PART I

Concepts
CHAPTER 1

Hiking in the Sierra Nevada

1.1 An imaginary outdoor adventure company: Buena Sierra

As is the case for any applied science, no geostatistical application is without context. This context matters; it determines modeling choices, parameter choices, and the level of detail required in such modeling. In this short first chapter, we introduce an imagined context that has elements common to many applications of geostatistics: sparse local data, indirect (secondary) or trend information, a transfer function or decision variable, as well as a specific study target. The idea of doing so is to remain general by employing a synthetic example whose elements can be linked or translated into one’s own area of application.

Consider an imaginary hiking company, Buena Sierra, a start-up company interested in organizing hiking adventures in the Sierra Nevada Mountains in the area shown in Figure I.1.1 (left). The company drops customers over a range of locations to hike over a famous but challenging mountain range and meets them at the other end of that range for pickup. Customers require sufficient supplies in what is considered a strenuous trip over rocky terrain, with high elevation changes on possibly hot summer days. Imagine, however, that this area lies in the vicinity of a military base; hence, no detailed topographic or digital elevation model from satellite observation is available at this point. Instead, the company must rely on sparse point information obtained from weather stations in the area, dotted over the landscape; see Figure I.1.1 (right). We consider that the exact elevation of these weather stations has been determined. The company now needs to plan for the adventure trip. This would require determining the quantity of supplies needed for each customer, which would require knowing the length of the path and the cumulative elevation gain because both correlate well with effort. The hike will generally move from west to east. The starting location can be any location on the west side from grid cell (100,1) to grid cell (180,1) (see Figure I.1.2).

To make predictions about path length and cumulative elevation gain, a small routing computer program is written; although it simplifies real hiking, the program is considered adequate for this situation. More advanced routing could be applied, but this will not change the intended message of this imaginary example.
Chapter 1

The program requires as input a digital elevation map (DEM) of the area gridded on a certain grid. The program has as input a certain point on the west side, then walks by scanning for the direction that has the smallest elevation change. The program simulates two types of hikers: the minimal-effort (lazy) hiker and the maximal-effort (achiever) hiker. In both cases, the program assumes the hiker thinks only locally, namely, follows a path that is based on where they

Figure I.1.1 (left) Walker Lake exhaustive digital elevation map (size: 260×300 pixels) grid; and (right) 100 extracted sample data. The colorbar represents elevation in units of ft.

Figure I.1.2 Visualization of the 80 paths taken by hikers of two types: (left) minimal effort; and (right) maximal effort. The color indicates how frequently that portion of the path is taken, with redder color denoting higher frequency.
Hiking in the Sierra Nevada

Figure I.1.3 Histograms of the cumulative elevation gain and path length for the minimal- and maximal-effort hiker. Cumulative elevation gain in units of ft, path length in units of grid cells.

are and what lies just ahead. The minimal hiker takes a path of local least resistance (steepest downhill or least uphill). The achiever hiker takes a path of maximal ascent (or minimal descent). Note that the computer program represents a deterministic transfer function: given a single DEM map, a single starting point, and a specific hiker type, it outputs a single deterministic hiking route. If the actual reference, Walker Lake, is used as input, then given starting locations from grid cell (100,1) to (180,1) on the west side, a total of 80 outcomes are generated. These 80 outcomes can be shown as a histogram; see Figure I.1.3. The resulting path statistics for both minimal effort and maximal effort are shown in Table I.1.1, which are summarized with quantiles (the eighth lowest, or P10; the 40th lowest, or P50; and the 72nd lowest, or P90).

1.2 What lies ahead

The problem evidently is that no DEM is available. How, then, would one proceed with forecasting path length and cumulative elevation change, and thereby make recommendations for Buena Sierra? We start in this Part I from very basic
Chapter 1

Table I.1.1 Summary statistics

<table>
<thead>
<tr>
<th></th>
<th>Minimum effort</th>
<th></th>
<th>Maximum effort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cumulative elevation gain (ft)</td>
<td>Path length (cell units)</td>
<td>Path length (cell units)</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>P10–P90</td>
<td>Median</td>
</tr>
<tr>
<td>Walker Lake reference</td>
<td>9862</td>
<td>9434–14,875</td>
<td>324</td>
</tr>
</tbody>
</table>

notions on how to formulate a theory for these kinds of problems, and then we present practical solutions based on that theory. This provides an opportunity to review important notions and assumptions that are common to most spatial prediction problems. The first such theory formulates spatial estimation, which in geostatistics is known as kriging. It is well known that kriging provides an overly smooth map, not reflecting the actual roughness of the terrain, and therefore any predictions of path length or elevation gain are biased. Such prediction would require stochastic simulation, which also allows statements of uncertainty about the calculated route statistics. Nevertheless, we will start with developing kriging because we will show how the traditional kriging (Chapter I.2) can be formulated without relying on the notions of expectation, probability, or random function theory, as long as a training image is available (Chapter I.3). The solution obtained is strikingly similar to traditional kriging, yet at no instance will we rely on random function theory.

Next, we will review stochastic simulation, which traditionally has relied on the same variogram and random function notions as kriging. In particular, we will review Gaussian theory and some popular methods that have been derived from this theory (Chapter I.4). Next, we will show, in a similar vein as for kriging, that the random function theory is not needed to perform stochastic simulation (Chapter I.5). We will present three alternative algorithms as an introduction to the many algorithms presented in Part II. These methods are compared in their ability to solve the practical problem discussed here (Chapter I.6).