Chapter III
Well Log Deglitching and Seismogram Inversion

This chapter uses the Markov model of impedance proposed in Chapter I to construct mathematical models for well log deglitching and seismogram inversion. In the latter case, a deconvolution, free of any source waveform and multiples, is converted into an impedance. In the former case, it is possible to reduce the effect of measurement noise in well logs.

The estimators, subsequently developed here, have memory, and in this aspect differ considerably from those of Chapter II. In addition, we adopt the maximum a posteriori estimator vs. the L-1 and L-2 estimators of the previous chapter. Physically, the first estimator corresponds to minimizing the entropy in a Markov chain, subject to the constraint that the chain must be consistent with some observed data. Before discussing the main application, that of seismogram inversion, the deglitching model will be presented. Most of the concepts in this chapter are covered in this example.

3.1. Well Log Deglitching

Assume that a well log, denoted by w_j, where j is the time index, is to be quantized into M states. Furthermore, assume that the well log contains noise glitches (measurement errors, background noise, etc.) and that these glitches are to be excluded in the quantization process. It seems reasonable that the quantized well log should resemble a Markov chain characterized by some probability transition matrix (PTM). Denoting the quantized log by z_j, we have the following model:

$$\text{Model: } W_j = Z_j + N_j \qquad (3.1)$$

Given: well log $\underline{w}$, PTM for Z_j, distribution of $\{N_j\}$

Desire: $\underline{\tilde{z}}$

As before, upper case refers to random variables (RVs), lower case to values, and underscoring represents a vector. In general, Z_j is a discrete RV whereas W_j is continuous. For the time being, assume W_j is also discrete (i.e. the original well log has already been quantized and a version free of noise glitches is desired). Temporarily replacing vectors by scalars, we face the problem of computing an estimate of z, denoted by $\tilde{z}$, given an observation W (=5), where the two RVs Z and W are related through the function Pr(Z,W). When W is arbitrary, Pr(Z,W) represents a two-dimensional surface, but fixing one of the variables (W=5) slices this surface and results in the one-dimensional function Pr(Z|W=5). Perhaps the most intuitive estimator of z is to search for the particular z that maximizes the latter function. Since Pr(Z|W=5) is the distribution ***after*** data has been observed, this estimator is called the ***maximum a posteriori*** (MAP) estimator. In the circumstance that Pr(Z|W=5) is symmetric, the MAP estimator is identical to both the conditional mean and median of the last chapter. We have a slightly more general problem in that a vector of observations ($\underline{w}$) is given, rather than simply a scalar (w). Extending the above analysis, the MAP estimator of $\underline{z}$ is defined by solving the optimization problem:

$$\underline{\tilde{z}}(\underline{w}) = \max_{\underline{z}} \Pr(\underline{Z}|\underline{W}=\underline{w}) \tag{3.2}$$

Using Bayes' rule, equation (3.2) can be rewritten as

$$\underline{\tilde{z}} = \max_{\underline{z}} \left[\frac{\Pr(W|Z)\Pr(Z)}{\Pr(\underline{W})}\right]$$

The maximization is accomplished by varying $\underline{z}$, hence $\Pr(\underline{W})$ can be ignored. Also, the estimator is unchanged if a monotone function is applied to the term in brackets. Choosing the logarithm, we have

$$\underline{\tilde{z}} = \max_{\underline{z}} \left\{\ln\left[\Pr(\underline{W}|\underline{Z})\right] + \ln\left[\Pr(\underline{Z})\right]\right\} \tag{3.3}$$

The reason for using Bayes' rule is now apparent, for

$$\Pr(\underline{Z}) = \Pr(Z_0) \prod_{1}^{N} \Pr(Z_j|Z_{j-1}) \tag{3.4a}$$

$$Pr(\underline{W}|\underline{Z}) = \prod_{0}^{N} Pr(W_j|Z_j) \qquad (3.4b)$$

Both the Markov property of $\{Z_j\}$ and the fact that given $\{Z_j\}$ the RVs $\{W_j\}$ are independent (assuming white noise) have been used in equation (3.4). Substituting equation (3.4) into (3.3) gives.

$$\tilde{\underline{z}} = \min_{\underline{z}} \left\{ \ln\left[\frac{1}{Pr(Z_0)}\right] + \sum_{0}^{N} \ln\left[\frac{1}{Pr(W_j|Z_j)}\right] + \sum_{1}^{N} \ln\left[\frac{1}{Pr(Z_j|Z_{j-1})}\right] \right\} \qquad (3.5)$$

The term $Pr(W_j|Z_j)$ in equation (3.5) represents the probability mass function of the noise. Physically, it corresponds to the reliability of a well-logging device. Different functions can be used depending on the type of rock (state) the device is operating in. The particular function we've chosen is illustrated in Figure 3.1. It means that either the correct or wrong measurement (with probability ratio S) is recorded. The term $Pr(Z_j|Z_{j-1})$ represents the PTM of the chain $\{Z_j\}$ and $Pr(Z_0)$ refers to the probability mass function.

