1 Introduction to seismic data and processing

Chapter contents

The discipline of subsurface seismic imaging, or mapping the subsurface using seismic waves, takes a remote sensing approach to probe the Earth's interior. It measures ground motion along the surface and in wellbores, then puts the recorded data through a series of data processing steps to produce seismic images of the Earth's interior in terms of variations in seismic velocity and density. The ground movements recorded by seismic sensors (such as geophones and seismometers onshore, or hydrophones and ocean bottom seismometers offshore) contain information on the media's response to the seismic wave energy that traverses them. Hence the first topic of this chapter is on seismic data and their acquisition, processing, and interpretation processes. Because nearly all modern seismic data are in digital form in order to be stored and analyzed in computers, we need to learn several important concepts about sampled time series such as sampling rate and aliasing; the latter is an artifact due to under-sampling. In exploration seismology, many useful and quantifiable properties of seismic data are called seismic attributes. Two of the most common seismic attributes are the amplitude and phase of seismic wiggles. They are introduced here together with relevant processing issues such as gain control, phase properties of wavelets, and the Hilbert transform, which enables many time-domain seismic attributes to be extracted. To process real seismic data, we also need to know the basic issues of data formats, the rules of storing seismic data in computers. To assure that the data processing works, we need to conduct many quality control checks. These two topics are discussed together because in practice some simple quality control measures need to be applied at the beginning stage of a processing project.

A newcomer to the field of seismic data processing needs to know the fundamental principles as well as common technical terms in their new field. In this book, phrases in **boldface** denote where special terms or concepts are defined or discussed. To comprehend each new term or concept, a reader should try to define the term in his or her own words. The subject of seismic data processing often uses mathematical formulas to quantify the physical concepts and logic behind the processing sequences. The reader should try to learn the relevant mathematics as much as possible, and, at the very least, try to understand the physical basis and potential applications for each formula. Although it is impossible for this book to endorse particular seismic processing software, readers are encouraged to use any commercially or openly accessible seismic processing software while learning seismic data processing procedures and exercises. An advanced learner should try to write computer code for important processing steps to allow an in-depth comprehension of the practical issues and limitations.

1.1 Seismic data and their acquisition, processing, and interpretation

As a newcomer, you first want to know the big picture: the current and future objectives and practices of seismic data processing, and the relationship of this field to other related disciplines. You will need to comprehend the meanings of the most fundamental concepts in this field. This section defines seismic data and a suite of related concepts such as signal-to-noise ratio (SNR or S/N), various seismic gathers, common midpoint (CMP) binning and fold, stacking, pre-stack versus post-stack data, and pre-processing versus advanced processing. The relationship between acquisition, processing, and interpretation of seismic data is discussed here, since these three processes interrelate and complement each other to constitute the discipline of subsurface seismic imaging.

1.1.1 Digital seismic data

Seismic data are physical observations, measurements, or estimates about seismic sources, seismic waves, and their propagating media. They are components of the wider field of geophysical data, which includes information on seismic, magnetic, gravitational, geothermal, electromagnetic, rock physics, tectonophysics, geodynamics, oceanography, and atmospheric sciences. The form of seismic data varies, and can include analog graphs, digital time series, maps, text, or even ideas in some cases. This book treats the processing of a subset of seismic data, those in digital forms. We focus on the analysis of data on body

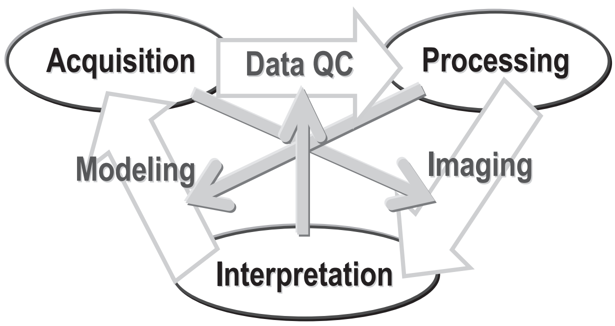

Figure 1.1 Relationship between data acquisition, processing, and interpretation.

waves, mostly P-waves, in their transmission, reflection, diffraction, refraction, and turning processes. The processing of other seismic data and many non-seismic data often follows similar principles.

