Projection before picking: a statistical approach to tomographic imaging

David C. Henley

ABSTRACT

The goal of tomographic imaging with seismic data is to create a velocity image of a portion of the earth from a set of transit time values for some seismic event, as measured along a multitude of diverse paths through the target medium. These transit time values, one for each seismic path, are subject to various estimation errors, which can be quite large in the case of mis-identified seismic events. Since the process of picking the transit times is usually fraught with error, we propose an imaging method wherein the transit time variable is first associated with a unique probability density function for each input trace, a 3-D volume of probability-weighted transit times is created, and the volume is back-projected to yield a probability-weighted image volume which is less noisy than the raw data volume. Each surface of constant probability within the image volume constitutes a potential tomographic transit time image. Interpretation, in this context, then becomes the extraction, by one of several possible processes, of an image surface from the back-projected volume. Hence, by interchanging the processes of back-projection and picking, we avoid explicit mis-picks on noisy data.

INTRODUCTION

Tomography

Conventional tomography constitutes a family of methods which image the variations of a physical property within a solid body from measurements of the integral of the property along various paths through the body. The simplest kind of tomography, which was first implemented with x-rays for medical purposes, is back-projection tomography, in which physical properties internal to a body can be mapped by combining measured "projections" of energy through the body at all azimuthal angles. When the variations of the imaged physical property within the body are small, it is assumed that the energy transmitted through the body follows straight line trajectories. In this case the imaging of the variations within the body can be done by a simple process called "back-projection", which is a simple algebraic method that can be iterated to improve accuracy. A typical seismic application of tomography is the computation of an image of velocity (or slowness) variations within a solid body from measurements of transit times of seismic energy through the body in all possible directions. Projection tomography requires redundancy in the measured data; the accuracy with which the local velocity of a particular parcel within the body can be estimated depends strongly upon the number of raypath segments traversing the parcel at significantly different angles.

Because of the physical constraints which usually control the acquisition of a tomographic velocity survey (sources and receivers confined to two bounding boreholes for a cross-well survey, for example), the various local regions within a solid body are rarely sampled uniformly by raypath segments. Usually, the regions nearest the centre of the body are the best sampled, and those nearest any boundaries are the least well sampled. Complicating the picture further is the fact that classical back-projection tomography is only suited for imaging bodies which exhibit only small variations from uniformity; strong velocity contrasts cause significant deviation of raypaths. Physically, this means that high velocity regions within the body are over-sampled, and low velocity regions are under-sampled, or even bypassed altogether by the probing seismic energy. A final consideration is the fact that the directional resolution of a tomographic image can often vary considerably due not only to non-uniform sampling imposed by survey parameters, but also due to the internal structure of the body itself. A good example is a cross-well survey in which sources and receivers can only be deployed in two boreholes bounding the surveyed earth section. Here, the vertical resolution is normally much better than the horizontal resolution, simply because there are few raypaths between the two boreholes that are steep enough to properly sample the earth laterally. In addition, since the velocity variations in the earth are usually greatest in the vertical direction due to layering, raypaths tend to align with the layers, further reducing lateral resolution.

To deal with the complications of tomography caused by non-ideal acquisition geometry and by violation of the “weak scattering” criterion (velocity variation of only a few percent) by most earth materials, many tomographic imaging techniques have been developed, usually involving repeated raytracing and modification of a model. Almost universally, however, these methods require input data with a very low level of noise; significant mis-picks of the energy arrival times are not well tolerated and at the very least manifest themselves as significant artefacts in the image. In some cases, a small percentage of significantly erroneous input data values can cause total failure of an imaging algorithm, particularly one which uses raytracing, since the raytracing simply cannot honour inconsistent transit time values. Because of the vulnerability of tomographic imaging to input data errors, the accurate “picking” of the input data is crucial for success.

Picking

Unlike other types of seismic imaging which use the entire set of wavefield amplitudes as recorded on seismic traces, conventional tomographic imaging uses only a single discrete time value selected from each seismic trace according to some specified criteria involving the trace amplitudes. The selected time value is usually the “arrival time” of a particular seismic wave mode, often the “direct arrival”, or the compressional wavefront which has traversed the most direct path through the section of earth being probed. In simple cases, the direct arrival is also the “first arrival” or the earliest seismic energy to appear on a trace, which makes its detection and subsequent picking easier. In many cases, however, the desired direct arrival follows the first arrival by a significant delay and is considerably harder to detect and to pick accurately. If a tomographic survey requires some other arrival (the direct shear mode, for example), none of the desired event arrivals will be first arrivals, and all of them will be hard to pick accurately.

