systems (GIS) allows such data to be gathered, processed, and displayed. The power and appeal of such systems often lie in their graphical output, the maps that they create. Users of GIS need to understand the quality of that output so that they can advise others on the integrity of their results. The issue is not a matter of which buttons to push but rather of the quality of the information that has been produced. Quality means "fitness for purpose" and "safety in use."

This book therefore looks at some of the fundamentals and provides an introduction to spatial data manipulation through which users of GIS may come to understand whether what they do results in what can genuinely be described as a "quality product." It has been copy edited for an American market, hence the spelling of words such as "meter" for the English "metre" and "center" for "centre."

The Author

Peter Dale trained as a land surveyor and worked for seven years in Uganda before entering the academic world. He ultimately became a professor in land information management at the University College London. He is an Honorary President of the International Federation of Surveyors and was awarded an OBE in recognition of his services to surveying. He is now retired and lives in a remote area of Scotland. Peter Dale can be contacted via e-mail at: peter.f.dale@btinternet.com.

List of Tables

List of Illustrations

List of Boxes

List of Examples

1 Characteristics of Geographic Information

1.1 GEOGRAPHIC INFORMATION AND DATA

It used to be said that geography was about “maps,” as distinct from “chaps.” Without doubt today it is about both, and a lot more besides. Ultimately, geography is about making sense of the world around us and this is done by observing, measuring, and processing data about the environment and then presenting the information either as text or pictorially. In particular, it is concerned with why things are where they are.

In recent years, much has been said and written about geographic information systems or GIS, which are tools that can help the process of understanding. Although there are various interpretations of what is meant by the acronym “GIS,” the majority of people would accept that it includes a computer system of hardware and software that can be used to record, manage, integrate, manipulate, analyze, and display data that are spatially referenced to the Earth. The term *spatially referenced* means that their location can be described by measured quantities. *Data* are basic facts that can somehow or other be measured and turned into information.

Information is the commodity that is used by people when they make decisions. Too many facts can be confusing—there may be many different possible routes from one’s home to the nearest shopping mall, each of which has its own quality of road surface, slopes, twists, turns and intersections, street lamps, drain covers, and so forth. All these facts about each route can be measured and recorded but all that the average user really wants to know is which is the shortest route. This is a piece of information that can be extracted from the basic data.

The term *shortest* is ambiguous since it could mean shortest in terms of time or shortest in terms of distance; these are not necessarily the same thing. The types of data that need to be collected depend on the use to which the information is to be put. The required output determines the required input and the manner in which the data may need to be processed. One can of course start with a set of data and see what sense can be made of all the facts and figures. Frequently, the most effective way to do this is through pictures, especially maps and graphs. Advocates of the use of GIS often quote the 19th century case in London where the location of cases of cholera were plotted on a map, which then showed clearly that they formed a cluster around an infected well whose water had become contaminated.

When processing data, two golden rules always apply.

1. Bad data plus good processing gives rise to unreliable information.
2. Good data plus bad processing also gives rise to unreliable information.

If data are to be converted into good quality information, then both the data and the way in which they are processed must be of good quality, that is, they must be "fit for purpose" or "safe in use." In the discussions that follow we will focus on the basic mathematical principles underlying how data are processed and not on the technical aspects of measurement or how the data are acquired.

1.2 CATEGORIES OF DATA

Data come essentially in two forms—*categorical* and *numerical.* As their name suggests, categorical data are those that are placed in a category or codified according to a classification system. Such data are sometimes referred to as *nominal data* and have no numerical value as such. Whether a piece of fruit is an apple or a pear or something else depends on the object itself and the way in which fruits have been classified. For many objects there are internationally and scientifically recognized standards for classification though even then there is the occasional dispute over whether some new discovery belongs as a subset of an existing class or whether it represents a totally new species.

With some data, categorization is less scientific, for instance, when designating the type of land use at a particular location. Although within each country there may be a national land use classification system, it does not mean that all those who record land use abide by it and it certainly does not follow that every country uses the same system. A building may be used in several different ways with, for example, the basement as a gymnasium, the ground floor as a shop, the next floor as commercial offices, and the top floor as a residential accommodation. In spite of national guidelines, investigators may still disagree as to how the use of the building should be categorized. It is, however, not the aim of this book to analyze the problems of data classification but rather to note that it is an issue that intimately affects the quality of data.