The purpose of acquiring and processing seismic data is to learn something about the Earth's interior. To understand certain aspects of the Earth, we initially need to figure out some specific relations between the intended targets and measurable parameters. Then our first step is to conduct **data acquisition** designed for the problem, our second step to use **data processing** to identify and enhance the desired signal, and our third step to conduct **data interpretations** based on the processed data. In reality, the processes of data acquisition, processing and interpretation are interconnected and complement each other; their relationship may be viewed as shown in Figure 1.1.

After data acquisition and before data processing, we need to conduct the process of data **quality control**, or **QC**. This involves checking the survey geometry, data format, and consistency between different components of the dataset, and assuring ourselves that the quality and quantity of the dataset are satisfactory for our study objectives. The data QC process is typically part of the **pre-processing**. After pre-processing to suppress various kinds of noise in the data, **seismic imaging** is conducted to produce various forms of imagery for the interpretation process. The seismic imaging methods include seismic migration, seismic tomography, and many other methods of extracting various seismic attributes. Some people call seismic imaging methods the advanced processing. The scope of this book covers the entire procedure from pre-processing to seismic imaging.

After data interpretation, we often conduct **seismic modeling** using the interpreted model and the real data geometry to generate predictions to compare with the real measurements, and hence further verify the interpretation. The three inner arrows shown in Figure 1.1 show how the interactions between each pair of components (namely the data QC, imaging, or modeling processes) are influenced by the third component.

1.1.2 Geometry of seismic data gathers

Seismic data acquisition in the energy industry employs a variety of acquisition geometries. In cross-section views, Figure 1.2 shows two **seismic acquisition spreads**, the arrangements of shots and receivers in seismic surveys. Panel (a) shows a **split spread**, using a shot located in the middle and many receivers spread around it. This spread is typical of onshore

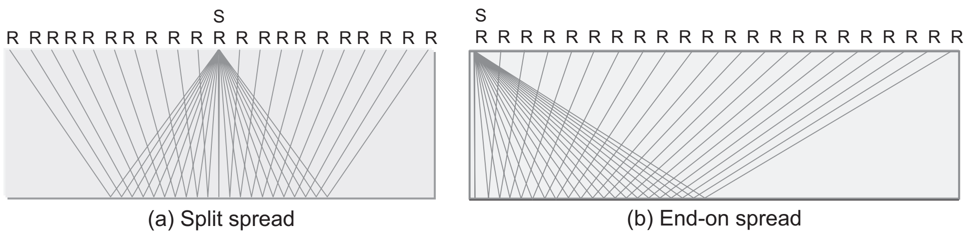

Figure 1.2 Cross-section views of two seismic data acquisition spreads and raypaths.

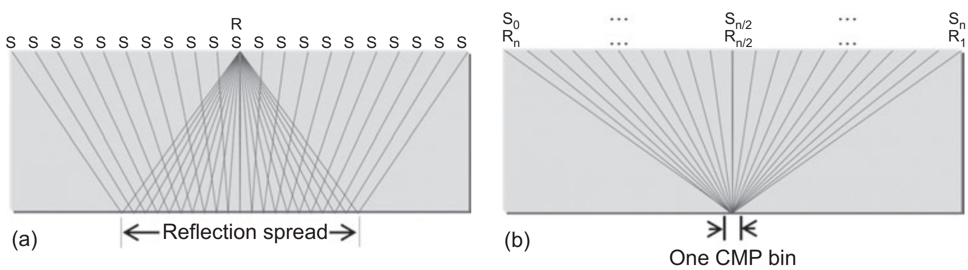

Figure 1.3 Cross-section views of (a) a common receiver gather and (b) a common midpoint (CMP) gather.

acquisition geometry using dynamite or Vibroseis technology as sources and geophones as receivers. The real-world situation is much more complicated, with topographic variations, irregular source and receiver locations in 3D, and curving raypaths. Panel (b) shows an **end-on spread**, with a shot located at one end and all receivers located on one side of the shot. This spread is the case for most offshore seismic surveys using airgun or other controlled sources near the boat and one or more streamers of hydrophones as receivers. In comparison with onshore seismic data, offshore seismic data usually have much higher quality because of a number of favorable conditions offshore, including consistent and repeatable sources, good coupling conditions at sources and receivers, and the uniform property of water as the medium. However, offshore seismic data may have particular noise sources, especially multiple reflections, and at present most 3D offshore seismic surveys have much narrower azimuthal coverage than their onshore counterparts.