Many algorithms have been developed to automatically pick events on seismic traces, but few of them are as accurate or consistent as the eye of an experienced geophysicist when picking arrival times for tomographic imaging. Only a geophysicist can reliably determine, for example, when a direct arrival is not a first arrival and subsequently find and pick the correct event. Nevertheless, even the relatively simple task of manually picking first arrivals can be challenging. Because of the variable quality of recording on individual seismic traces, and the different levels of background noise usually present, it is notoriously difficult to accurately and consistently locate the actual desired "first breaks" of arriving seismic energy on a set of seismic traces. Because of this, on some data sets, seismic events are picked on their first troughs, or first peaks. This introduces an error on each pick time, but one that is hopefully constant for a given data set.

The difficulty of picking seismic events is a consequence of the fact that all seismic wavefield data are bandlimited and invariably contaminated with noise. If there were some way of reducing noise, far more consistent sets of event arrival times could be obtained, thus improving the tomographic imaging of such data. What we propose in the following is a method wherein the "picking" operation *before* tomographic imaging is replaced by an "interpretation" operation *after* imaging. We use the averaging within a particular tomographic imaging algorithm to reduce the noise intrinsic to the input data. The act of picking seismic events on raw data, which is a fundamental interpretation step, is replaced by a somewhat different interpretation step following the imaging process.

PROBABILITY IMAGING

Statistics and the act of picking

The act of selecting a transit time for a discrete seismic event, or "picking" the event, consists of choosing a single time value from a range of possible values, based on an interpretation of the waveform of the event as it appears on a time series seismic trace. Whether the process is manual or automatic, various criteria applied to the shape of the waveform guide the process to select the most "likely" value for the transit time. Now suppose that the arrival time of a seismic event can be represented by a probability density function over some interval of possible transit time (Rothman, 1984a, 1984b), based on the event bandwidth and waveform, the level of additive noise, and any other known sources of uncertainty. In this context, the transit time is considered a random variable, and its expected value can thus be expressed using the standard statistical formula for the expected value of a discretely distributed random variable, T,

$$\overline{E}(T) = \sum_i T_i p_i(T_i), \tag{1}$$

where $\sum_i p_i(T_i) = 1$, and $p_i(T_i)$, the probability density function, is the probability that the transit time, T_i, is the true arrival time. Thus, if we assert that a picked transit time on a seismic trace can be represented by an *expected value* computed from a transit time probability density function, we can replace each picked time in subsequent arguments with the summation formula for its expected value:

$$T_{j,k} = \overline{E}(T_{j,k}) = \sum_i T_{i,j,k} p_{i,j,k}(T_{i,j,k}), \tag{2}$$

where $T_{j,k}$ is the *j*-th transit time pick on the *k*-th common offset (or common angle) gather, $T_{i,j,k}$ is a discrete transit time value in some time window (index *i*) that includes all possible transit time values, and $p_{i,j,k}(T_{i,j,k})$ is the probability that any individual time value, $T_{i,j,k}$, is the true transit time pick at position *j* on the *k*-th common offset gather. The probability density function $p_{i,j,k}\left(T_{i,j,k}\right)$ is normalised for each individual time pick:

$$\sum p_{i,j,k}\left(T_{i,j,k}\right) = 1 \tag{3}$$

Figure 1 is a schematic illustration of this concept. In this figure, a seismic event arrival is shown, with a low level of noise (or an earlier event) preceding it. If T_0 is the actual arrival time of the event, the discrete pick (Figure 1b) must align with it to be correct. In the presence of noise, or the gradual onset of the seismic event, however, a picking algorithm or interpreter might erroneously place the pick at an earlier or later zero crossing on the trace. Using a probability density function of arrival times, however, as in Figure 1a, allows for all possible picks, thus avoiding early commitment to a possibly erroneous pick which could degrade an image.