Once data have been categorized they can be subjected to comparison without being quantified. Thus, the data can be placed in a rank so that *a* is said to be more than *b*, which is more than *c*, and so forth. Such data are described as *ordinal,* an example of which is a list of preferences (area *a* is a nicer place to live than area *b,* etc.). Various statistical tests exist to process and analyze the differences between ranks or sequences of ordinal data but these too will not be discussed here.

Once the data have been categorized it is often necessary to indicate their magnitude. This may be done through the use of *discrete* or *continuous* variables. A discrete variable is one that can only take distinct values while a continuous variable is one that changes only gradually, allowing any intermediate values. Some data can only be measured in terms of whole numbers (called *integers*—such as the number of children in a family) while other items can be measured on a continuous scale (such as the height of each child). One can, of course, talk about the average family size being in the decimal system (see Chapter 2), 2.54 children, even though it is impossible to have 54 out of 100 parts of a child. Such a figure is useful for some practical purposes especially when associated with an estimate of its reliability, as discussed in Chapter 12.

Discrete variables are precise and are often expressed as whole numbers or integers (0, 1, 2, 3, etc.). More particularly, they can take a succession of distinct values

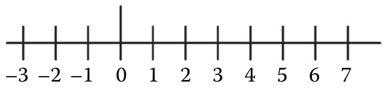

FIGURE 1.1 A scale bar.

at set intervals along a scale for which there are no intermediate values. Such data are often referred to as *interval data* (see Figure 1.1).

The data may be positive or negative but such items can only be compared quantitatively on the basis of the differences between them. Only when the values are *absolute* can valid conclusions be drawn about their relative sizes. One can say that a family with four children has twice as many youngsters as a family with two children because "zero children" is an absolute point of reference. One should not, however, say that a temperature of 16°C is twice as hot as a temperature of 8°C as zero on the centigrade scale is an arbitrarily chosen point.

The highest level of measurement is the *ratio scale*, which differs from the interval scale in that it relates to absolute zero (in the case of temperature this is approximately –273°C). Absolute temperature, length, and breadth are examples of measures on a ratio scale. They are *continuous variables* in that they are not restricted to integer forms but can take any value whatsoever from zero upward. The numerical quantity used to express the measurement of a continuous variable, such as the length of a line or the area of a field, presupposes a standard unit of measure. The numerical value represents the ratio between the quantity measured and the unit of measurement (e.g., the meter or "metre").

Geographical data have one particular characteristic that distinguishes them from all other forms of data, namely, location. Graphical data can be plotted on a map and be represented by points, lines, and areas. From a theoretical perspective, a point on its own has no dimension, a line has one dimension (length), and an area has two and a volume three. In practice, a point on a map is a blob or very small area while a line has thickness and also direction. Each has a category (the "what") representing some attribute or attributes associated with it, and each has a location ("where"). Examples of how the "what" may be categorized as points, lines, and areas when used by cartographers are shown in Table 1.1.

To define the location of any point there must be some reference to which the point can be related. The most common reference system uses a rectangular grid made up of squares of a standard size. For absolute position (as distinct from relative

TABLE 1.1
Points, Lines, and Areas on Maps

Feature	Points	Lines	Areas
Physical objects	Corner of building	Road network	Planning zone
Statistical values	Sampling point	Isoline	Layer tints
Areas	Central point	Boundary line	Polygon
Surfaces	Height point	Contour	Hill shading
Text	House numbers	Street names	District names

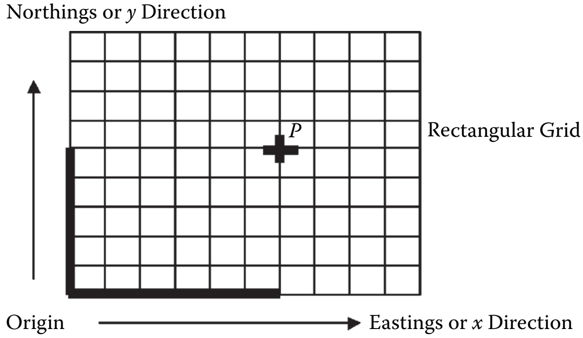

FIGURE 1.2 Rectangular Cartesian coordinates.

position), the grid must have a point of *origin* from which measurements are taken. Points may then be located so far east (or to the right) of the origin and so far north (or up the page), using a standard unit of measure. The two distances are called the *coordinates* or more particularly the *rectangular Cartesian coordinates* (named after the French mathematician Descartes).