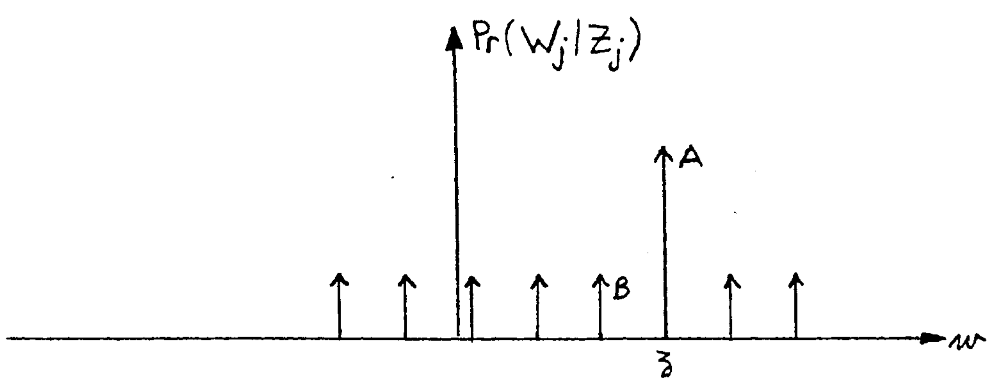

Figure 3.1. Given that the actual impedance is z, the probability that w is actually observed is shown above. The probability of making no error is large (=A) but the probability of making a large error or a small error is identical (=B). The signal to noise ratio, denoted by S, is defined as the ratio of A to B.

The algorithm developed to minimize equation (3.5) was invented by Viterbi [see Forney (1973)] and is commonly called the Viterbi Algorithm (VA). The VA uses dynamic programming to choose the sequence of states, such that the RHS of equation (3.5) is minimized. Equating inverse probability with cost (penalty), the VA finds the path having the minimum cost. Assuming diagonally dominant PTMs, changes of state are attained only at a high cost. Balancing this tendency to remain in the same state is the cost of choosing a state at time j that doesn't match the observation w_j. The VA steers a path through these conflicting requirements.

3.2. The Viterbi Algorithm

To understand the VA, we rewrite equation (3.5) in an explicit form involving the states $(s_1, s_2, \ldots, s_M)$:

$$\underline{\tilde{z}} = \min \sum_j (M_{kj} + T_{kij}) \qquad (3.6a)$$

$$M_{kj} = -\ln\left[Pr(W_j=w_j|Q_j=s_k)\right] - \ln\left[Pr(Q_j=s_k)\right]\delta_{j0} \qquad (3.6b)$$

$$T_{kij} = -\ln\left[Pr(Q_{j+1}=s_i|Q_j=s_k)\right] \qquad (3.6c)$$

Note that in equations (3.6b) and (3.6c), the RV Z_j has been replaced by the outcome function Q_j. The sequence of states must be chosen to minimize equation (3.6a). To facilitate the following discussion, consider the state-time lattice illustrated in Figure 3.2.

At time j=0, the cost function, c_{ij}, is initialized for each state i:

$$c_{i0} = M_{i0}$$

To update the cost function, isolate s_i at time j=1, and scan the entire state space at time 0 to find the path that minimizes the cost (minimal path):

$$c_{i1} = M_{i1} + \min_k (M_{k0} + T_{ki0}); \; i=1,2,\ldots,M \qquad (3.7)$$

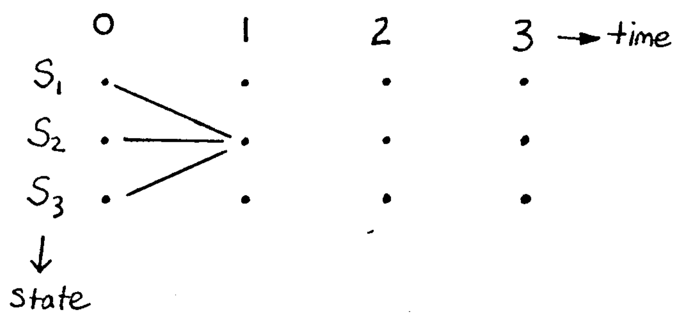

Figure 3.2. A state-time lattice for M=3. At time 1, state 2 has been isolated. The minimal path is found after scanning the entire state space at time zero.