The seismic data traces collected from many receivers that have recorded the same shot, such as that shown in Figure 1.2, produce a **common shot gather** (**CSG**). A **seismic gather** refers to a group of pre-stack seismic traces linked by a common threading point. The phrase “pre-stack traces” refers to data traces retaining the original source and receiver locations; they are in contrast to the “post-stack” or “stacked traces” that result from stacking or summing many traces together.

A **common receiver gather** (**CRG**) as shown in Figure 1.3a is a collection of traces recorded by the same receiver from many shots, and a **common midpoint** (**CMP**) **gather** (Figure 1.3b) is a collection of traces with their source-to-receiver midpoint falling within the same small area, called a **CMP bin**. Among the three common types of seismic gathers, the **reflection spread**, or the lateral extent of reflection points from a seismic gather across a reflector, is zero for the CMP gather in the case of a flat reflector beneath a constant

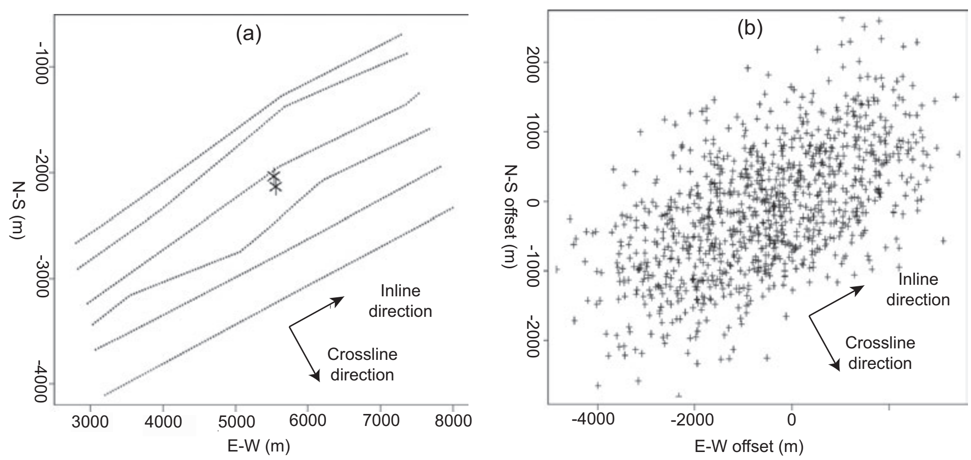

Figure 1.4 Map views of an acquisition geometry from the Canadian Rockies (Biondi, 2004). (a) Locations of shots (asterisks) and receivers (dots) for two consecutive shot gathers. (b) Offsets of 1000 traces, randomly selected.

velocity medium (Figure 1.3b). There are other gathers, such as a **common image-point (CIG) gather**, which is a collection of migrated traces at the same image bin location. Some people call a collection of traces with the same amount of source-to-receiver offset as a common offset gather, though it is logically a **common offset section**.

1.1.3 CMP binning and seismic illumination

Owing to the minimum spread of reflection points, traces of each CMP gather can be summed or stacked together to form a single **stacked trace**, A stacked trace is often used to approximate a zero-offset trace, which can be acquired by placing a shot and a receiver at the same position. The stacked trace has good signal content because the stacking process allows it to take all the common features of the original traces in the gather. Consequently, the CMP gathers are preferred to other gathers in many seismic data processing procedures. However, because the CSG or CRG data are actually collected in the field, a process of re-sorting has to be done to reorganize the field data into the CMP arrangement. This is done through a process called **binning**, by dividing the 2D line range or the 3D survey area into a number of equal-sized **CMP bins** and, for each bin, collecting those traces whose **midpoints** fall within the bin as the CMP gather of this bin. The number of traces, or midpoints, within each CMP bin is called the **fold**. As an important seismic survey parameter, the fold represents the multiplicity of CMP data (Sheriff, 1991).