In Figure 1.2, the rectangular grid coordinates of P are (x, y) relative to the origin and are shown here as (6, 5). The idea can be extended to three dimensions by adding the height above the origin. Although in some countries the x-direction is taken as being to the north or upward, here we will follow the convention that the direction across the page is the x-direction while up the page is described as the y-direction (x, y, or "in the door and up the stairs," which is the opposite of many computer software graphics packages that measure the position of points from the top of the screen downward). Height is then in the z-direction. For any point on a three-dimensional object the coordinates would be (x, y, z). For two-dimensional ("flat Earth") displays, $z = 0$.

Simple mathematical techniques can be used to analyze the locations of points that have been referenced to a rectangular grid. Sometimes it is useful to use a nonrectangular or skewed grid, for example, when trying to show three dimensions on a flat piece of paper (Figure 1.3). Data manipulation is slightly more complicated in these circumstances although the underlying principles are the same.

An alternative way of measuring the location of a point is through the use of *polar coordinates* (Figure 1.4). These describe points by their distance from an origin and their direction relative to some reference line. The direction is known as the *bearing*

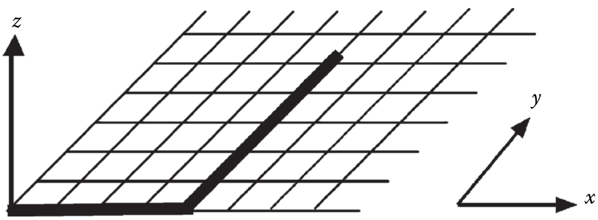

FIGURE 1.3 Nonrectangular or skewed grid.

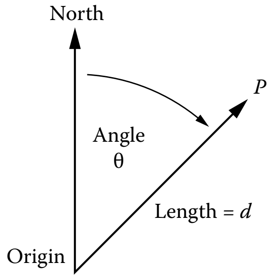

FIGURE 1.4 Polar coordinates.

and is normally measured clockwise from the north (or up the page). In Figure 1.4, P has polar coordinates (θ, d) relative to the origin, d being the distance, and θ (the Greek letter "theta") being the direction or bearing from the north.

Angles and distances are examples of measures on the ratio scale. Distances are normally expressed as a ratio, for instance, between the amount of space between two points in comparison with a standard length. Under the *Systeme International* (SI) the standard unit of length is known as the *metre* and was once defined as the distance between two marks on a bar of platinum kept at constant temperature in Paris. It is now defined by stipulating that the speed of light is 299,792,458 meters (or "metres") per second. As we will show later, trigonometrical formulae allow polar coordinates (θ, d) to be converted into Cartesians (eastings and northings or x and y) and vice versa.

Angles are a ratio between the amount of turning and a complete turn. They may be expressed as a proportion of either 360 degrees—written as 360° with each degree being subdivided into 60 minutes (60') and each minute into 60 seconds (60"); or 400 grads (where 100 grads equates with a quarter turn, with submeasurements being expressed as decimals) or 2π (two pi) radians where "pi" is the ratio between the diameter of a circle and its circumference.

Angular measures are important in surveying where positions may be expressed as if the Earth were a sphere using what are known as *spherical coordinates*. The *latitude* of a point is its angular distance north or south of the equator and is often represented by the Greek letter "phi" or ϕ. The *longitude* of a point is an angular measure east or west of the Greenwich standard meridian: it is normally represented by the Greek letter "lambda" or λ (see Figure 1.5).

The altitude or height of any point is measured as a distance above a reference level or surface, such as a mathematical shape that best approximates to the size and shape of the Earth. It is not normally given a Greek letter and hence the coordinates of points are either expressed as (ϕ, λ) or as (ϕ, λ, H). The use of Greek letters is common in mathematics. The full alphabet is given in Table 1.2.

For more accurate work, rather than assuming that the Earth is a perfect sphere, its shape is taken to be an *ellipsoid* (an ellipse rotated on its shorter axis creating a squashed sphere) as discussed in Chapter 4. However, for many practical purposes, the Earth can be regarded as a sphere. The word *accurate* as used here relates to nearness to the truth. The word *precision* will be used to refer to the exactness with