This step fixes the computational cost at M^2 per time step. Equation (3.7) is repeated at every time step, i.e.

$$C_{i,j+1} = M_{i,j+1} + \min_k (M_{kj} + T_{kij})$$

Note that the algorithm still hasn't chosen the optimal sequence of states. The optimal sequence is chosen when (a) the end of the log is reached, or (b) a "knot" is encountered. At the end of the well log, the state having the minimal cost is chosen (assuming a knot does not exist). A "knot" is specified by demanding that the state sequence terminate (and is then re-initiated) at a given state at a particular point in time. The remaining states in the optimal sequence are then determined by passing backwards through the lattice using the minimal paths determined in the forward pass. There is one other instance when a segment of the optimal state sequence can be chosen. Say at time j the minimal path to ***every*** state is initiated by a particular state (call it s_q) at time j-1. The lattice is said to have "merged" and the first j-1 states of the sequence are therefore fixed.

To illustrate the VA applied to deglitching, Figure 3.3 shows the result of applying the VA to the three quantized logs of Chapter I. The PTM used in each case was the appropriate telegraph PTM and S=50. Note that some of the noise glitches removed are actually quantization noise created by the initial quantization of the well logs.

3.3. *Seismogram Inversion - The Model*

This section formalizes the problem of seismogram inversion by introducing a mathematical model relating the deconvolved seismogram to impedance. A speculation is that the same technique can be used iteratively, using the original seismogram as input.

Denoting the deconvolved seismogram by x_j and the reflection sequence by c_j , we have the following equation relating x_j and c_j:

$$\beta_j x_j = c_j + n_j \qquad (3.8)$$

$$|c_j| \leq 1$$

In equation (3.8), n_j represents additive and convolutional noise and β_j is a scale factor, possibly time-varying. In practice, the scale factor is chosen by calibrating x_j against a nearby well log, and henceforth we assume that x_j has been pre-scaled, so that β_j can be ignored. The reflection sequence $\{C_j\}$ is related to the impedance sequence $\{Z_j\}$ through a differential (incorporating a factor of 2 in $\{C_j\}$):

$$C_j = Z_{j+1} - Z_j$$

In Chapter I, it was shown that if $\{Z_j\}$ is an M-state Markov chain, then $\{C_j\}$ is also Markov, provided that the state-space is expanded to M^2 states. This means that equation (3.8) is in the standard form for application of the VA:

$$\text{Model: } X_j = (Z_{j+1} - Z_j) + N_j \qquad (3.9)$$

Given: decon. $\underline{x}$, PTM for Z_j, Distribution of $\{N_j\}$

Desire: $\underline{z}$

Actually, one additional piece of information is usually provided,

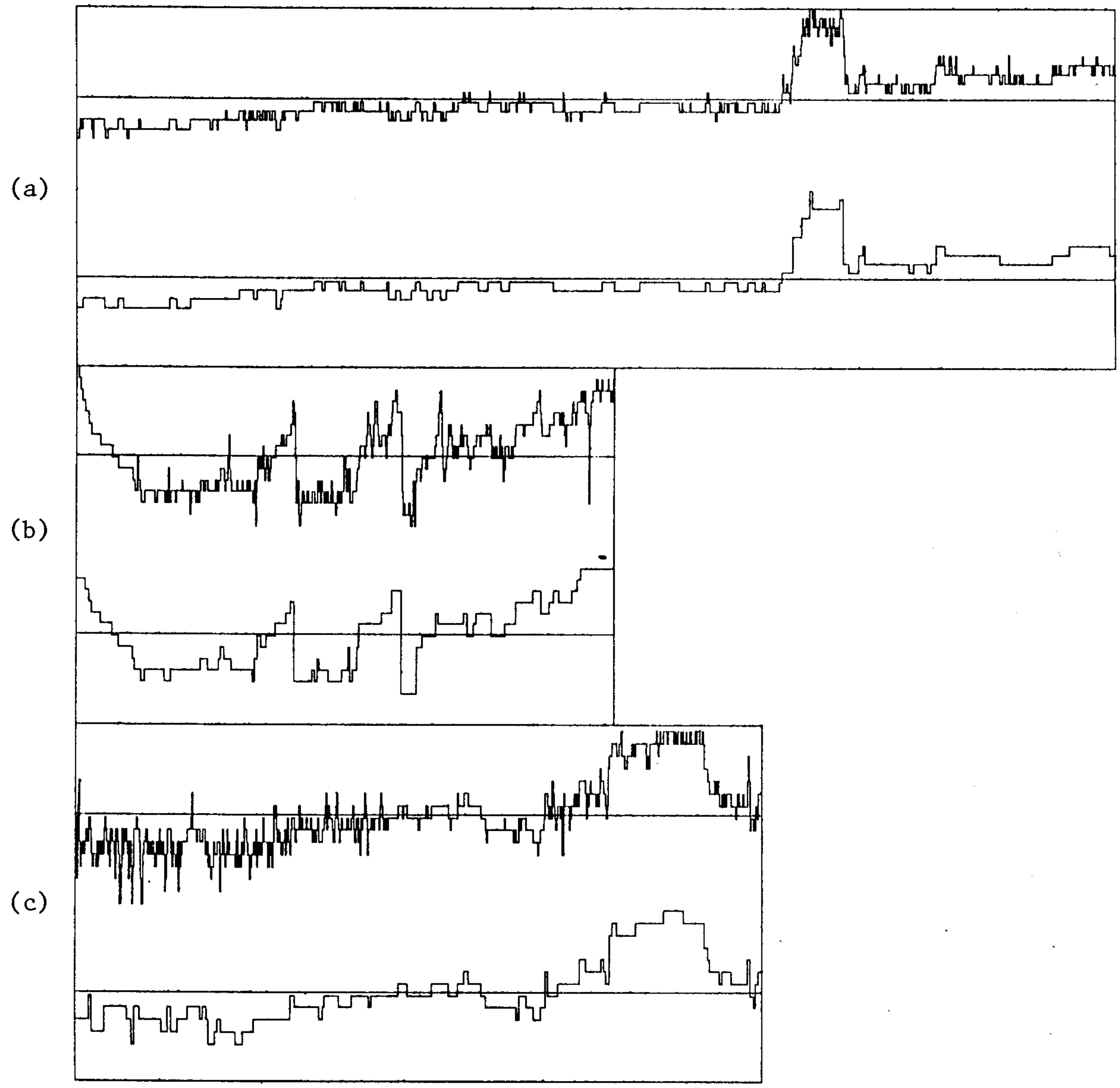

Figure 3.3. Deglitching the three logs of Chapter I. Note the removal of quantization noise. For this example, the PTM used was P_T and the noise parameter S=50.

namely the mean-square velocity, $\bar{v}_j^2$, is specified at certain time points or picks:

$$\bar{v}_j^2 = \frac{\sum_{1}^{j} v_i^2}{j} \qquad (3.10)$$

In equation (3.10), the interval velocity, v_j, is related to the log(imped), z_j, via

$$v_j = v_0 \exp(z_j) \qquad (3.11)$$

In equation (3.11), density is assumed constant. This assumption is removable if velocity-compaction laws are incorporated in equation (3.11). The VA as applied to equation (3.9) is similar to the deglitching problem, with a few important distinctions. The next section discusses these differences in detail.

3.4. *Seismogram Inversion - The Algorithm*

There are three major differences between the implementations of the VA for seismogram inversion and deglitching.

First, using the special structure of the reflectivity state space, it is possible to reduce the cost of the algorithm to a term proportional to M^2 vs. M^4 per time step. Figure 3.4 illustrates how the reduction is achieved. Essentially, the key observation is to recognize that the minimal path to each state in the q^{th} terminator group is identical and can be initiated only by a state within the q^{th} initiator group. Since there are both M initiator and terminator groups, the M^2 term is apparent.

Second, it is conceptually attractive to retain the continuous nature of the RV X_j, whereas in the deglitching problem we quantized the observations. This is feasible, provided the term $Pr(X_j|C_j)$ is defined by

$$Pr(X_j|C_j) = Pr(x-\frac{dx}{2} \le X_j \le x+\frac{dx}{2} \, |C_j=c)$$

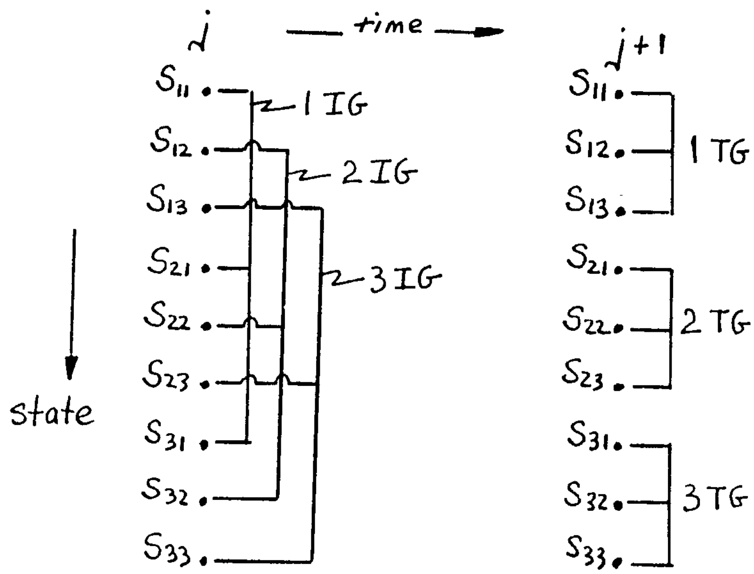

Figure 3.4. Two adjacent time steps (j,j+1) in the state-time lattice for the reflectivity state space (M=3). Each state within a terminator group (TG) has the same minimal path. This path can be initiated only by a state within the corresponding initiator group (IG). Using this structure, the algorithm cost is reduced to a term proportional to M^2 vs. M^4.

$$= f_N(x-c)dx$$

Since dx is a constant, not influencing the MAP estimator, it can be ignored:

$$Pr(X_j|C_j) = f_N(x-c)$$

In practice, a Gaussian distribution was used for $\{N_j\}$.