Figures 1.4 and 1.5, respectively, show the geometries of two 3D surveys onshore and offshore. In each of these figures the left panel shows the locations of the shots and receivers, and the right panel shows the midpoint locations of 1000 traces randomly selected from the corresponding survey. To maintain a good **seismic illumination**, the fold should be high enough and distributed as evenly as possible over the survey area. In practice, the desire for good seismic illumination has to be balanced against the desire to make the survey as efficient as possible to reduce the cost in money and time. In 3D onshore seismic surveys,

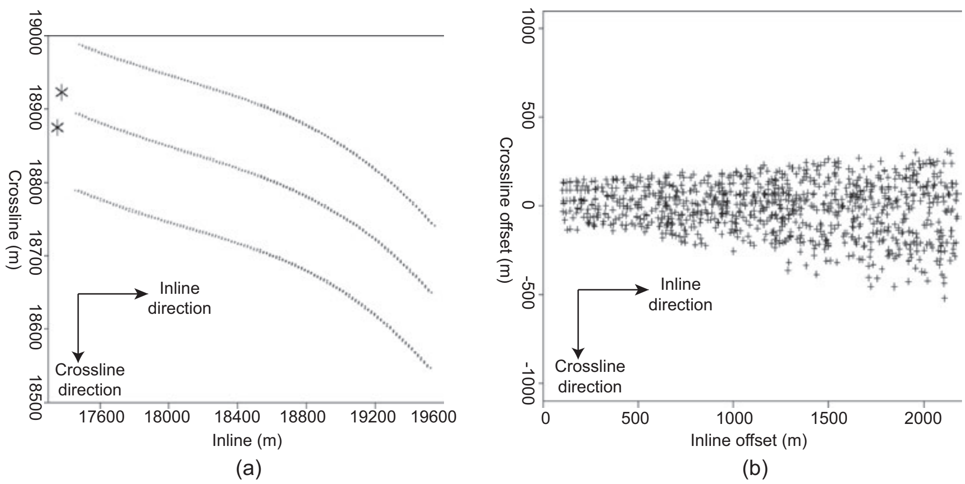

Figure 1.5 Map views of a three-streamer acquisition from the North Sea (Biondi, 2004). (a) Locations of shots (asterisks) and receivers (dots) for two consecutive shot gathers. (b) Offsets of 1000 traces, randomly selected.

the orientations of the shot lines are often perpendicular to the orientations of the receiver lines in order to maximize the azimuthal coverage of each **swath**, which is a patch of area recorded by an array of sensors at one time. Typically there is an **inline direction** along which the spatial sampling is denser than the perpendicular **crossline direction**. The inline is often along the receiver line direction, like that shown in Figure 1.4a, because the spacing between receivers is typically denser than the spacing between shots. In the case of irregular distributions of shots and receiver lines, however, the inline direction may be decided based on the distribution of midpoints of data, like that shown in Figure 1.5b.

Sometimes special layouts of shot and receivers are taken to optimize the seismic illumination. Figure 1.6 shows an example of a special 3D seismic survey geometry over the Vinton salt dome in southwest Louisiana. The survey placed receivers along radial lines and shots in circular geometry centered right over the subsurface salt diapir. In most applied sciences, quality and cost are the two main objectives that often conflict with each other, and the cost is in terms of both money and time. Because geophones today are connected by cables, they are most effectively deployed in linear geometry, such as along the radial lines in this example. The sources here were Vibroseis trucks which can easily be run along the circular paths. Similarly, in carrying out seismic data processing projects, we need to satisfy both the quality and cost objectives.

1.1.4 SNR and CMP stacking

With respect to the objectives of each project, geophysical data may contain relevant information – the **signal** – and irrelevant components – **noise**. A common goal for digital data processing in general and for seismic data processing in particular is to improve the **signal-to-noise ratio** or **SNR**. In seismology the SNR is often expressed as the ratio between the amplitude of the signal portion and the amplitude of the noise portion of seismic traces.