In this setting, a noiseless seismic event of infinite bandwidth would have a density function, $p_{i,j,k}\left(T_{i,j,k}\right)$, consisting of a single unit spike at the exact transit time, and computing the expected value from the formula will simply yield the single exact transit time. Seismic events with finite bandwidth and/or non-zero additive noise will have probability functions whose width reflects the uncertainty of the transit times. Figure 2 shows schematically the character of probability density functions associated with seismic events with various combinations of bandwidth and additive noise levels. The "discrete pick" function (Figure 2a) can only be justified by a noiseless event of infinite bandwidth, i.e. a "model" event. All other events imply some level of uncertainty and hence density functions of some finite width (Figures 2b-2d).

Back-projection imaging from transit times

Although other types of tomographic imaging algorithms should be considered in the context of probabilistic time picks, we start with the simplest algorithm of all, back-projection, since it makes our argument more straightforward and intuitive. The basic input data for tomographic imaging of any kind are transit time picks $T_{j,k}$, where the indices *j, k* correspond to the *j*-th position on the *k*-th "projection slice" of the input data. To relate this terminology to more familiar terms, a "projection slice" in a cross-well survey is any trace gather for which the raypaths are parallel (or share a common source-receiver offset). The index *j* thus usually identifies midpoint depth, and the index *k* identifies the common source-receiver offset, or, equivalently, the "projection angle" between the boreholes in the survey.

The Radon projection theorem states that an image can be reconstructed from "projection slices" by "back-projecting" them over the full range of projection angles. For a typical cross-well seismic survey, this means arranging the transit time picks for the survey in a 2-D array whose indices are source-receiver midpoint depth and source-receiver offset. The time picks for each common-offset gather (projection slice) are adjusted for moveout corresponding to a single "background" velocity. This step effectively equalizes the raypath lengths for the various projections and leaves only the local variations about the background velocity which constitute the image information of interest. After correction for moveout, the back-projection proceeds by summing the projection slice array along sets of parallel linear trajectories mirroring the various projection angles present in the initial survey. Each summation results in a "back-projection" slice; and these slices, ordered by projection angle, constitute a 2-D array that is the reconstructed image. The back-projection process is an algebraic analog of the physical "forward projection" process (seismic crosswell survey) that yielded the input data.

A typical back-projected image tends to under-estimate the true image contrasts; so the back-projection process is often iterated by forward-projecting the first estimated image, subtracting the resulting projection slices from the corresponding projection slices of the raw data, repeating the back-projection, and adding the incremental results to the first image estimate. This method, known as "iterated back-projection" is quite robust in the face of noisy or inconsistent input data. Like all methods which do not trace raypaths through the estimated image, iterated back-projection will show systematic errors for surveys in which the imaged properties vary from the background by more than a few percent; but, on the other hand, it cannot go wildly astray due to untraceable raypaths. For our purposes, we choose simple non-iterated back-projection. The *l*-th time value of the *m*-th back-projection slice, $T'_{l,m}$, is computed from the input projection slice data, $T_{j,k}$ as follows:

$$T'_{l,m} = \sum_{k=1}^{h} \frac{T_{i,l+(k-h/2).m.\delta,m}}{h} \quad , \tag{4}$$

where δ is the incremental crosswell source-receiver offset between successive back-projection slices and *h* is the number of projection slices in the input data set.

Interchanging picking and projection

As discussed earlier, the most difficult part of tomographic imaging is usually obtaining a set of input transit times that is both consistent and error-free. It is highly desirable to find a way to obtain more robust input values for the imaging process. In processing seismic data, we often use data redundancy to allow us to average noisy input data and thus improve results. The time picks included under the summation in equation (4) are usually considered to be the raw transit times picked on single traces, either manually, or by use of some type of threshold-based algorithm, and are affected by all the usual problems, including significant mis-picking. Consider what happens, however, if we replace these explicitly picked times with their expected values as defined by equation (2) above:

$$T'_{l,m} = \sum_{k=1}^{h} \overline{E}\,\frac{\left(T_{l+(k-h/2)\cdot m\cdot\delta,m}\right)}{h} = \frac{1}{h}\sum_{k=1}^{h}\sum_{i=1}^{n} T_{i,l+(k-h/2).m.\delta,m}\cdot p\left(T_{i,l+(k-h/2).m.\delta,m}\right), \tag{5}$$

where *i,* a time index, is summed over a discrete window interval of *n* values, within which the pick time is assumed to occur (this means that each sum over *n* transit time probabilities is unity). The choice of window length is the only subjective interpretive act in this process. Note that by expanding the back-projection formula in this fashion, we have added an extra dimension (time) to the computation, so that we are now dealing with a 3-D volume of (unpicked) transit times and their corresponding probability densities as the input data set for tomographic imaging.