The third difference is the most important. In the deglitching problem, if the states (or rocks) were known at a particular set of time points, a corresponding set of knots could be prescribed in the VA and the optimal path between knots could then be computed. Unfortunately, a $\bar{v}_j^2$ pick doesn't fix the interval velocity but rather the average square of the interval velocity up to the time pick. To incorporate these integrated measurements in the VA, an intuitive, sub-optimal solution can be effected by considering them to constitute a new set of "observations." It is necessary to interpolate between picks to extend the range of "observations." A new cost function can then be defined by prescribing a conditional probability function relating the actual mean square velocity, $\bar{v}_j^2$, with that obtained by integrating along each minimal path in the state-time lattice, $\tilde{v}_j^2$. The Gaussian function is convenient to use:

$$Pr(\tilde{V}_j^2=\tilde{v}_j^2|\bar{V}_j^2=\bar{v}_j^2) = \exp\left[\frac{-(\tilde{v}_j^2-\bar{v}_j^2)^2)}{2\sigma_{v_j}^2}\right]$$

or

$$-\log\left[Pr(\ |\ )\right] = \frac{(\tilde{v}_j^2-\bar{v}_j^2)^2}{2\sigma_{v_j}^2} \qquad (3.12)$$

By choosing $\sigma_{v_j}^2$ appropriately, a "tube" through which the mean-square velocity is allowed to range is defined. Conceptually, the tube is narrow at those points where the $\bar{v}_j^2$ picks are located and gradually widens to a maximum, midway between picks. Figure 3.5 illustrates this point.

To show how the low frequency component can be restored in an inverted deconvolution by using the mean square velocity picks as an additional set of "observations," a field recording was simulated. A low-cut filter was applied to the first reflectivity sequence of Chapter I (Figure 3.6). We assumed, however, that the mean square velocity (or equivalently, the mean square impedance in this example) was known at 7 points (shown in Figure 3.7). Interpolating linearly between the known

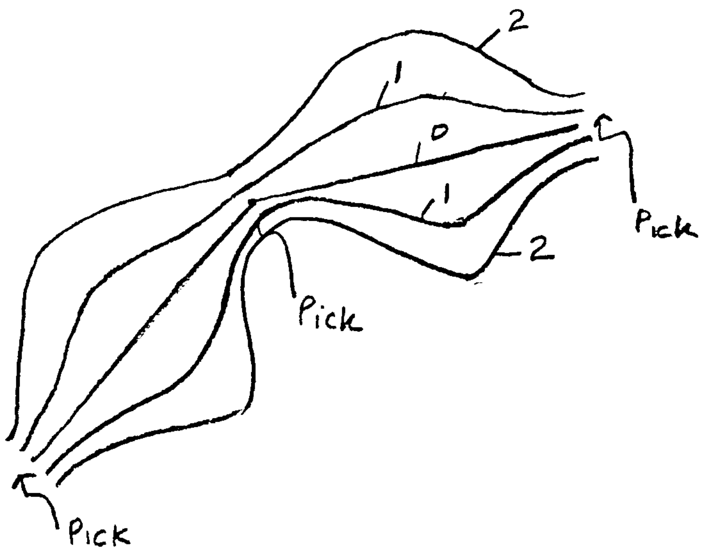

Figure 3.5. Contours of a hypothetical mean square velocity cost function. At the three picks, the tube narrows, forcing the mean square velocity to be close to the picked velocity. Midway between picks, the tube reaches its maximum width. A section perpendicular to the line joining picks has an inverted Gaussian profile if a Gaussian probability function is used to describe the cost function.

picks provided us with another set of "observations" (the filtered reflectivity constituting the first set). Using the technique outlined above, Figure 3.8 shows the output of the VA. It is evident that the trend has been restored. Of course, the success of the algorithm is correlated with the density of mean square velocity picks.

Finally, we note that the concept that impedance is physically confined to values within a maximum and minimum "corridor" is inherent in the VA. This prevents random walks from occurring outside the corridor when a deconvolution is inverted. Figure 3.9 illustrates the point with a synthetic. A telegraph wave was synthesized and differentiated to produce a reflectivity sequence. Gaussian noise was then added, and this provided the input to the VA. Note the absence of random walks in the computed impedance. The errors made using other techniques are contrasted with the VA in Figure 3.10.