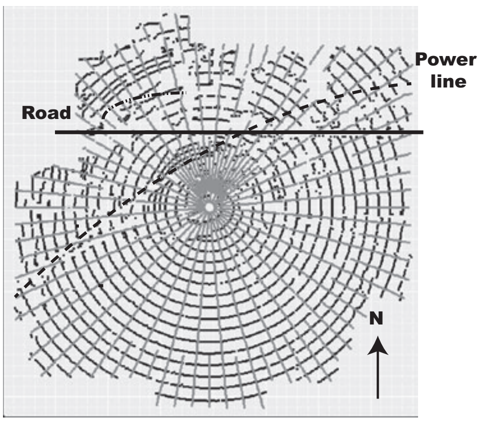

Figure 1.6 Map view of a 3D seismic survey over the Vinton salt dome in west Louisiana. The straight radial lines denote receiver positions, and the circular lines denote shot positions. The geometry of the shot and receiver layout is designed to optimize the coverage of reflection waves from the boundary of the underlying salt dome.

Box 1.1 Why use CMP stacking and what are the assumptions?

The main reason is to improve the SNR and focus the processing on the most coherent events in the CMP gather. CMP stacking is also a necessary step for post-stack migration where each stacked trace is regarded as a zero-offset trace. The assumption is there is a layer-cake depth velocity model, at least locally within each CMP gather.

In practice the meaning of signal versus noise is relative to the objectives of the study and the chosen data processing strategy. Similarly, the meanings of **raw data** versus **processed data** may refer to the input and output of each specific processing project. The existence of noise often demands that we treat seismic data from a statistical point of view.

Common midpoint (**CMP**) **stacking** (see Box 1.1) refers to summing up those seismic traces whose reflections are expected to occur at the same time span or comparable reflection depths. The main motivation for such stacking is to improve the SNR. In fact, stacking is the most effective way to improve the SNR in many observational sciences. A **midpoint** for a source and receiver pair is simply the middle position between the source and receiver. In a **layer-cake model** of the subsurface, the reflection points on all reflectors for a pair of source and receiver will be located vertically beneath the midpoint (Figure 1.7). Since the layer-cake model is viewed as statistically the most representative situation, it is commonly taken as the default model, and the lateral positions of real reflectors usually occur quite close to the midpoint. Consequently on cross-sections we usually plot seismic traces at their midpoints. Clearly, many traces share the same midpoint. In the configuration of CMP binning, the number of traces in each CMP bin is the fold.

It is a common practice in seismic data processing to conduct CMP stacking to produce stacked sections. Thus, reflection seismic data can be divided into pre-stack data and

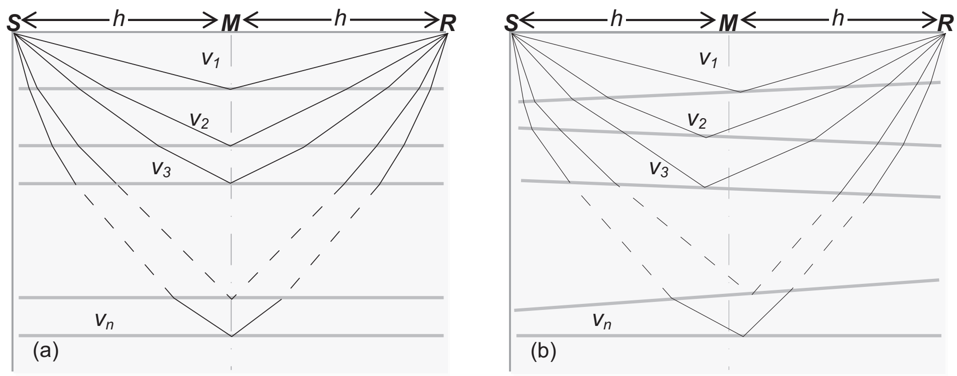

Figure 1.7 Reflection rays (black lines) from a source *S* to a receiver *R* in (a) a layer-cake model; and (b) a model of dipping layers. All reflection points are located vertically beneath the midpoint *M* in the layer-cake model.

post-stack data, and processing can be divided into pre-stack processing and post-stack processing. The traditional time sections are obtained through the process of stacking and then post-stack migration. Modern processing often involves pre-stack processing and migration to derive depth sections that have accounted for lateral velocity variations and therefore supposedly have less error in reflector geometry and amplitude. One can also conduct depth conversion from time section to depth section using a velocity–depth function. Post-stack seismic processing is cheaper and more stable but less accurate than pre-stack seismic processing. In contrast, pre-stack seismic processing is more costly, often unstable, but potentially more accurate than post-stack seismic processing.