There is nothing in equation (5) that dictates the order in which we carry out the double summation, so we interchange the order as follows:

$$T'_{l,m} = \sum_{i=1}^{n} \frac{1}{h} \sum_{k=1}^{h} T_{i,l+(k-h/2).m.\delta,m} . p\left(T_{i,l+(k-h/2).m.\delta,m}\right). \tag{6}$$

Now, since i is the index of transit times within a fixed window, $T_{i,j,k}$ values are all equal for a given i , so we can simply call the common value T_i and carry it outside the inner summation:

$$T'_{l,m} = \sum_{i=1}^{n} \frac{T_i}{h} \sum_{k=1}^{h} p\left(T_{i,l+(k-h/2).m.\delta,m}\right). \tag{7}$$

Now the inner summation describes just the back-projection of an array of probability densities, which will result in another array of probability densities which we shall call $p_i\left(T_{i,l,m}\right)$. Substituting into equation (7), we get

$$p_i\left(T_{i,l,m}\right) = \frac{1}{h} \sum_{k=1}^{h} p\left(T_{i,l+(k-h/2)\cdot m\cdot\delta,m}\right), \tag{8}$$

$$T'_{l,m} = \sum_{i=1}^{n} T_i . p_i\left(T_{i,l,m}\right) = \overline{E}\left(T_{l,m}\right), \tag{9}$$

which is in the form of an expected value. Hence, the substitution of the expected value expansion for each raw transit time pick has led to an expression for an “expected value” transit time image. By deferring the initial and often unreliable task of picking transit times and instead describing raw transit times with probability functions, we perform the imaging step using probabilities only and perform the interpretation or “picking” of the image only as the final step of evaluating the expected value of the transit time for each location, *l,m* in the image grid.

Probability functions

In the preceding section, we showed how to avoid picking transit times on raw data traces by assigning a probability density function to the range of possible transit times for each input trace and imaging the probabilities, relegating the interpretation/picking step to the end of the process in order to benefit from the smoothing inherent in the imaging process. While equation (9) expresses one way to obtain a final image (expected value computation), it is by no means the only possibility. Another perfectly valid method for extracting an interval transit time image from the back-projected 3-D probability array is to select the transit time corresponding to the maximum of each image probability density function. Other possibilities include choosing the transit time corresponding to the median value of each probability density function, or the transit time position of the mode of the probability density function. Which method is chosen will depend on the credibility of the probability density functions we choose to characterize the input transit times. In an earlier section we assumed without much discussion that a probability function could somehow be found to describe an uncertain transit time pick on a seismic trace. In fact, the success or failure of probabilistic imaging depends crucially upon this choice of probability density function and whether the function can be reliably created for each trace in a data set.

In a companion chapter (Henley, 2004), we demonstrate using the cross-correlation function on input seismic traces to help construct probability density functions for relative delay times of reflection events. The current paper, however, focuses specifically upon tomographic imaging from direct arrival events. These events are typically asymmetric in time, with the desired transit time lying distinctly to the early side of the event waveform. Furthermore, the event waveform, amplitude, and bandwidth often vary considerably from trace to trace, making it difficult to correlate consistently enough to create good probability density functions. One approach which has been tried with limited success is to compute various ad hoc envelope functions for each trace in order to remove some of the ambiguity of the direct arrival event. Empirical experience has shown that an envelope function which embodies both seismic event energy and frequency content can be consistent enough from trace to trace on a crosswell survey to yield useable probability functions. In the case of an envelope function, however, since time averaging of the input signal occurs, the resulting probability function resembles a probability *distribution* function or *cumulative distribution* more than an actual probability density function. This is not necessarily a problem, since cumulative probability functions can be imaged just as well as regular probabilities and have the added advantage that the intrinsic summation will further reduce noise. The only difference is in how we interpret a cumulative probability function. Instead of the function value being the probability that its corresponding transit time is the actual transit time of the seismic event, the cumulative probability value is the probability that the seismic event has a transit time less than or equal to the given transit time (see Figure 3). This suggests another way to perform an interpretation of the 3-D probability space resulting from the back-projection step. We can select from each trace the transit time corresponding to the same constant value of cumulative probability (for instance, 0.5). The choice of 0.5 would be equivalent to selecting the mode of the unsmoothed probability functions. One advantage of imaging with cumulative distribution functions rather than probability density functions is that any probability value can be chosen to select an image surface. The images for various probabilities can be compared to determine which seems most consistent, or even averaged (similar to computing the expected value).