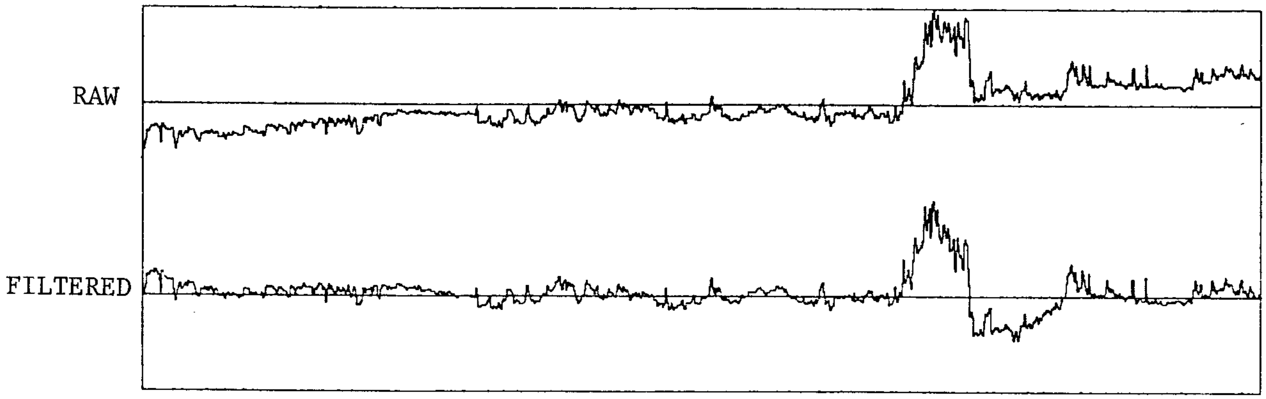

Figure 3.6. A 100-point running average of the original reflectivity series was subtracted from it to give a simulated field recording. Shown are the original impedance (top) and the "filtered" impedance (bottom).

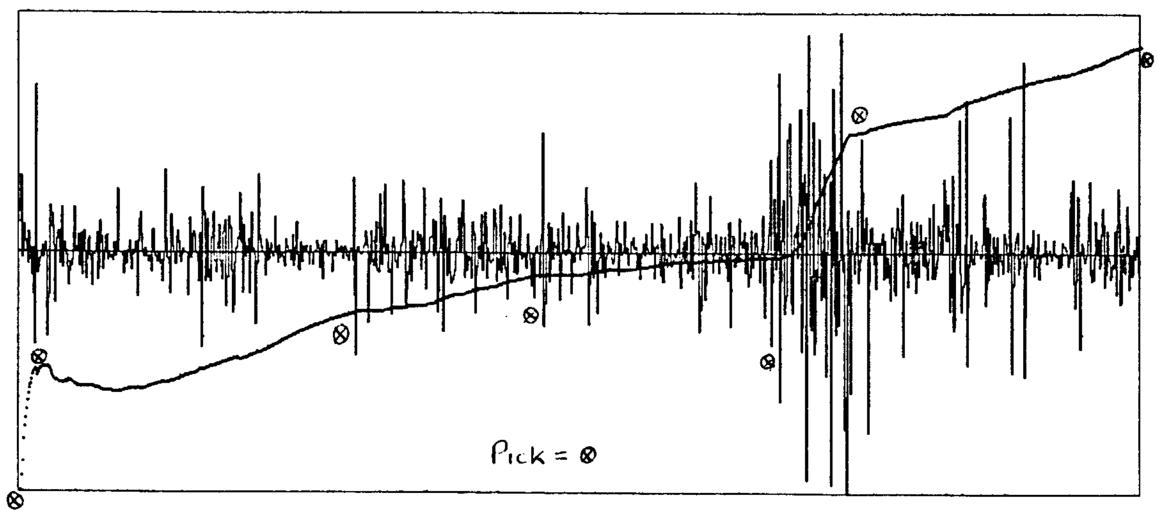

Figure 3.7. The mean square velocity curve (thick curve) superposed on the original reflectivity series. The 7 picks shown were used by the VA to restore the low-frequency component.