1.1.5 Data processing sequence

The primary objective of this book is to allow the reader to gain a comprehensive understanding of the principles and procedures of common seismic data processing and analysis techniques. The sequence of processing from raw seismic data all the way to final forms ready for interpretation has evolved over the years, and many general aspects of the sequence have become more-or-less conventional. It is a non-trivial matter to design a proper sequence of seismic data processing, called a **processing flow**. Figure 1.8 shows an example of a processing flow for reflection seismic data more than 30 years ago (W. A. Schneider, unpublished class notes, 1977). The general procedure shown in this figure still holds true for today's processing flow for making post-stack sections.

The goal of seismic data processing is to help interpretation, the process of deciphering the useful information contained in the data. The task is to transfer the raw data into a form that is optimal for extracting the signal. The word "optimal" implies making the best choice after considering all factors. Hence we need to make decisions in the process of seismic data analysis. All methods of seismic data analysis rely on physical and geological theories that tie the seismic data and the geological problem together. For instance, a problem of inferring aligned fractures may involve the theory of seismic anisotropy. The subsequent data processing will attempt to utilize this theory to extract the signal of the fracture

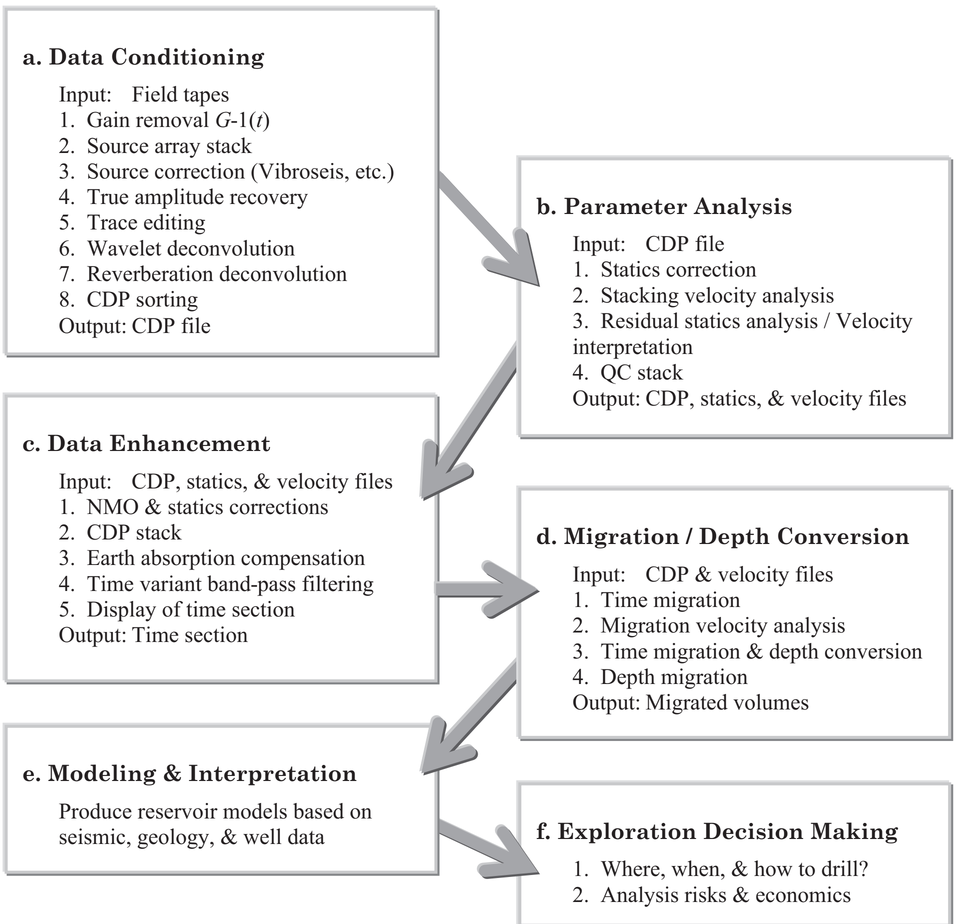

Figure 1.8 A general processing flow, after Schneider (unpublished class notes from 1977). Steps **c**, **d**, and **e** are usually iterated to test different hypotheses. Pre-stack processing is often conducted after a post-stack processing to help the velocity model building process. There are also reports of pre-stack processing using limited offsets to increase the efficiency.