A SUGGESTIVE EXAMPLE

We have not yet implemented a software procedure at CREWES for probability imaging; but Figure 4 shows a crosswell image created for Shell Canada using a method similar to that described above. For these data, an envelope function was constructed, which, when computed for each trace, yielded time series which bore a strong resemblance to cumulative probability distribution functions. The envelope functions for the entire data set, sorted as projection slices, were back-projected to create a 3-D cumulative probability space with coordinates of well depth, well-to-well distance, and transit time (actually interval transit time). Then a commercial rendering program was used to find and render a surface through all points of constant probability. In this image, the distance from the front plane to the rendered surface represents the interval transit time of the particular parcel of earth corresponding to the local coordinates. Though not shown, the sonic logs from the two wells involved in the survey correspond reasonably well to the transit times shown along both the right and left edges of the tomogram surface. The prominent diagonal feature is an artifact likely caused by a systematic error in receiver depth in one of the wells. Any number of similar tomograms can be created from the probability volume simply by choosing a different probability value for selecting the transit times to render.

The real potential of this method was suggested by another experiment with this same data set, which is not shown here. The presence of strong shear wave arrivals in the data set was noted early in its analysis; but the arrivals weren't clean enough to pick consistently, either manually or automatically. Fortuitously, however, it was possible to construct envelope functions which were consistent enough to image. Once again, the image was interpreted by rendering the transit time values at a specific value of cumulative probability. The image, while not as complete or consistent as that in Figure 4, was the only one which could be constructed for the transmitted shear waves in this cross-well survey.

EXTENSIONS

We have shown a suggestive example using probability as an imaging tool and justified it with some mathematics which pertain, strictly speaking, only to the back-projection method of tomographic reconstruction. Is the method more generally applicable than this? Can it be used, for example, when techniques other than back-projection are used for imaging? What functions are best for simulating probability functions or cumulative probability functions? What are the best methods for extracting a tomogram from an image volume? These and other questions remain to be investigated.

CONCLUSIONS

Tomographic imaging methods are sensitive to inaccurate input values, but it is notoriously difficult to obtain accurate transit time picks on seismic traces. One partial solution to this dilemma is to use probability density functions as the picks, generated from the data themselves, and to form an image space from these functions. The resulting image volume can be "interpreted" by one of a number of statistical methods, each aimed at extracting a surface from the image space. A statistical justification has been presented to support the method, and proof-of-principle demonstrated with a crosswell tomographic image created by a rudimentary form of the method.

REFERENCES

Henley, D.C., 2004, A statistical approach to residual statics removal, CREWES 2004 Research Report, Volume 16.

Rothman, D., 1984a, Probabilistic residual statics, Stanford Exploration Project Report SEP-37.

Rothman, D., 1984b, Nonlinear inversion by stochastic relaxation with applications to residual statics estimation, Stanford Exploration Project Report SEP-38.

ACKNOWLEDGEMENTS

The author gratefully acknowledges the support of CREWES sponsors and staff. Pat Daley was instrumental in discussing and reviewing the manuscript and its concepts, but any errors or misconceptions are solely those of the author. Thanks to Shell Canada Ltd. For use, in Figure 4, of an image which originally appeared in the final issue of Shell's Sphere magazine in 1990.