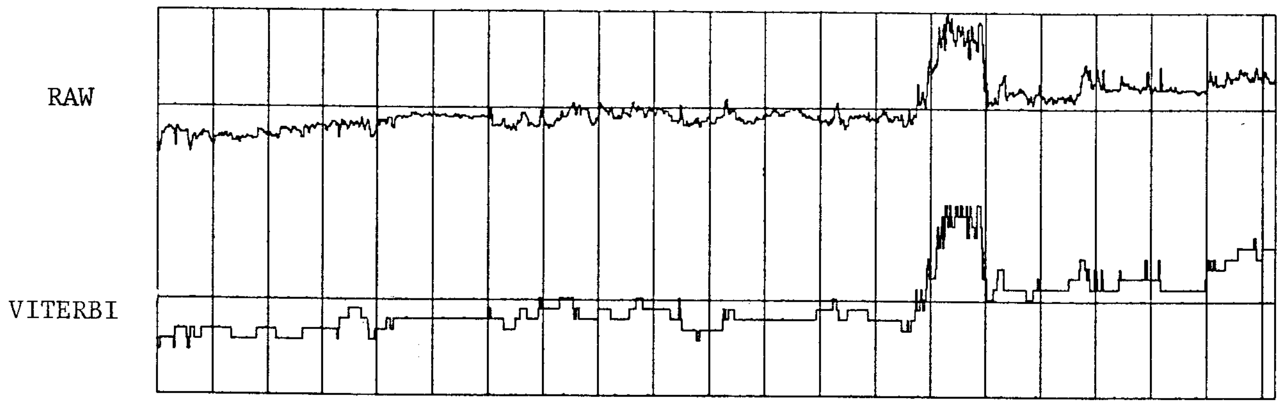

Figure 3.8. The restored impedance (bottom) contrasted with the original impedance (top). Agreement is generally good.

3.5. Conclusion

A few indications of the potential use of the stochastic model presented in Chapter I have been developed in this chapter. A number of areas are open to further research: (1) Only one well log was used in the deglitching application. If a suite of logs were available, the various logs could be analyzed together. This is a case of having vector vs. scalar observations. (2) An observation made by E. Eisner (personal communication) was that a major problem in well logs is cycle skipping. He suggested that a more reasonable noise function (Figure 3.1) would contain two peaks, thus incorporating the possibility of correcting for cycle skips in the VA. (3) Merging the mean square velocity curve into the VA required a sub-optimal scheme. We adopted the present one because of its intuitive appeal. Other more theoretically appealing schemes exist and these need to be explored.

A model's performance can be analyzed only after extensive tests on real data have been made. Ultimately, this is our goal. In this chapter, we have simply introduced the concepts of the Viterbi Algorithm as applied to some geophysical problems.