orientation according to the angular variation of traveling speed, and to suppress the noise that may hamper the signal extraction process.

Exercise 1.1

1. How would you estimate the fold, the number of the source-to-receiver midpoints in each CMP bin, from a survey map like that shown in Figure 1.6? Describe your procedure and assumptions.

2. As shown in Figure 1.7, the shapes of reflection raypaths tend to resemble the letter "U" rather than the letter "V". Explain the reason behind this phenomenon.

3. Update the processing flow shown in Figure 1.8 by finding and reading at least two papers published within the past 10 years. What happened to those processing steps in Figure 1.8 that are missing from your updated processing flow?

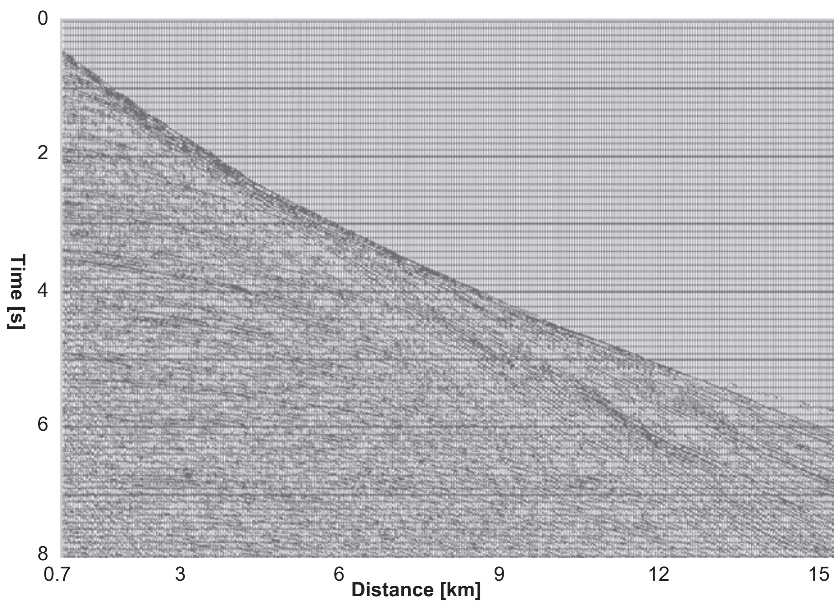

Figure 1.9 A common shot gather from an offshore seismic survey.

1.2 Sampled time series, sampling rate, and aliasing

Through their propagation history, seismic waves vary in a continuous manner in both temporal and spatial dimensions. However, measurements of seismic data need to be sampled into digital form in order to be stored and processed using computers. At the acquisition stage each trace of seismic wiggles has been digitized at a constant sample interval, such as 2 ms (milliseconds). The resulted string of numbers is known as a time series, where the number represents the amplitude of the trace at the corresponding sample points. In the following, some basic properties of the sampled time series are introduced.

1.2.1 Sampled time series

Figure 1.9 shows an example of offshore seismic data for which the streamer of hydrophones is nearly 20 km long. We treat each recorded seismic trace as a **time series**, which is conceptualized as an ordered string of values, and each value represents the magnitude of a certain property of a physical process. The word “time” here implies sequencing or connecting points in an orderly fashion. A continuous geological process may be sampled into a discrete sequence called a **sampled time series**. Although the length of the sample is usually finite, it may be extrapolated to infinity when necessary. All the data processing techniques discussed in this book deal with sampled time series. A 1D time series is usually taken to simplify the discussion. However, we should not restrict the use of time series to just the 1D case, because there are many higher-dimensional applications.