FIGURES

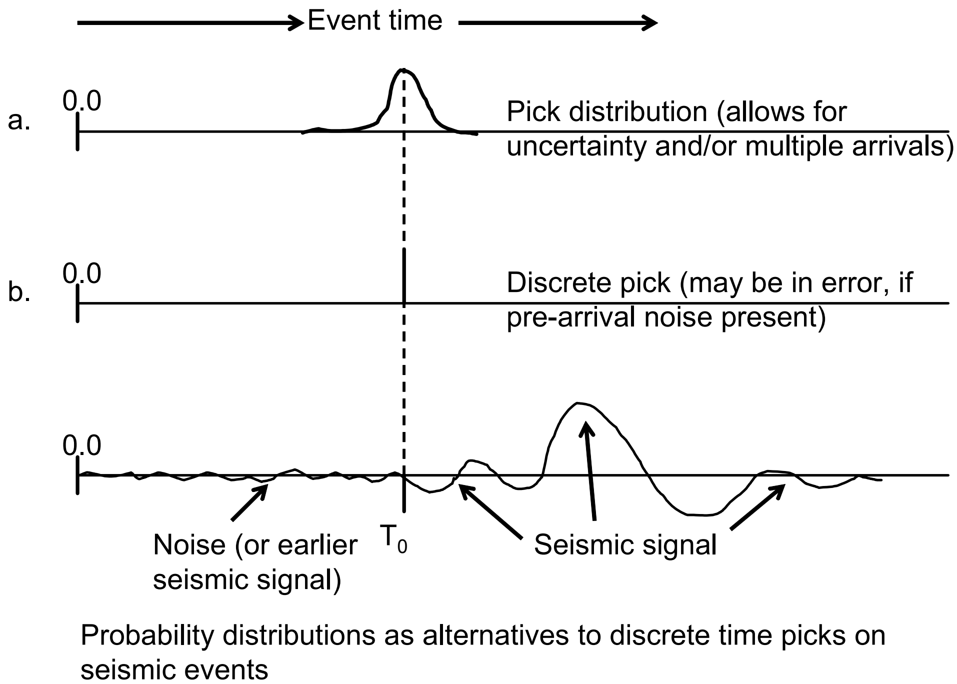

FIG. 1. Instead of associating a single discrete time pick (b), possibly erroneous, with the first break of a seismic event arrival, a distribution of picks (a), including all credible arrival times, can be used to represent the arrival.

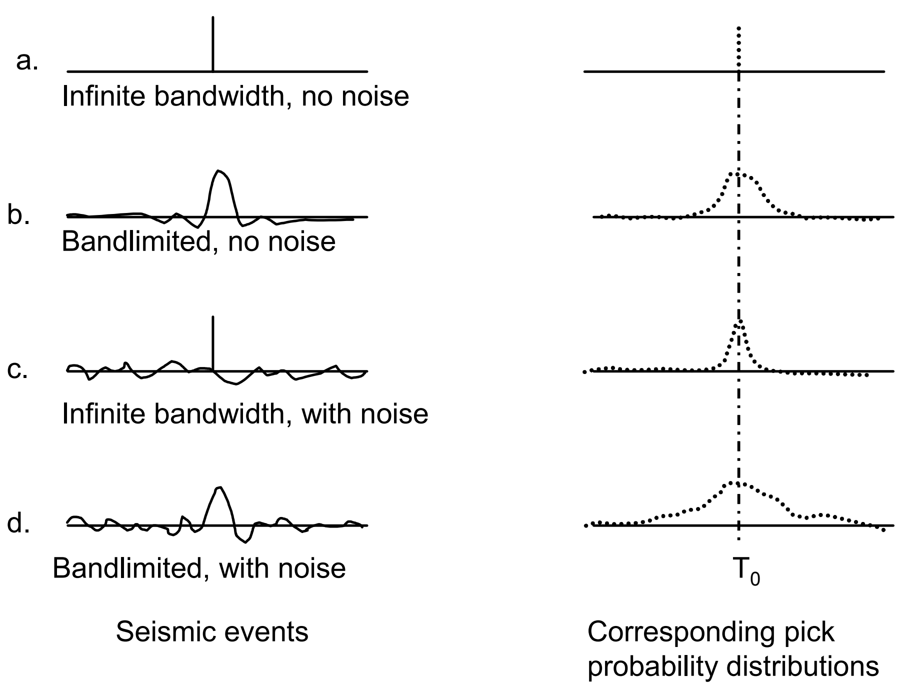

FIG. 2. Pick probability distributions associated with seismic events of varying bandwidth and noise level