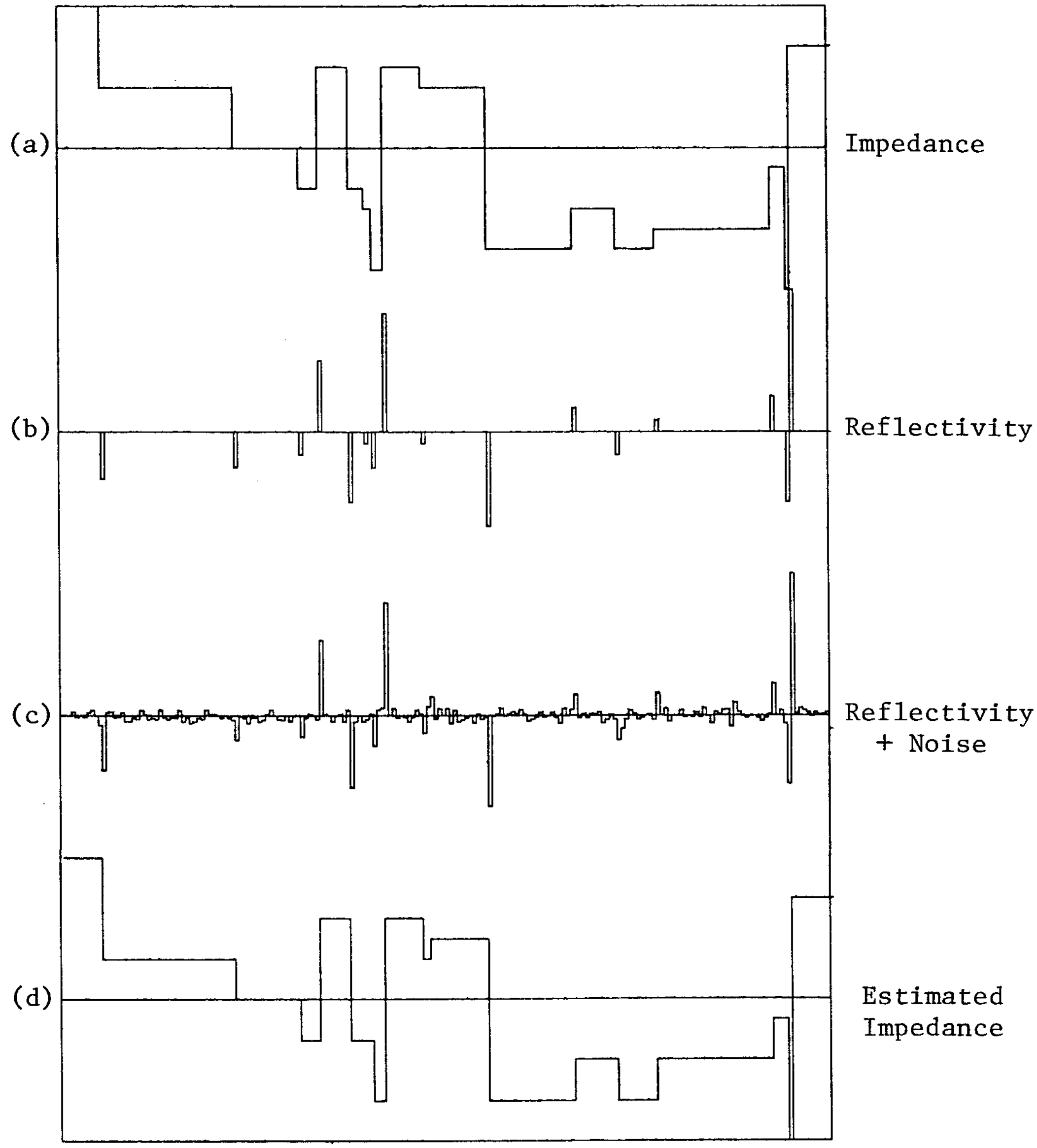

Figure 3.9. (a) A telegraph wave synthesized using M=15, λ=0.9, and a uniform probability mass function. (b) Differential of 9a. (c) Adding 10% noise to 9b. (d) The reflectivity in 9c was input to the VA and 9d resulted. Two knots were specified - the first and last point of the log - in the VA. Geologically, this corresponds to knowing the impedance at the surface and at the basement.

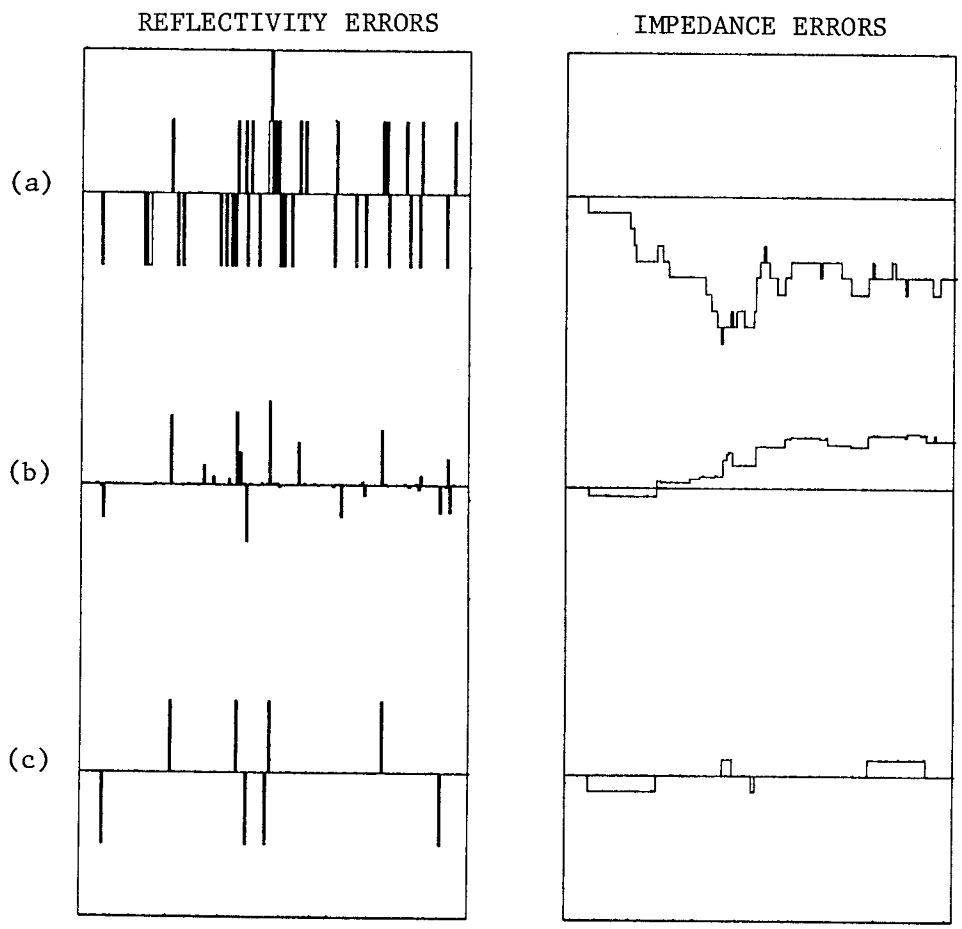

Figure 3.10. (a) The reflectivity in 9c was quantized into 15 states and subtracted from 9b to give the left-hand plot. The vertical scale is exaggerated. Integrating the left-hand plot gave the figure on the right. (b) The ZNL of the previous chapter was used to estimate reflection coefficients in 9c and these were subtracted from 9b to give the LHS of 10b. (c) Differentiating 9d and subtracting from 9b gave this result.