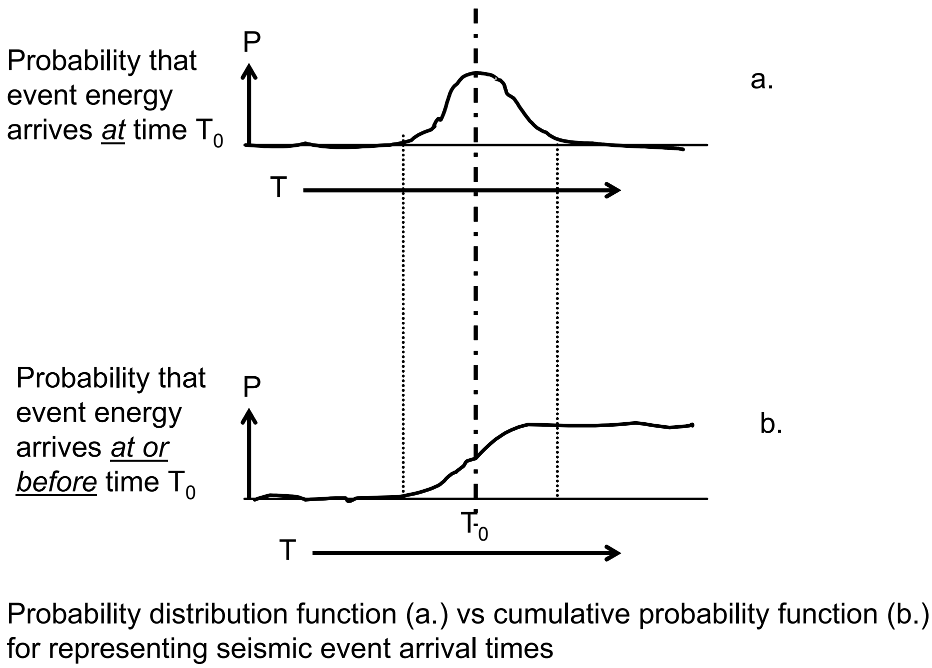

FIG. 3. The relation between a pick probability distribution (density) function (a) and a pick cumulative probability function (b).The latter is more useful for tomographic imaging and is just the integral (running sum) of the former.

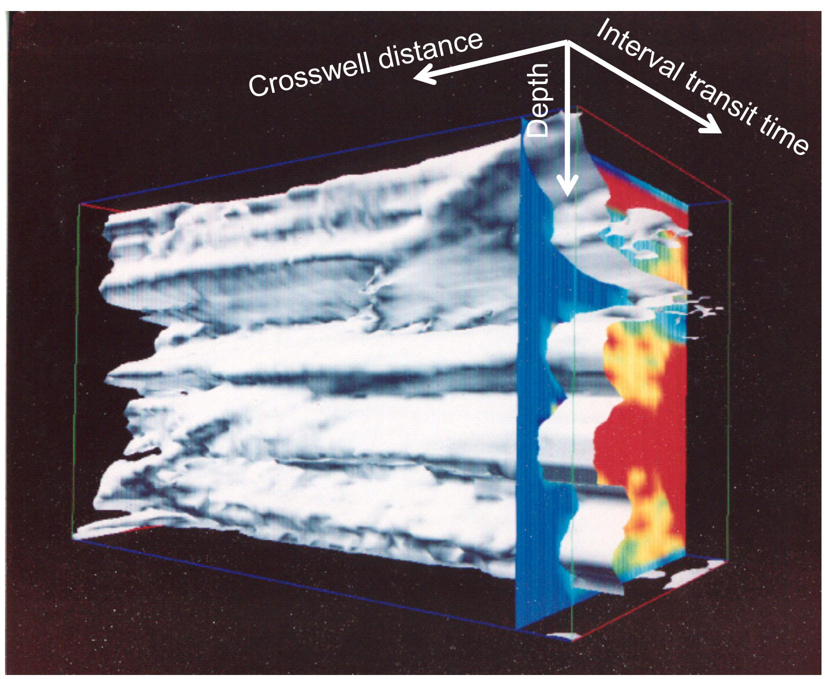

FIG. 4. A probability crosswell interval transit time tomogram. A surface of constant probability has been rendered from a transit time probability volume. (Image courtesy of Shell Canada Ltd.)